

LOGIC,
LANGUAGE,
AND MEANING

VOLUME 2
Intensional
Logic and
Logical
Grammar

L.T.F. Gamut

LOGIC, LANGUAGE, AND MEANING

L. T. F. Gamut

LOGIC, LANGUAGE, AND MEANING

VOLUME II
Intensional Logic
and Logical Grammar

The University of Chicago Press
Chicago and London

L. T. F. Gamut is a collective pseudonym for J. F. A. K. van Benthem, professor of mathematical logic, J. A. G. Groenendijk, associate professor in the departments of philosophy and computational linguistics, D. H. J. de Jongh, associate professor in the departments of mathematics and philosophy, M. J. B. Stokhof, associate professor in the departments of philosophy and computational linguistics, all at the University of Amsterdam; and H. J. Verkuyl, professor of linguistics at the University of Utrecht.

The University of Chicago Press, Chicago 60637

The University of Chicago Press, Ltd., London

© 1991 by The University of Chicago

All rights reserved. Published 1991

Printed in the United States of America

99 98 97 96 95 94 93 92 91 5 4 3 2 1

First published as *Logica, Taal en Betekenis*, (2 vols.) by Uitgeverij Het Spectrum, De Meern, The Netherlands. Vol. 1: *Inleiding in de logica*, vol. 2: *Intensionele logica en logische grammatica*, both © 1982 by Het Spectrum B.V.

Library of Congress Cataloging in Publication Data

Gamut, L. T. F.

[*Logica, taal en betekenis*. English]

Logic, language, and meaning / L. T. F. Gamut.

p. cm.

Translation of: *Logica, taal en betekenis*.

Includes bibliographical references.

Contents: v. 1. Introduction to logic — v. 2. Intensional logic and logical grammar.

ISBN 0-226-28084-5 (v. 1). — ISBN 0-226-28085-3 (v. 1, pbk.). —

ISBN 0-226-28086-1 (v. 2) — ISBN 0-226-28088-8 (v. 2, pbk.)

1. Logic. 2. Semantics (Philosophy) 3. Languages—Philosophy.

I. Title.

BC71.G33513 1991

160—dc20

90-10912

CIP

∞ The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

Contents

Foreword	ix
Preface	xi

Chapter 1 The Origins of Intensional Logic

1.1	Introduction	1
1.2	The Correspondence Theory of Meaning	1
1.3	Naturalism versus Conventionalism	2
1.4	Variants of the Correspondence Theory of Meaning	3
1.5	Logical Semantics as a Referential Theory of Meaning	4
1.6	Problems with the Referential Theory of Meaning	6
1.7	Frege's Theory of Meaning	7
1.8	Context Dependence	13

Chapter 2 Intensional Propositional Logic

2.1	Introduction	16
2.2	Possible Worlds Semantics	16
2.3	Modal Propositional Logic	19
	2.3.1 Historical Background	
	2.3.2 Syntax and Semantics	
	2.3.3 The Syntactic Approach to the Notion of Validity	
	2.3.4 Alethic and Epistemic Modalities	
	2.3.5 An Application	
2.4	Propositional Tense Logic	32
	2.4.1 Syntax and Semantics	
	2.4.2 'Now': an Extension	
	2.4.3 Other Approaches	
2.5	Tense and Modality Combined	40

Chapter 3 Intensional Predicate Logic

3.1	Opaque Contexts: Modalities de Dicto and de Re	45
3.2	Proper Names and Definite Descriptions: Rigid Designation	49
3.3	The Semantics of Modal Predicate Logic	53
	3.3.1 Formulas without Variables	
	3.3.2 Identity	
	3.3.3 Variables and Quantifiers	
	3.3.4 One Domain: The Existence Predicate	
3.4	Other Kinds of Contexts	66
3.5	A Methodological Note	70

Chapter 4 The Theory of Types and Categorical Grammar

4.1	Introduction	75
4.2	The Theory of Types	75
	4.2.1 Type Distinctions in Natural Language	
	4.2.2 Syntax	
	4.2.3 Semantics	
4.3	Categorical Grammar	92
	4.3.1 Introduction	
	4.3.2 Characteristics of Categorical Grammar	
	4.3.3 The Descriptive Adequacy of Categorical Grammar	
	4.3.4 Categorical Grammar and the Theory of Types	
4.4	λ -Abstraction	102
	4.4.1 The λ -Operator	
	4.4.2 λ -Conversion	
	4.4.3 The λ -Operator and Compositionality	

Chapter 5 The Intensional Theory of Types

5.1	Introduction	117
5.2	Intensional Constructions and Intensional Concepts	117
5.3	Syntax	118
5.4	Semantics	121
5.5	The Operators \wedge and \vee	128
5.6	λ -Conversion	131
5.7	Temporal Operators	132
5.8	Two-Sorted Type Theory	133

Chapter 6 Montague Grammar

6.1	Introduction	139
	6.1.1 Compositionality of Meaning and Syntax	
	6.1.2 Object Language and Metalanguage: Semantic Closure	
	6.1.3 Semantics and Truth Theory	
6.2	The Organization of a Montague Grammar	146
6.3	A Montague Grammar for a Fragment of English	150
	6.3.1 Categories and Basic Expressions	
	6.3.2 Terms, Intransitive Verbs, Sentences	
	6.3.3 The Organization of the Translation Process	
	6.3.4 The Translation of Terms	
	6.3.5 Transitive Verbs	
	6.3.6 The Function of Meaning Postulates	
	6.3.7 Meaning Postulates for the Fragment	
	6.3.8 Scope Ambiguities, de Re Readings, and Rules of Quantification	
	6.3.9 The Transitive Verb <i>Be</i>	
	6.3.10 Conjunction Rules, Disjunction Rules, and Negation Rules	
	6.3.11 Sentential and Infinitival Complements, Adjectives, Relative Clauses, and Adverbs	
6.4	Individual Concepts	203
	6.4.1 Arguments for the Introduction of Individual Concepts	
	6.4.2 Consequences of the Introduction of Individual Concepts	
	6.4.3 Some Examples	
	6.4.4 Meaning Postulates	
6.5	Compositionality, Logical Form, and Grammatical Form	214
6.6	Concluding Remarks	220

Chapter 7 Recent Developments

7.1	Introduction	222
7.2	The Theory of Generalized Quantifiers	223
	7.2.1 Principal Objectives	
	7.2.2 NPs as Generalized Quantifiers in Montague Grammar	
	7.2.3 Determiners: Two Perspectives	

7.2.4 Some Fundamental Properties of NPs
and Quantifiers

7.2.5 Global Constraints

7.2.6 Logical Determiners

7.2.7 Further Developments

7.3 Flexible Categorical Grammar and Type Theory 257

7.3.1 Category Change

7.3.2 A Logical Point of View

7.3.3 Further Developments

7.4 Discourse Representation Theory 264

7.4.1 Introduction

7.4.2 Some Problems with Anaphoric Relations
and Indefinite Terms

7.4.3 An Informal Introduction to DRT

7.4.4 Formal Definitions

7.4.5 DRT and Compositionality

7.4.6 Conclusion

Solutions to Selected Exercises 299

Bibliographical Notes 335

References 337

Index 345

Foreword

The Dutch not only have what must be the greatest number of linguists per capita in the world, they also have a very long and rich tradition of combining linguistics, logic, and philosophy of language. So it should not be a surprise that it is an interdisciplinary collaboration of Dutch scholars that has produced the first comprehensive introduction to logic, language, and meaning that includes on the one hand a very fine introduction to logic, starting from the beginning, and on the other hand brings up at every point connections to the study of meaning in natural language, and thus serves as an excellent introduction and logical background to many of the central concerns of semantics and the philosophy of language as well.

This book is pedagogically beautifully designed, with the central developments very carefully introduced and richly augmented with examples and exercises, and with a wealth of related optional material that can be included or omitted for different kinds of courses (or self-teaching) for which the book could very well be used: I could imagine tailoring very fine but slightly different courses from it for inclusion in a linguistics curriculum, a philosophy curriculum, a cognitive science curriculum, or an AI/computational linguistics program. It would be less suitable for a logic course within a mathematics department, since there is less emphasis on proofs and metamathematics than in a more mathematically oriented logic book. There is certainly no lack of rigor, however; I think the authors have done a superb job of combining pedagogical user-friendliness with the greatest attention to rigor where it matters.

One very noticeable difference from familiar introductory logic texts is the inclusion of accessible introductions to many nonstandard topics in logic, ranging from approaches to presupposition and many-valued logics to issues in the foundations of model theory, and a wide range of more advanced (but still very accessible) topics in volume 2. The book thereby gives the student an invaluable perspective on the field of logic as an active area of growth, development, and controversy, and not simply a repository of a single set of eternal axioms and theorems. Volume 2 provides an outstanding introduction to the interdisciplinary concerns of logic and semantics, including a good introduction to the basics of Montague grammar and model-theoretic semantics more generally.

I first became acquainted with this book in its Dutch version during a sabbatical leave in the Netherlands in 1982–83; it made me very glad to have learned Dutch to be able to appreciate what a wonderful book it was, but at the same time sorry not to be able to use it immediately back home. I started lobbying then for it to be translated into English, and I'm delighted that that has become a reality. I hope English-speaking teachers and students will appreciate the book as much as I anticipate they will. The authors are top scholars and leaders in their fields, and I believe they have created a text that will give beginning students the best possible entry into the subject matter treated here.

BARBARA H. PARTEE

Preface

Logic, Language, and Meaning consists of two volumes which may be read independently of each other: volume 1: *Introduction to Logic* and volume 2: *Intensional Logic and Logical Grammar*. Together they provide a survey of modern logic from the perspective of the analysis of natural language. They represent the combined efforts of two logicians, two philosophers, and one linguist. An attempt has been made to integrate the contributions of these different disciplines into a single consistent whole. This enterprise was inspired by a conviction shared by all of the authors, namely, that logic and language are inseparable, particularly when it comes to the analysis of meaning. Combined research into logic and language is a philosophical tradition which can be traced back as far as Aristotle. The advent of mathematical logic on the one hand and structuralist linguistics on the other were to give rise to a period of separate development, but as these disciplines have matured, their mutual relevance has again become apparent. A new interdisciplinary region has emerged around the borders of philosophy, logic, and linguistics, and *Logic, Language, and Meaning* is an introduction to this field. Thus volume 1 establishes a sound basis in classical propositional and predicate logic. Volume 2 extends this basis with a survey of a number of richer logical systems, such as intensional logic and the theory of types, and it demonstrates the application of these in a logical grammar.

Logic is introduced from a linguistic perspective in volume 1, although an attempt has been made to keep things interesting for readers who just want to learn logic (perhaps with the exception of those with a purely mathematical interest in the subject). Thus some subjects have been included which are not to be found in other introductory texts, such as many-valued logic, second-order logic, and the relation between logic and mathematical linguistics. Also, a first attempt is made at a logical pragmatics. Other and more traditional subjects like the theory of definite descriptions and the role of research into the foundations of mathematics have also been dealt with.

Volume 2 assumes a familiarity with propositional and predicate logic, but not necessarily a familiarity with volume 1. The first half of it is about different systems of intensional logic and the theory of types. The interaction between the origins of these systems in logic and philosophy and the part they have to play in the development of intensional theories of meaning is a common the-

matic thread running through these chapters. In the course of this exposition, the careful reader will gradually obtain a familiarity with logic and philosophy which is adequate for a proper understanding of logical grammar. Montague grammar, the most well known form of logical grammar, is described in detail and used on a fragment of the English language. Following this, attention is paid to some more recent developments in logical grammar, such as the theory of generalized quantification and discourse representation theory.

One important objective of this book is to introduce readers to the tremendous diversity to be found in the field of formal logic. They will become acquainted with many different logics—that is to say, combinations of formal languages, semantic interpretations, and notions of logical consequence—each with its own field of application. It is often the case in science that one is only able to see which of one's theories will explain what and how they might be modified or replaced when one examines the phenomena very closely. In this field too, it is the precise, formal analysis of patterns and theories of reasoning which lead to the development of alternatives. Here formal precision and creativity go hand in hand.

It is the authors' hope that readers will develop an active understanding of the matters presented, will come to see formal methods as flexible methods for answering semantic questions, and will eventually be in a position to apply them as such. To this end many exercises have been included. They should help make the two volumes suitable as texts for courses, the breadth and depth of which could be quite diverse. Solutions to selected exercises (marked with an asterisk) have also been included in order to facilitate individual study.

In order to underline their common vision, the authors of these two volumes have merged their identities into that of L. T. F. Gamut. Gamut works (or at least did work at the time of writing) at three different universities in the Netherlands: Johan van Benthem as a logician at the University of Groningen; Jeroen Groenendijk as a philosopher, Dick de Jongh as a logician, and Martin Stokhof as a philosopher at the University of Amsterdam; and Henk Verkuyl as a linguist at the University of Utrecht.

This work did not appear out of the blue. Parts of it had been in circulation as lecture notes for students. The exercises, in particular, derive from a pool built up through the years by the authors and their colleagues. The authors wish to express their thanks to all who have contributed in any way to this book. Special thanks are due to Piet Rodenburg, who helped write it in the early stages, to Michael Morreau for his translation of volume 1 and parts of volume 2, and to Babette Greiner for her translation of most of volume 2.

Summary of Volume 2

Chapter 1 provides a background to the systems of intensional logic presented in chapters 2 and 3. The nature and limits of the semantics of predicate logic

are discussed, and Frege's attempts to develop an intensional theory of meaning are sketched.

Chapter 2 is concerned with the propositional part of intensional logic. A general characterization of possible-worlds semantics is given, and it is then demonstrated with reference to modal propositional logic and propositional tense logic. Attention is paid not only to logical and philosophical matters but also to potential applications in the analysis of natural language.

The same applies to the treatment of intensional predicate logic given in chapter 3, which appears here almost exclusively as a modal predicate logic. Different alternative semantic options are compared. Matters like rigid designation are discussed, along with certain more general methodological questions which come up in connection with intensional logic.

Chapter 4 introduces and compares the theory of types and categorial grammar. One important reason for preferring type-theoretical languages is the syntactic and semantic diversity of natural language. Increasing the applicability of logical systems in the systematic analysis of natural language is also the most important reason for introducing λ -abstraction. Anticipating the discussion of Montague grammar in chapter 6, certain methodological requirements are discussed which must be met in such an application, together with the role which λ -abstraction can play in helping to meet these requirements. Chapter 4 also contains an exposition of the principles of categorial grammar.

In chapter 5, intensional logic and the theory of types are combined. The resulting intensional theory of types is the logical system exploited in Montague grammar in order to provide a logical semantics for (a fragment of) a natural language. A section on two-sorted type theory has been included in order to render certain formal properties of the intensional theory of types more comprehensible.

Chapter 6 begins with a discussion of a number of the assumptions made when systematically applying logical systems in the semantic analysis of natural language. This is followed by an exposition of the best-known model of logical grammar, viz., Montague grammar. The form and function of Montague grammar are demonstrated in detail with reference to the syntax and semantics of a fragment of English.

Chapter 7 is a survey of three recent developments in model-theoretic semantics of natural language. The first of these is the theory of generalized quantifiers which has been developed in the late seventies and which builds on the analysis of quantified expressions that can be found in Montague grammar. This development is particularly interesting because it brings logical grammar within the realm of empirical constraints. Next, some attention is paid to recent attempts to make 'classical' categorial grammar into a better tool for natural language description. The third subject is discourse representation theory, which was developed in the early eighties. This theory aims at

improving logical grammar with respect to problems of anaphoric relations and at extending it to the level of discourse.

Biographical notes and references to relevant literature conclude this volume, without any pretence at being exhaustive.

Required Background Knowledge and Notation

The reader is assumed to be familiar with the syntax and semantics of propositional logic and predicate logic and with elementary set theory, including the notion of a function. (Chapters 2 to 4 of volume 1 provide a suitable introduction here.) In particular, the reader should have an understanding of the notions of a formal language and a formula. As for the notations used in propositional logic: here we make use of the connectives \wedge (conjunction), \vee (disjunction), \neg (negation), \rightarrow ((material) implication), and \leftrightarrow ((material) equivalence). The letters p, q, r are used to refer to propositional letters; where necessary they are provided with primes or subscripts, as in p', p'', p_0, p_1 , and so on. These symbols, together with the brackets '(' and ')' enable us to introduce formulas like $\neg\neg\neg(p \rightarrow q)$ and $((p \wedge q) \vee r)$. In general, only the outermost brackets are left off, as in $(p \wedge q) \vee r$. The Greek letters $\phi, \psi, \chi, \phi', \phi''$, etc., are used as metavariables referring to formulas in general. Concepts like *propositional formula* and *formula of predicate logic* are introduced by means of inductive (that is to say, recursive) definitions. Such definitions always end with a so-called induction clause, decreeing that nothing is a formula that is not required to be such by the foregoing clauses. The notion of an inductive proof is also introduced in volume 1, but such mathematical proof techniques have been avoided in the text.

Unlike in some other texts, printed successions of symbols are here not to be thought of as the formulas themselves but as names which refer to these formulas. For example, the symbol (string of length 1) \wedge merely refers to the conjunction sign. It is not itself the conjunction sign. Thus sentences like \wedge *is the conjunction sign* do not constitute an abuse of notation; there are no quotation marks missing. (Incidentally, in both volumes italics are preferred to quotation marks for mention of an expression.) On the same basis, there is no unique language for propositional logic: any given set of propositional letters generates its own language, i.e., its own set of formulas built up from those propositional letters. That is, here 'p', 'q', and 'r' are not themselves propositional letters; they are metavariables referring to the propositional letters in any of a variety of different languages. Similar remarks hold for predicate logic.

With respect to predicate logic, the reader is assumed to be familiar with (individual) constants (notation: a, b, c, c_1, c_2 , etc.), with variables (notation: x, y, z, z_1, z_2 , etc.), with the distinction made between free and bound (occurrences of) variables in formulas, and with the notion of the scope of a quantifier \forall (universal quantifier) and \exists (existential quantifier). Formulas lacking

free variables are referred to as sentences. For the formula resulting from the substitution of y for (the free occurrences of) x in a formula ϕ , we have the notation $[y/x]\phi$.

In the semantics of propositional logic we make use of valuations, written V, V' etc. The notions of a tautology, a contradiction, and the concept of (logical) equivalence are presupposed, as is the notion of the validity of an argument schema $\phi_1, \dots, \phi_n/\psi$. For ' ϕ is a tautology' we write $\models \phi$, and for ' $\phi_1, \dots, \phi_n/\psi$ is a valid argument scheme', we write $\phi_1, \dots, \phi_n \models \psi$. The negations of these two are $\not\models \phi$ and $\phi_1, \dots, \phi_n \not\models \psi$, respectively.

The semantics of predicate logic is given in terms of models (M, M' , and so on), consisting of a domain D together with an interpretation function I_M , which assigns suitable values to the constants and predicate letters. Thus, by means of Tarski's truth definition, any given model M has its own valuation function V_M , which assigns to each sentence in the language in question either the truth value 1 (the sentence is true) or the truth value 0 (it is false), and a variety of valuation functions $V_{M,g}$ for the formulas. The latter valuations depend on which assignment g is chosen, these assignments being functions mapping the variables of the language in question into D . A sentence ϕ is said to be 'true in the model M ' just in case $V_M(\phi) = 1$. In connection with assignments, the following notation is useful: $g[x/d]$ refers to the assignment which assigns the value d to the variable x , and which otherwise agrees with g .

Given this semantics, notions like the universal validity of formulas, the validity of argument schemata, the equivalence of sentences, and (via assignments) the equivalence of formulas may be introduced. Principles like $\chi \leftrightarrow \chi' \models \phi \leftrightarrow [\chi'/\chi]\phi$, $s = t \models \phi \leftrightarrow [t/s]\phi$, and $\forall x_1, \dots, x_n(\chi \leftrightarrow \chi') \models \phi \leftrightarrow [\chi'/\chi]\phi$ are referred to as principles of extensionality.

Some familiarity with a syntactic derivability notion $\phi_1, \dots, \phi_n \vdash \psi$ (ψ is derivable from ϕ_1, \dots, ϕ_n) is useful but not essential. An axiomatic derivability notion would do just as well as the system of natural deduction which was introduced in volume 1. An understanding of the meaning of meta-logical theorems like the completeness theorem and its converse, the soundness theorem, is similarly desirable but not essential to the reader of this volume.

Among the set-theoretical notations used here, we have \emptyset for the empty set; \cap for intersection, and \cup for the union of sets; $\{1, 2\}$ for the set containing just 1 and 2, by way of example; and $\langle a_1, \dots, a_n \rangle$ for an ordered n -tuple. $A \times B$ refers to the Cartesian product $\{\langle a, b \rangle \mid a \in A \text{ and } b \in B\}$ of A and B ; $A \subseteq B$ means ' A is a subset of B ' (not necessarily a proper one). The set $\{A \mid A \subseteq B\}$ of all subsets of a set B is called the power set of B , and for this we have the notation $\text{POW}(B)$. Properties of relations like (ir)reflexivity, symmetry, and transitivity are assumed to be familiar.

Definite descriptions $\iota x \phi$ (*the x such that ϕ*) are usually analyzed in the Russellian way: a formula ψ containing a definite description $\iota x \phi$ is read as $\exists x(\forall y([y/x]\phi \leftrightarrow x = y) \wedge [x/\iota x \phi]\psi)$, assuming that y is free for x in ϕ .

Finally, mention should be made of the way in which translations from natural language into the formal languages of predicate and propositional logic are to be presented. An example will make this clear enough:

All teachers love Billy, but he does not love all teachers.

Translation into propositional logic: $p \wedge \neg q$

Key: p : All teachers love Billy; q : Billy loves all teachers.

Translation into predicate logic: $\forall x(Tx \rightarrow Lxb) \wedge \neg \forall x(Tx \rightarrow Lbx)$

Key: b : Billy; Tx : x is a teacher; Lxy : x loves y .

Such translations are of importance mainly in the exercises.

1

The Origins of Intensional Logic

1.1 Introduction

In this volume we shall go into intensional logic at some length. Intensional logic is probably the most important extension of standard logic, by which we mean propositional logic and predicate logic (see volume 1). It has a multitude of applications, both in the analysis of philosophical problems and in research into the semantics of natural language. So before we embark on an exposition of intensional logic itself, let us first sketch the logical and philosophical background against which it was developed.

Intensional logic has more than one root. One of these is an attempt to solve the problems that arise when one tries to extend semantic methods suited to the interpretation of the standard logical systems to the interpretation of languages which are 'richer' than those of propositional and predicate logic. Not all extensions of standard logic require new semantic methods. Some, like second-order logic (see vol. 1, chap. 5) require no more than an adaptation of the methods by means of which the semantics of propositional and predicate logic have been given. But for other extensions it is different. An example of such a system is modal propositional logic, which is the subject of chapter 2. There an expression \Box is added which is to be interpreted as *necessarily* or *it is necessary that*. An adequate semantics for this logical system requires a real extension of standard semantic methods. A natural language like English is of course a second example of such a richer language. Simply transferring the semantics of predicate logic to the semantics of English gives rise to all kinds of problems which indicate the need for richer semantic methods. In order to see this point clearly, it is important to understand just what kind of theory of meaning is inherent in the semantics of standard logic.

1.2 The Correspondence Theory of Meaning

There is a family of theories of meaning which all start out from the following principle: meaning is a relation between the symbols of a language and certain entities which are independent of that language. These theories may collectively be designated as *correspondence theories of meaning*. The 'inde-

pendence' of the entities in the above means, among other things, that the postulated entities are independent of whoever is using the language in question and of the circumstances under which it is used.

This point of departure can hardly be called universal. There are, for example, theories which say that the meaning of a symbol resides in the use which is made of that symbol. The 'meaning is use' theory defended by the later Wittgenstein is an example of one such theory. And then there are theories which identify the meaning of a symbol with the set of all stimuli which elicit the use of that symbol as a response. There, meaning is defined in terms of the disposition of language users to display certain kinds of behavior. As examples we have the behavioristic theories of meaning of Bloomfield, Morris, and Skinner. And finally, there are theories which accept the correspondence theory as a partial account of meaning, in the sense that correspondence to entities is thought to account for just one aspect of the total meaning of symbols. Grice's theory of implicatures (see vol. 1, chap. 6) is an example of such a theory.

The common point of departure of correspondence theories of meaning as formulated above can be expanded in divergent directions. These differ on two main points: the *nature of the relation* between symbols and entities and the *nature of the entities* themselves. The second point seems to be more controversial than the first, which now appears to have been settled to everybody's greater or lesser satisfaction. Nevertheless, we will go into the first very briefly.

1.3 Naturalism versus Conventionalism

The traditional debate on the relation between symbols and entities centers on whether this relation is a natural one or purely a matter of convention. It may not be a very lively debate these days, but the naturalism-conventionalism dispute persisted from classical times up until well into the eighteenth century.

In its most naive form, naturalism states that the meaning of a word is inherent in its sound. The relation between symbol and entity is in this case rather extremely 'natural'. It is clear that this naive form of naturalism is not viable. If it were true, for example, we would have no difficulty in learning a foreign language: presumably we would immediately understand it (at least in its spoken form). Another problem for naive naturalism is the existence of homonyms; words which sound the same may nevertheless have different meanings. And even the phenomenon of onomatopoeia, so dear to naturalism, presents its problems: must the difference between the French *cocorico* and the English *cock-a-doodle-do* be taken as an indication that French cocks and other members of their species on the other side of the channel crow differently at daybreak?

Plato defends a less naive form of naturalism in his dialogue *Cratylus*. He supposes that there is some affinity between certain sounds and properties. For example, he thinks that there is a special affinity between the sound of the

letter *r* and the property of motion. According to Plato, words characterize the essence of that to which they refer by virtue of these kinds of relationships. Some words obtain their meanings quite directly in this manner and others via composition of meanings, etymological relationships, or metaphorical transfer of meaning. But this form of naturalism, like the previous one, faces insuperable problems. For example, none of the natural affinities between sounds and properties is much in evidence if different languages in different language families are compared. And what is the theory to make of synonyms?

The debate between naturalism and conventionalism later developed into a controversy between *anomalists* and *analogists* as to whether language is regular or not. Irregularity was supposed to yield an argument in favor of naturalism. After all, if language is purely a matter of convention, then there would be no need for it to have any irregularities.

Ultimately the idea that meanings are conventions was to emerge victorious. The relation of meaning obtaining between a word and a thing is not natural but conventional. There are of course limits to conventionality, since one is not free to change the meanings of words at will. Such changes have to be in the interest of convenience, familiarity, or something of that kind.

1.4 Variants of the Correspondence Theory of Meaning

Within the correspondence theory, different variants have evolved with divergent ideas on the nature of the entities which form the second argument of the relation of meaning. Here we shall briefly consider three of them.

The first can be referred to as *conceptualism*. According to conceptualism, meaning is a relation between symbols and the contents of consciousness. Concepts, expressed by means of predicates, and propositions, expressed by means of sentences, are mental entities, with language functioning as a system of observable symbols which mediates between individuals, thus making communication possible. Locke has defended such a position: 'The use of words is to be sensible marks of ideas, and the ideas they stand for are their proper and immediate signification' (*Essay Concerning Human Understanding*, chap. 2, book 3). Conceptualistic conceptions of meaning may still be found in modern linguistics: 'Roughly, linguistic communication consists in the production of some external, publicly observable, acoustic phenomenon whose phonetic and syntactic structure encodes a speaker's inner, private thoughts or ideas' (Katz 1966). (Note that Katz changed to the position of a Platonist in his 1981 book.)

A second variant of the correspondence theory of meaning may be referred to as *Platonism*. According to Platonism, concepts and propositions are not mental entities but real things. Only they do not belong to the world of observable phenomena but to the world of ideas. Linguistic symbols refer to things in the observable world only in an indirect manner, via the reflections of the world of ideas in the observable world.

The third variant is what we may call *realism*. According to realism, the

entities to which linguistic symbols bear the relation of meaning all belong to the concrete, observable reality around us: they are individuals, properties, relations, and states of affairs. A typical example of this position is the 'picture theory of meaning', which was presented by Wittgenstein in the *Tractatus Logico-Philosophicus*. The relation between symbols and things is one of reference. This theory, which Wittgenstein was later to abandon, has the fundamental assumption that every symbol in an ideal language would refer to some unique thing, and that every thing would be the reference of some unique symbol.

Extracting from this theory just the idea that the meaning relation is one of reference, we arrive at what might be called a *referential theory of meaning*. It is readily seen to be compatible with any of the above three views on the nature of the entities, since it only states that the meaning of a symbol is that to which it refers. So that a theory of meaning is referential in itself says nothing about the nature of the entities to which symbols refer.

1.5 Logical Semantics as a Referential Theory of Meaning

The semantics of standard logic can be seen as a referential theory of meaning (and thus as a correspondence theory of meaning). Take, for example, the way in which the semantics of predicate logic is taught (see vol. 1, chap. 3). When defining a model for predicate logic, the first thing we do is choose some set of entities as our domain. The set is independent of the expressions which collectively form a predicate-logical language. We then specify a relation between the predicate-logical language in question and the domain. By means of an interpretation function, the constant symbols are assigned individual domain elements, and the predicate symbols are assigned sets of domain elements (or sets of ordered sequences of n domain elements in the case of n -ary predicate letters) as their references. With this as a basis, we are in a position to define the reference relative to this model of all sentences in our language (that is, their truth values), in the so-called truth definition. So this method of semantic interpretation evidently follows a principle which we said to be characteristic of the correspondence theory of meaning in §1.2: meaning is a relation between the symbols of a language and certain entities which are independent of that language.

The semantics of predicate logic is indifferent to the kinds of things we choose to put in the domains of our models. Sets of people, numbers, or mathematical points will all do equally well as domains. In fact, any set at all will do. And whatever the domain may be, the theory of meaning is always a referential one: the meanings of the symbols are always their references.

This identification of reference and meaning and the way in which semantic interpretation proceeds are sufficient to give standard logic a special property which we shall now go into briefly. One important characteristic of the semantic interpretation process, a characteristic which also happens to be shared by

the nonstandard systems we shall meet up with, is that a strict parallelism is maintained between the syntactic constructions and their semantic interpretations. The truth definition mirrors the syntactic definition of the formulas of the language in question. There is a methodological consideration underlying this practice, one which can be traced back to Frege. This German logician and mathematician gave the first satisfactory analysis of sentences with relational predicates and multiple quantification in 1879, in his *Begriffsschrift*. Now the fundamental insight behind his solution to these age-old problems is that every sentence, no matter how complex, is the result of a systematic syntactic construction process which builds it up step by step, and in which every step can receive a semantic interpretation. This is the well-known *principle of semantic compositionality*.

One consequence of all of this is that standard propositional and predicate logic are *extensional* logical systems. A logical system is said to be extensional if expressions with the same reference (or extension) may be freely substituted for each other. So the following theorem, known as the *principle of extensionality*, can be proved for propositional and predicate logic (see §4.2.2 in vol. 1 for a precise formulation of this principle and of the ones below):

$$\chi \leftrightarrow \chi' \models \phi \leftrightarrow [\chi'/\chi]\phi$$

This theorem states that if χ and χ' have the same truth values, then χ' may be substituted for χ in ϕ without a change of truth value. We say that χ and χ' are interchangeable *salva veritate* ('with conservation of truth value'). Predicate logic also satisfies some other principles of extensionality. So it is that we have, for example:

$$\forall x(Ax \leftrightarrow Bx) \models \phi \leftrightarrow [B/A]\phi$$

This theorem says that if A and B express properties which have the same extension (which are had by the same entities), then they may be interchanged in formulas ϕ *salva veritate*. Another example of the extensionality of predicate logic is what is referred to as *Leibniz's law of the indiscernibility of identicals*:

$$s = t \models \phi \leftrightarrow [t/s]\phi$$

in which s and t are terms, that is, if no function symbols occur in the language, individual constants or variables.

Its extensionality is both the strength and the weakness of standard propositional and predicate logic. It shows that in studying the validity of inferences in either of these systems, it suffices to consider the references of expressions and the principle of compositionality (which here amounts to this: the reference of a complex expression is a function of the references of its composite parts). On the other hand, as we shall see, there are also richer languages which are not extensional, for which the semantic methods for standard logical systems are not adequate. As we have already indicated, natural languages

are not the only languages which are essentially richer in this respect. Adding a single expression or construction to either of the standard systems can be enough to create a system which needs an essentially richer semantics. In the following chapters we will be dealing with several such systems. But before we turn to them, let us first briefly consider some of the difficulties which arise if a referential semantics is applied to natural language. For in the main, these problems are the ones which led to the development of intensional logic.

1.6 Problems with the Referential Theory of Meaning

The referential theory of meaning states roughly that the meaning of an expression is to be identified with its reference. As a theory of meaning for natural language, it is faced with insuperable difficulties. It forces us to conclude that since meaning is the same as reference, the proper name *Odysseus* has never meant anything, and that the proper name *Socrates* meant something once but no longer means anything. And a definite description like *the president of the United States of America*, we are forced to conclude, changes its meaning from time to time.

Application of the referential theory of meaning to natural language entails a certain measure of realism, since natural languages are used to say things about reality. But going all the way for a realistic referential theory of meaning would seem to forge too strong a link between meaning and reality. It would be preferable to have the meaning of symbols in some sense be more independent of reality without giving up the idea that there is a relation of reference which holds between symbols and entities. One famous example which illustrates this nicely is the *morning star/evening star paradox*, which was formulated by Frege in "On Sense and Reference" ("Über Sinn und Bedeutung" [1892]). So the man who may be considered to be the originator of the (extensional) standard systems of logic was well aware of the nonextensional character of natural language. As Frege pointed out, the following two statements have different cognitive contents:

- (1) The morning star is the morning star.
- (2) The morning star is the evening star.

Statement (1) is a tautology, an analytical a priori truth, but (2) expresses a significant astronomical discovery and as such is a synthetic a posteriori statement. But both expressions, *the morning star* and *the evening star*, refer to the same thing, namely, the planet Venus. So if meaning and reference were to coincide, then we would have to accept that these two statements have the same meaning, which is obviously not the case. This is a truly paradoxical situation. If (2) is true (which it is), and meaning coincides with reference, then (2) expresses the same thing as (1). But whereas (1) must always have been accepted by anyone who chose to consider the matter, (2) was considered untrue for a long time. The conclusion which Frege draws from this is that

meaning and reference are not the same but are to some extent independent of each other. It is quite possible to know what an expression means without being familiar with its reference, and vice versa. This is not to say that there is no relation at all between meaning and reference. Two expressions which have the same meaning must also have the same reference. So in this sense, meaning determines reference. But, as we have already seen, the reverse need not be true: two expressions, in our example *the morning star* and *the evening star*, may have the same reference without having exactly the same meaning. It is the difference in meaning between the two expressions which accounts for the difference between the meanings of (1) and (2). It would appear, then, that any semantic theory which is to be suited to natural languages will have to distinguish between meaning and reference.

Other problems having to do with the identification of meaning and reference arise in connection with what are called *intensional (sentence) constructions*. If the meaning of an expression is just its reference, then we would expect that expression B with the same reference (and thus meaning) as expression A may always be substituted for A in any sentence, without altering the meaning of that sentence. There are, however, sentences whose meaning is affected by such substitutions. Compare the following two statements:

- (3) John is looking for the supreme commander of the U.S. armed forces.
- (4) John is looking for the president of the United States of America.

The expressions *supreme commander of the U.S. armed forces* and *president of the United States of America* always refer to the same person, but (3) and (4) still do not have the same meaning: (3) can be true while (4) is false, and vice versa. Here, as in the previous example, the obvious solution is to distinguish between meaning and reference and to stipulate that only expressions with the same meaning (and not just the same reference) may be freely substituted for each other. Frege himself was the first to propose a solution along these lines. In a series of articles, of which the above-mentioned article "On Sense and Reference" is the best known, he developed a theory of meaning which has to a large extent been incorporated into modern intensional logic. We will discuss a few aspects of this theory in §1.7. It should be noted that neither the solution sketched here nor the problems which it was designed to solve are necessarily bound to realism, as is apparent from the fact that Frege is generally seen as a Platonist.

1.7 Frege's Theory of Meaning

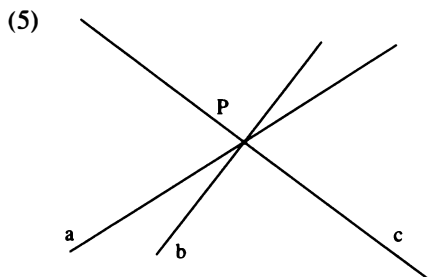
The fundamental distinction drawn in Frege's theory of meaning is that between *sense (Sinn)* and *reference (Bedeutung)*.

According to Frege, there is more to the full meaning of a sentence than just its *sense*; there is also *force (Kraft)* and *tone (Färbung)*. The force of a sen-

tence is something like what nowadays is called its 'illocutionary force'; it is that part of its meaning which determines what function it is to have. It indicates whether we are dealing with an assertion, a question, etc. By the tone of an expression, Frege means the *ideas* (*Vorstellungen*) which a language user associates with an expression. Frege emphasizes that these mental associations are subjective, and that they can therefore not play a part in communication. In communication we can only convey objective things, things which are common to everyone we can communicate with. It is the objective part of meaning that Frege calls *sense*.

So sense and reference are distinguished in Frege's theory of meaning. We have already seen an example by means of which this distinction can be motivated, for what the morning star/evening star paradox shows is that sense and reference are two different things. Not that reference is of no importance in Frege's theory of meaning. Sense and reference are to be distinguished from each other, but that does not mean that they have nothing to do with each other. Reference, we may say, is what explains the function of sense: expressions have a sense only by virtue of the fact that they also have a reference, and their sense is in fact nothing more than the way their reference is presented. Thus sense determines reference. Two expressions with the same sense have ipso facto the same reference, although this does not hold the other way around. Frege developed the distinction between reference and sense in a number of articles and attempted to say exactly what the senses and references of certain kinds of expressions are. In what follows, we shall briefly discuss his views on the senses and references of proper names and sentences.

In Frege's work names include not only proper names like *Amsterdam* and *Socrates* but also definite descriptions like *the morning star*, *the president of the United States of America*, and *the second power of 2*. They are roughly what are called terms in logic: expressions which refer to an entity. So, for Frege, the reference of a name is an entity. Its sense is what Frege calls *the mode of presentation* (*die Art des Gegebenseins*) of that entity. It is the way the reference is presented. This is illustrated in figure (5):



In (5), three lines, *a*, *b*, and *c*, intersect each other at a common point *P*. This point *P* may be characterized in a number of different ways: as *the intersection of a and b*, as *the intersection of a and c*, as *the intersection of b and c*, and finally as *the intersection of a, b, and c*. We see that a single entity, the point

P, is the reference of four different names. These four different descriptions have different senses but the same reference. Each presents the same entity in a different manner. It is of great importance that it is possible to be quite familiar with the sense of a name without knowing what its reference is. Anyone with a minimal competence in English understands the sense of *the richest citizen of the United States*, but that is not to say that he knows what individual is fortunate enough to have become the reference of the expression. Sense is 'the mode of presentation', but familiarity with the reference of any given expression is merely a possibility and may not be assumed. Sense is merely a criterion by means of which reference may be determined under various circumstances.

How does Frege's theory of meaning solve the *morning star/evening star paradox*? The senses of *the morning star* and *the evening star* consist in the two different ways in which the two expressions determine their reference. For the morning star this could be made explicit as 'the brightest heavenly body in the eastern skies at dawn', and for the evening star it could be 'the brightest heavenly body in the western skies at sunset'. Once again, it is quite possible to be familiar with the senses of the two names without knowing to what heavenly bodies they refer. So this gives us a simple explanation of the cognitive difference between (1) (The morning star is the morning star) and (2) (The morning star is the evening star). Sentence (1) is true just in case the reference of *the morning star* is the same as the reference of *the morning star*. And that this is so is quite clear independently of what the reference in fact is. (Strictly speaking, it is not even necessary to be familiar with the sense of *the morning star*, as long as one knows what it means). So to know that (1) is true is an a priori matter. Sentence (2) is true just in case the reference of *the morning star* is the same as the reference of *the evening star*. And whether this is the case cannot be determined solely on the basis of the meanings of the two names. It is necessary to know exactly what the references of the two expressions happen to be, so the truth of (2) was only apparent once astronomers discovered that they do both refer to the same celestial object, namely, the planet Venus.

It is Frege's view that names can have a sense without automatically having a reference. One example of this is the description *the first man on Mars*. At the time this book went to press, this expression did not have a reference, but it does have a sense. We all know what properties a thing would have to have in order to answer to this description. This is a special case of a more general phenomenon, namely, that the reference of a name can vary from situation to situation. One example of this is the expression *the queen of the Netherlands*, which changed reference as recently as 1980. But temporally different situations are not the only ones we have in mind. We can, for example, imagine situations in which the morning star has a different reference. There are imaginary situations in which not Venus but Mars is the brightest heavenly body at sunrise (Venus still being the brightest at sunset). In any such possible but not actual situation, the name *the morning star* would have a reference

differing from the reference of the name *the evening star* in that situation, and as a result, (2) would be false. That there are situations in which (2) can be false is the reason why (2) expresses not a necessary but a contingent proposition.

This last aspect of Frege's theory of the sense and reference of names has encountered a good deal of criticism in recent years, at least as far as real proper names like *Dukakis* and *Socrates* are concerned. Nowadays the prevalent view seems to be that proper names differ from definite descriptions in always referring to the same individual, under any circumstances whatever. Their reference is supposed to be absolute and unchangeable. The difference may be illustrated by means of 'counterfactual' statements like (6):

- (6) If Dukakis had won the presidential elections in 1988, then the president of the United States would have been a Democrat.

This statement introduces a situation other than the factual one, a situation which is at least partly determined by the condition 'Dukakis wins the 1988 elections'. The difference between the proper name *Dukakis* and the definite description *the president of the United States* is that the reference of the latter in the other situation is not the same as its reference in the actual situation; it is then Dukakis and not Bush, whereas the reference of the proper name *Dukakis* is in both situations the same: the man Dukakis. This thesis about the semantic behavior of proper names is known as *rigid designation*. We shall return to it at some length in §3.2. With this we conclude our discussion of Frege's theory of names and turn to his views on sentences.

According to Frege, a sentence has both a sense and a reference, just as names do. His analysis is restricted to sentences which express assertions, although, as we have seen, he was aware of other functions which language can fulfill. Every sentence, says Frege, corresponds to a certain thought (*Gedanke*). It expresses a thought or a proposition. Although it seems to us that the term *thought* carries a subjective 'tone', it follows from Frege's ideas on the nature of meaning that the *thought* expressed by a sentence is to be thought of as something objective. One and the same proposition is conveyed to all language users who understand a sentence.

Can we now state that the reference of a sentence is the proposition which it expresses? Frege does not think so, and his reasoning can be paraphrased as follows. Compare the following two sentences:

- (7) The supreme commander of the U.S. armed forces is a man.
(8) The president of the United States of America is a man.

Clearly these sentences express different propositions. But then, assuming the principle according to which the reference of a complex expression is a function of the references of its composite parts, the reference of a sentence cannot be the proposition which it expresses. For according to this principle, (7) and (8) have the same reference, whereas we agreed that they express different

propositions. Frege concludes that if the proposition expressed by a sentence is not its reference, then it must be its sense.

He then considers whether sentences must have references in addition to their senses. His reasoning that they do runs as follows. Consider the sentence:

- (9) Odysseus landed at Ithaca.

We know that the name *Odysseus* does not have a reference. Then in view of the principle of the compositionality of reference (see above), the whole sentence (9) can't have a reference either. Now Frege makes two points. First, the proposition expressed by (9) is independent of whether or not the name *Odysseus* has a reference. Second, if anyone wants to assert that (9) is a true sentence or a false one, then he will have to assume that there is something which is the reference of *Odysseus*. So the reference of *Odysseus* matters, even though it does not affect the sense of sentence (9). Apparently, concludes Frege, it is of importance in determining the reference of this sentence. And this must then be its truth value, since it is the truth or falsity of a sentence which is determined by the references of the expressions appearing in it.

The sense of a sentence is then the proposition which it expresses, and its reference is its truth value. As with names, a sentence may well have a sense without having a reference. Sentence (9), for example, expresses a proposition, but it does not have a truth value. And as with names, the sense of a sentence is a criterion for determining its reference. For a sentence is true just in case the proposition which it expresses holds. In other words, the sense of a sentence determines what must hold if the sentence is true. This is in agreement with the statement made by Wittgenstein in *Tractatus* 4.024: "To understand a proposition means to know what is the case if it is true" (transl. Pears and McGuinness).

Frege also had a theory about the sense and reference of predicate expressions. We shall not describe it here, since it is rather complicated and has not had the same influence on the semantics of intensional logic as the above.

We now conclude our discussion of Frege's theory of meaning with two principles which Frege uses in his reasoning and which are collectively referred to as Frege's principle. They may be formulated as follows:

- (10) The reference of a composite expression is a function of the references of its component parts.
- (11) The sense of a composite expression is a function of the senses of its component parts.

These two principles can also be presented as replacement principles:

- (12) If two expressions have the same reference, then substitution of one for the other in a third expression does not change its reference.

- (13) If two expressions have the same sense, then substitution of one for the other in a third expression does not change its sense.

We have already encountered the first of these principles a number of times in Frege's argumentation. There is some difference of opinion as to whether the second may be ascribed to Frege himself, but the general opinion is that it is at least Fregean, even if it is not Frege's. The two principles are also known as *principles of compositionality*, of reference and sense, respectively. The first, (12), is just the principle of extensionality that holds for standard logic.

As we saw in §1.6, the first principle does not hold for intensional constructions, as is apparent from the following sentence:

- (14) John said that Peter's barber is Mary's husband.

If John is not mistaken in his beliefs, and if Peter's barber and Mary's husband are in fact one and the same person, then *Peter's barber* and *Mary's husband* have the same reference. But (14) surely does not have the same reference (truth value) as (15):

- (15) John said that Mary's husband is Mary's husband.

This is not in accordance with principle (12).

This problem with intensional constructions could be tackled in various ways. One could, for example, try to restrict (12) to the extensional constructions in which substitution may take place freely. But Frege wished to maintain (10) and (12) unconditionally and therefore chose another solution. He proposed that expressions do not have their normal references in intensional constructions but refer instead to their senses. He says that in such cases expressions have an *indirect reference* (*ungerade Bedeutung*), which is then the same as what is normally their sense. A pair of sentences like (14) and (15) is then no longer a counterexample to (9) and (12). The expression *Peter's barber* can only be replaced by an expression with the same reference, and in the context of *say that*, this means having the same sense. Sentence (14) does have the same truth value as sentence (16):

- (16) John said that Peter's hairdresser is Mary's husband.

Modern intensional analysis is such that the two methods for dealing with the difficulties with the principle of the compositionality of reference can no longer be sharply distinguished. Facets of both approaches have found their way into modern analysis.

The reader should note that principles (10) and (11) implicitly presuppose a syntactic analysis. Whether an expression like *old men and women* refers to aged persons of either sex or to women and old men cannot be determined solely on the basis of the meanings of the lexical elements *old*, *men*, *and*, and *women*. And here we come up against one of the most important questions that must be answered if we wish to apply logical semantics to natural lan-

guage: what level of syntactic analysis do principles (10) and (11) presuppose? We shall return to this in chapter 6.

1.8 Context Dependence

As we have seen, one of the problems which led to the development of intensional logic was the unsuitability of the extensional semantics of standard logic for intensional constructions. Another problem was the fact that standard logic is restricted to propositions which are not context dependent, and this problem provides a nice illustration of the importance of the notion of context in intensional logic. This notion is of importance not only in the analysis of context-dependent propositions, for as we shall see, it also makes possible a formal account of the difference between sense and reference.

In logic, propositions are traditionally supposed to be independent of time and place, so that they may be said unconditionally to be either true or false. And philosophical propositions like *Knowledge implies belief*, mathematical ones like $5 + 7 = 12$, and theological propositions like *Belief implies knowledge* are clear examples of statements whose truth or falsity does not depend on the situation in which they are evaluated. But in this they are exceptional. Most propositions, like (17), for example, do not have this property:

(17) The queen is delivering an address.

Attempts have been made to adapt propositions like (17) in such a manner that their truth or falsity no longer changes from situation to situation. This has been done by building a specification of the situations in which sentences can be uttered into the sentences themselves. Sentence (17) might, for example, be expanded into something like (18):

(18) On June 9 at 8 P.M., the queen of the Netherlands is delivering an address.

It could still be argued that (18) is situational, since a time is mentioned but not a place. An elaboration like 'in Noordeinde Palace' would be needed in order to remedy this. But even that would presumably not be sufficient, for in which of the palace's many halls, chambers, or dungeons was the address actually given? And between which two exact points in time? Obviously we could go on elaborating this sentence beyond all recognition. Instead of doing so, it would seem much more natural to interpret it against the background of the context in which it is used. This context provides the here and now on which the truth of a situational sentence depends. So a sentence like *It is raining* will be true in a given situational context if it happens to be raining in that context. A sentence in a past tense, like *It rained*, refers to a moment in time before the now provided by the context in which it is uttered and is therefore a little more complicated. It requires not one but two contexts. And a sentence

like *Perhaps it is raining* introduces a conceivable state of affairs which certainly need not be present in the given context. So in interpreting a sentence in any given context, it is often necessary to take other contexts into consideration.

The name *intensional semantics*, which is given to a logical semantics in which the interpretation process is as sketched above, is derived from the distinction between *intension* and *extension*. The intension of an expression is something like its conceptual content, while its extension comprises all that exemplifies that conceptual content. Take the expression *digit*, for instance. The intension of the word (at least in the sense which it has in arithmetic) is the concept 'single symbol referring to a whole number', and its extension is the set of symbols {0, 1, 2, 3, 4, 5, 6, 7, 8, 9}. Extension is then what we have called reference until now.

The idea we have been introducing here is that expressions may have different references (extensions) in different contexts. The expression *the president of the United States*, for example, had Carter as its reference in 1979, Reagan in 1980, and Bush in 1989. Another example is *the morning star*, which under other conceivable astronomical circumstances might have referred to any of a number of heavenly bodies and not to Venus. Intensional logic has as its formal version of the traditional logical notion of *intension* the phenomenon of *multiple reference*. It is a function which gives the reference of an expression in each of the contexts under consideration. This formal notion of intension would appear to capture the essence of Frege's notion of *sense* as a criterion for determining the reference of expressions. The intension of the expression *the president of the United States* is, for example, the function which assigns to each context (moment in time) the person holding the office of president at that time. Such a function from contexts to individuals is also called an *individual concept*.

Predicates can also change their reference from context to context. For example, nowadays the reference of the predicate *American male* no longer includes the individual Elvis Presley, although it still did in the year 1976. The intension of a predicate is then a function which assigns to any given context the set of individuals forming the reference of that predicate in that particular context.

As was argued at length above, the reference of a sentence, like *It is raining*, can also change from context (time and place) to context. The intension of a sentence is then a function which assigns to any given context the truth value of that sentence in that particular context. The intension of a sentence is also called a *proposition*.

These are the two simple notions which form the foundations of modern intensional logic: *context* and *multiple reference*. We shall now demonstrate the technical apparatus of intensional logic in its simplest form: propositional logic with added intensional operators. Only after completing this in chapter 2

will we turn to the relatively complicated case of intensional predicate logic in chapter 3. There various problems will be seen to arise from interactions between quantifiers and intensional operators. After an introduction to the theory of types in chapter 4, we will move on to the intensional theory of types in chapter 5. This theory plays an important part in logical grammar, which is introduced in chapter 6.

2

Intensional Propositional Logic

2.1 Introduction

As we saw in chapter 1, there are many different kinds of intensional constructions. Of these, modal and temporal constructions have received the most attention in philosophical logic. It is for this reason that we have chosen to handle *modal propositional logic* and *propositional tense logic* here, and also the combination of the two. Other examples of intensional contexts which have been studied formally include *knowledge* and *belief* (*epistemic logic*) and *permission* and *obligation* (*deontic logic*). We begin with a section on the semantic approach which is common to all of these systems. It emerged in the fifties in the work of authors like Carnap, Kanger, Hintikka, and Kripke.

2.2 Possible Worlds Semantics

For the sake of convenience we begin with the following extremely simple language. An expression **O** is added to the vocabulary of propositional logic which when placed in front of a formula ϕ results in a new formula $O\phi$. Expressions like **O** are called *operators*. We thus obtain formulas like Op , $Op \rightarrow p$, $Op \rightarrow OOp$, $Op \rightarrow q$, $q \rightarrow Op$, $O(Oq \rightarrow Op)$, etc. The intuitive idea is that **O** stands for an intensional construction like *it ought to be the case that*, *I know that*, *it will always be the case that*, *it was once the case that*, *it is necessary that*, or *it is possible that*. Under the second interpretation, for example, the first three formulas say: *I know that p*, *if I know that p then p is the case*, and *if I know that p, then I know that I know this*. It should be clear that the extensionality of standard propositional logic is lost if such constructions are introduced. It clearly can be true that two propositions p and q have the same truth value, so that $p \leftrightarrow q$ is true, without it following from *I know that p* that *I know that q*, that is, without $Op \leftrightarrow Oq$ being true.

What aspects of the context must be taken into account will depend on the intended interpretation of the operator **O**. If we are only interested in temporal constructions like *it will always be the case that* and *it was once the case that*, then contexts reduce to moments in time. If we are only interested in modal

constructions like *it is necessary that* and *it is possible that*, then we can identify the contexts which must be taken into consideration with all possible situations. And if we are dealing with both temporal and modal constructions at the same time, then the contexts will be situations possible at moments in time. The point is that the set K of contexts which we will choose to work with depends very much on what the operator \mathbf{O} is supposed to mean.

In view of the above, it would seem that we need a context-dependent notion of meaning, that is to say, one in which the truth values of propositions are not absolute but are relative to the contexts in which their truth is evaluated. In formal terms, this will mean replacing the simple semantics of propositional logic, in which formulas receive absolute truth values, with a system in which evaluation functions assign truth values only relative to some context k (taken from the set K of such contexts). The clauses for the connectives of standard propositional logic then remain essentially the same. A formula $\neg\phi$, for example, will receive the truth value 1 in a given context just in case the formula ϕ receives the truth value 0 in that context. In fact, the set K of all contexts only comes into play when we start evaluating sentences of the form $\mathbf{O}\phi$ in a given context k . For the truth of any such formula in a context k is made to depend on the truth of ϕ , not only in that same context k but also in other contexts k' in K . This is what makes the system intensional. The truth of the construction *it was once the case that* p is, for example, dependent on there being some context (point in time) k' earlier than the present context (point in time) k at which p was true. And for *I know that* p to be true in a context k , it is necessary not only that p be true in k but also that it be true in all contexts k' which are compatible with the knowledge I have in k : with all of the so-called epistemic alternatives I have in context k .

Are the truth values of ϕ in all contexts k' within K relevant to the truth of $\mathbf{O}\phi$ in any one of these contexts? This all depends on which intensional concept \mathbf{O} is supposed to be modeling. If \mathbf{O} is to be interpreted as *it is logically necessary that*, then it is plausible to stipulate that $\mathbf{O}\phi$ is true in any context k just in case ϕ is true in every possible context k' . But if \mathbf{O} is, for example, to be interpreted as *it is a physical necessity that*, then it would seem more reasonable to have $\mathbf{O}\phi$ true in any given context k just in case ϕ is true in those contexts k' where the same physical laws hold as in k itself. And if \mathbf{O} stands for the temporal construction *it was once the case that*, then only the contexts (points in time) prior to k will matter. To put it generally then, what contexts need to be taken into account when evaluating a formula $\mathbf{O}\phi$ in some context k will depend on the interpretation given to \mathbf{O} . Depending on this interpretation, they may also depend on certain characteristics of the context k itself. The set of points in time prior to k will obviously be different for different points of time k . And much the same thing can be seen with epistemic constructions. What epistemic alternatives one has to consider in evaluating a sentence depends on one's context, as is apparent from the following concrete

example. Consider a chess player halfway through a game. He knows where all of the pieces stand on the board, and he is familiar with the rules of the game, so in principle at least, he is in a position to calculate all of his epistemic alternatives: those positions which can be reached by continuing the game. But the epistemic alternatives will vary with the stage the game has reached. In fact, the set of epistemic alternatives that this player has shrinks as the game progresses, for each move excludes whole branching trees of previously possible developments. Thus the statement *I know that black will not win* may be false in a given context, that is to say, at a given stage in the game, there being possible continuations of the game in which black checkmates white, while this statement at a later stage in the game becomes true, black having lost all of his pieces but the king.

So what contexts must be taken into account in evaluating a formula $\mathbf{O}\phi$ may depend not only on the intended interpretation of \mathbf{O} but also on the particular context in which the evaluation is to take place. Those contexts k' which are relevant when evaluating within a context k are said to be *accessible from* k . So the truth value of $\mathbf{O}\phi$ in k depends on the truth values taken on by ϕ in the contexts k' which are accessible from k . And the way it depends on these truth values depends in its turn on the intended interpretation of \mathbf{O} . If \mathbf{O} is supposed to mean *it is necessary that*, for example, then ϕ must be true in all contexts accessible from k if ϕ is to be true in k . But if \mathbf{O} is supposed to mean *it is possible that*, then it is sufficient that ϕ be true in any one of these contexts. Each interpretation of \mathbf{O} gives rise to some condition on the truth values of ϕ in accessible contexts which must be satisfied if $\mathbf{O}\phi$ is to be true.

The informal discussion above leads to the following formal definition:

Definition 1

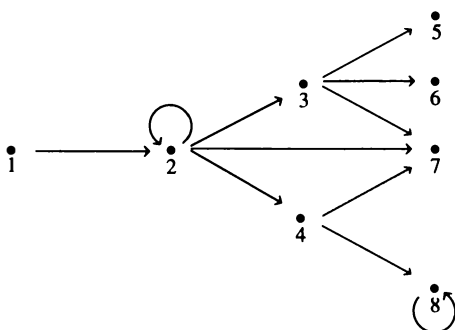
A *model* \mathbf{M} consists of:

- (i) a nonempty set K of contexts
- (ii) a binary relation R on K , the accessibility relation
- (iii) a valuation function V which assigns a truth value $V_k(p)$ to every proposition letter p in each context $k \in K$

(Models like these are often called 'Kripke models'.) Starting with this definition, a truth definition can be given which gives the truth value $V_{\mathbf{M},k}(\phi)$ of a formula ϕ in context k of a model \mathbf{M} . In this definition, the clauses for the standard connectives retain their usual form, while the clause for the intensional operator \mathbf{O} depends on its intended interpretation. We shall see precise formulations of the intensional clause for two different interpretations of \mathbf{O} in §§2.3 and 2.4.

In some cases it is convenient to give a diagram of the contexts and their accessibility relation. The contexts are then represented by means of points, with arrows indicating which contexts are accessible from which others. An example of one such diagram is given as figure (1):

(1)



Only context 2 is accessible from 1; from 2, 2 itself, 3, 4, and 7 are accessible; no context at all is accessible from 7, and so on.

It will be clear why extensionality fails in any such system, even without the concrete examples in what is to come. For that $p \leftrightarrow q$ is true in k is in itself no guarantee that $Op \leftrightarrow Oq$ will also be true in k , for the truth of this latter formula depends on the pattern of truth and falsity exhibited by p and q in contexts other than k itself. And the truth of $p \leftrightarrow q$ in the one context k does not tell us anything about this pattern.

2.3 Modal Propositional Logic

2.3.1 Historical Background

The modal concepts which are considered in modal propositional logic derive not so much from natural language as from philosophy. The modal constructions of natural language include all forms which contain elements such as *can*, *perhaps*, *must*, *certain*. But philosophy has its own traditional modalities: *it is necessary that*, *it is possible that*, and *it is contingent that*. We shall briefly return in §2.3.4 to the matter of whether modal constructions in natural language always express one of these philosophical modalities.

The philosophical modalities form one of the subject areas addressed in traditional logic. Aristotle considered modal syllogisms, and the Scholastics were concerned with the semantics of modal notions. And in a well-known table in his *Critique of Pure Reason*, Kant refers to the modalities as the fourth main category of propositions. But at least initially, there was no place for the modalities in modern logic. Frege discussed Kant's table in his *Begriffsschrift* and removed modalities from the logical agenda in a single sentence: "In saying that a proposition is necessary, I merely give an impression of the reasons for my judgment." The *content* of that judgment, Frege argues, is independent of those reasons, and it is only the content which matters to logic. But the modalities were able to penetrate into modern logic in disguise.

Around the turn of the century, certain overzealous supporters of modern logic pronounced material implication the *only* kind of implication. The rest of us would just have to learn to swallow the counterintuitive consequences of

this, including the fact that one of any two given propositions must always imply the other: for $(\phi \rightarrow \psi) \vee (\psi \rightarrow \phi)$ is a tautology. This and other *paradoxes of material implication*, like $\phi \rightarrow (\psi \rightarrow \phi)$ and $\neg\phi \rightarrow (\phi \rightarrow \psi)$, continued, however, to trouble some writers, notably C. I. Lewis. (The fact that the metanotation of derivability (\vdash) made every bit as good a claim as material implication to be an explication of implication seems to have been overlooked by believers and unbelievers alike. At least it is not true that for every two formulas ϕ and ψ , either $\phi \vdash \psi$ or $\psi \vdash \phi$.) Lewis introduced a *strict implication* \rightarrow as a complement to the material implication \rightarrow , which was supposed to formalize more (different) aspects of implication, and he attempted to capture the properties of strict implication in axiomatic systems. But it turns out to be rather difficult to judge the validity of his axioms. It is much simpler (if perhaps not in all respects entirely correct) to understand why strict implication is stricter than material implication as follows. A material implication $\phi \rightarrow \psi$ is equivalent to $\neg(\phi \wedge \neg\psi)$. It thus states that it is not the case that both ϕ and $\neg\psi$. But strict implication says much more than this—not only is this not the case, but it *couldn't* be the case: $\phi \rightarrow \psi$ is equivalent to $\neg\Diamond(\phi \wedge \neg\psi)$ (it is not *possible* that both ϕ and $\neg\psi$). And thus the modality \Diamond (*it is possible that*) once again turned up in logic. Its conceptual counterpart \Box (*it is necessarily so that*) then can not be far away, and via the equivalence of in turn $\neg\Diamond(\phi \wedge \neg\psi)$, $\Box\neg(\phi \wedge \neg\psi)$, and $\Box(\phi \rightarrow \psi)$, \rightarrow may be seen as *necessary material implication*. As weak as the grounds for introducing the new operator \rightarrow may seem to have been, its translation into modal terms returned modal notions to logic, and in the meantime, they have even managed to draw most of the attention (although the notion of strict implication has been revived a number of times in recent years, notably in 'relevance logic').

Right from the beginning, modal logic showed an uncertainty about the validity of its logical principles which is completely foreign to classical Fregean logic. Principles like (2) may have been clear enough, and with it (3):

- (2) $\neg\Diamond\phi \leftrightarrow \Box\neg\phi$ (The impossible is what is necessarily not the case.)
- (3) $\neg\Diamond\neg\phi \leftrightarrow \Box\phi$ (What cannot possibly not be the case is what is necessary.)

But (2) and (3) seem to be more definitions than principles. Formulas (4) and (5) also seem relatively unproblematic:

- (4) $\Box\phi \rightarrow \phi$ (What is necessarily true is true.)
- (5) $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$ (Strict consequences of necessary truths are themselves necessary truths.)

Principle (5), which is equivalent to $(\Box(\phi \rightarrow \psi) \wedge \Box\phi) \rightarrow \Box\psi$, may be considered a modal form of modus ponens. But the validity of principles becomes much harder to judge as soon as *stacked modal operators* start complicating things. Two such principles of disputable validity are (6) and (7):

(6) $\Box\phi \rightarrow \Box\Box\phi$ (If something is necessary, then it is necessarily so.)

(7) $\Diamond\Box\phi \rightarrow \phi$ (What is possibly necessary is true.)

Many different syntactic, axiomatic theories grew from different preferences for these and similar principles, and by the sixties they had turned into a tangled and increasingly impenetrable jungle. This uncertainty about the validity of certain principles may be considered as a sign that various (stronger and weaker) modal notions were interfering with each other in our intuitive judgments, and a semantics which would throw new light on all of the various syntactic theories was badly needed. That is why the idea of a possible worlds semantics was to make such an impact around 1960.

2.3.2 Syntax and Semantics

We discussed the basic idea behind possible worlds semantics in §2.2. As applied to modal propositional logic it amounts to the following. The operators \Box and \Diamond are added to propositional logic by means of the following addition to the definition of a propositional language L:

(8) If ϕ is a formula in L, then $\Box\phi$ and $\Diamond\phi$ are too.

According to (8), we now have $\Box p$, $\Box p \vee \Diamond q$, $\neg\Diamond(p \wedge q)$, $p \rightarrow \Box\Diamond p$, and $\Diamond p \rightarrow \Box\Diamond p$ as examples of formulas. Stacks of operators like those in examples (6) and (7) are also referred to as *iterations*.

Exercise 1*

Translate the following sentences into formulas of modal propositional logic. Represent the logical structure as well as you can and state the translation key you use.

- It is possible that you do not understand me, but it isn't necessary.
- If it may be raining, then it must be possible that it is raining.
- It is possible that if it may be raining, it is raining.
- If it may be necessary that it is raining, then it must be raining.
- Maybe it is raining, and perhaps this is necessary (try to find two translations).

The semantics of modal propositional logic is, as we have said, a concrete example of an intensional propositional logic as discussed in §2.2. The contexts we referred to there are now called possible worlds, a notion which goes

back to Leibniz. (Leibniz's idea that this world in which we happen to live is the best of all possible worlds was the target of Voltaire's sarcasm in *Candide*.) Leibniz distinguished between factual truths, which hold (only) for the world in which we happen to live, and rational truths, which hold in all worlds which God might have created. The latter clearly lie close to the idea of necessary truth as truth in all possible worlds.

The idea behind possible worlds semantics is that the truth of $\Box\phi$ and $\Diamond\phi$ in any given possible world depends on the truth of ϕ in other possible worlds. It may not be necessary to take all possible worlds into account; formally this is captured in an accessibility relation which says what worlds matter. We can now give:

Definition 2

A model \mathbf{M} for modal propositional logic consists of:

- (i) a nonempty set W of possible worlds
- (ii) a binary relation R on W , the accessibility relation
- (iii) a valuation function V which assigns a truth value $V_w(p)$ to every proposition letter p in each world $w \in W$.

Sometimes a special element w_0 of W is singled out as the actual world, but this is not really necessary. A set of possible worlds W together with a suitable accessibility relation R is referred to as a *frame*, or *structure*. So a model \mathbf{M} consists of a frame F together with a valuation function V . Any given frame F can be turned into a variety of different models, depending on the valuation function which is added. For a frame only fixes what possible worlds we are dealing with and which of these are accessible from which others. A valuation is needed to decide what facts obtain in each of the possible worlds, and in general there will be many different ways of doing this. Each corresponds to a different model \mathbf{M} . A model is an exact specification of a particular state of actual and possible reality. A frame provides, as it were, a structure, a framework that can form the basis of any one of a variety of such states.

The truth definition now tells us what formulas ϕ are true in what possible worlds w of any given model \mathbf{M} . The truth values of all of the proposition letters are fixed, for each possible world in \mathbf{M} , by \mathbf{M} 's valuation function V . What the truth definition does is determine what truth values must then be attributed to composite formulas in each of the possible worlds. In other words, the truth definition states, for given \mathbf{M} , how the valuation function available for the proposition letters can be extended to a valuation function $V_{\mathbf{M}}$ which applies to all formulas in the language in question. As usual, this is done by stating how the truth values of component formulas must depend on those of their component parts. The innovation of intensional propositional logic is that truth values are relative to the possible worlds in which evaluation takes place and may depend in part on truth values in other such possible worlds. The truth definition for modal propositional logic is now:

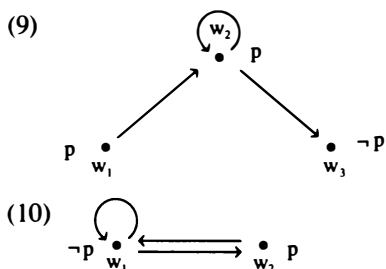
Definition 3

If \mathbf{M} is a model with W as its set of possible worlds, R as its accessibility relation, and V as its valuation, then $V_{\mathbf{M},w}(\phi)$, the truth value of ϕ in w given \mathbf{M} , is defined by the following clauses:

- (i) $V_{\mathbf{M},w}(p) = V_w(p)$, for all proposition letters p
- (ii) $V_{\mathbf{M},w}(\neg\phi) = 1$ iff $V_{\mathbf{M},w}(\phi) = 0$
- (iii) $V_{\mathbf{M},w}(\phi \rightarrow \psi) = 1$ iff $V_{\mathbf{M},w}(\phi) = 0$ or $V_{\mathbf{M},w}(\psi) = 1$
- (iv) $V_{\mathbf{M},w}(\Box\phi) = 1$ iff for all $w' \in W$ such that wRw' : $V_{\mathbf{M},w'}(\phi) = 1$
- (v) $V_{\mathbf{M},w}(\Diamond\phi) = 1$ iff for at least one $w' \in W$ such that wRw' : $V_{\mathbf{M},w'}(\phi) = 1$

Clearly the connectives \neg and \rightarrow have the same meaning here as in the truth tables of standard propositional logic. The same applies to \wedge , \vee , and \leftrightarrow , whose clauses have been skipped. It is only with the clauses for \Box and \Diamond that the whole apparatus of possible worlds gets into gear. According to (iv), *necessary* means *true in all accessible worlds*, while according to (v), *possible* means *true in at least one accessible world*. This clearly exposes the analogy between \Box and \forall as opposed to \Diamond and \exists which had already been noticed by so many authors. And it also makes clear why, as with the quantifiers, just one of the pair needs to be taken as a primitive, in terms of which the other can then be defined. The operator \Diamond may be defined as $\neg\Box\neg$, for example, just as \exists may be defined as $\neg\forall\neg$. The reader can easily check that this immediately seals the validity of principles (2) and (3).

In order to demonstrate how clauses (iv) and (v) work, we shall now turn to the two simple models given in (9) and (10):



The model \mathbf{M} depicted in (9) may be read off as follows. There are just three possible worlds, w_1 , w_2 , and w_3 , so that $W = \{w_1, w_2, w_3\}$. The arrows represent the accessibility relation between the worlds: w_2 is accessible from w_1 , w_2 itself and w_3 are accessible from w_2 , and no possible world at all is accessible from w_3 . Writing R as a set of ordered pairs, we obtain from (9): $R = \{\langle w_1, w_2 \rangle, \langle w_2, w_2 \rangle, \langle w_2, w_3 \rangle\}$. So now we have determined the frame for \mathbf{M} . Assuming we are dealing with a language with just a single proposition letter p , (9) also fixes the valuation function V (and thus, following the truth definition, the truth values of all of the formulas in all of these possible worlds): $V_{w_1}(p) = V_{w_2}(p) = 1$, and $V_{w_3}(p) = 0$. We have now fully specified \mathbf{M} . And

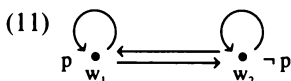
what about the truth values of $\Box p$ and $\Diamond p$ in the different possible worlds? Since $w_1 R w_2$ and $V_{M, w_2}(p) (= V_{w_2}(p)) = 1$, $V_{M, w_1}(\Diamond p) = 1$. And since w_2 is the only world accessible from w_1 , we also have $V_{M, w_1}(\Box p) = 1$. In w_2 , $\Diamond p$ is true, since p is true in w_2 itself, and $w_2 R w_2$. But $\Box p$ is false in w_2 , since $w_2 R w_3$, while p is false in w_3 . And finally, $V_{M, w_3}(\Diamond p) = 0$ and $V_{M, w_3}(\Box p) = 1$, since there are no worlds at all which are accessible from w_3 (so that p is true in all of the [nonexistent] worlds which are accessible). Formula $\Diamond \neg p$ is false in w_1 but true in w_2 , while both $\Diamond p$ and $\Diamond \neg p$ are false in w_3 , and $\Box p$ and $\Box \neg p$ are both true in w_3 .

Our second example of a model is represented as (10). Reading off its characteristic W , R , and V from that diagram, we have: $W = \{w_1, w_2\}$; $R = \{\langle w_1, w_1 \rangle, \langle w_1, w_2 \rangle, \langle w_2, w_1 \rangle\}$ and $V_{w_1}(p) = 0$, $V_{w_2}(p) = 1$. The truth values of various formulas may now be calculated just as in the first example. By way of illustration, we shall determine the truth values of two formulas with stacked modal operators, i.e., $\Box \Diamond p$ and $\Diamond \Box \neg p$.

Is $\Box \Diamond p$ true in w_1 ? That would be so just in case for all w' such that $w_1 R w'$ we have $V_{M, w'}(\Diamond p) = 1$. We know that $w_1 R w_1$ and $w_1 R w_2$, so $\Diamond p$ would have to be true in both w_1 and w_2 . Now $\Diamond p$ is true in w_1 , since $w_1 R w_2$ and p is true in w_2 . But $\Diamond p$ is not true in w_2 , since only w_1 is accessible from w_2 , and there p is false. So the answer is no: $\Box \Diamond p$ is false in w_1 . Is $\Box \Diamond p$ perhaps true in w_2 ? For this $\Diamond p$ would have to be true in all worlds accessible from w_2 . There is just one of these, w_1 , and $\Diamond p$ is indeed true there, since $w_1 R w_2$ and p is true in w_2 . So the answer is yes: $\Box \Diamond p$ is true in w_2 .

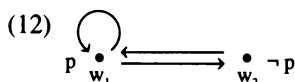
The second formula which we were going to consider is $\Diamond \Box \neg p$. Is this formula true in w_1 ? It would be just in case there is some w' with $w_1 R w'$ in which $\Box \neg p$ is true. Since both $w_1 R w_1$ and $w_1 R w_2$, both w_1 and w_2 are candidates. But $\Box \neg p$ happens to be false in w_1 , since $w_1 R w_2$ and $\neg p$ is false in w_2 . So w_1 is not the w' we were looking for. But $\Box \neg p$ is true in w_2 , since only $w_2 R w_1$ and $\neg p$ is true in w_1 . Thus w_2 is the w' we were looking for, and $\Diamond \Box \neg p$ is true in w_1 . That this formula is false in w_2 can be verified very quickly. For since only $w_2 R w_1$, $\Box \neg p$ would have to be true in w_1 . And that is not the case, since $w_1 R w_2$ and p is true in w_2 . Thus $\Diamond \Box \neg p$ is false in w_2 . (We might have anticipated this result, since in view of the equivalences given above, $\Diamond \Box \neg p$ is just the negation of $\Box \Diamond p$.)

For any given model M , there are always some formulas which are true in each of M 's possible worlds. By way of example, $\Diamond p \wedge \Diamond \neg p$ and $\Box p \rightarrow p$ are both true in every possible world of the model given in (11):

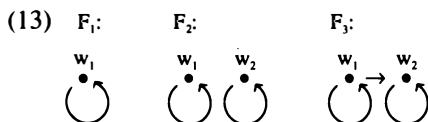


We say that the formulas that are true in each of a model's worlds are valid in that model, writing this as $V_M(\phi) = 1$. Among the formulas valid in M we may distinguish those whose validity is dependent on the particular valuation

V which M has from those whose validity is independent of this V . The former are valid in M by virtue of the facts which happen to hold there, but the latter would seem to be indifferent to these. Apparently they are valid purely in virtue of the basic structure of the model, i.e., its frame. The formula $\Diamond p \wedge \Diamond \neg p$ is an example of the former kind. If we were to change the model in (11) by having $V_{w_2}(p) = 1$ instead of 0, then this formula would no longer be valid. But $\Box p \rightarrow p$ would still be valid. In fact, no matter which V we choose in (11), $\Box p \rightarrow p$ is always true in all possible worlds. Clearly this is an example of the latter kind of formula. Its validity can be lost only if the underlying structure of the model is changed, as can be seen from the model given in (12), where $\Box p \rightarrow p$ is no longer valid.



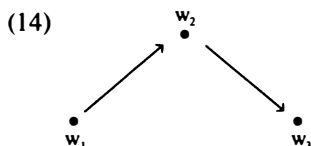
For while $\Box p$ is true in w_2 , p is false in that world, so that $\Box p \rightarrow p$ is false there too. We say that the model in (12) is a *counterexample* to (the validity of) $\Box p \rightarrow p$. What all of this means is that there is some relation between frames and the formulas valid on the models constructed on the basis of the frames. If a formula ϕ is valid in every model constructed on the basis of a frame F , then we say that ϕ is *valid on* F . There is a sense in which any such formula expresses a property of F ; often it turns out to be a property of a whole class of frames. Compare, for example, the frame of the model in (11) with the three frames in (13).



The formula $\Box p \rightarrow p$ is valid on all of these frames (and many more besides them). For they have a property in common which is responsible for its validity, namely, the *reflexivity* of their accessibility relations. Indeed, this is the property that is expressed by $\Box p \rightarrow p$, for it can be shown that $\Box p \rightarrow p$, *characterizes* the class of reflexive frames: $\Box p \rightarrow p$ is valid on any frame with a reflexive accessibility relation, and conversely, if $\Box p \rightarrow p$ is valid on a frame, then the frame must have a reflexive accessibility relation. This can easily be seen as follows. First we must show that $\Box p \rightarrow p$ is valid on any frame with a reflexive accessibility relation. So suppose M has a frame F with a reflexive accessibility relation R , and that in some w we have $V_{M,w}(\Box p) = 1$. Then in all w' such that wRw' , $V_{M,w'}(p) = 1$. Now since R is reflexive, we have wRw , so in particular we have $V_{M,w}(p) = 1$. This means that $V_{M,w}(\Box p \rightarrow p) = 1$, and since w was arbitrary, $\Box p \rightarrow p$ is true in every w in M , so $\Box p \rightarrow p$ is valid in M . And since M was an arbitrary model with a reflexive frame, $\Box p \rightarrow p$ is valid in any model M with an F with reflexive R . Now it only

remains to show that if $\Box\phi \rightarrow \phi$ is valid on a frame, then this frame must have a reflexive accessibility relation or, equivalently, that a counterexample to $\Box\phi \rightarrow \phi$ can be constructed on any frame of which this relation is not reflexive. So let F be a frame whose accessibility relation is not reflexive. This means that there is some w in F such that we do not have wRw . We now obtain our counterexample by constructing a model M with F as its frame, and with a valuation V such that $V_w(p) = 0$, while $V_{w'}(p) = 1$ for all other worlds w' in this frame. Then we have $V_w(\Box p) = 1$ and $V_w(p) = 0$, so that $V_w(\Box p \rightarrow p) = 0$. Thus $\Box\phi \rightarrow \phi$ is not valid in M . Figure (12) is an example.

This brings us to one of the main occupations of modal logicians, which is laying bare the relations between the validity of formulas and the properties of frames. Particular attention has been paid to modal principles like (4)–(7). It turns out that principle (5), $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$, is valid on every frame, no matter what accessibility relation it has. But many other modal principles turn out to correspond to particular characteristics of the accessibility relation. We have just seen that principle (4), $\Box\phi \rightarrow \phi$, corresponds to the reflexivity of R . Principle (6), $\Box\phi \rightarrow \Box\Box\phi$, corresponds to the *transitivity* of R . We will not give a full proof here but will make do with a demonstration of the fact that a counterexample to $\Box\phi \rightarrow \Box\Box\phi$ can always be constructed on a non-transitive frame. On any such frame there will always be three (not necessarily distinct) worlds w_1 , w_2 , and w_3 such that w_1Rw_2 , w_2Rw_3 , but not w_1Rw_3 . This situation may be represented as in (14).

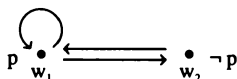


If we now choose V such that $V_{w_3}(p) = 0$ and $V_w(p) = 1$ for all other w , then we have $V_{w_1}(\Box p) = 1$ and $V_{w_1}(\Box\Box p) = 0$, since $V_{w_2}(\Box p) = 0$.

Principle (7), $\Diamond\Box\phi \rightarrow \phi$, expresses the *symmetry* of R . The proof of this fact is left to the reader (see exercise 3a). This is not the place to delve any deeper into these correspondences and the relations they bear to the numerous divergent axiomatizations which have been developed for modal propositional logic. We have illustrated these matters here in order to emphasize the flexibility of possible worlds semantics, which is certainly of some importance for applications in natural language research. It is this flexibility which enables us to represent different interpretations of the modal notions, by imposing different requirements on the accessibility relation R . Not that this flexibility is unlimited. It turns out that there are quite simple properties of frames which cannot be characterized by means of a formula. There is, for example, no formula which characterizes the irreflexivity of frames, which is a clear restriction on the expressive power of modal propositional logic.

Exercise 2*

(a) Consider this model:



Decide for each of the following four formulas whether it holds in w_1 ; in w_2 ; in the whole model.

(i) $\Box p \rightarrow \Box \Box p$

(ii) $\neg \Box p$

(iii) $p \rightarrow \Box \Diamond p$

(b) Consider the following model: $W = \{w_1, w_2, w_3, w_4\}$; $R = \{\langle w_1, w_2 \rangle, \langle w_2, w_3 \rangle, \langle w_3, w_1 \rangle, \langle w_3, w_4 \rangle, \langle w_4, w_2 \rangle\}$; $V_{w_1}(p) = V_{w_3}(p) = V_{w_1}(q) = V_{w_2}(q) = 1$, $V_{w_2}(p) = V_{w_4}(p) = V_{w_3}(q) = V_{w_4}(q) = 0$.

(i) Draw a picture of the model.

(ii) Determine:

1. $V_{w_1}(\Box q)$

2. $V_{w_2}(\Box \neg(p \rightarrow q))$

3. $V_{w_3}(\Box((p \wedge q) \vee (\neg p \wedge \neg q)))$

4. $V_{w_1}(\Diamond \Box p)$

5. $V_{w_1}(\Diamond p \wedge \Diamond q)$

(iii) Decide whether the following formulas are valid in the model:

1. $\Diamond \Box p \vee \Diamond \Diamond \Box p$

2. $\Box p \rightarrow \neg p$

3. $(p \rightarrow \Diamond p) \wedge (q \rightarrow \Diamond q)$

4. $\Diamond(p \vee \neg p) \rightarrow \Box(p \vee \neg p)$

(iv) Decide whether the following formulas are valid on the frame of the model:

1. $\Box p \rightarrow \Diamond p$

2. $\Diamond \Diamond \Box p \rightarrow p$

Exercise 3*

(a) Show that on every frame with symmetric R , $\Diamond \Box \phi \rightarrow \phi$ is valid and construct a counterexample for this formula on a nonsymmetric frame.

(b) Which class of frames is characterized by $\Diamond \Diamond \phi \rightarrow \phi$? And which class by $\Diamond \Diamond \Diamond \phi \rightarrow \phi$? And in general, for $\Diamond_1 \dots \Diamond_n \phi \rightarrow \phi$?

Exercise 4

(a) Interpret \Box as *I believe that*. What does \Diamond mean, given that we maintain principles (2) and (3) of §2.3.1? And which of the principles (4)–(7) stated there are plausible on this interpretation of \Box ?

(b) Now answer the same question with *it is obligatory that* as the interpretation of \Box .

- (c) What constraint on the accessibility relation will make *If I don't believe that ϕ , then I believe that I don't believe that ϕ* valid?
- (d) What property of frames is characterized by $\Box\phi \rightarrow \Diamond\phi$? Is this a reasonable property if we read \Box as *It is obligatory that*?

Exercise 5

We call a relation R *connected* if for all w, w' : if $w \neq w'$, then wRw' or $w'Rw$, and *universal* if for all w, w' : wRw' . Show that for every frame F whose relation R is reflexive and symmetric it holds that R is universal iff R is connected.

2.3.3 The Syntactic Approach to the Notion of Validity

Our approach to intensional logic here is wholly semantic. But it may be instructive to mention the syntactic side briefly. For this reason, we now give a short discussion of the way that the system of *natural deduction*, which was introduced in volume 1 as a syntactic explication of validity in propositional and predicate logic, may be extended to modal propositional logic. In the rest of our discussion of intensional logic and the theory of types we will not go into syntactic approaches to validity at all, so readers not familiar with natural deduction may skip this section without getting stuck later on.

The following introduction rule for \Box is quite acceptable: if ϕ can be derived without making any assumptions at all, then apparently ϕ is necessary, so we may draw the conclusion $\Box\phi$. This rule, cast in the form given below as $I\Box$, may now be added to our system of natural deduction for propositional logic:

1. .
- . .
- .
- . .
- m. ϕ
- . .
- .
- . .
- n. $\Box\phi$ $I\Box, m$

The restriction on this rule is that at step m there may be no standing assumptions. That this restriction is needed is immediately obvious from the fact that we could otherwise always derive $p \rightarrow \Box p$.

On the other hand, it is not possible to give a simple and intuitively appealing elimination rule for \Box . Instead, we can give axioms, which may be seen as background assumptions or as meaning postulates. Axioms may be introduced into natural deduction as formulas which may always be written down at any stage of a derivation without being defended in any way. If we take all

formulas with the general form of (5) (repeated here) as our axioms, we obtain the *minimal modal propositional logic* (sometimes referred to as K):

$$(5) \quad \Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$$

Every substitution of formulas for the ϕ 's and ψ 's in (5) results in an axiom. We say that (5) is an axiom schema. As an example, we give the following derivation of $\Box(p \wedge q) \rightarrow \Box p$ in the minimal modal propositional logic K:

1. $p \wedge q$	assumption
2. p	$E \wedge, 1$
<hr/>	
3. $(p \wedge q) \rightarrow p$	$I \rightarrow$
4. $\Box((p \wedge q) \rightarrow p)$	$I\Box, 3$
5. $\Box((p \wedge q) \rightarrow p) \rightarrow (\Box(p \wedge q) \rightarrow \Box p)$	axiom
6. $\Box(p \wedge q) \rightarrow \Box p$	$E \rightarrow, 5, 4$

Soundness and completeness proofs can also be given for this minimal logic K. There is a soundness proof, which states that if $\vdash \phi$ in K, then ϕ is valid in all models \mathbf{M} , and a completeness proof, which states that formulas which are valid in all models are derivable in K. (The proof of the soundness theorem actually amounts to little more than a comment already made, to the effect that (5) is valid independently of a model's accessibility relation.

As we have already mentioned, there are interpretations for \Box for which not all models are suited, but only those with accessibility relations satisfying certain requirements. It might, for example, be required that this relation be reflexive, transitive, and symmetrical. These requirements have already been linked to formulas (4), (6), and (7), which are repeated here:

$$(4) \quad \Box\phi \rightarrow \phi$$

$$(6) \quad \Box\phi \rightarrow \Box\Box\phi$$

$$(7) \quad \Diamond\Box\phi \rightarrow \phi$$

The system obtained by adding (4), (6), and (7) as axiom schemata to minimal modal propositional logic is known as S5. For this system too, soundness and correctness theorems can be proved. If there is a derivation of a formula ϕ in S5, then ϕ is valid in every model in which R is reflexive, transitive, and symmetrical, i.e., an equivalence relation (soundness), and if ϕ is valid in all such models, then it also has a derivation in S5 (completeness). Proving the completeness of all kinds of different axiom systems is something which modal logicians like even more than showing that given formulas characterize particular classes of frames, a matter briefly considered above. It is also a good deal more difficult. That things are not as simple as they perhaps seem is apparent from the following. It could be argued that if \Box is to be interpreted as *logically necessary*, then only those models must be taken into account in which every world is accessible to every other world (including itself), i.e., in which the accessibility relation R is universal. *Logically necessary* would then

mean *true in every possible world*. But it turns out that the system corresponding to this restriction on the models coincides with S5.

This should have illustrated the principles of the syntactic approach to intensional logic sufficiently. As we have mentioned, we will not return to it in what is to follow.

Exercise 6

Above it was claimed that ϕ is derivable in S5 iff ϕ is valid in all models in which R is an equivalence relation. Also, it was said that the system S5 is complete with respect to the class of all models in which R is universal. What can be deduced from these two facts concerning the modal characterizability of the property of connectedness?

2.3.4 Alethic and Epistemic Modalities

In this section we will briefly consider whether the modal expressions in natural language always correspond to the philosophical modalities of modal logic. These philosophical modalities are also referred to as *alethic* modalities (after the Greek word *alètheia* ‘truth’). They are concerned with the truth of sentences. And of course there are constructions in natural language in which modal expressions are alethically significant. Examples of these may be found in exercise 1 in §2.3.2. But there are also modal expressions like (15) and (16) which do not seem to be alethic:

(15) Perhaps it is raining in Southern California.

(16) John must be in his room.

These statements do not seem to concern the truth of *It is raining in Southern California* and *John is in his room* so much as the information that is available to whoever utters them. If the *perhaps* in (15) were an alethic modality, then (17) would make sense:

(17) Perhaps it is raining in Southern California, but it isn’t raining in Southern California.

Then we could represent (15) as $\Diamond p$, and (17) itself would correspond to the formula $\Diamond p \wedge \neg p$. But while this last formula makes sense, the same cannot be said of (17). Apparently (15) expresses not an alethic modality but what may be called an *epistemic* modality: what (15) expresses is the fact that the *information* that is available to the speaker does not enable him to decide whether it is raining in Southern California or not. The second part of (17) denies just this, which makes it impossible to interpret the sentence. The modality in sentence (16) would also seem to be epistemic: (16) does not mean that it is a necessary fact that John is to be found in his room but only that the information available to whoever is uttering the sentence would

suggest that he is there. A typical continuation of (16) would be something like (18):

- (18) John must be in his room; he is always there around this time of day.

If (16) were interpreted as an alethic modality and under the very reasonable assumption that the accessibility relation is reflexive, so that $\Box\phi \rightarrow \phi$ is valid, then it would follow from (16) that (19):

- (19) John is in his room.

But this is not in agreement with the meaning of (16). The epistemic modality expressed in (16) is essentially weaker than the alethic modality represented by \Box .

2.3.5 An Application

The true power of modal operators in arguments becomes apparent only when the operators are combined with predicate logic. But there are still some problems that can be clarified here at the elementary level of modal propositional logic. One example of this is Thomas Aquinas's discussion of the argument that God's Providence implies fatalism. According to this argument, that I am standing here right now is necessary (in other words, my free will doesn't have any say in the matter). It runs as follows. During the Creation, God saw everything, including me standing here. And if God saw me standing here during the Creation, then it is necessarily true that I am standing here. This argument, formalized into modal propositional logic, has the form of the following valid argument schema:

- (20) $p, p \rightarrow q, q \rightarrow \Box p / \Box p$

Key: p : I am standing here; q : God saw me standing here during the Creation.

Thomas notes that the last premise is the crucial one: to what does the qualification *necessarily* actually apply? In the formalization in (20), it only applies to the consequent of the implication in the last premise. But this premise would seem to be plausible only if the qualification were to apply to the implication as a whole. This would mean that it should read $\Box(q \rightarrow p)$ instead of $q \rightarrow \Box p$, in which case the argument has the form of the *invalid* argument schema (21):

- (21) $p, p \rightarrow q, \Box(q \rightarrow p) / \Box p$

So the argument collapses because of a logical subtlety concerning the representation of the scope of the modal expression *necessarily*. And there are several other well-known philosophical arguments which are subject to such an

analysis. Another example is Aristotle's sea battle argument, which attempts to demonstrate that $\Box\phi \vee \Box\neg\phi$, that is: everything is necessary in the sense that everything is either necessarily the case or necessarily not the case, from the law of the excluded middle, $\phi \vee \neg\phi$. It was this argument that first gave rise to many-valued logic, where truth values other than just 0 and 1 are introduced. (The reader is referred to chapter 5 of volume 1 for a discussion.) This approach enables one to escape fatalistic conclusions by pronouncing $\phi \vee \neg\phi$ invalid. These multiple truth values and extended truth tables were very popular in the twenties and thirties, and in them a semantics for intensional propositional logic was sought which would be analogous to the truth tables of propositional logic. It was never found. And it is not needed either, for the considerations based on the modal logic presented here are nearly always sufficient to defuse arguments like that in Aristotle's sea battle. Whatever virtues many-valued logic may have, its original motivation is not at all convincing.

2.4 Propositional Tense Logic

2.4.1 Syntax and Semantics

Technically speaking, *tense logic* is very closely related to modal logic. In tense logic contexts become moments in time, with *earlier than* as their accessibility relation. (This is the traditional way of doing things, but it is by no means the only one. These days many semanticists prefer to set things up with intervals of time as the contexts instead of moments.)

Tense logic originated with the observation that verb tenses display quite regular behavior, which seemed to lend itself to formalization. Two operators were introduced, **G** and **H**, as analogues of the \Box operator from modal logic. The **G** operator is interpreted as *it is always going to be the case that*, and the **H** operator as *it always has been the case that*. Now as such, **G** and **H** are hardly to be described as common tenses, but just as \Box is complemented by \Diamond , **G** and **H** have their own complements, **F** and **P**, which are to be read as *it will at some stage in the future be the case that* and *it was at some stage in the past the case that*, respectively. So the operator **F** serves as an acceptable formal pendant of at least some forms of the future tense, while **P** does the same for the past tense. The four usual tense operators have been summed up as (22):

- (22) **G** ϕ : it is always *going to be* the case that ϕ
H ϕ : it always *has been* the case that ϕ
F ϕ : it will at some stage in the *future* be the case that ϕ
P ϕ : it was at some stage in the *past* the case that ϕ

Adding these four operators **G**, **F**, **H**, and **P** to propositional logic, we obtain propositional tense logic. If p is now interpreted as *Mary is singing*, for ex-

ample, then (23) shows how a number of verb tenses may be represented in propositional tense logic.

- | | |
|--------|----------------------|
| (23) p | Mary is singing. |
| Fp | Mary will sing. |
| Pp | Mary sang. |
| PPp | Mary had sung. |
| FPp | Mary will have sung. |
| PFp | Mary would sing. |

It is clear that not every combination of **F** and **P** corresponds to a tense actually occurring in natural language. Nor is it possible to express all tenses by means of these operators. For example, the difference between the simple past and the present perfect cannot be accommodated. But it will do as a first start.

Exercise 7*

Translate the following sentences into formulas of propositional tense logic. Represent the logical structure as well as possible and state the translation key you use.

- Now you are still young, but one day you will no longer be.
- I am faithful to you, and I always will be.
- John has read *War and Peace*, and Charles has too.
- When Mary entered, John had put the whiskey bottle in the refrigerator.
- When Mary entered, John was about to put the whiskey bottle in the refrigerator.
- A sea battle will be fought or not. And if a sea battle will be fought, this has always been the case.
- Only if you will always stay with me will I really be happy.

A model **M** for propositional tense logic consists of a nonempty set **T** of moments in time, an *earlier than* relation **R**, and a valuation **V**, which for each proposition letter **p** and each moment in time $t \in T$ assigns a truth value $V_t(p)$ to **p** at time t . Just as in modal logic, **T** and **R** together form a frame, which in tense logic is sometimes referred to as a *time axis*. Here is that part of the truth definition which is concerned with the intensional operators:

Definition 4

Let **M** be a model which has **T** as its set of moments in time and **R** as its *earlier than* relation; then $V_{M,t}(\phi)$ is defined as follows:

- $V_{M,t}(G\phi) = 1$ iff for all $t' \in T$ such that tRt' : $V_{M,t'}(\phi) = 1$.
- $V_{M,t}(F\phi) = 1$ iff for at least one $t' \in T$ such that tRt' : $V_{M,t'}(\phi) = 1$.
- $V_{M,t}(H\phi) = 1$ iff for all $t' \in T$ such that $t'Rt$: $V_{M,t'}(\phi) = 1$.
- $V_{M,t}(P\phi) = 1$ iff for at least one $t' \in T$ such that $t'Rt$: $V_{M,t'}(\phi) = 1$.

Up until now we have not placed any special requirements on the time axis. Nor have we said anything about which properties R must have if it is to pass as an interpretation of the *earlier than* relation. One line of investigation in tense logic has been to think up plausible tense-logical principles and then to try and discover which requirements must be placed on the time axis in order to safeguard their validity. Another and complementary approach has been to sort out which principles become valid given particular restrictions on the time axis. As with modal logic, the notion of *validity on a frame* has played an important part in this.

We shall now discuss a few intuitively plausible tense-logical principles and see which properties of the time axis they express, if any. Clearly (24) and (25)

$$(24) \quad G(\phi \rightarrow \psi) \rightarrow (G\phi \rightarrow G\psi)$$

$$(25) \quad H(\phi \rightarrow \psi) \rightarrow (H\phi \rightarrow H\psi)$$

will be valid on any time axis, for G and H are both versions of the modal operator \Box , and the corresponding modal principle is valid independently of accessibility relations. It should be noted, however, that tense-logical principles (26) and (27) corresponding to modal principle (4) have none of the latter's intuitive plausibility:

$$(26) \quad G\phi \rightarrow \phi$$

$$(27) \quad H\phi \rightarrow \phi$$

These two principles are equivalent to $\phi \rightarrow F\phi$ (if ϕ is the case, then ϕ will be the case) and $\phi \rightarrow P\phi$ (if ϕ is the case, then ϕ was the case), respectively. Now if R is required to be *irreflexive*, which is a very reasonable restriction, since it means that no moment in time may be earlier than itself, (26) and (27) become invalid. But just as in modal logic, this requirement of irreflexivity cannot be expressed by means of a formula. Besides (24) and (25), there are the following intuitively sound principles:

$$(28) \quad \phi \rightarrow HF\phi$$

$$(29) \quad \phi \rightarrow GP\phi$$

$$(30) \quad P\phi \rightarrow H(F\phi \vee \phi \vee P\phi)$$

$$(31) \quad F\phi \rightarrow G(P\phi \vee \phi \vee F\phi)$$

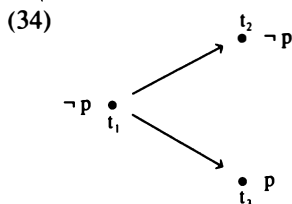
$$(32) \quad P\phi \rightarrow GP\phi$$

$$(33) \quad F\phi \rightarrow HF\phi$$

Principle (28) says that what is now the case has in the past always been something that would come to pass. And (29) says that what is now the case will always be something that has happened. Principle (30) states that if ϕ was once the case, then it has always been the case either that ϕ was yet to happen or that ϕ was happening or that ϕ had already happened. Formula (31) says

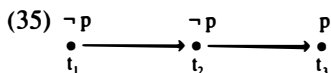
something analogous about the future. Principle (32) states that anything which has happened will always be something which has happened, and (33) makes the analogous statement about the future.

It is quite easy to check that principles (28) and (29) are true on every frame. They thus place no requirements at all on the time axis. Principles (30) and (31) are valid on all frames with a connected relation R (frames in which of any two different moments in time one is earlier than the other). If R is not connected, then a configuration like the one in (34) becomes possible.



Letting $V_{t_3}(p) = 1$ and $V_t(p) = 0$ for all other t , we have a counterexample to (31). For Fp is true at t_1 , because p is true at t_3 . But $G(P\phi \vee \phi \vee F\phi)$ is false at t_1 because $P\phi \vee \phi \vee F\phi$ is false at t_2 . This last is the case since neither t_2Rt_3 nor t_3Rt_2 nor $t_2 = t_3$, while t_3 is the only moment in time when p is true. As a result, neither Fp nor Pp nor p is true at t_2 . So the intuitively acceptable principles (30) and (31) rule out a forked or branched time axis, which would seem quite a reasonable thing to do (a possible application of a forked time axis is, however, given in §2.5).

Principles (32) and (33) are valid just in case R is *transitive*. If R is not transitive, then a configuration like the one in (35) becomes possible.



Here we do not have t_1Rt_3 . Letting $V_{t_3}(p) = 1$ and $V_t(p) = 0$ for all other t , we obtain a counterexample to principle (33): Fp is true at t_2 , because p is true at t_3 and t_2Rt_3 . But HFp is not true at t_2 , since Fp is false at t_1 because it is not true that t_1Rt_3 , and t_3 is the only moment at which p is true.

These and similar relations between principles like (24)–(33) and properties of the *earlier than* relation (and thus of different temporal structures) have been studied at length in tense logic. We have seen that a similar line of research has been followed in modal logic. One big difference between modal and tense logic is that with tense logic it seems more reasonable to begin by choosing a semantics. Unlike our intuitions about modalities, which concern the validity of various logical principles more than the relations between possible worlds, our temporal intuitions do seem to bear on the structure of time. So with tense logic it would seem reasonable to approach things back to front by first trying to formulate these intuitions and then trying to find the syntactic principles which they give rise to.

Actually there are different concrete conceptualizations of time which are considered in tense logic. They all, however, have one thing in common,

namely, the assumption that time is a *linear order*, that is, that the *earlier than* relation has the properties of a linear order: transitivity, asymmetry (and thus irreflexivity), and connectedness. Because of these properties, a linear *earlier than* relation is often represented as $<$. In all that is discussed below, the set of moments in time is ordered by $<$, and principles (24)–(33) are all valid. It is useful to represent the different conceptualizations of time with their divergent structures of the time axis as number lines. One intuitively appealing conceptualization is to take the *whole numbers* as our model of the time axis: $\dots, -n, \dots, -2, -1, 0, 1, 2, \dots, n, \dots$. Then time has no beginning and no end and runs in *discrete* steps. This is quite plausible if, for example, the days on the calendar are taken as the units of time. We can also conceive of time not as divided up into discrete steps, however small these might be (hours, minutes, seconds, nanoseconds), but as having a much finer structure, there always being a moment between any two other moments. This general property of relations, which is expressed by the formula $\forall x \forall y ((x \neq y \wedge Rxy) \rightarrow \exists z (z \neq x \wedge z \neq y \wedge Rxz \wedge Rzy))$, is called *density*. The idea that time is a dense order can be modeled by representing the moments in time as the *rational numbers*, which include the whole numbers. Then time has neither beginning nor end, and between any two moments in time there is always another to be found.

Determining the tense-logical principles which follow from a given choice of time axis is a rather complicated technical matter which is beyond the scope of this introduction. Taking the rational numbers as our model of the time axis, for example, we obtain the following three principles in addition to (24)–(33):

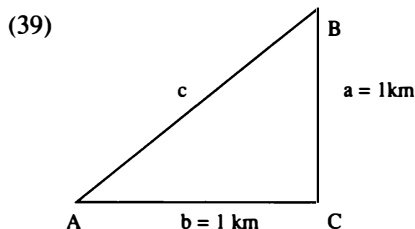
$$(36) \mathbf{F}\phi \rightarrow \mathbf{FF}\phi$$

$$(37) \neg \mathbf{G}(\phi \wedge \neg \phi)$$

$$(38) \neg \mathbf{H}(\phi \wedge \neg \phi)$$

Principle (36) says that if ϕ will happen, then it will be the case that ϕ will happen. That this principle is not valid on a discrete time axis like the whole numbers is apparent from the following example. Taking *A successor to the throne is born* as p and days as our units of time, $\mathbf{F}p$, *A successor to the throne will be born*, is true on Monday, given that a successor is brought into the world on Tuesday. But the truth of $\mathbf{F}p$ on Monday does not guarantee the truth of $\mathbf{FF}p$ on that same day. If there is a revolution on Wednesday, for example, or if the royal family dies out for any other reason, then $\mathbf{F}p$ need not be true on any day after Monday. It turns out that principle (36) corresponds to the density of the earlier than relation. Principle (37) says that time never stops, since $\mathbf{G}(\phi \wedge \neg \phi)$ would only ever be true at a very last moment t . For $\mathbf{G}(\phi \wedge \neg \phi)$ is true at t just in case $\phi \wedge \neg \phi$ is true at all instants t' which come after t , and these moments t' do not exist if t is the last moment. Principle (38) expresses the fact that time does not have a beginning in the same way.

A third conceptualization of time is found in physics, where the real numbers are taken as a model of the time axis. This conceptualization is needed in order to describe situations like that in (39).



Assuming you walk 1 kilometer per hour, it would take you $\sqrt{2}$ hours to walk from A to B. This follows from the Pythagorean theorem, which says that $c^2 = a^2 + b^2$. But it has been known since the ancient Greeks that $\sqrt{2}$ is not a rational number, which means that it cannot be expressed as a fraction. So we couldn't even be sure that there would be a moment for us to arrive at if the time axis were to consist only of rational numbers. The tense-logical principles corresponding to this third kind of temporal structure are rather more complicated than the other principles we have seen.

Exercise 8*

- (a) Consider the following model: $T = \{t_1, t_2, t_3, t_4, t_5, t_6\}$; $R = \{\langle t_1, t_2 \rangle, \langle t_2, t_3 \rangle, \langle t_3, t_6 \rangle, \langle t_1, t_4 \rangle, \langle t_4, t_5 \rangle, \langle t_5, t_6 \rangle\}$; $V_{t_1}(p) = V_{t_2}(p) = V_{t_3}(p) = V_{t_6}(p) = 1$; $V_{t_4}(p) = V_{t_5}(p) = 0$
- Draw a picture of the model;
 - Decide whether the following formulas are valid in the model:
 - $\neg p \rightarrow FGp$
 - $F\neg p \rightarrow FF\neg p$
 - $G(P\neg p \rightarrow \neg p)$
 - $(p \wedge Gp) \rightarrow Hp$
- (b) Decide which property of the time axis is characterized by each of the following principles:
- $FG\phi \rightarrow GF\phi$
 - $G(\phi \wedge \neg\phi) \vee FG(\phi \wedge \neg\phi)$
 - $PP\phi \rightarrow P\phi$

Exercise 9

Give a frame that consists of three points on which $Fp \rightarrow G(p \vee Pp \vee Fp)$ is valid, and one with the same number of points on which it is not.

2.4.2 'Now': An Extension

Much linguistically oriented research into tense logic is concerned with adapting and extending tense-logical languages and their semantics so as to bring

more of the behavior of verbs and other temporal expressions such as adverbs into the picture. By way of illustration, we will briefly describe one such extension of tense logic, the N-operator (N for *now*) originated by Kamp. On the face of it, it might seem that *now* is a redundant term. After all, (40) surely means the same thing as (41):

(40) Now John is asleep.

(41) John is asleep.

This would suggest the following definition as the interpretation of the N-operator:

(42) $V_{M,t}(N\phi) = 1$ iff $V_{M,t}(\phi) = 1$

If we let p stand for *John is asleep*, then (42) would ensure that Np is equivalent to p , which is just what (40) and (41) seem to suggest. But things are nowhere near as simple as this, as is apparent from (43):

(43) Someday you will be grateful for what I am doing now.

The *now* in (43) refers to the time when the sentence is uttered. The sentence is in the future tense, which means that its formal translation will be of the form $F(\text{you are grateful for what I am doing now})$. If we evaluate such a formula at a moment t , it will be true just in case there is some moment t' later than t at which *you are grateful for what I am doing now* is true. But then according to the definition of *now* given in (42), (43) will be true at a moment t just in case there is a moment t' later than t when it is true that you are grateful for what I am then, at t' , doing. But this is not what (43) means. So the interpretation given to *now* in (42) will not do. Apparently we need a way to refer back to the original moment of utterance, even if the interpretation process carries us to other moments. This can be done by adding a fixed moment t_0 to the structures: the moment *now*. The N-operator can then be interpreted as follows:

(44) $V_{M,t}(N\phi) = 1$ iff $V_{M,t_0}(\phi) = 1$

Given this interpretation, the *now* in (43) will indeed refer back to the moment of utterance t_0 , even though it occurs within the scope of the future tense.

Now we can also account for the difference between two well-known examples of Kamp:

(45) A child was born which would rule the world.

(46) A child was born which will rule the world.

Sentence (45) can be represented as (47),

(47) $P(\exists x(Cx \wedge Bx \wedge FRx))$

in which Cx stands for *x is a child*, Bx for *x is born*, and Rx for *x rules the world*. Note that the moment when x rules the world can be either before or

after the moment of utterance, since the tense operator **F** occurs within the scope of the operator **P**. This accords nicely with the *would* in (45). The *will* in (46), on the other hand, states that the moment at which *x* will rule the world comes after the moment of utterance. Thus (46) can be analyzed with the aid of the **N**-operator and can be represented as (48):

$$(48) \text{ P}(\exists x(Cx \wedge Bx \wedge \text{NFR}x))$$

Given the interpretation of **N** in (44), the **F** operator in (48) determines some moment in the future which we now have, even though it occurs within the scope of **P**.

Another application of Kamp's technique for dealing with the semantic behavior of *now* is given in §3.4.

2.4.3 Other Approaches

The above treatment of temporal expressions by means of tense operators is due to Prior. His is certainly the best-known approach, but it isn't the only one, and it may not even be the best. Therefore we will briefly mention two other approaches.

The first was developed by Reichenbach. He associates three contextual time points with each temporal expression: a *point of speech* *S* (which is comparable with Kamp's t_0), a *point of event* *E* (the time at which the event described in the expression takes place; in the above it is incorporated into the truth definition), and a *point of reference* *R*, which represents, as it were, the temporal vantage point adopted by the speaker. Then verb tenses may be represented by means of simple diagrams. By way of comparison, see (49):

(49) E, R, S	Mary sings.
S — E, R	Mary will sing.
E, R — S	Mary sang.
E — R — S	Mary had sung.
S — E — R	Mary will have sung.
R — E — S	Mary would sing.

This gives the representations of the same verb tenses as were treated by means of tense operators in (23). One interesting aspect of this approach is that it enables us to account for the difference between the simple past tense *Mary sang* and the present perfect *Mary has sung*, namely, as a difference in the temporal vantage point adopted by the speaker, which can be represented by the position of *R*:

(50) E — S, R	Mary has sung.
E, R — S	Mary sang.

That this theory does not offer a complete account of all temporal constructions either becomes apparent when we try to find a representation for *Mary would have sung*. This cannot be done with just a single point of reference *R*.

There is, however, an analysis in terms of tense operators, namely, **PFP**_p, and an integration of these two approaches is still being worked on.

The second alternative approach assumes that temporal constructions in natural language are so complex that more powerful logical machinery is needed: (many-sorted) predicate logic with explicit quantification over moments in time. (The reader is referred to §5.3 in volume 1.) We have already mentioned that tense operators behave much as quantifiers, and the same applies to temporal adverbs like *always*, *sometimes*, *never*, *frequently*, *rarely*, and so on. In this approach, the predicate letters are fitted out with an extra variable for moments in time (t , t' , t'' , etc.), and proposition letters become unary predicates. The language also includes the *earlier than* relation $<$. The constructions already treated in (23) and (49) are redone in (51):

(51) Pt_0	Mary sings.
$\exists t(t_0 < t \wedge Pt)$	Mary will sing.
$\exists t(t < t_0 \wedge Pt)$	Mary sang.
$\exists t \exists t'(t' < t_0 \wedge t < t' \wedge Pt)$	Mary had sung.
$\exists t \exists t'(t_0 < t' \wedge t < t' \wedge Pt)$	Mary will have sung.
$\exists t \exists t'(t' < t_0 \wedge t' < t \wedge Pt)$	Mary would sing.
$\exists t \exists t' \exists t''(t'' < t_0 \wedge t'' < t' \wedge t < t' \wedge Pt)$	Mary would have sung.

Here, we explicitly quantify over moments in time. A special moment t_0 figures once again as the representation of the current moment of evaluation ('now'). The formulas in (51) are relatively complex and difficult to interpret in comparison with the corresponding formulations in tense logic. They become somewhat more legible, however, if besides t_0 a second fixed moment in time t_1 is introduced as a representation of the point of reference, a move which also has the advantage of enabling us to express the difference between the simple past and the present perfect tenses. A theory along these lines has been developed by Paul Needham. Temporal predicate logics like this give rise to relatively complicated representations of the simpler constructions, but it must be said in their defence that at least they provide a technical apparatus which is capable of representing a tremendously wide variety of temporal constructions, constructions for which tense logic would need a whole new set of operators. (Two-sorted type theory, a logical system which will be discussed in §5.8, can be used as a formal tool in this kind of approach.)

2.5 Tense and Modality Combined

Counterfactual constructions like (52) seem to combine tense and modality:

(52) If I had gone I would have found happiness.

I did not in fact go, but I might have. Here we see different kinds of intentionality interfering with each other. Very often the resulting whole is no more

than the sum of its different parts, but in some cases new puzzles emerge from the combination which call for new and creative semantic solutions.

A combined modal and tense logic is obtained if not only \Box but also the operators G and H are added to propositional logic. Various semantic structures could be chosen, but to keep things from turning into science fiction, we will just take a set W of possible worlds, each with the same fixed time axis. We can then speak in terms of the truth value of a formula ϕ in a world w at time t . There is an *earlier than* relation $<$ on the set T of moments in time and an accessibility relation R on W . The key clauses in the truth definition are then:

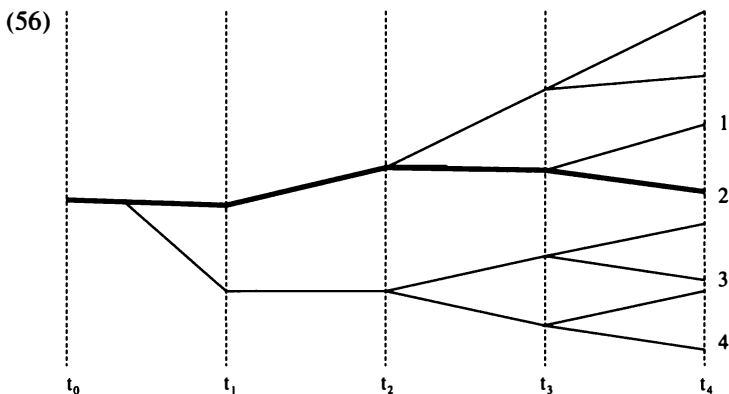
$$(53) V_{M,w,t}(\Box\phi) = 1 \text{ iff for all } w' \text{ such that } wRw': V_{M,w',t}(\phi) = 1$$

$$(54) V_{M,w,t}(G\phi) = 1 \text{ iff for all } t' \text{ such that } t < t': V_{M,w,t'}(\phi) = 1$$

The clause for $\Box\phi$ expresses a temporalized necessity: $\Box\phi$ is true in w at t iff ϕ is true at t in each world w' accessible from w . There is something to be said for allowing the accessibility of worlds to each other to change from time to time. This can be done by providing R with a temporal parameter, thus obtaining a set of accessibility relations R_t , one for each $t \in T$. The clause for \Box then becomes:

$$(55) V_{M,w,t}(\Box\phi) = 1 \text{ iff for all } w' \text{ such that } wR_t w': V_{M,w',t}(\phi) = 1$$

Paraphrasing, $\Box\phi$ is true in w at time t iff ϕ is true at t in each world which is accessible from w at that time t . This last option becomes a little more concrete if R_t is defined in the following manner: $wR_t w'$ holds just in case w and w' have the same history up until t (at which point they may or may not diverge). We then obtain an intuitively plausible *branching* time structure like that depicted in (56):

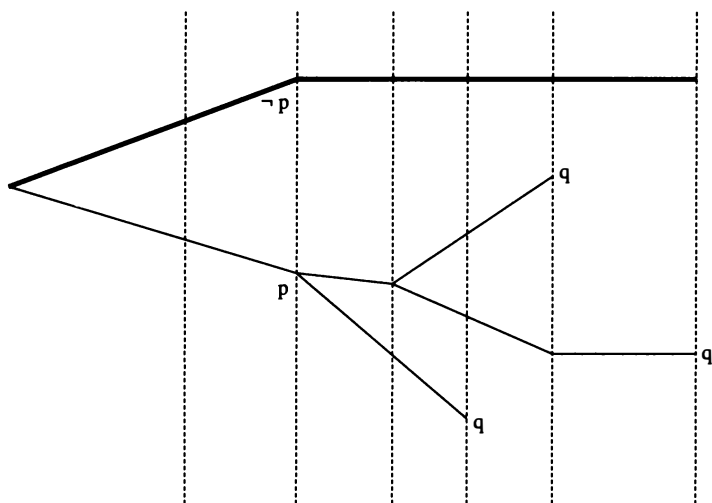


The bold line represents the actual history of the world. Let us suppose that it is 1978 and that we are watching the world cup soccer matches, which in that year were held in Argentina. The first rounds are played between times t_0 and

t_1 , the quarter-finals are played between t_1 and t_2 , the semifinals between t_2 and t_3 , and the finals at t_4 . The actual result of the whole train of events, 2, was that Argentina beat the Netherlands in the finals. Another possible result would have been 1, in which the Netherlands beat Argentina in the finals. This result was still possible at t_3 , the moment when the semifinals had just been played. The train of events resulting in 3 is one in which Brazil is not eliminated in the first rounds because of a lower goal average than Argentina's (which is what in fact happened), and subsequently goes on to win the cup. And 4 stands for a train of events which resembles 3 to the extent that Brazil makes the finals, but in which Brazil is then beaten by West Germany. The trains of events terminating at 3 and 4 were possibilities at the outset of the tournament but not after the first rounds.

This conception of time, in which every point in time is followed by a number of divergent possible 'futures', would seem the most appropriate for dealing with the counterfactual in (52): I did not in fact go, but it was at the time a necessary fact that if I had gone, I would have found happiness. The obvious formalization of (52) would then seem to be $P(\neg p \wedge \Box(p \rightarrow Fq))$, in which p stands for *I go* and q for *I find happiness*. This formula becomes true in situations like that in (57):

(57)

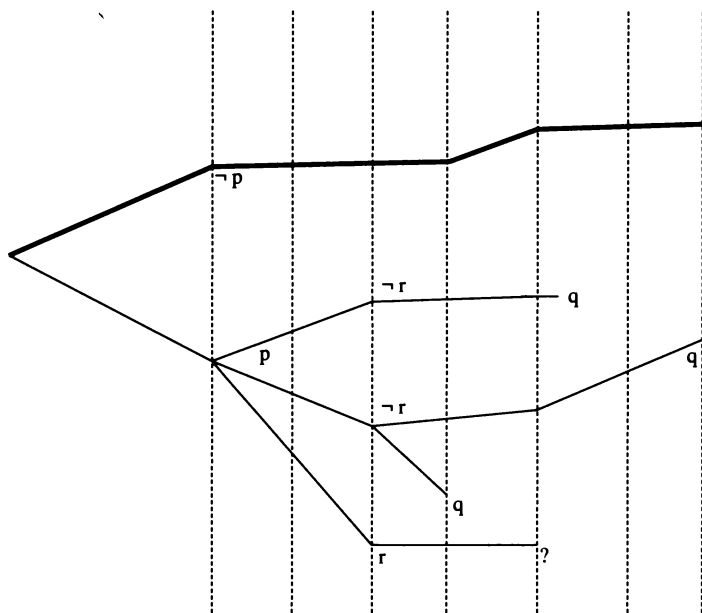


But things can't be as simple as this. For it is clear that $\Box((\phi \wedge \chi) \rightarrow \psi)$ always follows from $\Box(\phi \rightarrow \psi)$, which means that $P(\neg p \wedge \Box((p \wedge r) \rightarrow Fq))$ follows from $P(\neg p \wedge \Box(p \rightarrow Fq))$. Now let r stand for *I die*. Then (58) follows from (52):

(58) If I had gone and had died, I would have found happiness.

This inference is at least rather doubtful. But whether (58) is true or not, it seems strange that it should *follow* from (52), as is apparent from (59):

(59)



So this formalization of the counterfactual is not entirely satisfactory. It would seem that (52) should be interpreted as (60):

(60) If I had gone, then everything else being equal, I would have found happiness.

So we will have to find some semantic interpretation for the qualification *everything else being equal*. In this particular case, the qualification means something like: I *did* go, but apart from this, the world remained as close as possible to what it really is. In general then, in a counterfactual construction, $\Box(\phi \rightarrow \psi)$ is not to be interpreted as: ψ holds in every possible world where ϕ holds, but rather as: ψ holds in every possible world where ϕ holds, but which in all other respects resembles the real world as closely as possible. If we are to make this idea precise, then we will have to introduce some new element into semantics. One possible solution, that of Lewis, is to assume not only an accessibility relation between possible worlds but also a *similarity relation*. This relation would fix the extent to which different possible worlds resemble each other. Given such a similarity relation, it is not difficult to see that $\Box(\phi \rightarrow \psi)$, under the qualified interpretation given to it above, can be true without $\Box((\phi \wedge \chi) \rightarrow \psi)$ being true. For the possible worlds which resemble the real world as closely as possible apart from having $\phi \wedge \chi$ true will presumably be

less like the real world than the possible worlds which resemble the real world as closely as possible apart from having ϕ true.

Then we could go a step further and try to define the similarity relation in terms of the truth of formulas instead of assuming it as a given. The reader is referred to work done by Veltman and Kratzer for elaborations of this idea. The resulting theory can also be applied to normal indicative conditional sentences, and it provides an interpretation of *if . . . then* which seems more natural than material implication, at least for natural language.

3

Intensional Predicate Logic

3.1 Opaque Contexts: Modalities de Dicto and de Re

As an introduction to intensional predicate logic, we will return to a theme from §1.6. There we introduced intensional constructions as constructions for which certain substitution principles of predicate logic which are related to the principle of extensionality do not hold. Following Quine, we say that these intensional constructions create *opaque contexts*. These are to be distinguished from *transparent contexts*, for which these substitution principles do hold. Here are some constructions which give rise to opaque contexts, together with examples which show that they violate the principle of extensionality (1):

$$(1) \quad s = t \models \phi \leftrightarrow [t/s] \phi$$

Quotation. Sentence (4) does not follow from (2) and (3):

- (2) The gladiator spoke the words *Ave Caesar*.
- (3) Caesar is Gaius Julius.
- (4) The gladiator spoke the words *Ave Gaius Julius*.

Indirect speech. Sentence (7) does not follow from (5) and (6):

- (5) Harry said that John kissed Mary.
- (6) John is the smartest boy in the class.
- (7) Harry said that the smartest boy in the class kissed Mary.

Constructions with verbs expressing *propositional attitudes*, like *to discover*, *to believe*, *to suspect*, and *to know*. Sentence (10) does not follow from (8) and (9):

- (8) The detective knows that the thief entered through the skylight.
- (9) Biggles is the thief.
- (10) The detective knows that Biggles entered through the skylight.

Constructions with verbs expressing *intentions*, such as *to look for*, *to wish for*, and the like. We saw an example in §1.6:

(11) John is looking for the supreme commander of the armed forces of the United States of America.

(12) The president of the United States of America is the supreme commander of the United States armed forces.

(13) John is looking for the president of the United States of America.

Temporal designation. Sentence (16) does not follow from (14) and (15):

(14) George Bush is the president of the United States.

(15) In 1963, the president of the United States was assassinated in Dallas, Texas.

(16) In 1963, George Bush was assassinated in Dallas, Texas.

Modality. It is a necessary truth that nine exceeds seven: given the meanings of *nine*, *seven*, and *exceeds*, the sentence *Nine exceeds seven* is a necessary truth. But surely (17) is not a necessary truth:

(17) The number of planets exceeds seven.

The truth or falsity of (17) is not to be determined solely on the basis of the expressions it contains. In fact, (17) expresses a contingent astronomical fact. That there are more than seven planets is something which was discovered through observation and inference. So (20) does not follow from (18) and (19):

(18) Nine necessarily exceeds seven.

(19) Nine is the number of planets.

(20) The number of planets necessarily exceeds seven.

Besides these, there are many more constructions giving rise to opaque contexts. Just about every category of expressions contains elements which can create opaque contexts, for example adjectives like *suspected* and *alleged*, adverbs like *apparently*, and so on.

Philosophers have shown different reactions to the invalidity of the substitution principle (1) in the case of opaque contexts. Let us return to (20). It might be argued that there is a reading for (20) in which this sentence does indeed follow from (18) and (19). This reading can be paraphrased as follows: that number which is in fact the number of planets is necessarily greater than seven. This reading translates as (21), whereas the reading of (20) for which (20) does not follow from (18) and (19) may be rendered as (22):

(21) $\exists x(x = \text{the number of planets} \wedge \Box(x > 7))$.

(22) $\Box\exists x(x = \text{the number of planets} \wedge x > 7)$.

Reading (22) says that in every possible situation, the number of planets, whatever it happens to be, will exceed seven. The two readings (21) and (22) of (20)

lead to a distinction traditionally drawn in modal logic between modalities *de dicto* and *de re*. This distinction can be expressed in a predicate-logical language with an added \Box operator in terms of the scope of \Box . Let us consider the somewhat simpler examples (23) and (24) and their translations (25) and (26):

(23) Necessarily there is something which is greater than seven.

(24) There is something which is necessarily greater than seven.

(25) $\Box \exists x(x > 7)$

(26) $\exists x \Box(x > 7)$

In (25) the scope of \Box is $\exists x(x > 7)$, and in (26) it is $x > 7$. The scope of an occurrence of \Box may be considered to be the opaque context created by this operator. If all variables within the scope of \Box are bound by quantifiers likewise within its scope, then \Box is said to be a modality *de dicto*. As examples, then, we have (22) and (25). If, on the other hand, there is a free variable within the scope of \Box , that is to say, a variable bound by a quantifier outside the scope of \Box , then \Box is said to be a modality *de re*. As examples of this modality we have (21) and (26). Traditionally a modality *de dicto* was seen as an attribution of necessary (or possible) truth to a proposition (*dictum*), and a modality *de re* was seen as an attribution of a necessary (or possible) property to an entity (*res*). The traditional distinction corresponds to the formal one. In asserting the truth of (25), one asserts that the proposition $\exists x(x > 7)$ is necessarily true, while in asserting the truth of (26) one asserts the existence of an entity which necessarily has the property of being greater than seven.

Some philosophers have objected to the latter. For them, recognition of modalities *de re* amounts to a revival of *essentialism*, a philosophical position which distinguishes between *accidental* properties of things and *essential* properties. They have their objections to any such position and therefore reject modalities *de re* as meaningless and thus useless; at best they suggest reducing modalities *de re* to modalities *de dicto*. One such vigorous opponent of modalities *de re* has been the philosopher and logician Quine. Even leaving aside the question of whether recognizing modalities *de re* really leads to essentialism, it would seem to us that a position like his is particularly unsuited to our purposes. In our opinion, philosophical objections should never be allowed to weigh heavily if the aim is the description of natural language. We want descriptions of how we speak, not of how we would have to speak in order to earn the approval of philosophers. It is quite possible that speakers of natural languages make philosophically dubious assumptions, but that is a fact of life which should not be swept under the rug of some philosophically more sophisticated reformulation. But that modalities *de re* occur in natural language seems to us indisputable. An example is (27):

(27) Each of those present may have committed the murder.

It is clear what this means. Sentence (27) may be formalized as $\forall x \Diamond Mx$. It certainly does not mean the same thing as $\Diamond \forall x Mx$, which is the translation of (28):

- (28) It is possible that each of those present has committed the murder.

It is not at all clear how a de re modality like (27) could be reduced to a de dicto modality. And besides, we are of the opinion that possible-worlds semantics provides a clear interpretation for modalities de re.

Adding modal and/or tense operators to predicate logic, we obtain a system of intensional predicate logic. Tense operators lead to the same kinds of ambiguities as we have seen with modal operators. Sentence (29), for example, has two distinct readings, (30) and (31). The key to the translations is given as (32):

- (29) The president received a decoration.

- (30) $\mathbf{P} \exists x (\forall y (Py \leftrightarrow y = x) \wedge Dx)$

- (31) $\exists x (\forall y (Py \leftrightarrow y = x) \wedge \mathbf{P} Dx)$

- (32) Px : x is president; Dx : x receives a decoration.

The same distinction can be drawn as with modalities: (30) is said to be the de dicto reading and (31) the de re reading. The reader can practice working with the scopes of quantifiers, modal operators, and tense operators in the following translation exercises.

Exercise 1*

Translate the following sentences into formulas of intensional predicate logic.

- Lendl may win the Wimbledon tournament one day.
- Perhaps everybody has always been aware of something.
- Perhaps there is something that everybody has always been aware of.
- If anybody can be smarter than anybody else, then everybody can be the smartest.
- You can fool some of the people all of the time, and all of the people some of the time, but you can't fool all of the people all of the time.
- The president will always be a democrat (try to find two translations for this sentence).
- Every schoolboy believes that a mathematician wrote *Through the Looking Glass* (try to find three translations for this one, representing 'x believes that ϕ ' as $B(x, \phi)$).

Exercise 2

Try to find examples of intensional expressions (i.e., expressions which do not allow substitution of materially equivalent expressions in their scope) in each of the following categories:

(a) adjectives; (b) adverbs; (c) conjunctions; (d) prepositions; (e) determiners.

The exposition of the semantics of predicate logic given in §3.2 focuses almost exclusively on the special case of modal predicate logic. Modal predicate logic is the oldest and best understood member of the intensional family, and it remains instructive of what goes on in temporal predicate logic, epistemic predicate logic, and other related systems.

3.2 Proper Names and Definite Descriptions: Rigid Designation

As we will see in the rest of this chapter, there is no single preferred semantics for modal predicate logic. Different choices of domains and the interpretations of constants and quantifiers give rise to divergent semantic interpretations. A number of different truth definitions arising from such choices will be presented below, and their advantages and disadvantages will be compared. Besides a frame, which consists of a set W of possible worlds and an accessibility relation R , a model for modal predicate logic will need a domain which decides the range of the quantifiers. And here we face the first choice: (1) each world w gets its own domain D_w ; the model as a whole will then contain a set $\{D_w | w \in W\}$ of domains; or (2) we consider a fixed domain D shared by all of the possible worlds. Since (2) is a special case of (1) (namely, that in which for all $w, w' \in W, D_w = D_{w'}$), we will begin with a discussion of (1), but we will also see that various considerations may ultimately lead us to prefer approach (2). But before giving a general definition for the interpretation of a modal predicate logical language, we will first concentrate on the interpretation of constants, this with reference to the debate on the meanings of proper names and definite descriptions.

In interpreting constants we once again face a choice between two alternatives. One way would be to choose some fixed entity as the interpretation of a constant c ; in this case we would end up with an interpretation function I which assigns entities to constants. The other way would be to make the interpretation of the constants world-dependent: for every $w \in W, I_w(c)$ will then be some member of D_w . We say that c is interpreted as an *individual* (entity) in the first case and as an *individual concept* in the second. So, as we noted in §1.8, an individual concept is a function from worlds to individuals. Individual concepts open all kinds of interesting possibilities. A person, for example, in view of individual concepts, needs no longer be identified with some element of a domain but may (in tense logic) be seen as a function from moments in time to biological entities, a conceptual transition which may shed some light on the fact that individuals can 'change'. We cannot pause to follow this line of thought here, so the interested reader is referred to Stevenson 1886, for example.

It is perhaps something of a surprise that we prefer the first way of interpreting constants. We prefer to interpret them as individuals. The reason for this is

that natural language contains expressions which function in this manner, namely, *proper names*. This point will be argued below. Natural language also contains expressions of another kind, which refer to entities but may have different references in different possible worlds: *definite descriptions*. This is, for example, apparent from sentence (20), in which *the number of planets* functions as a definite description. If the reference of *the number of planets* were to be the same in every possible world, it would in all cases be nine, and sentence (22) would be true, contrary to our conclusions about (20). A description can function in many different ways in the context of a sentence, depending on its scope. Is it necessary that the winner wins? That all depends on the scope of the operators in $\Box W(\exists x Wx)$. The description $\exists x Wx$ can be rewritten in Russellian form in two different ways, inside and outside the scope of \Box . We then obtain (33) and (34):

$$(33) \quad \Box \exists x (\forall y (Wy \leftrightarrow x = y) \wedge Wx)$$

$$(34) \quad \exists x (\forall y (Wy \leftrightarrow x = y) \wedge \Box Wx)$$

Now (33) is always true (if the game is played), but presumably (34) is false: the actual winner might have been beaten. Note that the first reading is *de dicto*, and that the second is *de re*: there is a free variable x within the scope of \Box . In the above example with the planets, things were the other way around; the *de dicto* reading was false, and the much less obvious *de re* reading was true. The distinction between *de dicto* and *de re* would disappear if definite descriptions were to have a fixed interpretation, that is to say, the same interpretation in every world.

Philosophers have devoted a considerable amount of energy to the differences and similarities between proper names and definite descriptions. One question of central importance has always been whether or not proper names have meanings, and if they do, how they are to be represented. Frege, as we saw in §1.7, was of the opinion that every name, including every proper name, has a *Sinn*, which may be expressed as a definite description. According to Frege, it is a lamentable shortcoming of natural language that not everyone associates the same definite description with a proper name. One sees *Aristotle* as *the discoverer of syllogistic logic*, and another sees him as *the Stagirite tutor to Alexander the Great*. According to Frege, this would never be allowed to occur in a logically ideal language: each proper name would be introduced explicitly by means of a single definite description.

Taking definite descriptions as the meanings of proper names solves a number of difficult problems. First, there is the problem of how proper names refer to individuals, or of how we identify the individual to which a given proper name refers. We need to know what we mean by a name, and in order to single out some individual in particular as the reference of a name, we have to exploit any properties which distinguish this individual from others in the domain. If proper names are really no more than a shorthand for definite descrip-

tions, then the problem disappears: the meaning of the definite description indicates properties which distinguish the individual in question from all others. Second, there are problems with proper names which are like the paradox of the morning star and the evening star, which was discussed at length in §1.6. *Hesperus* and *Phosphorus* are two proper names which, like *the morning star* and *the evening star*, refer to the same object, viz., the planet Venus. If the meanings of these proper names *Hesperus* and *Phosphorus* are descriptions, then it is clear why *Hesperus is Hesperus* and *Hesperus is Phosphorus* have two entirely different meanings. A third problem concerns statements like *Pegasus does not exist*. A formal representation like $\neg \exists x(p = x)$ could never be true in the above kind of model, assuming that *Pegasus* is represented in the language by means of a constant p . Furthermore, there is the philosophical argument according to which anyone who asserts the truth of *Pegasus does not exist* is forced to concede that there is a nonexistent individual, namely the reference of the proper name *Pegasus* (this philosophical puzzle is known as 'Plato's beard'). If names are analyzed as descriptions, then both problems disappear. Understanding *Pegasus* to mean *the flying horse* and rewriting this description along Russell's lines, *Pegasus does not exist* reduces to *There is no unique flying horse*, which has $\neg \exists x(\forall y((Hy \wedge Fy) \leftrightarrow x = y))$ as its formal representation. This formula may very well be true in a model, and it does not contain names which may be exploited in a Plato's beard type argument.

These ideas on proper names solve certain problems, but they also raise a few of their own. We have seen that in natural language, not everyone associates the same descriptions with the same proper names. This implies that people may assign different meanings to a given proper name, and also that some may assign a mistaken meaning to it. It might happen that a description is associated with a name which does not apply to the individual to which that name refers, and this is a somewhat counterintuitive consequence.

It has been suggested that we get a better account of the meanings of proper names if we regard them as complex descriptions. For the name *Agamemnon* we might have something like: *the king of Mycenae who led an expedition of all Greeks against Troy, who destroyed that city after ten years at war, and who on his return was murdered by his wife and her lover*. Now suppose there had been a king of Mycenae who had achieved all of what Agamemnon is supposed to have achieved, but who on returning home had lived happily ever after. Then on the above view we would have to say that there never was an Agamemnon. The normal reaction, however, would be to say that Agamemnon certainly existed, but that he is wrongly assumed to have been murdered. And the same applies to all other information we have about Agamemnon's personal history: it may yet turn out to be mistaken. This difficulty could perhaps be avoided in the following manner. Let ϕ_1, \dots, ϕ_n be formulas each of which has a free variable x , and which express all properties which we believe Agamemnon to have had. Now of each of these properties individually it may

be said that Agamemnon might not have had it. But then Agamemnon could still answer to the following disjunctive description:

$$(35) \quad \lambda x((\neg\phi_1 \wedge \phi_2 \wedge \dots \wedge \phi_n) \vee (\phi_1 \wedge \neg\phi_2 \wedge \dots \wedge \phi_n) \vee \dots \\ \vee (\phi_1 \wedge \phi_2 \wedge \dots \wedge \neg\phi_n) \vee (\phi_1 \wedge \phi_2 \wedge \dots \wedge \phi_n))$$

This, however, is not at all satisfactory. Perhaps we would be prepared to add disjuncts to the formula in which more than just one of the formulas ϕ is negated, but where would the limit lie? Would we say that Agamemnon is the unique individual who has, say, 50 percent of the listed properties ϕ ? And shouldn't some of the properties weigh more heavily than others? It would seem that we must conclude that at least in its present form, this theory runs into more problems than it solves, and that it does not offer a satisfactory interpretation for proper names.

The philosopher and logician Kripke introduced a quite different approach to the semantics of proper names and definite descriptions (see Kripke 1972). His position is that there is the following fundamental difference between proper names and definite descriptions. While the latter can change their references from world to world, according to Kripke the former refer to the same thing in every world in which they have a reference. They are what he calls *rigid designators*. We have already illustrated this difference in the behavior of proper names and definite descriptions in §1.7, with a *counterfactual*:

- (36) If Dukakis had won the presidential elections in 1988, then the president of the United States would have been a Democrat.

Dukakis refers to the same person as in the actual world in each of the possible worlds introduced by the antecedent of this implication, while the person referred to as *the president of the United States* in each of these worlds is different from the person this expression refers to in the actual world. Thus Kripke reduces the so-called transworld identity problem to the status of a pseudo-problem. This is a well-known problem raised by possible worlds semantics: what does it mean to say that an individual in a given world is the same as an individual in another world? According to Kripke, it is senseless to try and determine whether two entities in two different worlds are in fact one and the same by comparing their properties. Instead this is something which is given in advance. In expressing a sentence like (36), we introduce other possible worlds in which Dukakis's presence is unproblematic, although the Dukakis in these worlds may differ from the real one in various respects.

So how does Kripke's position fare as far as the other three problems we have mentioned are concerned? Kripke's general explanation of how proper names relate to descriptions is that names are often assigned to individuals by means of descriptions. This occurs, for example, in *I will call the littlest goat Jenny*. Once its reference has been established in this manner, the name *Jenny* remains applicable to the same individual, even if the original description no

longer fits (because Jenny has grown up). The goat in question will still be called Jenny when she gives birth to her own kids, and the discovery that she wasn't actually the smallest goat at the time of her 'baptism' will not change the fact that this is her name. This also explains how proper names come to refer. For the reference of a name is determined not by the meaning of the name but by the description by means of which it was originally *fixed* in the *initial baptism*. The second problem will be discussed in §3.3.2, when we come to identity. A solution to the third problem which is compatible with Kripke's position will have to wait until the notion of the existence predicate is introduced in §3.3.4.

Kripke's ideas about the meaning and the roles of proper names are quite attractive. And if they are to be reflected in the formal semantics of modal predicate logic, then individual constants, as the formal versions of proper names, will have to be interpreted as individuals rather than individual concepts.

3.3 The Semantics of Modal Predicate Logic

3.3.1 *Formulas without Variables*

Initially it is helpful to leave out quantifiers when explaining the semantics of modal predicate logic, postponing the complications resulting from interference between quantifiers and modal operators until later.

We have decided that the interpretations of the constants are to be independent of the worlds. But since we certainly want the truth values of sentences and the references of definite descriptions to be relative to possible worlds, we clearly must interpret the predicate letters in each world separately. So an n -ary predicate letter P is interpreted in each world w as a subset $I_w(P)$ of $(D_w)^n$. To recapitulate:

Definition 1

A *model* M for a modal predicate logical language L consists of:

- (i) a nonempty set W of possible worlds
- (ii) an accessibility relation R on W
- (iii) a domain function D which assigns a domain D_w to each world $w \in W$
- (iv) an interpretation function I which assigns an entity $I(c)$ to each constant c of L , and for every world $w \in W$ a subset $I_w(P)$ of $(D_w)^n$ to each n -ary predicate letter P of L .

Now we want to know what it means for a formula lacking variables to be true in a world w in a model M . As we shall soon see, there is no single truth definition for modal predicate logic that is clearly superior to all the rest. At various points we will have to choose between different alternatives, and the choices we make will be guided by the applications we have in mind in natural

language. Readers with other applications in mind may agree to differ. The important thing is that the motives behind our choices are revealed.

The first problem we are faced with in giving a truth definition is that the entities which the interpretation function I assigns to the different constants need not exist in every possible world where we evaluate. What, for example, is the truth value of sentence (37)

(37) Eve was blonder than Adam.

in this our real world (assuming that neither has ever existed)? It would seem that the most natural thing to do is not to have (37) false but to leave its truth value undefined. We must, however, be careful when writing up the truth definition: $\neg\phi$ may only be allowed a truth value if ϕ gets one, for example, for otherwise a sentence like (38):

(38) Eve was not blonder than Adam.

might end up with a truth value and (37) without one. We will also deny $\phi \rightarrow \psi$ a truth value if either ϕ or ψ is lacking one. The definition which these considerations lead to, given below as definition 2, still has a few snags. We will turn to them shortly and will suggest some improvements.

Definition 2

Let \mathbf{M} be a model, ϕ a formula of modal predicate logic lacking variables, and $w \in W$. Then $V_{\mathbf{M},w}(\phi)$, the truth value of ϕ in w given \mathbf{M} , is defined as follows:

- (i) $V_{\mathbf{M},w}(Pc_1 \dots c_n) = 1$ iff $I(c_1) \in D_w, \dots, I(c_n) \in D_w$
and $\langle I(c_1), \dots, I(c_n) \rangle \in I_w(P)$
 $= 0$ iff $I(c_1) \in D_w, \dots, I(c_n) \in D_w$,
and $\langle I(c_1), \dots, I(c_n) \rangle \notin I_w(P)$
- (ii) $V_{\mathbf{M},w}(\neg\phi) = 1$ iff $V_{\mathbf{M},w}(\phi) = 0$
 $= 0$ iff $V_{\mathbf{M},w}(\phi) = 1$
- (iii) $V_{\mathbf{M},w}(\phi \rightarrow \psi) = 0$ iff $V_{\mathbf{M},w}(\phi) = 1$ and $V_{\mathbf{M},w}(\psi) = 0$
 $= 1$ iff $V_{\mathbf{M},w}(\phi) = 1$ and $V_{\mathbf{M},w}(\psi) = 1$,
or $V_{\mathbf{M},w}(\phi) = 0$ and $V_{\mathbf{M},w}(\psi) = 1$,
or $V_{\mathbf{M},w}(\phi) = 0$ and $V_{\mathbf{M},w}(\psi) = 0$
- (iv) $V_{\mathbf{M},w}(\Box\phi) = 1$ iff for every $w' \in W$ such that wRw' :
 $V_{\mathbf{M},w'}(\phi) = 1$
 $= 0$ iff there is a $w' \in W$ such that wRw'
and $V_{\mathbf{M},w'}(\phi) = 0$

The clauses for the other connectives and for \Diamond follow from the above clauses together with the definitions of those connectives in terms of \rightarrow and \neg , and the definition of \Diamond in terms of \neg and \Box .

The comments made above about the interpretations of constants and predi-

cates and about some formulas having undefined truth values have all found their way into this definition. But there still are some problems with it. It turns out that clause (iv), which deals with $\Box\phi$, is rather too strict. The requirements it places on sentences of the form $\Box\phi$, if these are to be true, are too stringent. According to this clause, any such sentence can be true in a world only if all of the constants occurring in it refer to entities which are present in all worlds accessible from that world. For otherwise the truth value of $\Box\phi$ will be undefined, since the truth value of ϕ will be undefined in some accessible world. This means, for example, that in our world it is necessary neither that Amsterdam is Amsterdam, nor that the sun rises if the sun rises—for Amsterdam and the sun might just as well not exist. That is to say, other possible worlds are accessible from our actual world in which Amsterdam and the sun do not exist. One possible remedy would be to relax clause (iv) so that $\Box\phi$ is true in a world w just in case ϕ is true in all worlds accessible from w in which ϕ has a truth value (that is to say, in all worlds which contain the entities referred to by the constants in ϕ). Clause (iv) for the truth value of $\Box\phi$ then becomes (iv'):

$$\begin{aligned} \text{(iv')} \quad V_{M,w}(\Box\phi) &= 1 \text{ iff } V_{M,w'}(\phi) = 1 \text{ for every } w' \in W \text{ such that } wRw' \\ &\quad \text{for which } V_{M,w'}(\phi) \text{ is defined} \\ V_{M,w}(\Box\phi) &= 0 \text{ iff } V_{M,w'}(\phi) = 0 \text{ for at least one } w' \in W \\ &\quad \text{with } wRw' \end{aligned}$$

Replacing (iv) in definition 2 with (iv') removes the above complications, but new ones rush in to take their places. For example, (39) becomes true in the actual world without (40) becoming true:

(39) It is necessary that Adam is a mortal.

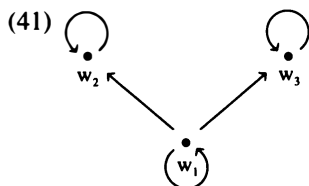
(40) Adam is a mortal.

Clause (iv') has the effect that not all worlds accessible from w need to be taken into account in assessing the truth value of a sentence of the form $\Box\phi$ there. This can lead to w itself not being taken into account, even if w is in fact accessible from itself. In other words, the reflexivity of R is no longer sufficient to guarantee the validity of the logical principle $\Box\phi \rightarrow \phi$. Ad hoc adjustments could be made in order to get this guarantee back, by explicitly requiring ϕ to be true in w if $\Box\phi$ is to be true in w (assuming reflexive R). We would then replace clause (iv') by (iv''):

$$\begin{aligned} \text{(iv'')} \quad V_{M,w}(\Box\phi) &= 1 \text{ iff } V_{M,w}(\phi) = 1 \text{ and } V_{M,w'}(\phi) = 1 \text{ for every } w' \in W \\ &\quad \text{such that } wRw', \text{ for which } V_{M,w'}(\phi) \text{ is defined} \\ V_{M,w}(\Box\phi) &= 0 \text{ iff } V_{M,w}(\phi) = 0 \text{ for some } w' \in W \text{ with } wRw' \end{aligned}$$

But this remains an ad hoc solution which leaves many other problems unsolved. Various other principles which were valid in modal propositional logic are in danger. One example: $\Box(\phi \wedge \psi) \rightarrow \Box\phi$ is no longer valid. For

$\Box(\phi \wedge \psi)$ to be true it is only required that $\phi \wedge \psi$ be true in all worlds which have in their domains the references of the constants occurring in either ϕ or ψ . There may well be a lot fewer of these worlds than there are worlds containing the references of the constants occurring in just ϕ . This can also be illustrated with reference to a concrete model. \mathbf{M} has as its set of possible worlds the set $\{w_1, w_2, w_3\}$, with the accessibility relation given in (41):



Let the domain function be as follows: $D_{w_1} = \{a, b\}$; $D_{w_2} = \{a, b\}$; $D_{w_3} = \{a\}$. The interpretation of a predicate letter A in this model is defined by: $I_{w_1}(A) = \{a, b\}$, $I_{w_2}(A) = \{a, b\}$, $I_{w_3}(A) = \emptyset$; and that of the constants c_1 and c_2 is defined by $I(c_1) = a$; $I(c_2) = b$. Now in \mathbf{M} we have $V_{w_1}(\Box(Ac_1 \wedge Ac_2)) = 1$, according to (iv'') (and in fact according to (iv') as well), since $V_{w_1}(Ac_1 \wedge Ac_2) = V_{w_2}(Ac_1 \wedge Ac_2) = 1$, while $V_{w_3}(Ac_1 \wedge Ac_2)$ is undefined. But on the other hand, $V_{w_1}(\Box Ac_1) = 0$, since $V_{w_3}(Ac_1) = 0$.

Problems like these would seem to suggest that a 'strong' interpretation of the connectives would be preferable to the 'weak' interpretation given them in definition 2. Under the weak interpretation, the truth value of a conjunction is undefined if that of either of its conjuncts is undefined, even if the other conjunct is false. Under the strong interpretation, on the other hand, a conjunction is false if either of its conjuncts is, even if the other conjunct is neither true nor false. Thus, we are in effect back with the problems of many-valued logic, as discussed in §5.5 in volume 1. This is not a coincidence. Although we dismissed many-valued logic earlier on as a serious modeling of *modality*, it is becoming increasingly important in current research on *partiality* of interpretation.

Returning to the example and recalculating the truth values under the strong interpretation of the connectives, we see that because $V_{w_3}(Ac_1) = 0$, $V_{w_3}(Ac_1 \wedge Ac_2) = 0$ instead of being undefined. As a result, we have $V_{w_1}(\Box(Ac_1 \wedge Ac_2)) = 0$ instead of $V_{w_1}(\Box(Ac_1 \wedge Ac_2)) = 1$. So this model is no longer a counterexample to the validity of $\Box(\phi \wedge \psi) \rightarrow \Box\phi$. It is quite possible that the approach sketched above, given these adjustments, results in a satisfactory semantics. But the truth definition still seems to lack a coherent intuitive backing. So we will mention a few of the other possibilities. One entirely different response to the problems with definition 2 is to impose the following restriction on the domain function:

$$(42) \text{ if } wRw', \text{ then } D_w \subseteq D_{w'}$$

During transitions to other accessible worlds, in other words, nothing at all may get lost. The domain may only grow larger. This restriction has the effect that the reference of any constant in a given possible world is also to be found in the domain of any world accessible from that world. So given (42), we can stick to the original clause (iv) in definition 2 without running into the complications we have just discussed. For given restriction (42), we can be sure that Amsterdam, and the sun are to be found in any possible world which is accessible from this one, so that the facts that Amsterdam is Amsterdam and that the sun rises if the sun rises all become necessary truths. Furthermore, (39) and (40) both become undefined, so that they no longer constitute a counterexample to $\Box\phi \rightarrow \phi$. And it can easily be seen that the problems we had with principles like $\Box(\phi \wedge \psi) \rightarrow \Box\phi$ are also to be solved in this manner.

It seems likely, however, that the *increasing domains requirement* has ramifications which are less than entirely acceptable. One of these is that it is no longer true that Amsterdam and the sun might just as well not have existed: they exist in every possible world which is accessible from this one. It should be noted that a related objection can be made to the clauses (iv') and (iv''): given that Amsterdam is present in our world, they imply that it is necessary that Amsterdam exists in our world, and so on.

3.3.2 Identity

Our decision to treat constants as rigid designators has quite spectacular consequences for the validity of certain principles concerning identity. We don't have much freedom in interpreting the identity predicate:

$$(43) \text{ in every world } w, I_w(=) = \{\langle d, d \rangle \mid d \in D_w\}$$

Given the truth definition and either the increasing domains requirement (42) or the modified clause (iv''), (44) will be a valid principle:

$$(44) b = c \rightarrow \Box(b = c)$$

As we mentioned in §3.2, the planet Venus happens to be the reference for two other proper names: *Hesperus* (the evening star) and *Phosphorus* (the morning star). It would seem that treating proper names as rigid designators inevitably leads to the conclusion that (45) is a necessary truth:

$$(45) \text{ Hesperus is Phosphorus.}$$

Kripke is willing to accept this conclusion. For him, the status of a proposition as necessary has nothing to do with how we come to recognize its truth but only follows from the fact that it could not have been otherwise than true. He distinguishes between necessary truths and a priori ones. A proposition is a priori if its truth status may be established purely by reasoning, indepen-

dently of (sensory) experience. This makes *a priori* an epistemological notion. A proposition is necessary, on the other hand, if it describes a situation which could not have been different from how it is. And this makes *necessary* an ontological notion. It has often been argued that the two notions coincide, but according to Kripke this is a mistake: there are in any case examples of necessarily true propositions which are not *a priori*. Sentence (45) is an example of just this. Although (45) is a necessary truth, its truth cannot be established independently of sensory experience. This distinguishes (45) from (46):

(46) Hesperus is Hesperus.

Sentence (46) is both necessary and *a priori*. Thus Frege's original problem with the difference between these two sentences can be adequately dealt with within the theory of rigid designation. As another example of necessarily true propositions which need not be *a priori*, Kripke mentions true but as yet unproved mathematical propositions. Either Goldbach's conjecture ("every even number larger than 2 is the sum of two primes") is an example of such a proposition, or its negation is, depending on which of the two is true. Whichever of the two happens to be true, it is a necessary truth. But since the truth of Goldbach's conjecture has not yet been decided either way, it is not a truth about which we (now) have any *a priori* knowledge. Perhaps we will have *a priori* knowledge about it some day, if someone ever succeeds in proving or disproving Goldbach's conjecture. But we have no guarantee that this can be done. So while it is a necessary truth, we may not conclude from this that it is also an *a priori* truth.

It seems that Kripke has given a satisfactory solution to the problems with proper names which we discussed in §3.2 (and see §3.3.4 for the problem about Pegasus not existing). It should, however, be noted that complications arise in treating belief contexts in a possible worlds semantics with rigid designation. If belief is analyzed as a relation between individuals and propositions, as has been proposed, and if proper names are interpreted as rigid designators, sentence (48) will, in view of (45), follow from (47):

(47) The Babylonians believed that Hesperus is Hesperus.

(48) The Babylonians believed that Hesperus is Phosphorus.

It is not at all clear whether we have here a problem with the theory of rigid designation or a problem with the proposed analysis of belief in possible world semantics (also see the comments at the end of §3.5).

Exercise 3

Use definition 2, with (iv'') instead of (iv). Consider a world in which Adam exists, but Eve does not. Determine the truth value in such a world of the following formulas:

(a) $\Box \exists x(x = \text{Adam})$

(b) $\Box \exists x(x = \text{Eve})$

What is the relationship between clause (iv'') and the increasing domains requirement?

3.3.3 Variables and Quantifiers

We encounter the same problem in the semantics of quantification as with constants: are quantifiers to range over individuals or over individual concepts? Once again our preference is for individuals. The reason for this is that natural language contains a kind of expression, namely, definite descriptions, which may be thought of as designating individual concepts. Now if the quantifiers were to range over individual concepts in modal predicate logic, principles (49) and (50) would end up valid:

(49) $\forall x\phi \rightarrow [\lambda y\psi/x]\phi$

(50) $[\lambda y\psi/x]\phi \rightarrow \exists x\phi$

If ϕ is true of all individual concepts, then it will most certainly be true of the concept $\lambda y\psi$ in particular, and if ϕ is true of the particular concept $\lambda y\psi$, then there is at least one individual concept for which ϕ is true. But these principles do not seem to hold for natural language. Here are some examples to illustrate this. Sentence (52) does not follow from (51):

(51) Everyone can lose this game.

(52) The winner can lose this game.

Nor does (55) follow from (53) and (54):

(53) All numbers greater than seven are necessarily greater than seven.

(54) The number of planets is a number greater than seven.

(55) The number of planets is necessarily greater than seven.

And finally yet another example which really belongs to deontic logic. Sentence (57) does not follow from (56):

(56) The president of the United States must have been born in the United States.

(57) There is someone who must have been born in the United States.

In the truth definition, then, we will be dealing with assignments g which assign an individual to each variable. The quantifiers will then range over

individuals all right, but it would seem only reasonable that in evaluating quantified sentences in any given world w , we take into account only the individuals in (the domain of) that world. We shall now replace definition 2 by a definition of valuations $V_{M,w,g}$ based on models M , worlds w in those models, and assignments g for the full language of modal predicate logic. First we define the interpretation of terms:

$$(58) \llbracket t \rrbracket_{M,g} = \begin{array}{ll} I(t) & \text{if } t \text{ is a constant} \\ g(t) & \text{if } t \text{ is a variable} \end{array}$$

The full definition now reads as follows:

Definition 3

Let M be a model, ϕ a formula of modal predicate logic, and $w \in W$. Then $V_{M,w,g}(\phi)$, the truth value of ϕ in w given M , is defined as follows:

- (i) $V_{M,w,g}(Pt_1 \dots t_n) = 1$ iff $\llbracket t_1 \rrbracket_{M,g} \in D_w, \dots, \llbracket t_n \rrbracket_{M,g} \in D_w$
and $\langle \llbracket t_1 \rrbracket_{M,g}, \dots, \llbracket t_n \rrbracket_{M,g} \rangle \in I_w(P)$
 $= 0$ iff $\llbracket t_1 \rrbracket_{M,g} \in D_w, \dots, \llbracket t_n \rrbracket_{M,g} \in D_w$
and $\langle \llbracket t_1 \rrbracket_{M,g}, \dots, \llbracket t_n \rrbracket_{M,g} \rangle \notin I_w(P)$.
- (ii) $V_{M,w,g}(\neg\phi) = 1$ iff $V_{M,w,g}(\phi) = 0$
 $= 0$ iff $V_{M,w,g}(\phi) = 1$.
- (iii) $V_{M,w,g}(\phi \rightarrow \psi) = 0$ iff $V_{M,w,g}(\phi) = 1$ and $V_{M,w,g}(\psi) = 0$
 $= 1$ iff $V_{M,w,g}(\phi) = 1$ and $V_{M,w,g}(\psi) = 1$,
or $V_{M,w,g}(\phi) = 0$ and $V_{M,w,g}(\psi) = 1$,
or $V_{M,w,g}(\phi) = 0$ and $V_{M,w,g}(\psi) = 0$.
- (iv) $V_{M,w,g}(\Box\phi) = 1$ iff for every $w' \in W$ such that wRw' :
 $V_{M,w',g}(\phi) = 1$
 $= 0$ iff there is a $w' \in W$, such that wRw'
and $V_{M,w',g}(\phi) = 0$.
- (v) $V_{M,w,g}(\forall x\phi) = 1$ iff for every $d \in D_w$: $V_{M,w,g[x/d]}(\phi) = 1$
 $= 0$ iff there is a $d \in D_w$ such that
 $V_{M,w,g[x/d]}(\phi) = 0$

The clauses for the other connectives and for \Diamond and \exists follow from the above clauses together with the definitions of those connectives in terms of \rightarrow and \neg , and the definition of \Diamond in terms of \Box and \neg , and of \exists in terms of \forall and \neg . As in standard logic, the truth values of sentences, i.e., formulas lacking free variables, are independent of the assignment g . So for sentences ϕ we can just write $V_{M,w}(\phi)$. If this truth definition were to be applied without assuming growing domains (or without replacing clause (iv) by a suitably adapted version of (iv'')), then we would run into even stickier problems with formulas with undefined truth values than we have already encountered. For if any of the entities in a world w is lacking in any world accessible from w , then $V_{M,w}(\forall x\Box\phi)$ would always be undefined there. We will therefore assume that

either the requirement of increasing domains has been met, or that clause (iv) has been replaced by a suitably adapted version of (iv").

In the semantics of modal predicate logic, a good deal of attention has been paid to the interactions between modal operators and quantifiers. This has brought out in full relief the distinction between modalities *de dicto* and *de re*. In studying the interactions between \Box and the quantifiers, the validity of the following four principles can be checked:

$$(59) \Box \forall x \phi \rightarrow \forall x \Box \phi$$

$$(60) \forall x \Box \phi \rightarrow \Box \forall x \phi$$

$$(61) \Box \exists x \phi \rightarrow \exists x \Box \phi$$

$$(62) \exists x \Box \phi \rightarrow \Box \exists x \phi$$

The most well known of these schemata is (60), the *Barcan formula*. It is named after Ruth Barcan, who pointed out that it is problematic. An equivalent with \Diamond instead of \Box can be obtained for each of these formulas by means of contraposition (that is, by means of the equivalence of $\psi \rightarrow \chi$ and $\neg \chi \rightarrow \neg \psi$):

$$(63) \exists x \Diamond \phi \rightarrow \Diamond \exists x \phi$$

$$(64) \Diamond \exists x \phi \rightarrow \exists x \Diamond \phi$$

$$(65) \forall x \Diamond \phi \rightarrow \Diamond \forall x \phi$$

$$(66) \Diamond \forall x \phi \rightarrow \forall x \Diamond \phi$$

As an example we will show how (63) may be derived from (59). $\Box \forall x \phi \rightarrow \forall x \Box \phi$ implies $\neg \forall x \Box \phi \rightarrow \neg \Box \forall x \phi$, and thus $\exists x \neg \Box \phi \rightarrow \Diamond \neg \forall x \phi$, and thus $\exists x \Diamond \neg \phi \rightarrow \Diamond \exists x \neg \phi$. But if the schema $\Box \forall x \phi \rightarrow \forall x \Box \phi$ holds for all formulas ϕ , then it holds in particular for all formulas $\neg \phi$. So $\exists x \Diamond \neg \neg \phi \rightarrow \Diamond \exists x \neg \neg \phi$, which immediately implies $\exists x \Diamond \phi \rightarrow \Diamond \exists x \phi$.

Unlike (60) and (64), (59) and (63) would seem to be valid. It is clear, for example, that (67), which may be symbolized as $\exists x \Diamond \phi$, is a much stronger statement than (68), which may be symbolized as $\Diamond \exists x \phi$; in other words, it is clear that (67) implies (68):

(67) There is someone who will possibly do better than I.

(68) It is possible that someone will do better than I.

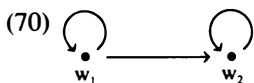
That (68) is weaker than (67) is, for example, apparent from the fact that (68) can be followed by (69), whereas this is absurd for (67):

(69) But it is improbable that there is someone who will do better than I.

This comes out quite nicely in the semantics which we have given: (63), and thus (59), is valid, while (64), and thus the Barcan formula, (60), is invalid.

The validity of $\exists x \Diamond \phi \rightarrow \Diamond \exists x \phi$ may be proved as follows: If $V_{M,w,g}(\exists x \Diamond \phi) = 1$, then $V_{M,w,g[x/d]}(\Diamond \phi) = 1$ for some $d \in D_w$. So then $V_{M,w',g[x/d]}(\phi) = 1$ for some w' with wRw' . In this w' , we then have $V_{M,w',g}(\exists x \phi) = 1$, so that $V_{M,w,g}(\Diamond \exists x \phi) = 1$.

That $\Diamond \exists x \phi \rightarrow \exists x \Diamond \phi$ is not valid for every formula ϕ is clear from the counterexample given to $\Diamond \exists x Ax \rightarrow \exists x \Diamond Ax$ in (70), in which $D_{w_1} = \{a\}$, $D_{w_2} = \{a, b\}$; $I_{w_1}(A) = \emptyset$, $I_{w_2}(A) = \{b\}$.



For while, on the one hand, it is clear that $V_{M,w_2}(\exists x Ax) = 1$, and thus $V_{M,w_1}(\Diamond \exists x Ax) = 1$, on the other hand, taking $g(x) = a$, we see that $V_{M,w_1,g}(\Diamond Ax) = 0$, and thus $V_{M,w_1}(\exists x \Diamond Ax) = 0$, since a is the only value of $g(x)$ which is to be found in w_1 .

Principle (61) and its equivalent (65) are extremely implausible. According to (65), for example, (72) follows from (71):

(71) Everyone can win in this game.

(72) It is possible that everyone wins in this game.

This is clearly not as it should be, and (61) and (65) do indeed turn out to be invalid in the semantics we have given. At first sight, (62) and (66), the converses of (61) and (65), would seem much more plausible. It is worth noting that accepting the validity of (62) and (66) practically amounts to accepting the increasing domains requirement. This can be shown quite easily. Suppose that the domains in a model M all increase, and that $V_{M,w,g}(\exists x \Box \phi) = 1$. Then there must be a $d \in D_w$ such that $V_{M,w,g[x/d]}(\Box \phi) = 1$. Suppose now that wRw' , that is to say, that w' is accessible from w . Then because increasing domains are required, we have $d \in D_{w'}$ and $V_{M,w',g[x/d]}(\phi) = 1$, so that $V_{M,w',g}(\exists x \phi) = 1$. As this argument applies to all worlds w' which are accessible from w , we now have $V_{M,w,g}(\Box \exists x \phi) = 1$. Thus (62) has been shown to be valid on this model.

That (62) and (66) are not valid if the increasing domains requirement is not satisfied may be seen as follows. Let M be a model which does not satisfy the increasing domains requirement. Then there are two possible worlds, w and w' , such that wRw' , and an entity d such that $d \in D_w$ and $d \notin D_{w'}$. Now (62) can be falsified in w by letting ϕ be such that d satisfies ϕ in w and in every world accessible from w and having d in its domain, while no entity at all in w' satisfies ϕ . For given any such ϕ and assuming it has a free variable x , we have $V_{M,w,g}(\exists x \Box \phi) = 1$, since according to clause (iv''), $V_{M,w,g[x/d]}(\Box \phi) = 1$, but $V_{M,w,g}(\Box \exists x \phi) = 0$, since $V_{M,w',g}(\exists x \phi) = 0$. That is, we have $V_{M,w,g}(\exists x \Box \phi \rightarrow \Box \exists x \phi) = 0$.

Examples such as these point at a more general correspondence between proposed principles of a modal predicate logic and structural conditions on

frames of possible worlds with domains of individuals attached. This theme can be pursued just as in propositional modal logic (cf. §2.3.2), but we shall not do so here.

As plausible as (62) and (66) might seem at first sight, it still is possible to imagine situations in which one, in violation of (66), would want to accept (72) without automatically having to accept (71). It is conceivable that there is a game in which everyone can win (that is, in which everyone can end with exactly the same score), at least under the proviso that Jones, who is a particularly poor player, does not participate. But this need not mean that everyone can win the game, because Jones will always be beaten. It will be clear that this matter can only be settled by a thorough analysis in which the modal, temporal, epistemic, and deontic aspects of intensionality are all taken into account.

By way of conclusion, it is worth returning briefly to an option we rejected earlier on, namely, letting the quantifiers range over individual concepts instead of over individuals. This option does extremely badly in the test posed by the principles discussed above: they all turn out to be valid, which is enough to remove completely the distinction between modalities *de dicto* and *de re*.

Exercise 4*

- (a) Prove that (65) is invalid.
- (b) Show that accepting the Barcan formula (64) implies assuming decreasing domains, i.e., assuming that if wRw' , then $D_{w'} \subseteq D_w$.

3.3.4 One Domain: The Existence Predicate

In the semantics for modal predicate logic given in the last few sections, we have not assumed that all formulas must have truth values in all possible worlds. This gave rise to various problems with the validity of modal principles, some of which we have seen are solvable. But it is not yet clear whether any such problems still remain, and that is a rather unsatisfactory state of affairs. Furthermore, arguments of a principled nature may be leveled against the idea of formulas which are neither true nor false, since it violates the principle of bivalence.

It is possible to set things up in such a way that the principle of bivalence is maintained, which means that every atomic formula $Pc_1 \dots c_n$ must be assigned a truth value in every possible world. One obvious way of doing this is to make sure that all individual constants have references in all possible worlds. This amounts to creating a single, common domain for all possible worlds. From the point of view of any one of these worlds, then, there are two kinds of things: there are the things which really exist in that world, but there are also the things which really exist in any of the other worlds. In each world, the membership of predicates will have to be defined for all entities in that

world (even for the ones which don't really exist there). This would seem to be a fairly radical solution to all of the problems with undefined truth values.

Within this approach, it would seem desirable to have some way of distinguishing between individuals which really exist in any given world and all of the other things there. We do this by means of an *existence predicate* E . This predicate singles out in every world the possible individuals which really exist there. Introducing this predicate provides a ready solution to some problems with Pegasus's nonexistence which we mentioned earlier on. It is not at all clear how a sentence like (73):

(73) Pegasus does not exist.

could ever be true in the rigid designators approach. The natural translation of (73), $\neg\exists x(p = x)$, turns out to be false or undefined in all variants of the semantics given above. But using the existence predicate, (73) can be rendered as $\neg Ep$, a formula which is true in just those worlds in which p refers to a nonexistent individual. Nor does a semantics with a common domain and existence predicate suffer from the handicap of all the other systems we have seen, that an object exists necessarily if it exists at all.

It is quite possible to find philosophical objections to a domain containing all possible individuals, or to analyzing existence as a predicate. But as we pointed out in §3.1, philosophical considerations may not be allowed to have the last say if our aim is the application of logical methods in the description of natural language. And sentences like (73), in which the existence of certain kinds of individuals is asserted or denied, are not the only ones for which the availability of nonexistent individuals would seem to be an advantage. Consider (74), for example:

(74) John is talking about Pegasus.

Obviously the truth of this sentence may not be allowed to depend on the existence of Pegasus. Now the natural analysis of (74) is one in which *is talking about* is seen as a relation between John and Pegasus. But this means that nonexistent entities must be allowed to enter into relations, if (74) is ever to be true.

This approach leads to the following alternative to definition 1 as the definition of a model:

Definition 4

A *model* for a modal predicate-logical language L consists of:

- (i) a nonempty set W of possible worlds
- (ii) an accessibility relation R on W
- (iii) a domain D
- (iv) an interpretation function I which assigns
 - (a) an element $I(c)$ of D to each constant c in L

- (b) a nonempty subset $I_w(E)$ of D to E , for each world $w \in W$
- (c) the same set $\{\langle d, d \rangle \mid d \in D\}$ to $=$, for each world w
- (d) a subset $I_w(P)$ of D^n to each n -ary predicate letter P in L , for each $w \in W$

This definition leads to the following alternative to the truth definition given as definition 3 in §3.3.3:

Definition 5

Let M be a model, $w \in W$, and g an assignment. Then $V_{M,w,g}(\phi)$, the truth value of ϕ in w given M , is defined as follows:

- (i) $V_{M,w,g}(Pt_1 \dots t_n) = 1$ iff $\langle \llbracket t_1 \rrbracket_{M,g}, \dots, \llbracket t_n \rrbracket_{M,g} \rangle \in I_w(P)$
- (ii) $V_{M,w,g}(\neg\phi) = 1$ iff $V_{M,w,g}(\phi) = 0$
- (iii) $V_{M,w,g}(\phi \rightarrow \psi) = 1$ iff $V_{M,w,g}(\phi) = 0$ or $V_{M,w,g}(\psi) = 1$
- (iv) $V_{M,w,g}(\Box\phi) = 1$ iff for each w' such that wRw' : $V_{M,w',g}(\phi) = 1$
- (v) $V_{M,w,g}(\forall x\phi) = 1$ iff for all $d \in D$: $V_{M,w,g[x/d]}(\phi) = 1$

The Barcan formula is true in any such model with a common domain, since it satisfies the decreasing domains requirement mentioned in exercise (4b). The objections which were originally mentioned in connection with the Barcan formula, however, are no longer valid, since now the quantifiers range not only over existing individuals but also over possible individuals. The original reading of the Barcan formula can, with the help of the existence predicate, be reconstructed as:

$$(75) \quad \forall x(Ex \rightarrow \Box\phi) \rightarrow \Box\forall x(Ex \rightarrow \phi)$$

Restricting the quantifier to E in (75) has the effect of having it range over existing individuals. But unlike the original Barcan formula, (75) is not universally valid under truth definition 5.

In fact, the quantifiers in most representations of natural language sentences of the form *all A are B* and *some A are B* will end up being restricted to E . That is, these sentences will be represented as $\forall x(Ex \rightarrow (Ax \rightarrow Bx))$ and $\exists x(Ex \wedge (Ax \wedge Bx))$, respectively. So the validity of the Barcan formula has no adverse affects on the applicability of this semantics. Another rather slippery principle which no longer holds if its quantifiers are restricted to E is (59), the converse of the Barcan formula. And rightfully so. For (76) would otherwise lead to (77) after substitution of Ex for ϕ , which would in turn imply (78):

$$(76) \quad \Box\forall x(Ex \rightarrow \phi) \rightarrow \forall x(Ex \rightarrow \Box\phi)$$

$$(77) \quad \Box\forall x(Ex \rightarrow Ex) \rightarrow \forall x(Ex \rightarrow \Box Ex)$$

$$(78) \quad \forall x(Ex \rightarrow \Box Ex)$$

And with (78) we would once again be confronted with the specter of all the individuals that exist in a world existing necessarily.

Exercise 5*

- (a) Show that (75) is not valid.
 (b) Under what condition will (78) be valid in a model?

Exercise 6

Give an example of an expression which can be viewed as a one-place predicate P for which the requirement for all w : $I_w(P) \subseteq I_w(E)$ is not correct. And can you think of an expression which must be considered a two-place relation Q for which the requirement for all w : $\{x \mid \text{there is a } y \text{ such that } \langle x, y \rangle \in I_w(Q)\} \subseteq I_w(E)$ is not correct?

3.4. Other Kinds of Contexts

In the above exposition of intensional predicate logic, we have concentrated almost exclusively on modal predicate logic. But other intensional operators, for example, tense operators, can also be added to predicate logic. The whole above discussion about how constants and quantifiers are to be interpreted and on the choice of domains may then be repeated more or less as it is.

Principles (59)–(62) discussed in §3.3.3 all have tense-logical variants. The following two formulas are, for example, the tense-logical versions of the Barcan formula (60), for future and past, respectively:

$$(79) \quad \forall x G\phi \rightarrow G\forall x\phi$$

$$(80) \quad \forall x H\phi \rightarrow H\forall x\phi$$

As they stand, we are not inclined to accept the validity of (79) and (80) any more than that of the Barcan formula. And the same applies, *mutatis mutandis*, to the other tense-logical principles corresponding to (59), (61), and (62). In adopting moments in time as our contexts, we once again encounter various alternative ways of setting up the semantics. We can give each moment in time its own domain, or we can introduce a single domain which is common to all moments in time. The advantages and disadvantages of the different alternatives are the same as with modal logic. And the status of principles like (79) and (80) in the various alternatives is also analogous to the status of the corresponding modal principles. The considerations with respect to how individual constants and variables are to be interpreted are also the same as in modal logic.

Exercise 7

- (a) Translate the following two sentences into the language of temporal predicate logic:
 (i) One day everybody will be happy forever after.
 (ii) There is always someone who is happy only when someone else isn't.
 Show that if we assume that each point in time is succeeded by another

one and that the 'earlier than' relation is transitive, then there cannot be a model *M* and a point in time *t* in which (i) and (ii) are both true.

- (b) Consider the existence predicate *E*. Is it reasonable to lay down special requirements on *I(E)* in a temporal predicate logic? If so, what should these be?

When dealing with both modal and tense operators at the same time, our contexts will become worlds at moments in time, as was explained in §2.5. Different alternative interpretations for the modal operators are then available; as far as tense operators are concerned, the combination with modal operators does not seem to introduce any new possibilities into the discussion. In chapter 5 we will encounter an intensional logical system with both modal and tense operators and in which quantification ranges over a single domain. It is, then, a semantics along the lines of §3.3.4.

Other extensions of intensional logic may be obtained by including yet more different factors in the contexts. The contexts we have seen until now—possible worlds, moments in time, and the combination of the two—may also be used in the analysis of epistemic and deontic expressions. But these are not the only kinds of expressions with context-dependent interpretations. Personal pronouns like *I*, *he*, *we*, and *you*, locative expressions like *here* and *there*, and deictic pronouns like *this* and *that* are other examples. Let us consider the pronouns *I* and *you*. Just as the truth status of (81) depends on the time of the context, the truth values of (82) and (83) are dependent on the context:

(81) It is raining.

(82) I live in Amsterdam.

(83) You work in Paris.

Sentence (82) is true in a given context just in case the individual who is speaking in that context does in fact live in Amsterdam. And (83) is true in a given context just in case whoever is being addressed in that context does in fact work in Paris. The point is that in order to represent this, the contexts need more than just a world and a time: they need a speaker and someone who is being spoken to, an addressee. Locative expressions like *here* and *there* require further extensions to the notion of context: in order to interpret (84) the contexts will also have to specify a place:

(84) I live here.

(Actually a sentence like (81) is just as dependent on place.) We can imagine a context as a sequence of parameters: a possible world *w*, a moment *t* in time, a speaker *s*, an addressee *a*, and a place *p*. Of course there are always other specifications which could be added to this list. The parameters speaker, addressee, and place are often treated within the framework of worlds at moments in time. A distinction is then drawn between *indices*, or worlds at moments in time, and *contexts of use*, which specify speakers, addressees,

and places. It thus becomes possible to draw a line between expressions whose interpretation differs from index to index and those whose interpretation depends on the context of use. The first to systematically explore this distinction was Kaplan (1978, 1979). Some authors, notably Cresswell, object to the 'open' character of the context notion. According to them, there is no end to the possible parameters. They therefore prefer an approach in which contexts are unanalyzed concepts which can have certain kinds of properties. Instead of specifying a context as (among other things) an individual *a* who is the speaker and an individual *b* who is the addressee, they would just say that there is a context which has the property of containing a speaker *a* and an addressee *b*. The truth status of a sentence like (82) in any such context is then no longer dependent on any parameter in that context but rather on the properties of that context. It is not yet clear whether this approach leads to results other than the original one.

By way of illustration, we will spend the rest of this section on different ways of representing first- and second-person pronouns. A first attempt to account for the contextual character of *I* and *you* might go something like this. A model *M* consists of a set *W* of possible worlds with an accessibility relation *R*, a set *T* of moments in time with an *earlier than* relation $<$, a domain *D* of entities, and an interpretation function *I*. A context *k* is to be considered as a sequence $\langle w, t, s, a \rangle$ in which $w \in W$, $t \in T$, and $s, a \in D$. (There is no need to ensure that *s* and *a* are different elements of *D*.) So a context *k* fixes a possible world *w*, a moment in time *t*, a speaker *s*, and an addressee *a*. The symbols *s* and *a* must not be thought of as the names of individuals; they are 'metanames'. A context would look something like $\langle w_1, t_8, d_2, d_9 \rangle$. We wish to represent the expressions *I* and *you* in our formal language by means of individual constants. We reserve the constants *i* and *y*, respectively, for this purpose. These constants can, of course, not be treated as rigid designators, for the idea was precisely the opposite: their references were to essentially depend on the context in which they are interpreted. The expressions *I* and *you* are not names for particular individuals anyway. So we will make an exception to the rule that constants are rigid designators for *i* and *y* by requiring the interpretation function *I* in a model to comply with the following:

$$I_k(i) = s_k \text{ and } I_k(y) = a_k, \text{ for all } k$$

So given a context *k*, *I* assigns the speaker in *k*, *s_k*, to the constant *i*, and the addressee in *k*, *a_k*, to the constant *y*. The interpretation function *I* interprets *i* and *y*, in other words, as *individual concepts*. An individual concept, we recall, is a function from contexts to individuals. The interpretation of *i* may be seen as the concept of a speaker and that of *y* as the concept of an addressee. Instead of $I_k(i) = s_k$, we could also write $I(i)(k) = s_k$, which is to be read as: The individual concept *I(i)* which *I* assigns to *i* assumes, when applied to the context *k*, the value *s_k*.

This approach accounts for the contextual character of *I* and *you*, and thus of sentences like (82) and (83). But it is still rather naive in a number of respects. Consider sentences (85) and (86):

(85) I am the speaker.

(86) I shall always be the speaker.

Even if these sentences are a little strange, they illustrate a general point which we want to make. Sentence (85) expresses a necessary truth: *I* refers, in every context, to whoever is the speaker in that context. So (85) can never be false. Sentence (86), on the other hand, can most certainly be false. Even though (85) is always true, (86) is certainly false in any realistic situation. But in the approach we have just sketched, they both turn out to be true in all contexts. Sentences (85) and (86) may be represented as (87) and (88), respectively:

(87) Si

(88) GSi

Here the predicate *S* functions as a representation of *is the speaker*. The interpretation of *S* in this approach is of course fixed: *S* has just a single individual in its extension in any context *k*, namely s_k . That is:

$I_k(S) = \{s_k\}$, for all *k*

Now it can easily be checked that (85) is true in every context *k*: we have $I_k(i) = s_k$ for all *k*, and so $I_k(i) \in I_k(S)$ for all *k*. This means that $\forall_{M,k}(Si) = 1$, for all *k*. But it also means that (88) is true in every context *k*. For (88) can only be false in a context *k* if there is some context *k'* in the future of *k* (that is, for which $t_k < t_{k'}$) in which *Si* is false. But we have seen that *Si* is true in every context, so there never can be any such *k'*. So the results this approach gives are consistent with the meanings of sentences like (85) but certainly not with those of sentences like (86). And these are not the only problematic sentences. The request expressed by sentence (89), for example, becomes extremely difficult to comply with:

(89) Would you remind me tomorrow that I must phone Mary?

Possibly a solution is to be found along the following lines. What (86) expresses is that whoever is now the speaker (*I*) will at all stages in the future also be the speaker. The personal pronoun *I* thus exhibits the same kind of behavior as the temporal modifier *now*, which refers to the time at which the sentence was uttered even if it occurs in an embedded clause (see §2.4.2). In exactly the same way, *I* refers back to the individual who is now, at the time of the utterance, the speaker. This is why (86) can be false even though (85) is always true. There are different ways of accounting for this aspect of the semantic behavior of *I*, (and of course of *you*). One possibility is the following. We restrict our contexts to moments in time (the possible worlds parameter is of no importance in this context). Instead of including *s* and *a* as parameters in our contexts, we add two functions *s* and *a* to the model. These functions say who the speaker is and who the addressee is at each moment in time. And then we introduce a fixed moment in time t_0 , just as we did when dealing with *now*

in §2.4.2. The only problem is that various moments in time may function as the moment of utterance, so it would seem advisable to generalize the method given in §2.4.2. We then obtain the method of *double indexing* developed by Kamp (1971). The idea is that sentences are to be interpreted with respect to two moments in time, t and t' . The first is called the moment of utterance and the second is called the moment of evaluation. We begin to evaluate a statement made at time t with respect to this time t . But the evaluation process may lead us to consider moments in time other than t . An example: in order to determine whether $P\phi$ is true in a given model at time t , we must try and find a moment earlier than t at which ϕ is true. Now if ϕ contains expressions like *now* and *I* which necessarily refer back to the original t , then we will need some way of keeping track of this t . We do this by always considering two moments in time. Valuations are then of the following form: $V_{M,t,t'}(\phi)$ is the truth value of ϕ in M at t' , given t as the moment of utterance. We will not develop all of the details of this method of interpretation here but will make do with an indication of how it solves the above problem with (85) and (86). The interpretation function I will also work with two moments t and t' . For almost all expressions, only the moment of evaluation is of any importance, but for some, like i and y , it is only the moment of utterance which counts:

$$\begin{aligned} I_{t,t'}(i) &= s(t) \\ I_{t,t'}(S) &= \{s(t')\} \end{aligned}$$

The interpretation of the personal pronoun *I* is then: the speaker at the moment of utterance; and the interpretation of the predicate S is: the set of all individuals who are the speaker at the moment of evaluation. Under this interpretation, (87), the representation of (85), is, as required, always true. But this no longer implies that (88), the representation of (86), must always be true. We can construct a model in which (87) is always true but in which (88) is false at at least one moment in time.

We have by no means exhausted the interesting puzzles arising from the semantic behavior of *I* and *you*. The above was only intended as an illustration of how one of these may be dealt with. But it also shows the importance of a certain flexibility with respect to the available semantic apparatus. It occasionally happens that ideas, like our original ideas about the structure of contexts, translate directly to other realms and new phenomena. But it doesn't happen very often.

3.5 A Methodological Note

In spite of its comparatively recent origins in the late fifties and early sixties, intensional logic has become one of the most important fields in philosophical logic. It has also proved a worthy tool in the semantics of natural language. And yet the status of intensional logic is not unassailable. We wish to con-

clude this chapter with some of the objections which have been made to it, and which are still being made.

These objections are of two main kinds. Some are objections of *principle*, objections of a philosophical and methodological nature. Others are of an *empirical* nature and concern the limits to the applicability of intensional semantics. We will not be able to do full justice to all of the objections we shall mention in this short section, so the reader is referred to the literature for a more complete account.

The objections of principle leveled against intensional logic are usually aimed at its conceptual apparatus. Possible worlds and possible individuals are the main targets. It is argued that these concepts are fundamentally obscure. We do not know exactly what a possible world is, and we have no way of finding out. The notion of a possible world is purely metaphysical, it is stated, and completely lacks an empirical content. Everything which is analyzed in terms of it—the concept of intensionality, the concepts of necessity and possibility, the modalities *de dicto* and *de re*, and many more—consequently remain as obscure as they were in the first place. Although it may seem as though they have been clarified, it is argued, intensional logic really only succeeds in substituting one murky notion for another. Moreover, intensional notions are supposed to lead to dubious philosophical positions. For some, as we mentioned in §3.1, recognizing modalities *de re* amounts to embracing essentialism. And intensional semantics is sometimes thought to be accompanied by more problems than it solves. Of these, the ‘transworld identity problem’ is one of the more notorious (see §3.2). This problem of determining whether an entity in one world is the same as an entity in another is supposed to overshadow the problems intensional logic solves. Similar doubts are held about possible worlds. How many of these are there? What counts as a possible world and what does not?

The reactions to these objections in the literature are quite divergent. Kripke’s reaction to the problem about what possible worlds are is, for example, as follows. Kripke argues that it is incorrect to think of possible worlds as things which we can discover anything about by observation. They must rather be thought of in epistemological terms, as being determined by the descriptive terms we associate with them. Among these terms we have, for example, counterfactuals like *If Dukakis had won the presidential elections in 1988, then the president of the United States would have been a Democrat* (cf. (6) in §1.7). Such sentences have the effect of introducing a possible world, one in which Dukakis wins the presidential elections in 1988. So according to Kripke, possible worlds are not things which we can discover; they are things which must be introduced, which must be *stipulated*. From this point of view, the transworld identity problem is not a genuine problem: we never need to find out whether an individual in a given world is the same as an individual in another world (so it doesn’t matter that we can’t perform this feat). We never need to find out because the only worlds we ever deal with are

ones which we have introduced as containing one or more of the individuals that are to be found in this world or in another possible world already introduced. The individuals may be quite different in the stipulated worlds from what they are in the real world, but we never need to worry about their identities.

Lewis's ontological interpretation is diametrically opposed to this epistemic interpretation of Kripke. Lewis defends a purely realistic conception of possible worlds. For him, possible worlds exist just as much as the actual world does. The actual world occupies no exceptional place in Lewis's view; it is just one of many possible worlds, just as this instant is one of many possible moments in time. This throws a wholly different light on the transworld identity problem. That individuals in different possible worlds may be identical to one another is out of the question. Instead of the relation of identity, then, Lewis introduces his *counterpart* relation. Two individuals in different worlds can resemble each other so closely that they are each other's counterparts in their respective worlds.

Another way of dealing with the objections to intensional logic is of a more methodological nature. Every theory, it is argued, makes use of concepts which it leaves unanalyzed as primitives, and the primitive concepts used in intensional semantics are the notions of a possible world and a possible individual. The content of the primitive concepts of theory is wholly determined by the role which they play in that theory. In the case of intensional logic, this amounts to the following. Take modal propositional logic, for example. One of the things we want in semantics is a way of accounting for the intuitive validity of $\Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi)$ as a principle of inference. It turns out that this can be accounted for by assuming a set W and then setting up truth definitions such that truth is relative to the elements of this set. For with the usual truth definition for formulas of the form $\Box\phi$, the principle is indeed valid. And imposing a particular structure on this set W by means of a relation R renders yet other principles valid as well. At no point in this whole story do we ever have to take into account the nature of W (or R). We needn't even think of them as a set of possible worlds and an accessibility relation. Any set of objects and any relation with the right properties would do just as well. So as far as the semantic theory of intensional logic is concerned, on this view, the meaning of the elements of W need go no further than the role which they play as parameters in the truth definition given by this theory. The philosophical discussion of the nature of possible worlds may then be seen as a part of the general debate in the philosophy of science about the role of theoretical terms in science.

From the point of view of the applications of intensional semantics in natural language research, these objections may or may not be correct, but they are irrelevant either way. For as we argued in §3.1, philosophical argumentation is out of place if the main interest is the description of natural language. After all, what we are trying to describe is not how we should speak if we want to

please philosophers but how we do in fact speak. The objections made to intensional logic are not so much refuted from this linguistic perspective as simply laid aside.

A final comment which can be made in connection with these objections of principle is that it is at least debatable whether positions like essentialism are really as objectionable as they have been alleged to be by some philosophers. Others are of the opinion that they are interesting and fertile philosophical doctrines and see intensional logic as a suitable aid to further investigations into them.

So much for the philosophical and methodological objections. The other objections which have been made are of a more empirical kind, and we will briefly discuss two of them. The first concerns whether a complete and empirically adequate semantic theory can be developed on the basis of intensional logic. One serious problem in this connection is the existence of *hyperintensional contexts*. These differ from normal intensional contexts in that even logically equivalent expressions cannot always be substituted into them *salva veritate*. The context created by \Box is an example of a normal intensional context. If ϕ and ψ are logically equivalent formulas (and thus have the same truth values in all worlds), then $\Box\psi$ follows from $\Box\phi$. The context created by *belief*, on the other hand, may be said to be hyperintensional. Even if ϕ and ψ are logically equivalent, the truth of *John believes that ϕ* does not imply the truth of *John believes that ψ* , since it may well be that John is not aware of their equivalence. The same point can be made using rigid designators. Assuming that the verb *believe* creates a normal intensional context, and given that *Hesperus* and *Phosphorus* are rigid designators which refer to the same entity, it would follow from *John believes that Hesperus is Hesperus* that *John believes that Hesperus is Phosphorus*, which is clearly unrealistic. So intensional semantics clearly runs into complications when applied to the verb *believe*.

No consensus has been reached on how to get around this. It has been proposed that the solution lies in a more refined intensional semantics. The above examples indicate that more than just logical equivalence, that is to say, equality of intension, is required for interchangeability *salva veritate* in hyperintensional contexts. Apparently expressions need to have more semantic properties in common than just the property of having the same reference in all possible worlds. Perhaps the ways in which the intensions of expressions are built up from the intensions of their composite parts should also be taken into account.

It has also been proposed that the hyperintensional contexts lie beyond the limits of (intensional) semantics and that a satisfactory solution will mean getting beyond these limits. It is argued that semantics must join forces with pragmatics in order to give an adequate treatment of hyperintensional contexts like that created by the verb *believe*. The relations between language and language users can to a large extent be abstracted away in semantics, but not

entirely, and the analysis of belief contexts is thought to be one area in which the semantic interpretation must take language users into account.

Be that as it may, if hyperintensional contexts should lie at or beyond the limits of intensional semantics, that would in no way diminish its utility in research into the semantics of natural language. Moreover, even with the proposed refinements or extensions added, intensional semantics would still have an essential part to play.

Other doubts about the empirical adequacy of intensional semantics concern the extent to which the notion of intension is an adequate explication of the concept of meaning. Given a mentalistic approach to meaning, it would seem to be a mistake to equate intension and meaning. The intension of an expression is a function which indicates its reference in various contexts. But familiarity with the meaning of an expression is not always enough to enable one to determine its reference. So intension and meaning cannot simply be equated. Therefore some see in the notion of intension an explication of the semantic competence of an ideal language user. Others distinguish between an individual psychological component of the concept of meaning and a social one. In this way, the notion of intension that originated in intensional semantics can function as an abstract explication of the function of language as the communal instrument by means of which a language community can speak about the world. These remarks do not diminish the usefulness of intensional semantics any more than the first lot did. They do, however, call for a little modesty. The phenomenon of language has facets which lie beyond the reach of intensional semantics.

4

The Theory of Types and Categorical Grammar

4.1 Introduction

The subject of this chapter is the *theory of types*, a more powerful logical system than standard predicate logic. It can be considered a further extension of second-order logic, which is a logical system in which the quantifiers are allowed to range not only over individuals, but also over properties (see vol. 1, chap. 5 for an introduction). Section 4.2 contains an exposition of the theory of types and comments on its applicability in linguistics. Section 4.3 is devoted to *categorical grammar*, a model for syntactic description which fits in nicely with the theory of types and which is often exploited in grammar models which use logical techniques, so-called logical grammars. One such model is Montague grammar, which is the subject of chapter 6. The λ -operator is added to the theory of types in §4.4. This extension greatly increases the suitability of the theory of types as an aid to the description of the semantics of natural language. In chapter 5, the theory of types and intensional logic are merged into the *intensional theory of types*.

4.2 The Theory of Types

4.2.1 Type Distinctions in Natural Language

Besides connectives and quantifiers (and sometimes identity and function symbols), languages for predicate logic contain just two kinds of symbols. There are individual constants and variables, expressions which refer to entities in some given domain. And besides these there are the predicate constants, expressions which refer to sets of entities in the case of one-place predicate constants, and to sets of ordered sequences of n entities in the case of n -ary predicate constants.

This means that in predicate logic one can say things only about the properties which entities have and the relations they bear to other entities. In a natural language like English, on the other hand, we can talk about much more than those kinds of things. We can say things about properties, for example. The point is that any logical system which is appropriate as an instrument for the analysis of natural language needs a much richer structure than predicate

logic. We shall now consider various kinds of sentences lacking direct translations into predicate logic and try to think of the new types of expressions which we would need in order to express them adequately in a logical formalism.

Our first examples concern sentences with quantification over properties. Besides sentences in which some fixed property is attributed to one or more entities, there are, for example, sentences which say that there is some property which two entities have in common without explicitly saying what this common property is. Consider (1):

- (1) If John is self-satisfied, then there is at least one thing he has in common with Peter.

Sentence (1) contains quantification over properties. Another example of this is (2), which asserts that a particular entity has all of the properties which typify a particular kind of thing:

- (2) Santa Claus has all the attributes of a sadist.

Sentence (2) says of every property that if it is a typical property of sadists, then it is a property of Santa Claus. If we are to quantify not only over entities but also over properties of entities, then we need to extend predicate logic by introducing variables other than the ones we already have, which only range over entities. Besides predicate letters, we need *predicate variables*, so that we can quantify over this kind of variable in the syntax. Letting X be such a variable, (1) and (2) may be represented as in (3) and (4):

$$(3) \quad Z_j \rightarrow \exists X(X_j \wedge X_p)$$

$$(4) \quad \forall X(\forall x(Sx \rightarrow Xx) \rightarrow Xs)$$

The logical system with quantification over both entities and properties of entities is called *second-order predicate logic*. Standard predicate logic is then sometimes referred to as *first-order predicate logic*. Because it lacks a completeness theorem, second-order predicate logic has been less intensively studied by logicians (see vol. 1, chap. 5). This would, however, not appear to have much bearing on its utility in linguistics.

But second-order predicate logic does not exhaust the expressive power of natural language any more than first-order logic does. For not only are there natural language sentences which quantify over properties of entities, but there are also sentences which attribute properties to these properties of entities in turn. The predicate *red*, by way of illustration, expresses a property of individuals, so the predicate *color* expresses a property of properties of individuals. So in a sentence like *Red is a color*, which we represent as $\mathcal{C}(R)$, the second-order predicate *color* is applied to the first-order predicate *red*. We can also quantify over these properties of properties, as in *Red has something (a property) in common with green*. This sentence can be represented as $\exists \mathcal{X}(\mathcal{X}(R) \wedge \mathcal{X}(G))$. So not only must we introduce predicate constants of this

kind, like \mathcal{C} , but we must also introduce predicate variables like \mathcal{X} in order to be able to quantify over the properties of properties of entities which are represented by these predicates. In principle there is no limit to this hierarchy of increasingly higher order predicates. This results in predicate logic being extended with predicate constants and predicate variables of arbitrary order. In natural language, the fourth and higher orders in the hierarchy are seldom if ever used.

Besides higher-order predicates, there are other kinds of expressions which for linguistic purposes may usefully be added to predicate logic. The following examples illustrate this.

One first class of examples is formed by expressions with *predicate adverbials*.

(5) John is walking quickly.

In sentence (5), the expression *quickly* is, from a linguistic perspective, a modifier acting on the verb *is walking*. From a logical perspective, the property of walking quickly is attributed to an entity, John, in sentence (5). This property cannot be seen as a conjunction of two properties, 'being quick' and 'walking'. For sentence (5) does not mean the same thing as sentence (6):

(6) John is walking and John is quick.

In logical terms, *quickly* is an expression which when applied to the first-order predicate *walking* results in a new first-order predicate *walking quickly*. From a logical point of view, the *relative adjectives* are expressions of the same kind. Sentence (7) may be represented in first-order predicate logic as formula (8):

(7) Jumbo is a pink elephant.

(8) $E_j \wedge P_j$

The adjective *pink* may, in other words, be represented as a standard first-order predicate. But the same does not apply to relative adjectives like *small*. Sentence (9) is the same kind of sentence as (7):

(9) Jumbo is a small elephant.

But sentence (9) cannot be analyzed as a conjunction of two first-order predicates. The formula (10) which we would then obtain:

(10) $E_j \wedge S_j$

expresses something which is generally false. It may well be that Jumbo is small (for an elephant), but even small elephants are creatures of considerable size. The relative adjective *small* works the same way as the predicate adverbial *quickly*. When applied to the predicate *elephant*, it results in a new predicate *small elephant*.

Predicate adverbials and relative adjectives are not the same kinds of expressions as second-order predicates like *color*. One of the latter, when applied to a first-order predicate, results not in a new first-order predicate but in a sentence. *Color*, when applied to *red*, results in the sentence *Red is a color*. But *quickly* when applied to *walking* results in a new predicate *walking quickly*.

Expressions which in turn modify these predicate adverbials and adjectives form yet another distinct category of expressions. As examples we have *very*, as in *John is walking very quickly*, and *terribly*, as in *Jumbo is a terribly small elephant*. An expression like *terribly*, when applied to a relative adjective like *small*, results in a new, composite relative adjective *terribly small*. Yet other new kinds of expressions are to be found in *prepositions*. In the sentence *Mary is sitting next to John*, the preposition *next to* is an expression which when applied to the term *John* results in the predicate adverbial *next to John*.

These examples should make it clear that a logical language has to contain a great diversity of expressions if it is to be a useful instrument in the analysis of natural language. One logical system which satisfies this requirement was developed at the beginning of the twentieth century, although its original motivation was not linguistic but purely logical. This system, which is called *the theory of types* or *the theory of finite types*, was developed by Russell as a response to the paradoxes which had been discovered in set theory. One of the best known of these is the *Russell paradox*, which he himself had discovered. This paradox arises as soon as we assume that for every property P , there is a set $\{x|Px\}$ consisting of all and only those entities which have P . Under this assumption, for example, there must be a set $\{x|x = x\}$. This set is the universal set that contains everything there is, for everything is equal to itself. And since this set contains everything, it must also contain itself as a member: $\{x|x = x\} \in \{x|x = x\}$. Now consider this property of self-membership. Some special sets like $\{x|x = x\}$ have this property $x \in x$, but most familiar entities do not have it. The number 0, for example, is not a member of itself, since it isn't even a set. And $\{0\} \notin \{0\}$, since $\{0\}$ has just a single element, the number 0, and $0 \neq \{0\}$. Nor is the set N of natural numbers a member of itself. $N \notin N$, since this set only contains numbers, while N is not itself a number but a set of numbers. So let us consider the set R of all entities which are not members of themselves: $R = \{x|x \notin x\}$. The Russell paradox now turns up if we attempt to find out whether the set R is a member of itself or not. Suppose, to begin with, that $R \in R$. R must satisfy its own requirement for membership $x \notin x$, so that $R \notin R$. So $R \in R$ is impossible. But if, on the other hand, $R \notin R$, then R satisfies its own requirement for membership, so that $R \in R$. We see that $R \notin R$ is impossible too.

The theory of types gets around this paradox by locating entities at sharply distinguished levels. The membership relation is then allowed to obtain only between entities which are exactly one level apart. This distinction between levels is paralleled in the language of the theory of types by a distinction be-

tween different types of expressions. Two expressions a and B which refer to entities at different levels are said to be of different types. The symbol \in , which expresses the relation of set membership, can apply to two symbols a and B only if B is an expression of a type which refers to sets of the entities referred to by expressions of the type of a . This makes it impossible to set up a Russell paradox. For $\alpha \in \alpha$ is not a well-formed expression. This explicit, type-theoretical solution to the problem of the paradoxes is not the most highly favored one nowadays. There are axiomatic formalizations of set theory into which the theory of types has been integrated but in which it is no longer explicitly present in the language used. These formalisms not only avoid the Russell paradox but also have the advantage of being easier to work with than the theory of types. But as far as the applications in linguistic analysis are concerned, the theory of types remains a useful tool.

4.2.2 Syntax

We shall now turn to the manner in which the languages used in the theory of types are built up. We will begin by determining which types of expressions any such language may have. It turns out that we can make do with just two basic types in the analysis of the examples from natural language presented above. All of the remaining types that are needed can be constructed from these. As our two basic types we have e , which is the type of those expressions which refer to entities, and t , the type of those expressions which refer to truth values. As examples of expressions of type e we have the individual constants and variables familiar from standard predicate-logical languages, and formulas are examples of expressions of type t . The set of all types may be defined in terms of these two basic types in the following manner:

Definition 1

T , the set of types, is the smallest set such that:

- (i) $e, t \in T$
- (ii) if $a, b \in T$, then $\langle a, b \rangle \in T$

The requirement that T be the smallest set satisfying (i) and (ii) has the same effect as the usual closure or induction clause: '(iii) nothing is an element of T except on the basis of (i) and (ii)'. Clause (ii) in this definition generates, starting with e and t , a supply of types unlimited in principle. The general idea behind a type $\langle a, b \rangle$ is the following: an expression of type $\langle a, b \rangle$ is an expression which when applied to an expression of type a results in an expression of type b . In other words, if α is an expression of type $\langle a, b \rangle$ and β is an expression of type a , then $\alpha(\beta)$ will be an expression of type b . This process of applying an α of type $\langle a, b \rangle$ to a β of type a is called (*functional*) *application of α to β* .

As an example of a derived type we have $\langle e, t \rangle$. An expression of this type

results, when applied to an expression of type e , in an expression of type t . One-place predicate letters belong to this type. For the result of applying a one-place predicate to an individual constant or variable, since these are expressions of type e , is a formula, and these are of type t .

As a second example of a derived type we have $\langle\langle e, t \rangle, t\rangle$. The expressions in this type are those which result in formulas when applied to one-place predicates. So these expressions are predicates of one-place predicates over individuals, that is to say, second-order predicates.

The type $\langle\langle e, t \rangle, \langle e, t \rangle\rangle$ contains expressions which when applied to a one-place predicate result again in a one-place predicate. Predicate adverbials and relative adjectives both correspond to expressions of this type.

One type which deserves special mention is $\langle e, \langle e, t \rangle \rangle$. An expression of this type results, when applied to an expression of type e , in a one-place predicate. Two-place predicates will be considered to be expressions of type $\langle e, \langle e, t \rangle \rangle$ in the theory of types. A sentence like *John loves Mary* translates into predicate logic as the formula Ljm , in which the two-place predicate L is an expression which in combination with the two individual constants j and m results in the formula Ljm . In a type-theoretical language, L is treated as an expression which, when applied to an individual constant m , results in a one-place predicate $L(m)$, an expression of type $\langle e, t \rangle$. This one-place predicate expresses the property of 'loving Mary'. And this predicate can in turn be applied to the individual constant j , as a result of which we obtain the formula $(L(m))(j)$. The formula says that the individual John has the property of loving Mary. The proposition is equivalent to the proposition that John bears the relation of 'loving' to Mary, as will become apparent from the semantic interpretation of the theory of types. This treatment of two-place predicate letters like L generalizes very easily to n -place predicate letters. Table 4.1 sums up a few types by way of illustration, together with glosses and examples. We now have a definition fixing the types dealt with in the theory of types and an indication of how expressions of derivative types are to function. So we can go on to define the languages of the theory of types, as follows.

The *vocabulary* of a type-theoretical language L contains some symbols which are shared by all such languages and a number of symbols which are characteristic of L . The shared part consists of:

- (i) For every type a , an infinite set VAR_a of variables of type a
- (ii) The usual connectives $\wedge, \vee, \rightarrow, \neg, \leftrightarrow$
- (iii) The quantifiers \forall and \exists
- (iv) Two brackets (and)
- (v) The symbol for identity =

The part of the vocabulary which is characteristic of L contains:

- (vi) for every type a , a (possibly empty) set CON_a^L of constants of type a

Obviously constants and variables of the various types must be kept apart. We will write v_a for variables of type a and c_a for constants of type a (though the

Table 4.1 Types and Expressions

Type	Kind of expression	Example
e	Individual expression	John
$\langle e, t \rangle$	One-place first-order predicate	Walks, red, loves Mary
t	Sentence	John walks, John loves Mary
$\langle t, t \rangle$	Sentential modifier	Not
$\langle e, e \rangle$	Function from individuals to individuals	The father of
$\langle \langle e, t \rangle, \langle e, t \rangle \rangle$	Predicate modifier	Quickly, beautifully
$\langle e, \langle e, t \rangle \rangle$	Two-place first-order relation	Loves, lies between Amsterdam and
$\langle e, \langle e, \langle e, t \rangle \rangle \rangle$	Three-place first-order relation	Lies between (and)
$\langle \langle e, t \rangle, t \rangle$	One-place second-order predicate	Is a color
$\langle \langle e, t \rangle, \langle \langle e, t \rangle, t \rangle \rangle$	Two-place second-order relation	Is a brighter color than
$\langle e, \langle \langle e, t \rangle, t \rangle \rangle$	Two-place relation between individuals and first-order predicates	Is a property of
$\langle \langle \langle e, t \rangle, t \rangle, t \rangle$	One-place third-order predicate	Is a second-order predicate

subscripts will be dropped where this does not lead to confusion). In practice we will tend to revert to the more convenient notations familiar from standard predicate logic wherever possible. Thus wherever possible we will represent the distinctions between types typographically, writing x, y, \dots for variables over entities, X, Y, \dots for first-order predicate variables, and $\mathcal{X}, \mathcal{Y}, \dots$ for second-order predicate variables. Individual constants are to be written as the lower-case letters a, b, c, \dots, l, m, n , and so on; first-order n -place predicate constants as uppercase letters A, B, C, \dots, L, M, R , and so on; and second-order predicate constants as $\mathcal{C}, \mathcal{H}, \mathcal{M}, \mathcal{D}, \dots$. Any deviations from this practice will be mentioned explicitly.

The inductive definition of the formulas is more complicated than in predicate logic. For what we have to give is a general definition of what it is to be an expression of a type $a \in T$; the *formulas* are then those expressions which are of the particular type t . One characteristic thing about expressions of this type, however, is the ways they may be formed. Initially, expressions are formed from constants and variables by means of application. Then from expressions of type t , new expressions of this type may be formed by means of the connectives and quantifiers, while insertion of the symbol $=$ for identity between any two expressions of the same type also results in an expression of type t . Here is the precise definition which fixes the syntax of type-theoretical languages:

Definition 2

- (i) If α is a variable or a constant of type a in L , then α is an expression of type a in L

- (ii) If α is an expression of type $\langle a, b \rangle$ in L , and β is an expression of type a in L , then $(\alpha(\beta))$ is an expression of type b in L
- (iii) If ϕ and ψ are expressions of type t in L (i.e., formulas in L), then so are $\neg\phi$, $(\phi \wedge \psi)$, $(\phi \vee \psi)$, $(\phi \rightarrow \psi)$, and $(\phi \leftrightarrow \psi)$
- (iv) If ϕ is an expression of type t in L and v is a variable (of arbitrary type a), then $\forall v\phi$ and $\exists v\phi$ are expressions of type t in L
- (v) If α and β are expressions in L which belong to the same (arbitrary) type, then $(\alpha = \beta)$ is an expression of type t in L
- (vi) Every expression in L is to be constructed by means of (i)–(v) in a finite number of steps.

We refer to the set of all expressions in L of type a as WE_a^L or, if it is clear which L is meant, as WE_a . In this terminology, the formulas are the elements of WE_t . Clause (ii) in the above definition may seem rather liberal with the brackets, since it puts brackets both around the argument and around the entire expression. However, this often helps to keep the expressions legible. Where it does not, we will tend to leave off the superfluous brackets. This applies, for example, to all outer brackets. But brackets ‘inside’ can also often be left out, especially when the type of the expressions is clear. By way of example, according to definition 2, our translation of *John loves Mary* is the formula $((L(m))(j))$. Leaving off the outer brackets gives $(L(m))(j)$. The pair of outer brackets in $(L(m))$, however, serves no disambiguating purpose here, and so these too may be left off, resulting in $L(m)(j)$.

Exercise 1*

Let j be an expression type e ; M of type $\langle e, t \rangle$; S of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$; and \mathcal{C} of type $\langle \langle e, t \rangle, t \rangle$.

- (a) Determine whether the following sequences are well-formed expressions of the theory of types:
 - (i) $j(M)$
 - (ii) $S(M(j))$
 - (iii) $S(M)$
 - (iv) $(S(M))(j)$
 - (v) $\mathcal{C}(M)$
 - (vi) $(\mathcal{C}(M))(j)$
 - (vii) $\mathcal{C}(S(M))$
 - (viii) $(S(S(M)))(j)$
- (b) Determine the types a and b in the following expressions, given that the latter are of type t :
 - (i) $(c_a(M))(j)$
 - (ii) $c_a(M(j))$
 - (iii) $(S(M))(c_a)$
 - (iv) $c_b((S(M))(c_a))$
 - (v) $c_b((c_a(S))(M))$

Exercise 2*

Translate the following sentences into formulas of the theory of types. Represent the descriptions *the salami* and *the couch* by constants of type e . State the translation key and give the types of the various expressions.

- (a) John sleeps soundly.
- (b) John sleeps, presumably.
- (c) Harry slices the salami carefully.
- (d) If you do nothing, you don't do anything wrong.
- (e) If you do nothing, you do nothing wrong.
- (f) Mary is sitting on the couch.
- (g) Mary is sitting between John and Harry.
- (h) Mary is sitting on the couch between John and Harry.

Exercise 3

Is it possible to assign types to α , β , and γ in such a way that both $(\alpha(\beta))(\gamma)$ and $\alpha(\beta(\gamma))$ are well-formed expressions?

4.2.3 Semantics

As usual, the semantic interpretation of the well-formed expressions of a language for the theory of types follows the definition of those well-formed expressions given in the syntax. For example, let W be a well-formed expression of type $\langle e, t \rangle$ which stands for *walks*, and let j be a well-formed expression of type e . Then $W(j)$ is a well-formed expression of type t . Now given that type t expressions refer to truth values and that type e expressions refer to entities, the interpretations of type $\langle e, t \rangle$ expressions will have to be things which yield a truth value when applied to an entity. This means that the interpretation of the one-place predicate W is a function from entities to truth values, namely, a function which yields the truth value 1 when applied to an entity d which has the property of walking, and 0 when applied to an entity d which does not have this property. That is to say, given a domain consisting of people, the predicate W is no longer interpreted as the set of all people who walk but as the function which assigns the truth value 1 to the elements of that set (the people who walk), and the truth value 0 to everyone else in the domain. Any such function that assigns 1 to those elements of a set A which are members of a subset B of A , and 0 to the rest, is called the *characteristic function of B (over A)*. An example. If $A = \{a, b, c, d\}$ and $B = \{a, b\}$, then the characteristic function of B over A , written as f_B , may be defined: $f_B(a) = f_B(b) = 1$, and $f_B(c) = f_B(d) = 0$. This function f_B has as its domain the set A , its range is the set $\{0, 1\}$ of truth values, and it assigns the truth value 1 to all elements of A which are in B and the truth value 0 to all elements of A which are not.

Sets and their characteristic functions really amount to the same thing. Given any subset X of a set Y , we can construe X 's characteristic function

over Y , and conversely, given any function f from Y to $\{0, 1\}$, we can determine the subset X of Y of which that function is the characteristic function. Let $X \subseteq Y$. Then f_X is that function from Y to $\{0, 1\}$ such that for all $y \in Y$: $F_X(y) = 1$ iff $y \in X$. And if f_X is a function from Y to $\{0, 1\}$, then $X = \{y | f_X(y) = 1\}$. In other words, the assertions ' $y \in X$ ' and ' $f_X(y) = 1$ ' are equivalent, so that the concepts *characteristic function of a set X* and *set X* are interchangeable.

Given a domain D , then, one-place predicates are interpreted as the characteristic functions of subsets of that domain. In view of the interchangeability mentioned above, this is no different from the situation in standard predicate logic. Different one-place predicates may be interpreted as different characteristic functions. The interpretation of a one-place predicate is an element of the set of all characteristic functions of subsets of the domain D . The notation for this set of functions is $\{0, 1\}^D$. So for every subset X of D , $\{0, 1\}^D$ contains the characteristic function f_X . Because sets and their characteristic functions are interchangeable, the set of all characteristic functions of subsets of D and the set of all subsets of D are interchangeable too. In short, $\{0, 1\}^D$ is interchangeable with $\text{POW}(D)$, the *power set* of D .

The interpretation of expressions of type theory follows this functional pattern quite generally. An expression of type $\langle a, b \rangle$ is treated as a function from 'type a things' to 'type b things'. The general notation for the set of all functions mapping a set X into a set Y is Y^X . So this set contains all functions which assign an element of Y to each element of X . The set of functions $\{0, 1\}^D$ mentioned above is a special case of this. X and/or Y themselves may be sets of functions, so we also have, for example, $(\{0, 1\}^D)^D$, the set of all functions from D to the set of all functions from D to $\{0, 1\}$, and $\{0, 1\}^{(D^D)}$, the set of all functions from functions from D to D , to the set $\{0, 1\}$. Notice how the brackets play a disambiguating role here.

By means of this notation, we can represent for every type a the domain within which any expression of that type is to be interpreted. Expressions of type e are to be interpreted as elements of D . Expressions of type t , formulas, are to be interpreted as elements of the set $\{0, 1\}$ of truth values. And just as in the definition of the types, all derived types were constructed from the basic types e and t , we will now construct the domains in which these derived types are to be interpreted from D and $\{0, 1\}$. The domain of interpretation of expressions of type a , given a domain D , is written as $D_{a,D}$, and is defined as follows:

Definition 3

- (i) $D_{e,D} = D$
- (ii) $D_{t,D} = \{0, 1\}$
- (iii) $D_{\langle a,b \rangle,D} = D_{b,D}^{D_{a,D}}$

We will drop the subscript D whenever this does not give rise to confusion. Here are two worked-out examples by way of illustration. In the theory of

types, a two-place predicate L (*loves*) is an expression of type $\langle e, \langle e, t \rangle \rangle$. The corresponding interpretation domain $\mathbf{D}_{\langle e, \langle e, t \rangle \rangle}$ is the set of functions $\mathbf{D}_{\langle e, t \rangle}^{\mathbf{D}_{\langle e, t \rangle}}$, that is, $(\mathbf{D}_{\langle e, t \rangle}^{\mathbf{D}_{\langle e, t \rangle}})^{\mathbf{D}_{\langle e, t \rangle}}$, that is, $(\{0, 1\}^{\mathbf{D}_{\langle e, t \rangle}})^{\mathbf{D}_{\langle e, t \rangle}}$. The interpretation domain of two-place predicates is thus the set of functions mapping entities onto (characteristic functions of) sets of entities. In concrete terms, this means the following. The two-place predicate L is interpreted as a function from entities to sets of entities. This function maps an entity d onto the set of all entities which bear L to d , that is, onto the set of all entities which love d . The characteristic function of this set can then be applied to an entity in order to obtain a truth value. As was already hinted at above, there is no essential difference between this way of interpreting two-place predicates and the standard way: there is a one-to-one correspondence between functions in $(\{0, 1\}^{\mathbf{D}_{\langle e, t \rangle}})^{\mathbf{D}_{\langle e, t \rangle}}$ and subsets of the set \mathbf{D}^2 , the set of all ordered pairs of elements of \mathbf{D} . For if f is an element of $(\{0, 1\}^{\mathbf{D}_{\langle e, t \rangle}})^{\mathbf{D}_{\langle e, t \rangle}}$, then for every $d_1 \in \mathbf{D}$, $f(d_1) \in \{0, 1\}^{\mathbf{D}_{\langle e, t \rangle}}$ and so for every $d_2 \in \mathbf{D}$ we have $(f(d_1))(d_2) \in \{0, 1\}$. The set A corresponding to f can be defined as: $A = \{\langle d_2, d_1 \rangle | (f(d_1))(d_2) = 1\}$. And conversely, if A is a subset of \mathbf{D}^2 , then the function f corresponding to A may be defined by means of: for every $d_1 \in \mathbf{D}$, $f(d_1)$ is the function for which, for every $d_2 \in \mathbf{D}$, $(f(d_1))(d_2) = 1$ iff $\langle d_2, d_1 \rangle \in A$.

As a second example, consider the second-order predicate \mathcal{C} (*color*), which is of type $\langle \langle e, t \rangle, t \rangle$. The interpretation domain $\mathbf{D}_{\langle \langle e, t \rangle, t \rangle}$ is the set of functions $\{0, 1\}^{(\{0, 1\}^{\mathbf{D}_{\langle e, t \rangle}})^{\mathbf{D}_{\langle e, t \rangle}}}$. So the interpretation domain of second-order predicates is the set of functions mapping (characteristic functions of) sets of entities onto truth values, that is, the set of characteristic functions of sets of characteristic functions of sets of entities. Any such function characterizes a subset of the power set of the domain \mathbf{D} . In second-order logic, such subsets are the interpretations of second-order predicates. In this, too, the interpretation of the theory of types introduces nothing novel. Table 4.2 sums up the interpretations of the expressions included in table 4.1. In the interest of legibility, the phrase 'characteristic function of a set' has been replaced by 'set' in all but its first occurrences.

A model \mathbf{M} for a language L for the theory of types consists of a nonempty domain set \mathbf{D} together with an interpretation function I . By means of \mathbf{D} and the truth values, the domains of interpretation for expressions of arbitrary type are defined as in definition 3. The interpretation function I assigns to each constant in L some element of the interpretation domain which corresponds to the type of that constant. That is to say, for each type a , I is a function from CON_a^L into $\mathbf{D}_{a, \mathbf{D}}$. Besides this, we need assignments which will interpret the variables. Once again, if v is a variable of type a , the interpretation of v is an element of the corresponding domain. In other words, for every type a , these assignments are functions from VAR_a into $\mathbf{D}_{a, \mathbf{D}}$. We have thus defined the interpretation of the noncomposite expressions of L , given a model \mathbf{M} and an assignment g . It remains to provide a definition of the interpretation of the composite expressions in L . This is done, in accordance with the principle of compositionality, in terms of the expressions of which these are composed.

Table 4.2 Types and Interpretations

Type	Interpretation
e	Entity
t	Truth value
$\langle e, t \rangle$	Function from entities to truth values, i.e., (characteristic function of) a set of entities
$\langle t, t \rangle$	Function from truth values to truth values
$\langle e, e \rangle$	Function from entities to entities
$\langle \langle e, t \rangle, \langle e, t \rangle \rangle$	Function from sets of entities to sets of entities
$\langle e, \langle e, t \rangle \rangle$	Function from entities to sets of entities, i.e., set of ordered pairs of entities
$\langle e, \langle e, \langle e, t \rangle \rangle \rangle$	Function from entities to functions from entities to sets of entities, i.e., set of ordered triples of entities
$\langle \langle e, t \rangle, t \rangle$	(Characteristic function of) a set of sets of entities
$\langle \langle e, t \rangle, \langle \langle e, t \rangle, t \rangle \rangle$	Function from sets of entities to sets of sets of entities, i.e., set of ordered pairs of sets
$\langle e, \langle \langle e, t \rangle, t \rangle \rangle$	Function from entities to sets of sets of entities
$\langle \langle \langle e, t \rangle, t \rangle, t \rangle$	(Characteristic function of) a set of sets of sets of entities

The situation is different, however, from that in predicate logic, since not all of these composite expressions are formulas. For this reason we need more than just a truth definition, that is, a definition of $V_{M,g}$. We must define the general concept of *the interpretation of α with respect to a model M and an assignment g* , to be written as $\llbracket \alpha \rrbracket_{M,g}$. Just as a valuation $V_{M,g}$ can be regarded as a function from formulas into truth values, the interpretation function $\llbracket \cdot \rrbracket_{M,g}$ can be seen as a function which, for all types a , maps WE_a^L into $D_{a,D}$. For $a = t$, then, $\llbracket \cdot \rrbracket_{M,g}$ functions as a valuation. The definition of $\llbracket \alpha \rrbracket_{M,g}$ runs, as usual, parallel to the syntactic construction of the well-formed formulas of L , which was given in definition 2.

Definition 4

- (i) If $\alpha \in CON_a^L$, then $\llbracket \alpha \rrbracket_{M,g} = I(\alpha)$
if $\alpha \in VAR_a$, then $\llbracket \alpha \rrbracket_{M,g} = g(\alpha)$
- (ii) if $\alpha \in WE_{(a,b)}^L$, $\beta \in WE_a^L$, then $\llbracket \alpha(\beta) \rrbracket_{M,g} = \llbracket \alpha \rrbracket_{M,g} (\llbracket \beta \rrbracket_{M,g})$
- (iii) if $\phi, \psi \in WE_t^L$, then
 - $\llbracket \neg \phi \rrbracket_{M,g} = 1$ iff $\llbracket \phi \rrbracket_{M,g} = 0$
 - $\llbracket \phi \wedge \psi \rrbracket_{M,g} = 1$ iff $\llbracket \phi \rrbracket_{M,g} = \llbracket \psi \rrbracket_{M,g} = 1$
 - $\llbracket \phi \vee \psi \rrbracket_{M,g} = 1$ iff $\llbracket \phi \rrbracket_{M,g} = 1$ or $\llbracket \psi \rrbracket_{M,g} = 1$
 - $\llbracket \phi \rightarrow \psi \rrbracket_{M,g} = 0$ iff $\llbracket \phi \rrbracket_{M,g} = 1$ and $\llbracket \psi \rrbracket_{M,g} = 0$
 - $\llbracket \phi \leftrightarrow \psi \rrbracket_{M,g} = 1$ iff $\llbracket \phi \rrbracket_{M,g} = \llbracket \psi \rrbracket_{M,g}$
- (iv) if $\phi \in WE_t^L$, $v \in VAR_a$, then
 - $\llbracket \forall v \phi \rrbracket_{M,g} = 1$ iff for all $d \in D_{a,D}$: $\llbracket \phi \rrbracket_{M,g[v/d]} = 1$
 - $\llbracket \exists v \phi \rrbracket_{M,g} = 1$ iff there is at least one $d \in D_{a,D}$ such that:
 $\llbracket \phi \rrbracket_{M,g[v/d]} = 1$
- (v) if $\alpha, \beta \in WE_a^L$, then $\llbracket \alpha = \beta \rrbracket_{M,g} = 1$ iff $\llbracket \alpha \rrbracket_{M,g} = \llbracket \beta \rrbracket_{M,g}$

The truth definition can now be extracted from this general definition of the interpretation of an arbitrary expression with respect to a model \mathbf{M} and an assignment g . A sentence ϕ is said to be *true with respect to \mathbf{M}* just in case $\llbracket \phi \rrbracket_{\mathbf{M}} = 1$. (As in predicate logic, because sentences lack free variables, their truth values are independent of assignments. As a result, these can be left out.) A sentence ϕ is said to be *universally valid* just in case $\llbracket \phi \rrbracket_{\mathbf{M}} = 1$ for every appropriate \mathbf{M} , and once again the notation is $\models \phi$. We also say that two sentences ϕ and ψ are *equivalent* iff $\models \phi \leftrightarrow \psi$, that is to say iff $\llbracket \phi \rrbracket_{\mathbf{M}} = \llbracket \psi \rrbracket_{\mathbf{M}}$ for every \mathbf{M} . More generally: formulas ϕ and ψ are said to be *equivalent* iff for every \mathbf{M} and g , $\llbracket \phi \rrbracket_{\mathbf{M},g} = \llbracket \psi \rrbracket_{\mathbf{M},g}$. In the theory of types there is no reason to restrict the notion of equivalence to formulas. Any two expressions α and β (of the same type) are said to be equivalent iff for every \mathbf{M} and g , $\llbracket \alpha \rrbracket_{\mathbf{M},g} = \llbracket \beta \rrbracket_{\mathbf{M},g}$. For expressions lacking free variables, this amounts to α and β being equivalent just in case $\models \alpha = \beta$. It should be noted at this point that for formulas ϕ and ψ , $\phi \leftrightarrow \psi$ and $\phi = \psi$ mean exactly the same thing. They both say that ϕ and ψ refer to one and the same truth value.

We shall now illustrate the truth definition by applying it to some concrete examples. Clause (ii) assigns, among other things, the truth definition for formulas of the form $\alpha(\beta)$. The formula $W(j)$, our representation of the sentence *John is walking*, is an example. Here W is a constant of type $\langle e, t \rangle$, while j is a constant of type e . The formula $W(j)$ itself is of type t . Clause (ii) defines the interpretation of this formula with respect to a model \mathbf{M} and an assignment g as follows: it is the result of applying the interpretation with respect to \mathbf{M} and g of the predicate W to the interpretation with respect to \mathbf{M} and g of the constant j . More succinctly: $\llbracket W(j) \rrbracket_{\mathbf{M},g} = \llbracket W \rrbracket_{\mathbf{M},g}(\llbracket j \rrbracket_{\mathbf{M},g})$. According to clause (i), we have $\llbracket W \rrbracket_{\mathbf{M},g} = I(W)$ and $\llbracket j \rrbracket_{\mathbf{M},g} = I(j)$. And according to the definition of the interpretation function I , we know that $I(W)$ is an element of the set of functions $D_t^{D_e} = \{0, 1\}^{D_e}$. This is the characteristic function of that set of entities within D which possess the property expressed by W , which is of course the set of things which walk. The interpretation of the constant, $I(j)$, is an element of D_e , which means that it is an element of the domain D . It is, then, an entity. The interpretation, that is, the truth value, of the formula $W(j)$ is now to be obtained by applying $I(W)$ to $I(j)$. It is 1 if $I(j)$ is an element of the set characterized by $I(W)$, and 0 if this is not the case.

Clause (ii) gives the truth definition for sentences which attribute properties to properties in just the same way. As an example, take the sentence $\mathcal{C}(R)$, which is the representation of the sentence *Red is a color*. Clause (ii) defines the interpretation of $\mathcal{C}(R)$ as follows: $\llbracket \mathcal{C}(R) \rrbracket_{\mathbf{M},g} = \llbracket \mathcal{C} \rrbracket_{\mathbf{M},g}(\llbracket R \rrbracket_{\mathbf{M},g})$. According to clause (i), this is equal to $I(\mathcal{C})(I(R))$. The interpretation function I then says that $I(\mathcal{C}) \in D_{t^{(D_e, D_e)}}^{D_t} = \{0, 1\}^{\{0, 1\}^{D_e}}$. The interpretation of the second-order predicate \mathcal{C} is then (the characteristic function of) a set of (characteristic functions of) sets of entities. The latter are the denotations of those properties which themselves have the property of 'being a color'. Furthermore, we have $I(R) \in D_t^{D_e} = \{0, 1\}^{D_e}$: the interpretation of R is (the characteristic function of)

a set of entities, namely, those which have the property of 'redness'. The second-order predication $\mathcal{C}(R)$ is true in a model \mathbf{M} just in case the result of applying $I(\mathcal{C})$ to $I(R)$ is 1. That is, just in case that set of entities which forms the interpretation of R in \mathbf{M} is itself an element of the set of sets forming the interpretation of \mathcal{C} in \mathbf{M} .

Clause (ii) of definition 4 is also responsible for interpreting atomic sentences with relations. As an example, consider the formula $L(m)(j)$, which is our representation of the sentence *John loves Mary* in the theory of types. Applying clause (ii), we obtain $\llbracket L(m)(j) \rrbracket_{\mathbf{M},g} = \llbracket L(m) \rrbracket_{\mathbf{M},g}(\llbracket j \rrbracket_{\mathbf{M},g})$. Note that this is quite analogous to the earlier examples: the interpretation of an atomic sentence is obtained by applying the interpretation of the predicate to the interpretation of the constant. Clause (i) defines $\llbracket j \rrbracket_{\mathbf{M},g}$ as $I(j)$, that is, as one or another entity in the domain D . But how are we to determine the interpretation of the predicate in this case? $L(m)$ is a composite expression and not a constant, so clause (i) will not be of much help. The answer is that clause (ii) of definition 4 also determines the interpretation of this kind of expression. In fact, it is responsible for the interpretations of all expressions obtained by means of functional application, that is, the application of an expression of type $\langle a, b \rangle$ to an expression of type a . Such expressions include not only formulas but also expressions of types other than t . The composite predicate $L(m)$ is a case in point. It is an expression of type $\langle e, t \rangle$, formed by applying the expression L , which is of type $\langle e, \langle e, t \rangle \rangle$, to m , which is of type e . Its interpretation is to be obtained, once again, by means of clause (ii): $\llbracket L(m) \rrbracket_{\mathbf{M},g} = \llbracket L \rrbracket_{\mathbf{M},g}(\llbracket m \rrbracket_{\mathbf{M},g})$. We have once again arrived at constants: $\llbracket L \rrbracket_{\mathbf{M},g}(\llbracket m \rrbracket_{\mathbf{M},g}) = I(L)(I(m))$. $I(m)$ is an element of D , and $I(L)$ is an element of the set of functions: $(D^D)^D = (\{0, 1\}^D)^D$. As such it is a function which, when applied to an entity, yields (the characteristic function of) some set of entities. Applying $I(L)$ to an entity d , we obtain the set of entities that love d . So applying $I(L)$ to $\llbracket m \rrbracket_{\mathbf{M},g}$, we obtain (the characteristic function of) the set of entities that love Mary. This function is $\llbracket L(m) \rrbracket_{\mathbf{M},g}$, the interpretation of the composite predicate $L(m)$. When applied to $\llbracket j \rrbracket_{\mathbf{M},g}$ the result is the truth value 1 if John does indeed happen to love Mary, and otherwise it is 0. We see that $L(m)(j)$, the representation of the sentence *John loves Mary* in the theory of types, has just the same truth conditions as Ljm , the standard representation in predicate logic. From the perspective of the semantics of natural language, however, the theory of types has an advantage compared to predicate logic. The latter interprets the proper names *John* and *Mary* and the verb *to love* and combines the three into the interpretation of the whole sentence in one step. The verb phrase *loves Mary* receives no independent interpretation. This is different in the theory of types. There, the interpretation of the complex predicate $L(m)$ is a representation of the interpretation of the verb phrase *loves Mary*. To this extent, the theory of types does more justice to the syntactic and semantic structure of the sentence *John loves Mary*.

We now return to the truth definition given in definition 4. Clause (ii) gives

the truth definition of atomic sentences. We have discussed three different examples of these. There are of course many other kinds of atomic sentences, but they are interpreted in an analogous fashion.

The truth conditions for negations, disjunctions, conjunctions, and so on are given in clause (iii). These are as in standard logic and need no further elaboration.

Existential and universal quantification are dealt with in clause (iv). In particular this of course includes the quantification over entities which is familiar from predicate logic and which we now call *first-order quantification*. This kind of quantification is interpreted exactly as in the semantics of predicate logic. But besides first-order quantification, clause (iv) also deals with quantification over other types of things besides entities. As an example of this *higher-order quantification*, consider the formula $\exists \mathcal{X}(\mathcal{X}(R) \wedge \mathcal{X}(G))$, the representation of the sentence *Red and green have something in common*. R and G are constants of type $\langle e, t \rangle$, and \mathcal{X} is a variable of type $\langle \langle e, t \rangle, t \rangle$. The interpretation of $\exists \mathcal{X}(\mathcal{X}(R) \wedge \mathcal{X}(G))$ runs as follows: $\llbracket \exists \mathcal{X}(\mathcal{X}(R) \wedge \mathcal{X}(G)) \rrbracket_{M,g} = 1$ if and only if there is a $d \in D_{\langle \langle e, t \rangle, t \rangle}$ such that $\llbracket \mathcal{X}(R) \wedge \mathcal{X}(G) \rrbracket_{M,g[\mathcal{X}/d]} = 1$. $D_{\langle \langle e, t \rangle, t \rangle}$ is the set of functions $D_i^{D^D} = \{0, 1\}^{\{0,1\}^D}$. This is the set of properties of properties of entities. So the formula $\exists \mathcal{X}(\mathcal{X}(R) \wedge \mathcal{X}(G))$ is true just in case there is some property d of properties of entities such that $\llbracket \mathcal{X}(R) \wedge \mathcal{X}(G) \rrbracket_{M,g[\mathcal{X}/d]} = 1$. This is the case if both $\llbracket \mathcal{X}(R) \rrbracket_{M,g[\mathcal{X}/d]} = 1$ and $\llbracket \mathcal{X}(G) \rrbracket_{M,g[\mathcal{X}/d]} = 1$. The interpretation of these two atomic formulas is once again the business of clause (ii). So we have $\llbracket \mathcal{X}(R) \rrbracket_{M,g[\mathcal{X}/d]} = 1$ iff $\llbracket \mathcal{X} \rrbracket_{M,g[\mathcal{X}/d]} (\llbracket R \rrbracket_{M,g[\mathcal{X}/d]}) = 1$.

Because of (i) we know that $\llbracket R \rrbracket_{M,g[\mathcal{X}/d]} = I(R)$ and that $\llbracket \mathcal{X} \rrbracket_{M,g[\mathcal{X}/d]} = g[\mathcal{X}/d](\mathcal{X}) = d$. So $\mathcal{X}(R)$ is true with respect to M and $g[\mathcal{X}/d]$ iff $I(R)$ has d . The whole formula $\exists \mathcal{X}(\mathcal{X}(R) \wedge \mathcal{X}(G))$, then, is true if there is some property of properties of entities which is shared by both red and green. Such a property is in fact easily found: the second-order property 'color' is an example, so that sentence (10) is trivially true. In *uttering* (10) it will not in general be one's intention to express this trivial truth. It is more likely that (10) is uttered in order to state some such thing as that red and green are both to be found on traffic lights, to take just one possibility. This and similar phenomena are dealt with not in semantics but in pragmatics.

The last section of the truth definition is to be found in clause (v). This clause deals with the truth conditions of expressions with identities. Presumably (v) speaks for itself. Suffice it to say that identity can be expressed not only of entities in the theory of types but also of other types of things.

We have now completed our discussion of the truth definition for the theory of types, which is implicit in the more general definition of interpretation. For definition 4 is more than just a truth definition. It is a comprehensive definition of the interpretation of all expressions in a type-theoretical language. The interpretation of all basic expressions (constants and variables) is defined in clause (i), and that of the composite expressions comes in the remaining clauses. But not all composite expressions are formulas; the composite predi-

cate $L(m)$ which appears in the formula $L(m)(j)$, for example, is a composite expression but not a formula. The interpretation of composite expressions of this sort is in all cases laid down in clause (ii). This is apparent from the fact that all such expressions, whatever their type, result from functional application. By way of illustration, consider the formula $(Q(W))(j)$, which is the representation of the sentence *John walks quickly* in the theory of types. As we argued in §4.2.1, the adverbial *quickly* should be treated as an expression whose application to a predicate results in another predicate. It is represented by means of a constant Q , which is of type $\langle\langle e, t \rangle, \langle e, t \rangle\rangle$. Application of this to the constant W of type $\langle e, t \rangle$ results in the expression $Q(W)$, which is of type $\langle e, t \rangle$. The interpretation of this composite expression is now laid down in clause (ii): $\llbracket Q(W) \rrbracket_{M,g} = \llbracket Q \rrbracket_{M,g}(\llbracket W \rrbracket_{M,g})$, and that is $I(Q)(I(W))$; $I(Q)$ is an element of $(D_t^D)^{(D_e^D, D_t^D)} = (\{0, 1\}^D)^{(\{0, 1\}^D)}$, which is the set of functions which map (characteristic functions of) sets of entities onto (characteristic functions of) sets of entities.

Note that the above only says what *kind* of thing the interpretation of an adverbial like *quickly* is: it is a particular kind of function, namely, one which takes the interpretations of one-place predicates as its arguments and yields things of the same kind as values, the interpretation of the composite predicate. This says nothing at all about the relation between the interpretation of the predicate to which Q is applied (W in the example) and the interpretation of the composite predicate which is the result ($Q(W)$ in the example). For example, the validity of the argument:

(11) John walks quickly.

John walks.

is then not yet guaranteed: in the theory of types, $(Q(W))(j)/W(j)$ is not a valid argument schema. In order to account for the validity of this kind of argument, it will be necessary to go beyond the interpretation of the theory of types as given in definition 4. We have already encountered a similar state of affairs in predicate logic. Consider the following argument:

(12) Albert is taller than Bert.

Bert is taller than Charley.

Albert is taller than Charley.

The predicate logic schema corresponding to this is $\text{Tab}, \text{Tbc}/\text{Tac}$. It is not valid. The validity of (12) depends essentially on the transitivity of the relation *is taller than*. Adding a premise expressing this fact, that if x is taller than y , and y in turn is taller than z , then x is taller than z , we obtain the valid argument schema $\text{Tab}, \text{Tbc}, \forall x \forall y \forall z ((\text{Txy} \wedge \text{Tyz}) \rightarrow \text{Txz})/\text{Tac}$.

Something similar applies to (11). There too an additional premise is needed. In this case it would say that whenever x does X quickly, x does X . This premise can be expressed, in the formalism of the theory of types, as

$\forall x \forall X ((Q(X))(x) \rightarrow X(x))$. The argument schema (13) is indeed valid in the theory of types:

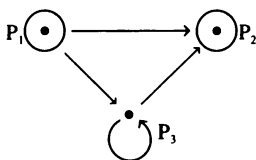
$$(13) \quad \forall x \forall X ((Q(X))(x) \rightarrow X(x)), (Q(W))(j)/W(j)$$

Such extra premises are known as 'meaning postulates', and we will return to the part they play in the description of natural language in §6.3.6.

We will not develop a syntactic notion of derivability for the theory of types. The way this could be done and the kinds of problems which would have to be overcome have been discussed at length in chapter 5 of volume 1, in connection with second-order logic. The relation between the syntactic notion of derivability and semantics is also discussed there. Among other things, it is explained why there can be no completeness theorem for second-order logic. Much the same argument applies to the theory of types, which includes second-order logic. In view of the fact that none of these matters bears directly on the applications of logical systems in linguistics, which is our primary concern, we will not go into them in any greater depth here.

Exercise 4*

Consider the model shown in the figure. Its domain consists of three points, P_1 , P_2 , and P_3 , which are denoted by the individual constants e_1 , e_2 , and e_3 . The property of being encircled is expressed by the predicate constant M , which is of type $\langle e, t \rangle$; the relation which holds between two points if there is an arrow pointing from the first to the second is represented by the two-place predicate constant A , of type $\langle e, \langle e, t \rangle \rangle$. The second-order property of being a property which applies to encircled points is represented by the constant \mathcal{E} , which is of type $\langle \langle e, t \rangle, t \rangle$. Finally, the operation which maps properties onto their complements (the operation of 'predicate negation') is denoted by T , a constant of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$.



- (a) Write out the interpretation function I .
- (b) Determine for each of the following formulas whether it is true in this model and work out its interpretation:
 - (i) $\exists x \exists y \exists z (A(y)(x) \wedge M(y) \wedge A(z)(x) \wedge \neg M(z))$
 - (ii) $\forall x (A(x)(x) \leftrightarrow \neg M(x))$
 - (iii) $\forall x (A(x)(x) \rightarrow \exists y (A(y)(x) \wedge M(y)))$
 - (iv) $\forall X \exists x X(x)$
 - (v) $\forall X (\forall y (M(y) \rightarrow \neg X(y)) \rightarrow (\exists y (X(y) \wedge A(y)(y)) \vee \neg \exists y X(y)))$
 - (vi) $\exists X (\mathcal{E}(X) \wedge \mathcal{E}(T(X)))$

4.3 Categorical Grammar

4.3.1 Introduction

The syntax of the theory of types closely resembles *categorical grammar*. There are also some differences, which we will return to in §4.3.4.

The original formulation of a categorical grammar was done by the Polish logician Lesniewski, who developed what he called 'a theory of semantic categories' in 1929. His system was further developed by another Polish logician, Ajduiewicz. The link between categorical grammar and the systematic description of the syntax of natural languages was forged by Bar-Hillel in the fifties. For the analysis of the syntax of natural languages, he proposed exploiting a categorical grammar as a mechanical, accepting system.

In the fifties and the sixties, this idea was forced out of the limelight by the mushrooming growth of transformational-generative grammar. John Lyons was the first to propose adding a transformational component to a categorical grammar, but this proposal was not welcomed in linguistic circles either. Things didn't change much until the early seventies, which saw a growing interest in semantic questions and in the application of logical methods to semantics. Around this time, the idea of employing a categorical grammar was adopted by a number of philosophers and logicians who were interested in the analysis of natural language, among them Lewis, Montague, Cresswell, Bartsch and Vennemann, and Geach. The reason for this renewed interest was that categorical grammar lends itself naturally to the kind of semantics which is done in logic, and especially to that in the theory of types.

4.3.2 Characteristics of Categorical Grammar

A *pure categorical grammar* has the following four characteristics:

- (a) There is a finite (and in practice small) set of *basic categories*.
- (b) From these basic categories, a set of *derived categories* is constructed.
- (c) There are either one or two *syntactic rules* describing the one syntactic operation of concatenation and determining the category of the result of this operation.
- (d) Every lexical element is assigned to a category.

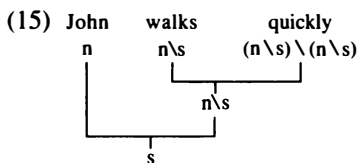
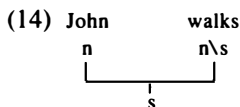
Here is a very simple example of a categorical grammar:

- (i) The basic categories are *n* (for 'noun') and *s* (for 'sentence').
- (ii) The derived categories may be obtained as follows: If *A* and *B* are categories, then $(A \setminus B)$ is a category too.
- (iii) The syntactic rule is: If α is an expression of category *A*, and β is an expression of category $(A \setminus B)$, then $\alpha\beta$ is an expression of category *B*.
- (iv) *John* is of category *n*; *walks* is of category $n \setminus s$; and *quickly* is of category $((n \setminus s) \setminus (n \setminus s))$

(As usual, we leave off outer brackets.) An expression of a category $A \setminus B$ is one which, together with an expression of category A , forms an expression of category B . It is the syntactic rule which is responsible for this: it decrees that if we first write down an expression of category A and then follow it with an expression of category $A \setminus B$, then the whole thing is an expression of category B .

Note that the category of a composite expression $\alpha\beta$ is obtained by 'erasing' the category of α in the (derived) category of β . If this is not possible, then the composite expression is said to be ungrammatical. The business of 'erasing' is reminiscent of the way the denominator gets erased in simplifying an expression like $2 \times \frac{1}{2}$. In fact, other notations for derived categories which bring out this analogy with fractions are to be found in the literature, like A/B , or $B:A$.

According to this rudimentary categorial grammar, the expressions in (14) and (15) are of category s ; they are, in other words, sentences.



In general, then, a categorial grammar enables us to determine whether the result of combining expressions of any given categories is itself a grammatical expression, and if so, to determine its category. Conversely, it also enables finding out whether a given composite expression is in category s or not. That is to say, a categorial grammar furnishes an automatic procedure for determining which expressions are sentences and which are not.

The above example is of a *unidirectional grammar*. You can only work in one direction, in the sense that if you have an expression of category $A \setminus B$, then you have to write an expression of type A on the *left-hand* side in order to obtain an expression of type B . Of course, there are many expressions which would result in a new expression if something were written to their right. Take, for example, an adjective like *poor*. Together with *John*, obtained from category n , this gives us *poor John*, likewise in category n . The definition of derived categories can be modified in the following manner so as to allow for expressions like this:

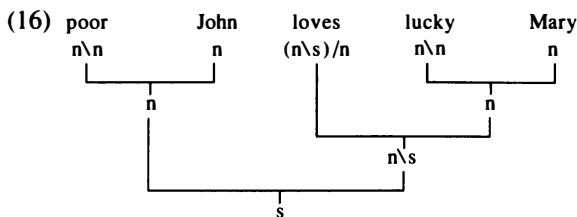
(v) If A and B are categories, then both $(A \setminus B)$ and (A / B) are categories.

Thus we obtain a *bidirectional* categorial grammar. Here the direction of the slash indicates whether concatenation is to occur on the left- or the right-hand side. Such a categorial grammar needs two syntactic rules:

- (vi) (1) If α is in category A and β is in category $A \setminus B$, then $\alpha\beta$ is in category B ;
 (2) If α is in category A/B and β is in category B , then $\alpha\beta$ is in category A .

Both rules deal with the same syntactic operation, namely, concatenation.

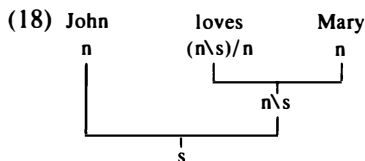
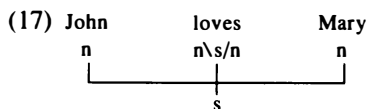
In the example given in (16), both kinds of derived categories are involved:



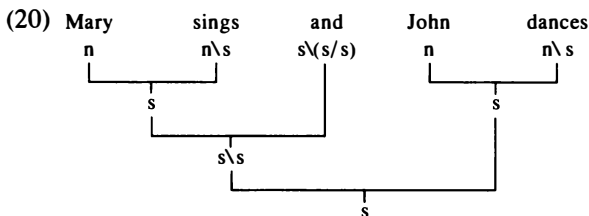
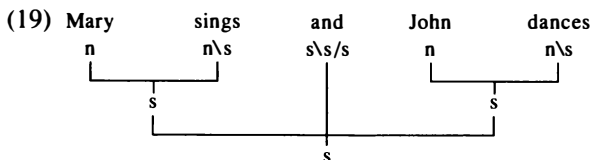
In another variant of categorial grammar, expressions in derived categories may be concatenated with several other expressions simultaneously. In such a categorial grammar, there is a new rule, (vii), for constructing derived categories. The corresponding syntactic rule is then (viii).

- (vii) If A , B , and C are categories, then $A \setminus B/C$ is a category.
 (viii) If α is in category A , β is in category $A \setminus B/C$, and γ is in category C , then $\alpha\beta\gamma$ is in category B .

In this way, the transitive verb *loves* may be categorized as $n \setminus s/n$, instead of as $(n \setminus s)/n$. This means analyzing the sentence *John loves Mary* in the manner indicated in (17) instead of as in (18).



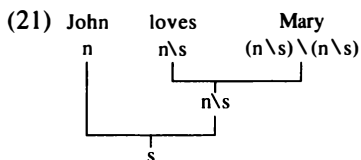
The analysis depicted in (17) attributes less structure to this example sentence than the analysis in figure (18). In (18), *loves Mary* is treated as a single constituent, which is not so in (17). Generally speaking, an analysis like that in (18) will be preferred, so that categories of the form $A \setminus B/C$ will not be needed. An exception is, for example, formed by coordinative conjunctions, for which the categorization $s \setminus s/s$ is to be preferred over $s \setminus (s/s)$ or $(s \setminus s)/s$. Compare the analyses given in (19) and (20).



The analysis of *and* given in (20) predicts that *Mary eats and* is a constituent of the whole sentence, while according to (19) the sentence has *Mary eats*, *and*, and *John drinks* as its three constituents. The latter analysis is certainly the more realistic.

It should be noted that our intuitions about the constituents which an expression has and the categories to which expressions belong are not independent of the semantic functions attributed to the expressions in question. And this is precisely why categorial grammars figure, in particular, in semantically oriented models of grammar.

In 1960 it was proved by Bar-Hillel, Gaifman, and Shamir that a pure categorial grammar is equivalent to a context-free grammar (for the definition of context-free grammar, see vol. 1, chap. 7). This applies to both variants, unidirectional and bidirectional. The equivalence here is a *weak equivalence*: all three types of grammar can be used to generate one and the same language. They do not, however, all do this in the same manner, they are not *strongly equivalent*. Thus they do not all attribute the same structures to every expression in that language. In the cases of unidirectional and bidirectional grammars, this is easily seen. A unidirectional categorial grammar will always distinguish more lexical elements than a bidirectional grammar, since any single lexical element which can appear in different positions will necessarily have to be placed in more than one category. This is illustrated in (21), which represents an analysis of *John loves Mary* in a (left) unidirectional grammar.



Proper names can appear both to the left and to the right of a transitive verb. In a bidirectional grammar, the names in both positions may belong to the same category, as is apparent from (17) and (18). In a unidirectional grammar, on

the other hand, proper names appearing in these positions must be of different categories, since we are only able to do the analysis in one direction. So it is that in figure (21), for example, *Mary* is in category $(n \setminus s) \setminus (n \setminus s)$, while *John* is in category n . But as the name *Mary* can of course also appear to the left of the verb, as in *Mary loves John*, it is clear that *Mary* will have to be placed in at least two different categories. The same applies to all expressions which can appear at different places in a sentence.

The difference between a bidirectional categorial grammar and a context-free grammar is in essence the following: A bidirectional categorial grammar always indicates which of a given pair of constituents is dependent on the other, whereas a context-free grammar need not always provide this information.

Below is an example of a simple context-free grammar and a bidirectional categorial equivalent:

Context-free grammar

$S \Rightarrow NP VP$

$NP \Rightarrow N$

$VP \Rightarrow V NP$

$Adj \Rightarrow poor, lucky$

$N \Rightarrow John, Mary$

$N \Rightarrow Adj N$

$V \Rightarrow kisses, loves$

Bidirectional categorial grammar

Expressions of category n : *John, Mary*

Expressions of category $(n \setminus s) / n$: *kisses, loves*

Expressions of category n/n : *poor, lucky*

Exercise 5*

- (a) Convert the following context-free grammar ('|', distinguishes, as usual, different options for rewriting a symbol) into a bidirectional categorial grammar which has as its basic categories S (for sentences), CN (for common noun phrases), and T (for full noun phrases):

$S \Rightarrow NP VP$

$VP \Rightarrow V_{intrans} | V_{trans} NP$

$NP \Rightarrow PropN | Det N$

$N \Rightarrow Adj N$

$Det \Rightarrow the, a(n)$

$PropN \Rightarrow John, Peter$

$N \Rightarrow man, horse$

$V_{intrans} \Rightarrow walks, swears$

$V_{trans} \Rightarrow eats, makes, curses$

$Adj \Rightarrow green, big, honest$

(b) Try to find English expressions which may serve as examples of expressions of the following categories:

- (i) $(T \setminus S) \setminus (T \setminus S)$
- (ii) $((T \setminus S) \setminus (T \setminus S)) / T$
- (iii) $T \setminus (T / CN)$
- (iv) $(T \setminus S) / (CN / CN)$

What rules need to be added, or modified in the context-free syntax given in (a), for such expressions to be incorporated?

Exercise 6

Of what category are expressions such as *almost* and *at most* as they occur in phrases such as *almost all* and *at most two*? Specify the corresponding type and give a description of the kind of semantic object denoted by these expressions.

Exercise 7

Suppose that every lexical element of a certain language belongs to only one category. Is it possible for a unidirectional categorial grammar to produce complex expressions of the language which are structurally ambiguous? And how about a bidirectional one?

4.3.3 The Descriptive Adequacy of Categorical Grammar

As was mentioned above, pure categorial grammars are equivalent to context-free grammars. This was proven in the beginning of the sixties, during the heyday of transformational grammar, when the belief that we need grammars of greater generative capacity than context-free ones for the description of natural language was firmly established. This is certainly one of the reasons why categorial grammar has had a hard time acquiring a respectable status among syntactically oriented linguists. This has changed, but only recently. Some new developments that helped to bring about this change will be introduced in chapter 7.

Exactly why were context-free grammars, and in their wake pure categorial ones, believed to be descriptively inadequate? The reasons mainly concern various constructions which such grammars are not supposed to be able to deal with, at least in an intuitively adequate way. By way of illustration we discuss three of these phenomena very briefly.

First, there are discontinuous constituents. Compare sentences (22) and (23):

(22) The job was quickly finished.

(23) The job was finished quickly.

In (22), the constituent *was finished* occurs discontinuously, that is, it is interrupted by another expression. This contrasts with its continuity in (23). In a categorial grammar in its pure form, this presents a problem. The rules do not allow any part of a constituent to be separated from the rest, since the only operation they use is the simple concatenation of two strings of symbols. Hence, we are forced to consider *was* and *finished* as separate lexical items and to place each of them in (at least) two different categories, so that they can form both continuous and discontinuous constituents. Sentence (24) gives another example of this phenomenon:

(24) John never calls up Mary, so she calls him up instead.

Here we have both a continuous and a discontinuous occurrence of the constituent *calls up*, in the first and second clause, respectively. Here too, in a categorial grammar there is no choice but to classify *calls up*, *calls*, and *up* separately, as three distinct lexical items. While this does not constitute any formal problem, it certainly seems not to accord with the intuition that it is one and the same constituent which appears in the two clauses of (24), even if it is continuous in the first and discontinuous in the second.

A second phenomenon which presents problems for pure categorial grammars and context-free grammars centers around the intuition that sentence (25) means the same as sentence (26):

(25) John loves Mary, and Jack, Jill.

(26) John loves Mary, and Jack loves Jill.

This relation between (25) and (26) could be accounted for in a number of different ways. One view is that (25) is derived from (26) by leaving out the word *loves* in the second conjunct. Another conjecture is that the 'missing part' of (25) gets filled in during the process of interpretation. Either way, leaving out a constituent or filling one in introduces context-dependency into the picture, since the piece to be left out or filled in must always be present somewhere else in the structure (of course, this is not the only condition). Such straightforward context-dependent processes clearly fall outside the scope of context-free grammars and pure categorial ones. Notice, however, that these views about what goes on are not the only ones that are possible: there is no proof that an adequate description of the phenomena in question must involve context-dependency of this kind.

A third phenomenon which illustrates that the limited generative capacity of context-free grammars and pure categorial grammars may lead to unintuitive results is that of *word order*. Both kinds of grammar decree a fixed word order and hence seem to fail as adequate descriptive tools for languages in which it is not word order but, for example, a case system which determines grammatical relations, or even for languages where the word order of

main clauses differs from the word order of subordinate clauses, as in some of the Germanic languages.

Objections such as these have convinced people for a long time that context-free grammars are inadequate for giving a descriptively adequate description of natural languages which accords with our intuitions about constituency and the like. This verdict has been extended to pure categorial grammars, the grammars of the kind introduced in §4.3.2.

For quite some time these conclusions have been unchallenged. Yet categorial grammars remain attractive for those people who are interested in semantics, since they provide a simple account of the correlation between the syntactic categories of expressions and their semantic functions. It is for this reason that categorial grammar functions prominently in various models of 'logical grammar' (see chapter 6), the first of which were developed in the seventies. A number of different ways to deal with those problems that were mentioned above have been proposed.

One way, which followed the lead of transformational grammar, is to include a categorial grammar as the input of a transformational component. The base component of a transformational grammar is generally a context-free grammar, and the idea is to replace it by a categorial one. This suggestion was made by John Lyons (1968), and David Lewis (1972) has elaborated on it. The idea is that the phenomena discussed above can be accounted for by means of transformations. But note that if we do this, all of the objections which may be made about the excessive generative power of transformational grammars apply here as well.

A second possibility is to increase the generative power of categorial grammar itself. In pure categorial grammars, concatenation is the only syntactic operation which may be applied. It is possible to increase the production, as it were, of the syntactic rules by including other syntactic operations. This second option was the one followed by Richard Montague (1970, 1973). We shall return to this in chapter 6.

These two approaches accept traditional wisdom concerning the descriptive inadequacy of context-free grammars. But lately, people have begun to question the alleged facts: are natural languages really not context-free, that is, can natural languages not be described by means of a context-free mechanism? This discussion has unmasked many a proof as unsound, and at the moment the question is generally thought to be open (see Pullum and Gazdar 1982 and Savitch et al. 1987, for an overview of various arguments, both old and new). Moreover, the general feeling among quite a number of linguists nowadays is that whatever the final answer to this question may be, trying to define a context-free grammar for a natural language is in itself a significant and worthwhile undertaking, if only to pin down where exactly the non-context-freeness, if such there be, comes in. Also, the increasing interest in computationally adequate and efficient models of language and the significant lack of success of

transformational grammarians in finding any real constraints on the generative power of transformational grammars has been a stimulating influence. Various models of grammar have been devised in this spirit, the most influential being that of 'generalized phrase structure grammar' (Gazdar 1982; Gazdar, Klein, Pullum, and Sag 1985) and that of 'lexical functional grammar' (Kaplan and Bresnan 1982).

As for categorial grammar, similar developments have taken place there. In this context it is worth pointing out that arguments against the descriptive adequacy of pure categorial grammars often involve an appeal, implicit or explicit, to intuitions about the constituent structure of expressions, about categories of lexical expressions, and so on. It turns out that categorial grammars can be enriched in certain ways which allow them to deal with the phenomena mentioned above, but often in a rather unorthodox manner. According to some, this only shows that, for example, the notion of constituent structure is much more of a theoretical notion than has been acknowledged, and that hence arguments appealing to it are theoretically biased. Some of the techniques that have been introduced in categorial grammar over the last few years will be discussed in detail in §7.3. We refer the reader for references to the relevant literature in the same section.

4.3.4 Categorial Grammar and the Theory of Types

In this section we will make a few observations on the relationship between the syntax of type-theoretical languages and categorial grammar.

First, we draw attention to the similarity between the definition of types and the way categories are defined in categorial grammar. A finite number of basic types is specified, and there is a rule saying how derived types may be built up from these. The notation for derived types differs from what we gave for derived categories, but as is apparent from (27), the difference is not essential.

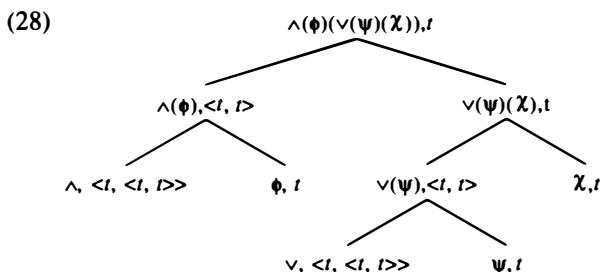
(27)	<div style="display: inline-block; vertical-align: middle; text-align: center;"> John n └──────────┘ s └──┘ n\s </div> <div style="display: inline-block; vertical-align: middle; text-align: center; margin-left: 20px;"> swims n\s </div>	$(S_{(e,t)}(j_e)),$
	<div style="display: inline-block; vertical-align: middle; text-align: center;"> Swimming n\s └──────────┘ s └──┘ (n\s)\s </div> <div style="display: inline-block; vertical-align: middle; text-align: center; margin-left: 20px;"> is healthy (n\s)\s </div>	$(H_{((e,t),t)}(S_{(e,t)})),$
	<div style="display: inline-block; width: 45%;">Categorial grammar</div> <div style="display: inline-block; width: 45%;">Theory of types</div>	

Essential differences are, however, to be found for the other two principles of categorial grammar.

Thus the theory of types makes use of another syntactic operation besides concatenation, namely, an operation introducing brackets. The latter serve to fix the scope relations between various kinds of expressions. As a result of

this bracketing, expressions of type-theoretical languages lack the ambiguity which is to be found in many expressions in natural languages.

There is another respect in which the syntax of the theory of types differs from that produced by a categorical grammar. Not all expressions in the language are placed in a particular category, in this case a particular type. The quantifiers, the connectives, and the relation of identity are introduced *syn-categorematically*. That is to say, they are not treated as lexical items of a particular type. In the case of connectives, it should be noted that this is not necessary. (This matter is discussed in §2.7 of volume 1.) The negation \neg might, for example, have been introduced as an expression of type $\langle t, t \rangle$, since it results, when placed before a formula, in an expression of type t in a new formula. Analogously, the conjunction \wedge might have been introduced as an expression of type $\langle t, \langle t, t \rangle \rangle$. Thus expressions would be generated in a notation with the connectives written as operators in front of their arguments instead of between them, as is more usual. An example: the formula $\phi \wedge (\psi \vee \chi)$ would appear in this notation as $\wedge(\phi)(\vee(\psi)(\chi))$. The construction tree for the latter formula is given in (28) (each expression is followed by its type).



In this notation, the brackets are in fact superfluous. We might just as well write $\wedge \phi \vee \psi \chi$. This notational variant without brackets is known as *Polish notation*. Things become a little more complicated than this in the case of identity. Since two expressions of any type a may be linked up by means of the identity relation $=$ as a formula, the symbol $=$ would have to be treated as an expression of every type of the form $\langle a, \langle a, t \rangle \rangle$. What this means is that for every type $\langle a, \langle a, t \rangle \rangle$, a separate identity relation of this type must be introduced. In the case of the quantifiers, a categorematic introduction runs up against considerable difficulties. The obvious thing would be to treat quantifiers as expressions which turn formulas into other formulas. This approach fails, however, as we shall now see.

According to this idea, quantifiers would be treated as expressions of type $\langle t, t \rangle$. In this case, however, the correspondence between the type of an expression and its semantic interpretation leads to insurmountable difficulties. As we have already seen, an expression of type $\langle a, b \rangle$ is interpreted as a function from D_a to D_b . In the special case of expressions of type $\langle t, t \rangle$, the seman-

tic interpretation is thus a function from truth values to truth values. Now here is a list of all four functions from $\{0, 1\}$ into $\{0, 1\}$:

- (i) The function mapping both truth values onto 1
- (ii) The function mapping each truth value onto the other
- (iii) The function mapping both truth values onto 0
- (iv) The function mapping each truth value onto itself

The interpretation of a formula like $\forall x P(x)$ would then be the result of applying the interpretation of $\forall x$ to the interpretation of $P(x)$. Now the interpretation of $P(x)$ is a truth value. The interpretation of $\forall x$ would, in view of its syntactic type, have to be one of the above four functions from truth values into truth values. What this means is that the truth value of a formula $\forall x P(x)$ would depend only on the truth value of $P(x)$. The truth value of $P(x)$, a formula with a free variable x , is determined by whether or not $g(x)$, the particular entity which the assignment assigns to x , has the property expressed by the predicate P . But that is, of course, not what the truth of the statement that everything has the property should depend on. In order to determine the truth value of $\forall x P(x)$, we need to know whether or not all of the entities in the domain have the property expressed by P . We need to know, in other words, whether or not $P(x)$ is true with respect to every assignment g . The truth value of $P(x)$ with respect to a single assignment is not enough. The conclusion to be drawn from this is that the semantic interpretation of a quantifying expression like $\forall x$ cannot be identified with one of the four possible functions of type $\langle t, t \rangle$. And this means that syntactically the expression $\forall x$ cannot be treated as an expression of type $\langle t, t \rangle$. So the obvious way of introducing quantifiers categorically is unworkable. This is not to say, however, that there is no way at all of introducing the quantifiers syncategorematically. We will return to this matter in §4.4.3.

4.4 λ -Abstraction.

4.4.1 *The λ -Operator*

We shall now extend the theory of types as described in §4.2 by adding a new kind of expression, the lambda operator λ . This operator enables us to form new expressions from expressions by abstracting over variables. In doing so, we increase the expressive power of the theory of types in a way which will prove to be of particular interest in the analysis of natural language. Before we go on to introduce the λ -operator, however, let us first briefly consider some of the forms and constructions of a natural language like English which make such an operator desirable.

Let us begin by discussing the translation of the following sentence into the theory of types:

- (29) Jogging is healthy.

A translation of this sentence into the theory of types may be obtained as follows. Given that *jogging* expresses a property of individuals, the expression may be translated as a predicate constant J of type $\langle e, t \rangle$. *Healthy* expresses, at least in this context, a property of properties of individuals and is as such to be rendered as a constant \mathcal{H} of type $\langle \langle e, t \rangle, t \rangle$. The whole of (29) is then to be translated as the formula $\mathcal{H}(J)$, which expresses the proposition that jogging is a healthy activity.

Let us now turn to an apparently quite analogous sentence and see how we would translate it:

(30) Not smoking is healthy.

Once again we have a sentence in which a property, that of being healthy, is attributed to a property of individuals, that of not smoking. The translation of *healthy* presents no new problems; as with (29), it is translated as a constant \mathcal{H} of type $\langle \langle e, t \rangle, t \rangle$. The problem lies with the translation of *not smoking*. Let us begin by translating *smoking* as a constant S of type $\langle e, t \rangle$. Now the problem is that negation cannot be applied to this constant S , since the negation symbol \neg may only be applied to formulas and not to expressions of other types. The expression *not smoking* could, of course, be translated as a whole as a constant N of type $\langle e, t \rangle$. But this would not do justice to the fact that the meaning of *not smoking* is composed of the meanings of the words *not* and *smoking*. In the English language, there is a productive process enabling the word *not* to be combined with expressions of various types, thus giving rise to new, composite expressions. Parallel to this, the meanings of these composite expressions are built up from the meanings of the expressions of which they are composed. We now wish to build a similar process into the theory of types, so as to obtain a better correspondence with natural language.

Here is a second example of a productive process of this kind working in natural language: the coordination of predicates. Consider example (31).

(31) Drinking and driving is unwise.

In (31) we find a composite expression, *drinking and driving*, which expresses a property of individuals. This expression is formed from the conjunction *and* together with the predicates *drinking* and *driving*. In translating (31), we encounter difficulties similar to those we found with (30). Since the conjunction \wedge may be used only to conjoin formulas, it cannot be used on predicates. We could, of course, just decide to translate *drinking and driving* as a single unanalyzed constant of type $\langle e, t \rangle$, but once again this would be to ignore the fact that the meaning of the composite expression is built up from the meanings of the expressions *drinking*, *and*, and *driving* of which it is composed. And the same remarks apply to *unwise* as to *not smoking* in (30). *Unwise* is a composite second-order predicate which is the result of a productive process being applied to the second-order predicate *wise*.

Yet another example is to be found in the reflexive predicates, like *to ad-*

mire oneself. If this predicate were to be dealt with independently of the predicate *to admire*, then we would be at a loss to explain the close links between the properties they express. We would, for example, not be able to account for the equivalence of (32) and (33):

(32) John admires John.

(33) John admires himself.

The process by which the composite predicate *to admire oneself* is obtained from the predicate *to admire* is apparently such that *admiring oneself* may be predicated of an entity d just in case *admiring* may be predicated of the pair $\langle d, d \rangle$.

These, then are just a few of the examples which suggest extending the theory of types.

In order to account for constructions like these, the following new rule is added to the syntax of the theory of types as given in definition 2 in §4.2.2:

(vii) If α is expression of type a in L , and v is a variable of type b , then $\lambda v \alpha$ is expression of type $\langle b, a \rangle$ in L .

Let us turn to an example. Let W be a constant of type $\langle e, t \rangle$, and let x be a variable of type e . Then $W(x)$ is a formula in which x appears as a free variable. According to clause (vii), we may form the expression $\lambda x(W(x))$ from $W(x)$. Since $W(x)$ is of type t and x is of type e , this new expression $\lambda x(W(x))$ is of type $\langle e, t \rangle$. We say that the expression $\lambda x(W(x))$ has been formed from the expression $W(x)$ by *abstraction over* the free variable x . We say that the free occurrences of the variable x in α are *bound* in $\lambda x \alpha$ by the λ -operator λx .

A word about brackets is in order here. Recall that the brackets around $W(x)$ in $\lambda x(W(x))$ are introduced by the application of W to x (see clause (ii) of definition 2). Here they cannot be left out, since the result of doing so, $\lambda x W(x)$, is an expression with a different syntactic structure. It is what we get if we first abstract (vacuously) over x in W , and next apply the result of that, $\lambda x W$, to x . Now it happens that in this particular case leaving the brackets off would not be harmful semantically, since as the interpretation of λ -abstraction will make clear, $\lambda x W(x)$ and $\lambda x(W(x))$ happen to be equivalent. However, it is not generally the case that if we apply λ -abstraction to a complex expression formed by functional application, the outer brackets around the latter can be left off without any semantic harm being done. A case in point is $\lambda x(A(x)(x))$, to be discussed later on. Leaving out the brackets around $(A(x)(x))$ would yield $\lambda x A(x)(x)$, and this expression has a different meaning.

What interpretation is now to be given to expressions formed in this manner by means of (vii)? Consider the example of $\lambda x(W(x))$ once again. As we have already seen, this composite expression is of type $\langle e, t \rangle$. An expression of this type is interpreted as a function from entities into truth values. So the interpretation of $\lambda x(W(x))$ is of the same type as the interpretation of the constant

W. It will become apparent from the clause defining the interpretations of expressions formed by λ -abstraction that the interpretations of $\lambda x(W(x))$ and W not only are of the same type but are identical. To put it generally, then, the interpretation of a λ -abstraction $\lambda x_b \alpha_a$ (the subscripts refer to the types of the expressions) is a function, namely, one belonging to the set of functions $D_a^{D_b}$. For this reason, λ -abstraction is also referred to as *functional abstraction*. Note that if α is of type t , then the interpretation of $\lambda x_b \alpha$ is an element of the set $\{0, 1\}^{D_b}$ of functions. It is, in other words, the characteristic function of a set. Given that sets can be identified with their characteristic functions, $\lambda x_b \alpha$ can serve as notation for a set. In these cases, the λ -operator is also referred to as the *set abstractor*.

Now we add the following clause to definition 4 in §4.2.3, which lays down how expressions of a type-theoretical language L are to be interpreted with respect to a given model M and assignment g :

- (vi) If $\alpha \in WE_a^L$ and $v \in VAR_b$, then $\llbracket \lambda v \alpha \rrbracket_{M,g}$ is that function $h \in D_a^{D_b}$ such that for all $d \in D_b$: $h(d) = \llbracket \alpha \rrbracket_{M,g[v/d]}$

By way of illustration, let us turn back to the example of $\lambda x(W(x))$, in which W is a constant of type $\langle e, t \rangle$ and x is a variable of type e . Clause (vi) then says that $\llbracket \lambda x(W(x)) \rrbracket_{M,g}$ is the function h derived from the set $D_t^{D_e}$ of functions, such that for all $d \in D_e$ we have: $h(d) = \llbracket W(x) \rrbracket_{M,g[x/d]}$. We know that $D_t^{D_e} = \{0, 1\}^{D_e}$ = the set of characteristic functions of sets of entities. So h is the characteristic function of some set of entities. This function is defined as follows: for all $d \in D$ we have $h(d) = \llbracket W(x) \rrbracket_{M,g[x/d]}$. Thus $h(d) = 1$ iff $\llbracket W(x) \rrbracket_{M,g[x/d]} = 1$. We know that $\llbracket W(x) \rrbracket_{M,g[x/d]} = 1$ iff $\llbracket W \rrbracket_{M,g[x/d]}(\llbracket x \rrbracket_{M,g[x/d]}) = 1$ iff $I(W)(d) = 1$. Thus: $h(d) = 1$ iff d has the property expressed by W . This means that h is the characteristic function of the set of all entities having the property expressed by W . And this means that the interpretation of $\lambda x(W(x))$ is just the same as the interpretation of W . For $\llbracket W \rrbracket_{M,g} = I(W)$, while for all d in D we have $h(d) = 1$ iff $I(W)(d) = 1$, so that $h = I(W)$. We can therefore say that $\lambda x(W(x))$ and W are equivalent.

Let us now consider how the examples of composite expressions referred to at the beginning of this section can be analyzed in the theory of types enriched with the λ -operator. Sentence (30), *Not smoking is healthy*, can be translated as follows. Let S be the translation of *smoking*. This is a constant of type $\langle e, t \rangle$. And let x be a variable of type e . Now $S(x)$ is a formula, an expression of type t . Applying negation to this formula, we obtain a second formula: $\neg S(x)$. And applying λ -abstraction over the variable x , we obtain the expression $\lambda x \neg S(x)$. The latter is of type $\langle e, t \rangle$, since we abstracted over a variable of type e in an expression of type t . The expression $\lambda x \neg S(x)$ is therefore a predicate of entities, just like S . In order to find out which property is expressed by this predicate, we must determine its interpretation: $\llbracket \lambda x \neg S(x) \rrbracket_{M,g}$ is the function $h \in D_t^{D_e}$ such that for all $d \in D_e$ we have $h(d) = \llbracket \neg S(x) \rrbracket_{M,g[x/d]}$. That is: $h(d) = 1$ iff $\llbracket \neg S(x) \rrbracket_{M,g[x/d]} = 1$ iff $\llbracket S(x) \rrbracket_{M,g[x/d]} = 0$ iff $\llbracket S \rrbracket_{M,g[x/d]}(\llbracket x \rrbracket_{M,g[x/d]}) = 0$

iff $I(S)(d) = 0$. Thus: $h(d) = 1$ iff d does not have the property expressed by S . This means that $\llbracket \lambda x \neg S(x) \rrbracket_{M,g}$ is the characteristic function of the set of non-smokers, that is, it expresses the property 'not smoking'. As such, it is a suitable formal translation of the expression *not smoking*. The expression of the whole of (30) may now be obtained by applying the second-order predicate \mathcal{H} , the translation of the word *healthy* as it appears in (29) and (30), to $\lambda x \neg S(x)$. The result is then $\mathcal{H}(\lambda x \neg S(x))$.

In (31), *Drinking and driving is unwise*, there are two composite predicates: *drinking and driving*, a first-order predicate, and *unwise*, in this case a second-order predicate. The first of these composite predicates may be translated as follows. Let D_1 and D_2 be the translations of *drinking* and *driving*, respectively. Both of these are constants of type $\langle e, t \rangle$. And let x be a variable of type e . Applying λ -abstraction over x in the conjunction $D_1(x) \wedge D_2(x)$, we obtain $\lambda x(D_1(x) \wedge D_2(x))$. This expression is once again of type $\langle e, t \rangle$. That $\lambda x(D_1(x) \wedge D_2(x))$ expresses the property of drinking and driving is apparent from its interpretation: $\llbracket \lambda x(D_1(x) \wedge D_2(x)) \rrbracket_{M,g}$ is the function $h \in D_1^{D_2}$, such that for all $d \in D$, we have $h(d) = 1$ iff $\llbracket D_1(x) \wedge D_2(x) \rrbracket_{M,g[x/d]} = 1$ iff $\llbracket D_1(x) \rrbracket_{M,g[x/d]} = 1$ and $\llbracket D_2(x) \rrbracket_{M,g[x/d]} = 1$ iff $I(D_1)(d) = I(D_2)(d) = 1$. A translation of the composite second-order predicate *unwise* is obtained by λ -abstraction over a variable of the same type as first-order predicates. As in the above, let W be the translation of *wise*. This constant is of type $\langle \langle e, t \rangle, t \rangle$. Let X be a variable of type $\langle e, t \rangle$. Then $\neg W(X)$ is a formula expressing that X does not have the property of properties of 'being (a) wise (property)'. Abstraction over X results in the expression $\lambda X \neg W(X)$. This expression is of type $\langle \langle e, t \rangle, t \rangle$, since we have abstracted over a variable of type $\langle e, t \rangle$ in a formula. This composite second-order predicate expresses a property of properties, namely, the property of 'being an unwise property'. For $\llbracket \lambda X \neg W(X) \rrbracket_{M,g}$ is that $k \in D_1^{D_2} = \{0, 1\}^{\{(0,1)^D\}}$ such that for all functions $h \in D_1^{D_2}$: $k(h) = 1$ iff $\llbracket \neg W(X) \rrbracket_{M,g[X/h]} = 1$ iff $I(W)(h) = 0$, which is to say, iff h does not belong to the set of wise properties. Now the translation of the whole of (31) may be obtained by applying the translation of *unwise* to the translation of *drinking and driving*: $\lambda X \neg W(X)(\lambda x(D_1(x) \wedge D_2(x)))$.

A general strategy for translating predicates composed by means of *not*, *and*, *or*, and the like may be extracted from these examples. Since in the theory of types, connectives and negation may only be applied to formulas, formulas are first formed by applying the original predicates to one or more variables. Then connectives and/or negation can be applied to these formulas, and abstraction over the variables in question then results in the required composite predicate.

Now it is, of course, also possible to treat the composition of predicates without a λ -operator. We shall indicate briefly how this may be done in the case of negation. First the restriction in definition 2 that negation may only be applied to an expression of type t , a formula, must be done away with. Instead we stipulate that if α is an expression of type $\langle a, t \rangle$, then $\neg \alpha$ is an expression

of this type too. (For example, *smoking* and *not smoking* are of type $\langle e, t \rangle$; *wise* and *unwise* are of type $\langle \langle e, t \rangle, t \rangle$). Then definition 4 must be extended too, defining $\llbracket \neg \alpha \rrbracket_{M,g}$, the interpretation of $\neg \alpha$, as that function in $D_{\langle a, t \rangle} = \{0, 1\}^{D_a}$ such that for all $d \in D_a$: $\llbracket \neg \alpha \rrbracket_{M,g}(d) = 0$ if $\llbracket \alpha \rrbracket_{M,g}(d) = 1$, and $\llbracket \neg \alpha \rrbracket_{M,g}(d) = 1$ if $\llbracket \alpha \rrbracket_{M,g}(d) = 0$. Thus, by way of example, $\neg S$ is interpreted as that function which gives value 1 when applied to d just in case the function interpreting S gives 0 as value for d , which is to say, just in case d does not smoke; this is the right result. Similar modifications to definitions 2 and 4 would account for composition of predicates with the other connectives. So why do we need a λ -operator? The advantage of a λ -operator is that it provides a *uniform* treatment not only of these examples but of many others too. For this reason, we will not go any further into the possibility of doing without a λ -operator.

Another example which the λ -operator deals with nicely is that of reflexive predicates, like the *admires oneself* in (33) *John admires himself*. Here too one must first go back to formulas. Let A be a constant of type $\langle e, \langle e, t \rangle \rangle$, the translation of *admires*, and let x be a variable of type e . $A(x)(x)$ is then a formula. From it, the expression $\lambda x(A(x)(x))$ may be obtained by abstraction over x . Here both occurrences of the variable x are bound by the λ -operator. This new expression is of type $\langle e, t \rangle$; hence it is not of the same type as A . Its interpretation is as follows: $\llbracket \lambda x(A(x)(x)) \rrbracket_{M,g}$ is the function $h \in D_t^{D_e}$ such that for all $d \in D$ we have $h(d) = 1$ if and only if $\llbracket A(x)(x) \rrbracket_{M,g[x/d]} = 1$ iff $\llbracket A(x) \rrbracket_{M,g[x/d]}(\llbracket x \rrbracket_{M,g[x/d]}) = 1$ if and only if $\llbracket A \rrbracket_{M,g[x/d]}(\llbracket x \rrbracket_{M,g[x/d]})(\llbracket x \rrbracket_{M,g[x/d]}) = 1$ iff $I(A)(d)(d) = 1$. That is, an entity d has the property expressed by $\lambda x(A(x)(x))$ just in case the pair of entities $\langle d, d \rangle$ stand in the relation expressed by the predicate A .

As was remarked above, leaving off the outer brackets of $(A(x)(x))$ would result in an expression, $\lambda x A(x)(x)$, which not only has a different syntactic structure but also a different interpretation. For $\lambda x A(x)(x)$ is equivalent to $A(x)$, and the meaning of $A(x)$ and that of $\lambda x(A(x)(x))$ are obviously not identical.

Not only the variables over which we abstract but also the expressions in which these occur may be of any type. Let A be as above, and let x and y be variables of type e . We may now abstract over the variable y in the formula $A(y)(x)$. The result, $\lambda y(A(y)(x))$, is of type $\langle e, t \rangle$ and expresses the property of 'being admired by x '. Then we can abstract over x in this new expression so as to obtain the formula $\lambda x \lambda y(A(y)(x))$, which is of type $\langle e, \langle e, t \rangle \rangle$, since it is formed by abstraction over a variable of type e in an expression of type $\langle e, t \rangle$. We thus obtain a two-place predicate expressing the relation of 'being admired by'.

It is important to note that if there are several variables, the *order* in which abstraction takes place makes a difference. If we abstract over the x in $A(y)(x)$ first, we obtain the expression $\lambda x(A(y)(x))$, which expresses the property of 'admiring y '. If we then abstract over the variable y , we obtain the two-place

predicate $\lambda y \lambda x (A(y)(x))$, which, just like the constant A , expresses the relation 'admires'. But if we start with the same formula $A(y)(x)$ and abstract first over y and then over x , what we get is $\lambda x \lambda y (A(y)(x))$, which, as we saw above, expresses a different relation, that of 'being admired by'.

That $\lambda y (A(y)(x))$ and $\lambda x (A(y)(x))$ express different properties is apparent from a comparison of their interpretations: $\llbracket \lambda y (A(y)(x)) \rrbracket_{M,g}$ is the function $h \in D_I^{D_I}$ such that for all $d \in D$: $h(d) = 1$ iff $\llbracket A(y)(x) \rrbracket_{M,g[y/d]} = 1$ iff $(I(A)(d))(g(x)) = 1$, that is, iff $g(x)$ admires d ; $\llbracket \lambda x (A(y)(x)) \rrbracket_{M,g}$ is the $h \in D_I^{D_I}$ such that for all $d \in D$: $h(d) = 1$ iff $\llbracket A(y)(x) \rrbracket_{M,g[x/d]} = 1$ iff $(I(A)(g(y)))(d) = 1$, that is, iff d admires $g(y)$.

Exercise 8*

Let j be a constant of type e ; M of type $\langle e, t \rangle$; S of type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$; and \mathcal{C} of type $\langle \langle e, t \rangle, t \rangle$. Furthermore, x is a variable of type e , and Y a variable of type $\langle e, t \rangle$. Determine which of the following sequences of symbols are well-formed expressions. If an expression is well-formed, give its type.

- (i) $\lambda x (M(x))(\mathcal{C})$
- (ii) $\lambda x (M(x))(j)$
- (iii) $\lambda x M(j)$
- (iv) $S(\lambda x M(x))$
- (v) $\lambda Y (Y(j))(M)$
- (vi) $\lambda x \lambda Y (Y(x))$
- (vii) $\lambda x (M(x)) \wedge M(j)$
- (viii) $\lambda x (M(x) \wedge M(j))$
- (ix) $(S(\lambda Y (Y(x))))(M)$
- (x) $\lambda Y (\mathcal{C}(\lambda x (Y(x))))(M)$
- (xi) $\lambda x (\lambda Y (Y(x))(M))(j)$
- (xii) $\lambda x (\lambda Y (Y(x))(j))(M)$
- (xiii) $\lambda x \lambda Y (Y(x))(j)(M)$
- (xiv) $\lambda Y ((S(\lambda x (M(x))))(j) \wedge \mathcal{C}(Y))(M)$

Exercise 9*

Translate the following expressions into the theory of types. State the translation key.

- (a) To wash yourself properly is important.
- (b) It is healthy to love somebody.
- (c) Forwards or backwards
- (d) To put the queen forwards or backwards leads to checkmate.
- (e) Everything that grows and glows and always restores us again.
- (f) To have forgotten something is to have known something but to not know it now.
- (g) Always to be oneself is impossible.
- (h) To be perfect is to have all good properties.
- (i) To share all your bad properties with Mary
- (j) To be or not to be

4.4.2 λ -Conversion

A λ -abstract like $\lambda x \neg(S(x))$ of type $\langle e, t \rangle$ behaves exactly like any other expression of the same type. This means that this predicate can be used not only in order to represent higher-order predication, as in $\mathcal{H}(\lambda x \neg(S(x)))$, but also in representing first-order predication. Thus it may be applied to a constant of type e , for example j , thus resulting in the formula $\lambda x \neg(S(x))(j)$, which expresses the proposition that the entity to which j refers has the property of 'not smoking'. This is the same as the proposition that this entity does not smoke, which may be represented by means of the equivalent formula $\neg(S(j))$. What we have done here is to leave out the λx in $\lambda x \neg(S(x))$ and to replace free occurrences of x in the remaining part, $\neg(S(x))$, by j . The general notation for the result of replacing all free occurrences of a variable v in an expression β by an expression γ is $[\gamma/v]\beta$. So in this terminology, what we have done is to form $[j/x]\neg(S(x))$ from $\lambda x \neg(S(x))(j)$.

The question now arises as to whether an expression of the form $\lambda v \beta(\gamma)$ can always be reduced to $[\gamma/v]\beta$ with conservation of meaning. This reduction process is known as λ -conversion. So the question to be addressed here is whether λ -conversion always results in an equivalent expression, that is, whether or not (34) can be answered affirmatively:

- (34) *Question:* Does it hold for all v , β , and γ that $\lambda v \beta(\gamma)$ is equivalent to $[\gamma/v]\beta$?

In the above we saw a number of instances confirming (34). In illustrating the clause dealing with the interpretation of λ -abstraction, for example, we saw that $\lambda x(W(x))$ and W are equivalent. This implies that $\lambda x(W(x))(c_e)$ and $W(c_e)$ are equivalent for an arbitrary individual constant c_e . And indeed, $[c_e/x]W(x)$ and $W(c_e)$ are one and the same expression. And it also follows from the interpretation of $\lambda x(A(x)(x))$ that $\lambda x(A(x)(x))(c_e)$ is equivalent to $A(c_e)(c_e)$, that is, to $[c_e/x](A(x)(x))$.

And yet (34) cannot be affirmed unconditionally. This can be made clear by means of a similar situation which turns up in predicate logic. There it is not true that for every ϕ , $\models \forall x \phi \rightarrow \forall y [y/x]\phi$. The reason for this is that it can happen that the variable y which replaces x in ϕ can be bound, in ϕ , by a quantifier $\forall y$ or $\exists y$ already present in ϕ . As an example of a ϕ in which this happens, consider $\exists y Rxy$. It can easily be seen that $[y/x]\phi$ is the formula $\exists y Ryy$, whence $\forall y [y/x]\phi$ is $\forall y \exists y Ryy$, which is equivalent to $\exists y Ryy$. That $\forall x \exists y Rxy$ and $\exists y Ryy$ are not in general equivalent formulas is obvious from the particular case in which R is the relation \neq . The formula $\forall x \exists y (x \neq y)$ is clearly not equivalent to $\exists y (y \neq y)$, since the first is true in any model whose domain contains more than just a single element, while the second sentence is true in no model whatsoever: $\exists y (y \neq y)$ is a contradiction. If we now take $\exists y (x \neq y)$ for our formula β , then $\lambda x \beta(y)$ is $\lambda x \exists y (x \neq y)(y)$ and $[y/x]\beta$ is $\exists y (y \neq y)$. And $\lambda x \exists y (x \neq y)(y)$ and $\exists y (y \neq y)$ are certainly not equivalent expressions. The expression $\lambda x \exists y (x \neq y)$ expresses the property 'there is some-

thing besides x' (the property of not being unique), since $\llbracket \lambda x \exists y (x \neq y) \rrbracket_{M,g}$ is the function which assigns the value 1 to d just in case there is some $d' \neq d$ in the domain, and the value 0 if this is not so. So if D contains more than one element, then the function $\llbracket \lambda x \exists y (x \neq y) \rrbracket_{M,g}$ assigns to each of the elements of D the value 1. Now $\llbracket \lambda x \exists y (x \neq y)(y) \rrbracket_{M,g}$ is the value assigned by $\llbracket \lambda x \exists y (x \neq y) \rrbracket_{M,g}$ to $g(y)$, so this value will also be 1 if the domain contains more than a single element. This contrasts with $\llbracket \exists y (y \neq y) \rrbracket_{M,g}$, which will certainly be 0. From this it follows that $\lambda x \exists y (x \neq y)(y)$ and $\exists y (y \neq y)$ are not equivalent formulas.

In predicate logic there is, however, a condition under which $\forall x \phi$ and $\forall y[y/x]\phi$ turn out to be equivalent, namely, the condition that y is *free for x in ϕ* . This condition may also be turned to our present purposes, although here it must be generalized because of the capacity of the λ -operator to bind variables. See definition 5:

Definition 5

A variable v' is *free for v in the expression β* iff no free occurrence of v in β is within the scope of a quantifier $\exists v'$ or $\forall v'$, or a λ -operator $\lambda v'$.

It turns out that $\lambda v \beta(v')$ and $[v'/v]\beta$ are indeed equivalent if v' is free for v in β . As we have seen, however, some of the expressions γ that we have to deal with are more complex than just a single variable v' . In such cases, λ -conversion is allowed only if all of the free variables in γ are free for v in β . Thus we have theorem 1.

Theorem 1

If all variables which occur as free variables in γ are free for v in β , then $\lambda v \beta(\gamma)$ and $[\gamma/v]\beta$ are equivalent.

We will not prove this theorem here. The proof is by induction on β .

Theorem 1 is particularly useful in that it enables long and complex expressions to be reduced to shorter, more readable ones which mean the same thing. It also facilitates a simple demonstration of how the order in which an expression is applied to two others can matter. If, for example, $\lambda x \lambda y (A(y)(x))$ is first applied to j and then to m , then we obtain $\lambda x \lambda y (A(y)(x))(j)(m)$. This determines the order in which the λ -conversion must take place. The functor $\lambda x \lambda y (A(y)(x))$ is first applied to the argument j , and the result, $\lambda x \lambda y (A(y)(x))(j)$, which is again a functor, to the argument m . (Writing the outer brackets around $\lambda x \lambda y (A(y)(x))(j)$ would give $(\lambda x \lambda y (A(y)(x))(j))(m)$, which makes this more clear, though given that $A(y)(x)$ is of type t , it is not necessary to do so.) The first reduction step, then, is to apply λ -conversion in $\lambda x \lambda y (A(y)(x))(j)$, which reduces the entire expression $\lambda x \lambda y (A(y)(x))(j)(m)$ to $\lambda y (A(y)(j))(m)$. The next step is to substitute m for y , as a result of which $A(m)(j)$ is obtained. This agrees with the meaning of $\lambda x \lambda y (A(y)(x))$. This expression is, as we saw above, to be interpreted as the relation 'is admired by'.

Application to the constant j results in $\lambda y(A(y)(j))$, which expresses 'being admired by j '. And application of this to m results in $A(m)(j)$, which means the same as ' j admires m '. If, on the other hand, $\lambda x\lambda y(A(y)(x))$ is first applied to m and then to j , the result is $\lambda x\lambda y(A(y)(x))(m)(j)$. A first λ -conversion has $\lambda y(A(y)(m))(j)$ as its result, and a second results in $A(j)(m)$.

There are cases in which theorem 1 is not directly applicable, but even then the situation is not hopeless. Let us return, for example, to the formula $\lambda x\exists y(x \neq y)(y)$, on which λ -conversion is not directly sanctioned by theorem 1. The problem lies with the quantifier $\exists y$. We can get rid of this quantifier by replacing $\exists y(x \neq y)$ with the formula $\exists z(x \neq z)$, which has the same meaning. This is because z is free for y in $x \neq y$. Now we have obtained $\lambda x\exists z(x \neq z)(y)$, and y is free for x in $\exists z(x \neq z)$, so that λ -conversion now results in $\exists z(y \neq z)$. It is clear that the latter formula has the right meaning.

Two theoretical questions may be asked concerning λ -conversion in logical systems like this. First, does one, in applying successive λ -conversions to a complex formula, always reach a stage at which λ -conversion is no longer possible? The answer is that one does; λ -conversion cannot be applied indefinitely. This is by no means a trivial result. There are expressions—though as it happens they cannot be formed in the version of the theory of types with λ -abstraction dealt with here—in which unlimited λ -abstraction would be possible. One example of such an expression is $\lambda x(x(x))(\lambda x(x(x)))$, which is a kind of Russell paradox in lambda form: $[\lambda x(x(x))/x](x(x))$ is $\lambda x(x(x))(\lambda x(x(x)))$.

The second question arises from the fact that in complex expressions, λ -conversion can often be applied at any of a number of different points. So the question is this: from the result reported above, we know that λ -conversion must at some point come to a halt. Is this point independent of where we begin? Do different series of λ -conversions always give the same result? The answer to this question is affirmative, but qualified. One must allow bound variables to be replaced by other variables, just as we have done in the above by substituting z for y in $\exists y(x \neq y)$. And the final result is unique only to the extent that no notice is taken of such variations in bound variables. This result, too, is anything but trivial. It does not hold, for example, for the intensional theory of types to be discussed in chapter 5.

Exercise 10*

Let j be a constant of type e ; M of type $\langle e, t \rangle$; and A of type $\langle e, \langle e, t \rangle \rangle$. Furthermore, x and y are variables of type e , and Y is a variable of type $\langle e, t \rangle$. Reduce the following expressions as much as possible by means of λ -conversion:

- | | |
|--|--|
| (i) $\lambda x(M(x))(j)$ | (v) $\lambda x\forall y(A(x)(y))(y)$ |
| (ii) $\lambda Y(Y(j))(M)$ | (vi) $\lambda Y(Y(j))\lambda x(M(x))$ |
| (iii) $\lambda x\lambda Y(Y(x))(j)(M)$ | (vii) $\lambda Y\forall x(Y(x))\lambda y(A(x)(y))$ |
| (iv) $\lambda x\forall y(A(x)(y))(j)$ | |

Exercise 11

It is possible to define the usual connectives and quantifiers using just identity and the λ -operator. As an auxiliary notion we give the definition of the tau-ology T :

$$T =_{\text{def}} \lambda p(p) = \lambda p(p)$$

where p is a variable of type t . Give the definitions of negation, conjunction, and the universal quantifier.

4.4.3 The λ -Operator and Compositionality

We have seen how the λ -operator provides us with a formal translation for various expressions and constructions from natural language. Now we shall briefly indicate how it enables the *process* of translation to proceed more systematically, in a sense which will soon become clear. The main purpose of translating a natural language into a formal language is to obtain a semantic interpretation of the former via the semantics of the latter. For the meaning of a correct translation is the same as the meaning of what is translated. In order for the semantic interpretation to be satisfactory, however, it is necessary that the process of translation comply to certain requirements. Among these, two important requirements are that the process be *explicit* and that it can be *specified in a finite manner*. Just as in syntax, the requirement that it be explicit means that it may not in any way rely on the knowledge or creativity of the translator: it must be such that it could, at least in principle, be automated. Furthermore, the translation process, though essentially finite, must translate a potentially infinite number of sentences.

One way of doing this is to stay close to the syntactic rules of the natural language in question, which are finite in number. Here we assume that translations are available for all of the lexical elements of the language, which are finite in number. Then for each syntactic rule saying how expressions may be combined to form composite expressions we formulate a parallel rule, which says how the translations of these expressions may be combined to give the translations of the composite expressions. This is strongly reminiscent of the principle of the compositionality of meaning, and in fact the resemblance is not just a coincidence. For given that translations form semantic representations of the translated, the compositionality of meaning brings the *compositionality of translation* in its train.

Now it should be clear that the way we have translated natural language sentences into standard predicate logic up until now is neither explicit nor compositional. Consider sentence (35) for example, which would most naturally be translated as (36):

(35) John smokes and drinks.

(36) $S_j \wedge D_j$

This way of translating is not explicit, in that essential use is made of our knowledge of the meaning of (35), in particular our knowledge of the fact that (35) expresses a conjunction of two sentences. And it is not compositional, in that no account is given of how the translation (meaning) of (35) is built up from the translations (meanings) of *John* and *smokes and drinks*, or of how the translation of the *smokes and drinks* is built up from the translations of *smokes* and *drinks*. Indeed, if all of the translations have to be put into a predicate logical language, it is not possible to account for this, since the phrase *smokes and drinks* cannot be translated into such a language. As we have seen, however, adding a λ -operator makes such a translation possible. Given that the lexical elements of (35) are once again rendered as follows: *John*: j , *smokes*: S , *drinks*: D , the phrase *smokes and drinks* can be rendered as (37), while the whole of (35) translates as (38).

$$(37) \lambda x(S(x) \wedge D(x))$$

$$(38) \lambda x(S(x) \wedge D(x))(j)$$

Applying λ -conversion to (38) the familiar sentence (36) may be obtained (give or take a few brackets). So the result is the same, but the way it was reached is preferable.

As a second example, take the translation of sentences containing quantified terms. A sentence like (39), for instance, is translated as (40)

$$(39) \text{Every man walks.}$$

$$(40) \forall x(M(x) \rightarrow W(x))$$

This translation into standard predicate logic is not compositional either. We do not have a separate translation for *every man*. Using the λ -operator, however, it is possible to give the phrase a translation of its own (and thus its own meaning) in the theory of types. *Every man* is then translated as (41):

$$(41) \lambda Y \forall x(M(x) \rightarrow Y(x))$$

Here Y is a variable of type $\langle e, t \rangle$, so that (41) is an expression of type $\langle \langle e, t \rangle, t \rangle$. As such, its interpretation is (the characteristic function) of some set of sets of individuals. In other words, (41) is a second-order predicate expressing 'Y is a property which is true of all men'. Applying (41) to the predicate W , the translation of *walk*, we obtain (42) as our translation of (39):

$$(42) \lambda Y \forall x(M(x) \rightarrow Y(x))(W)$$

This formula expresses the proposition that the property of being something which walks is among the properties of all men. This of course means 'every man walks'. Once again, (42) reduces by λ -conversion to the more familiar (40). So it is not the result of the translation which is better but the way we arrive at it. The λ -operator enables us to translate (39) compositionally, from the translations of *every man* and *walk*.

Even the determiner *every* can be given a separate translation, namely:

$$(43) \lambda X \lambda Y \forall x (X(x) \rightarrow Y(x))$$

The determiner thus translates as an expression of type $\langle\langle e, t \rangle, \langle\langle e, t \rangle, t \rangle\rangle$. Its interpretation may be considered a two-place second-order relation, a relation between sets of individuals. A set *A* bears this relation to a set *B* just in case all elements of *A* are elements of *B*. Applying (43) to a one-place first-order predicate, for example, *M*, we obtain the one-place second-order predicate (44):

$$(44) \lambda X \lambda Y \forall x (X(x) \rightarrow Y(x))(M)$$

This is then the compositional translation of the quantified term *every man*. Expression (41) may be recovered by applying λ -conversion to (44).

The procedure sketched above may also be applied to other quantified terms, like *a woman*, *three boys*, and *the king of France*, and to other determiners like *a*, *the*, *all*, *two*, etc. We shall return to these questions at greater length in chapter 6. There a translation procedure for a fragment of the English language is given, working within the framework of *Montague grammar*.

The λ -operator also makes it possible to introduce quantifiers *categorically* instead of *syncategorematically*. Here we shall restrict ourselves to first-order quantifiers. The categorematic treatment of higher-order quantifiers is completely analogous. We shall introduce the first-order quantifiers \exists and \forall categorically by treating them as second-order predicates, that is to say, as expressions of type $\langle\langle e, t \rangle, t \rangle$. When applied to a first-order predicate, then, quantifiers result in a formula. Quantifiers remain logical constants. We must therefore add to the definition of the interpretation function clauses that state which elements of $\mathbf{D}_{\langle\langle e, t \rangle, t \rangle}$ are to be taken as the interpretations of the quantifiers:

$I(\exists)$ is that function $f_{\exists} \in \{0, 1\}^{\{0, 1\}^D}$ such that if $h \in \{0, 1\}^D$, then $f_{\exists}(h) = 1$ iff there is a $d \in D$ such that $h(d) = 1$.

$I(\forall)$ is that function $f_{\forall} \in \{0, 1\}^{\{0, 1\}^D}$ such that if $h \in \{0, 1\}^D$, then $f_{\forall}(h) = 1$ iff for all $d \in D$: $h(d) = 1$.

In other, simpler, words: $I(\exists)$ is (the characteristic function of) the set of non-empty subsets of *D*, that is, (the characteristic function of) $\{A \mid A \subseteq D \ \& \ A \neq \emptyset\}$. And $I(\forall)$ is (the characteristic function of) the set of subsets of *D* containing all elements of *D*, that is, the set with *D* as its only element: $\{D\}$. For a one-place first-order predicate *P*, the formula $\exists(P)$ is true just in case at least one thing in the domain is a *P*. And the formula $\forall(P)$ is true just in case everything in the domain is a *P*.

The λ -operator is now needed in order to turn formulas with a free variable *v* (that is, just the kinds of formulas to which we would normally apply a quantifier $\exists v$ or $\forall v$) into one-place predicates to which the categorically introduced quantifiers \exists and \forall may be applied. So we write $\forall \lambda x (A(x)(x))$

instead of $\forall x(A(x)(x))$, and $\exists \lambda x \forall \lambda y(A(x)(y))$ instead of $\exists x \forall y(A(x)(y))$. To put it generally: $\exists \lambda x \phi$ is to be written instead of $\exists x \phi$, and $\forall \lambda x \phi$ instead of $\forall x \phi$.

We have interpreted \exists as the set of all nonempty subsets of D . This means that the quantifier \exists is equivalent to the expression $\lambda Y \exists x(Y(x))$ in the theory of types, which contains the normal quantifier $\exists x$. In just the same way, \forall is equivalent to $\lambda Y \forall x(Y(x))$. The expression \exists and the equivalent expression $\lambda Y \exists x(Y(x))$ may be seen as logical representations of the quantified term *something* (or *someone*, if the domain under consideration happens to consist only of people). The expression \forall and thus $\lambda Y \forall x(Y(x))$ similarly represent the quantified term *everything* (or *everyone*). The link with, for example, the way the term *every man* was represented in (41) will be obvious. It is also possible to give a categorematic treatment of the expressions in the theory of types which correspond to determiners. This raises no new issues, however, so we will not pursue it here. It is perhaps important to stress that although we now have two apparently different ways of translating *something*, namely, \exists and $\lambda Y \exists x(Y(x))$, the two are in fact equivalent, so there is no reason to prefer one to the other. Both representations represent the same meaning, so that the difference is purely one of notation.

A last point which we wish to make here concerns representing quantified expressions like *many*, *most*, *few*, *more than half*, etc. These expressions cannot be represented by means of the quantifiers familiar from standard predicate logic. They can, however, be represented in the theory of types with λ -abstraction. But even then there is still an essential difference between, on the one hand, those quantifying expressions which may be represented in terms of the standard quantifiers, such as *all*, *one*, *exactly one*, *at most three*, *more than four*, and the like, and on the other hand, the expressions just mentioned, which cannot be represented in this manner. This difference may be illustrated as follows. Sentences with restricted quantification like *a man* or *all boys* can always be paraphrased using the corresponding unrestricted quantifiers:

- (45) A man is walking.
- (46) All boys are sleeping.
- (47) Something (is a man and walking).
- (48) Everything (sleeps, if it is a boy).

In the case of restricted quantifiers like *many men* and *most boys*, however, paraphrases of this sort are quite unsuitable. This is apparent from the following example:

- (49) Many millionaires are happy.
- (50) Many (are millionaires and happy).

A paraphrase along the lines of the universal quantifier does no better:

(51) Many millionaires are poor.

(52) Many (are poor, if they are millionaires).

Whereas (51) is simply false, (52) is true given that there are relatively few millionaires (and given the meaning of the material implication). The point is this: sentences with restricted standard quantifiers may always be reduced to sentences in which a quantifier, this being a one-place second-order predicate, is applied to some composite first-order predicate. This does not, however, apply to sentences in which quantifiers like *many*, *most*, and the like are restricted. Such sentences may not be reduced to sentences in which the quantifier becomes a one-place second-order predicate which is applied to a composite predicate. Quantifiers like *many* and *most* are essentially two-place. This can be proved rigorously, but we will not do so here. Such expressions must, then, be interpreted as expressions of type $\langle\langle e, t \rangle, \langle\langle e, t \rangle, t \rangle\rangle$, that is, as two-place second-order relations. Let us take an example: a set A of individuals bears the relation *more than half* to a set B of individuals iff more than half of the elements in A are also in B. An extra complication turns up with expressions like *many*, *few*, and *most*. These expressions must likewise be interpreted as two-place second-order relations, but it is not obvious exactly which two-place second-order relations they refer to. The proportion of A's which must be B's in order for the sentence *many A's are B's* to be true is a highly context-dependent matter.

The analysis of quantified expressions outlined above plays an important role in the framework of Montague grammar, which will be introduced in detail in chapter 6. Recently it has inspired a lot of research into the nature of quantified expressions which is known as 'generalized quantifier theory'. This theory is introduced in chapter 7.

5

The Intensional Theory of Types

5.1 Introduction

This chapter is mainly taken up with an exposition of the *intensional theory of types*, a system obtained by providing the theory of types with an intensional semantics. The particular intensional semantics used here differs in several respects from that given for modal predicate logic in §3.3.4. We will return to these differences, and to the similarities, presently. One reason for dealing with this particular intensional theory of types is that we will need it later when we come to Montague grammar (see chap. 6). The last section of this chapter contains some observations on two-sorted type theory. Like the intensional theory of types, this is an extension of the extensional theory of types dealt with in chapter 4.

5.2 Intensional Constructions and Intensional Concepts

There is much to be said for the intensional theory of types as an intermediary between natural language and its semantic interpretation. As was argued in §4.2.1, natural languages contain expressions of very many divergent types. Thus a logical language with a type structure is indicated. Furthermore, natural languages are intensional, containing expressions and constructions which create opaque contexts. We saw many examples of these in chapter 3, in particular in §3.1. Opaque contexts are also known as *intensional contexts*, and the expressions and constructions they give rise to are likewise said to be intensional.

The intensionality of natural language is relevant at several different points. To begin with, natural languages contain temporal, modal, and deontic expressions, all of which involve intensionality. An adequate logical system would need to contain expressions corresponding to these. Besides these, however, it would also need expressions which refer directly to intensional entities like propositions, individual concepts, and properties. For natural languages clearly contain such expressions too. As an example of such an expression, consider

- (1) John asserts that the Dutch queen resides in the Hague.

Now the expression *assert* in (1) cannot stand for a relation between an individual, in this case John, and a *sentence*, in this case:

(2) The Dutch queen resides in the Hague.

For (1) may well be true without John bearing any special relation to this or any other English sentence. He need not have written it, pronounced it, or anything else along these lines. John might be an illiterate Welshman who wouldn't know an English sentence if he saw one and still assert that the Dutch queen resides in the Hague. He might do so, for example, by pronouncing the words *Mae brenhines yr Isalmaen yn byw'n yr Hâg*, this being the Welsh translation of (2). This strongly suggests that *assert* is a relation, not between individuals and sentences, but between individuals and *propositions*. In particular, (1) says that John bears this relation to the proposition expressed not only by (2) but also by its Welsh translation and by the following sentence

(3) The female monarch of Holland lives in the Hague.

In sentence (1), then, the expression *John* refers to an individual and *that the Dutch queen resides in the Hague* refers to a proposition, while *assert* refers to a relation between individuals and propositions. This is of course not to say that the truth of (1) may not involve John in certain relations to particular sentences. Typically, in asserting a proposition, one must say, write, shout, or otherwise utter some sentence which expresses that proposition. The point is just that (1) itself is not about uttering but about asserting.

If a logical theory is to provide representations of sentences which refer to intensional entities like propositions, then it will need expressions which stand for such entities. The expression *that the Dutch queen resides in the Hague* refers to a proposition, and a compositional rendering of the meaning of (1) will require a logical theory with the capacity to refer to propositions. Examples similar to (1) may be constructed to show that logical expressions referring to individual concepts and to properties will also be needed.

We saw in §1.8 that propositions, individual concepts, and properties are the intensions of sentences, terms, and predicates, respectively. The concept of intension, which may be defined in terms of multiple reference, is the formal pendant of Frege's notion of *Sinn*. It lies at the very core of the concept of meaning. Where intensional semantics enables us to define intensional concepts, the intensional theory of types now provides us with expressions by means of which we may refer to these.

5.3 Syntax

In defining the syntax of the intensional theory of types, we begin, as usual, by spelling out the possible types:

Definition 1

T , the set of types in intensional type theory, is the smallest set such that

- (i) $e, t \in T$.
- (ii) If $a, b \in T$, then $\langle a, b \rangle \in T$.
- (iii) If $a \in T$, then $\langle s, a \rangle \in T$.

Just as in extensional type theory, we have e and t as our two basic types, e once again being the type of those expressions that refer to entities, and t being the type of those expressions that refer to truth values. Clause (ii), too, is as usual; what is new is clause (iii). This clause enables us to form a new type $\langle s, a \rangle$ from an arbitrary type a . Note that s is itself not a type; its only purpose is to enable us to form new composite types. There are, then, no expressions of 'type' s . This restriction is done away with in the two-sorted theory of types, which is dealt with briefly in §5.8. A type $\langle s, a \rangle$ can of course be used in the construction of yet other composite types, according to clauses (ii) and (iii). Expressions of type $\langle s, a \rangle$ will be seen to refer to functions from possible worlds to entities of type a . Such expressions thus refer to intensional entities.

The vocabulary of any particular intensional, type-theoretical language L consists once again of a part shared by all such languages, together with a number of symbols which are peculiar to it. The shared part is:

- (i) for every type a , an infinite set VAR_a of variables of type a
- (ii) the connectives $\wedge, \vee, \rightarrow, \neg, \leftrightarrow$
- (iii) the quantifiers \forall and \exists
- (iv) the identity symbol $=$
- (v) the operators $\Box, \Diamond, \wedge, \vee$
- (vi) the brackets (and)

The part which is peculiar to L consists of:

- (vii) for every type a , a (possibly empty) set CON_a^L of constants of type a

Just as in ordinary type theory, we must take care not to confuse constants and variables of different types. To this end we will observe notational conventions already introduced and introduce new ones where needed.

The syntax may now be defined along the lines of ordinary type theory. Novel in comparison with definition 2 of §4.2.2 are clauses (vi), (vii), (viii), and (ix):

Definition 2

- (i) If $\alpha \in VAR_a$ or $\alpha \in CON_a^L$, then $\alpha \in WE_a^L$.
- (ii) If $\alpha \in WE_{(a,b)}^L$ and $\beta \in WE_a^L$, then $(\alpha(\beta)) \in WE_b^L$.
- (iii) If $\phi, \psi \in WE_t^L$, then $\neg\phi, (\phi \wedge \psi), (\phi \vee \psi), (\phi \rightarrow \psi)$, and $(\phi \leftrightarrow \psi) \in WE_t^L$.

- (iv) If $\phi \in WE_t^L$ and $v \in VAR_a$, then $\forall v\phi, \exists v\phi \in WE_t^L$.
- (v) If $\alpha, \beta \in WE_a^L$, then $(\alpha = \beta) \in WE_t^L$.
- (vi) If $\alpha \in WE_a^L$ and $v \in VAR_b$, then $\lambda v\alpha \in VAR_{(b,a)}$.
- (vii) If $\phi \in WE_t^L$, then $\Box\phi, \Diamond\phi \in WE_t^L$.
- (viii) If $\alpha \in WE_a^L$, then $\wedge\alpha \in WE_{(s,a)}^L$.
- (ix) If $\alpha \in WE_{(s,a)}^L$, then $\vee\alpha \in WE_a^L$.
- (x) Every element of WE_a^L for any a is constructed in a finite number of steps using (i)–(ix).

Clause (vii) introduces the modal operators \Box and \Diamond in the usual manner. We will not go to the trouble of also introducing temporal operators, since this would complicate things unnecessarily. We will briefly return to them later on. Clause (viii) serves to introduce an operator \wedge , read 'cap' or 'up', which when applied to an expression of any type a results in a new expression of type $\langle s, a \rangle$. As will become apparent when we get around to the semantics, an expression $\wedge\alpha$ refers to the intension of α . For example, if $\phi \in WE_t^L$, then $\wedge\phi \in WE_{(s,t)}^L$. Thus where ϕ refers to a truth value, $\wedge\phi$ refers to a function from possible worlds into truth values, that is, to a proposition. The operator \vee introduced in clause (ix), read 'cup' or 'down', may only be applied to an expression if it is of some intensional type $\langle s, a \rangle$. If $\alpha \in WE_{(s,a)}^L$, then $\vee\alpha \in WE_a^L$. We will see that in any given possible world w , $\vee\alpha$ refers to the result of applying α 's reference in w to w . Obviously this only makes sense if α is of some intensional type and as such refers to a function from possible worlds into some domain or other. Two examples: if $\alpha \in CON_{(s,e)}^L$, then $\vee\alpha \in WE_e^L$ (and $\wedge\vee\alpha \in WE_{(s,e)}^L$). Furthermore, if $\phi \in WE_t^L$, then $\wedge\phi \in WE_{(s,t)}^L$, and thus $\vee\wedge\phi \in WE_t^L$.

Exercise 1*

- (a) Let $j \in WE_e^L$, $j \in WE_{(s,e)}^L$, $M \in WE_{(\langle s,e \rangle, t)}^L$, and $M \in WE_{(e,t)}^L$. Which of the following sequences of symbols are well-formed expressions of the language of intensional type theory?
- (i) $M(\wedge j)$
 - (ii) $M(\wedge j)$
 - (iii) $M(\wedge j)$
 - (iv) $M(\vee j)$
 - (v) $M(\wedge M)$
 - (vi) $M(\wedge \vee j)$
 - (vii) $M(\vee \wedge j)$
 - (viii) $M(\wedge \vee j)$
 - (ix) $\vee(M(j))$
 - (x) $\vee M(j)$
 - (xi) $\wedge(M(j))$
 - (xii) $\wedge(M(j))$
 - (xiii) $\wedge \wedge j$
 - (xiv) $M(\vee \wedge \wedge j)$

(b) What is the type of α in each of the following four cases:

- (i) $p \in WE_t^L$ and $\alpha(\wedge p) \in WE_t^L$
- (ii) $j \in WE_e^L$ and $\alpha(\wedge j) \in WE_t^L$
- (iii) $j \in WE_e^L$ and $\lambda\alpha(\vee\alpha(j))(\alpha) \in WE_t^L$
- (iv) $p \in WE_{(s,t)}^L$ and $\vee\alpha(\wedge\vee p) \in WE_e^L$

5.4 Semantics

The first step in providing the intensional theory of types with a semantics again consists of specifying domains of interpretation appropriate to the different types we have at our disposal. What is new is that we are now dealing with intensional types, that is to say, with types of the form $\langle s, a \rangle$. An expression of any intensional type $\langle s, a \rangle$ is to be interpreted as a function mapping possible worlds to elements of the interpretation domain corresponding to type a . Hence, we define the interpretation domains for expressions of the various types on the basis of some domain of individuals D and a set of possible worlds W . This definition of the interpretation domain of expressions of type a with respect to a domain D and a set W of possible worlds, which we write as $D_{a,D,W}$, runs as follows:

Definition 3

- (i) $D_{e,D,W} = D$
- (ii) $D_{t,D,W} = \{0, 1\}$
- (iii) $D_{\langle a,b \rangle,D,W} = D_{b,D,W}^{D_{a,D,W}}$
- (iv) $D_{\langle s,a \rangle,D,W} = D_{a,D,W}^W$

Wherever possible, the subscripts D and W will be left out. The first three familiar clauses state that here too expressions of type e refer to entities, while expressions of type t refer to truth values, and expressions of functional types $\langle a, b \rangle$ refer to functions from things of type a into things of type b . It is only in the new clause (iv) that the set W of possible worlds is really involved. $D_{\langle s,a \rangle} = D_a^W$, which is the set of all functions with W as their domain and D_a as their range. An example: $D_{\langle s,t \rangle} = D_t^W = \{0, 1\}^W$ = the set of all functions from possible worlds to truth values. An expression of type $\langle s, t \rangle$ thus refers to a function from possible worlds to truth values. Functions of this kind will be called *propositions*. Thus $D_{\langle s,t \rangle}$ is the set of propositions. A second example: $D_{\langle s, \langle e,t \rangle \rangle} = D_{\langle e,t \rangle}^W = (\{0, 1\}^D)^W$: the set of functions from possible worlds to (characteristic functions of) sets of individuals. An expression of type $\langle s, \langle e, t \rangle \rangle$ thus refers to a function from possible worlds to sets of individuals. Now sets of individuals serve as the interpretations of predicates, and a predicate refers in different worlds to different sets. This multiple reference of a predicate may be seen as a function from possible worlds to sets of individuals, and this function may be thought of as the predicate's intension. Any such intension will be called a *property*. Thus $D_{\langle s, \langle e,t \rangle \rangle}$ is the set of properties of individuals, and expressions of type $\langle s, \langle e, t \rangle \rangle$ refer to properties of individu-

Table 5.1 Intensional Types and Interpretations

Type	Interpretation
$\langle s, e \rangle$	Function from worlds to entities, that is, an <i>individual concept</i>
$\langle s, t \rangle$	Function from worlds to truth values, that is, a <i>proposition</i>
$\langle s, \langle e, t \rangle \rangle$	Function from worlds to sets of entities, that is, a <i>first-order property</i>
$\langle s, \langle e, \langle e, t \rangle \rangle \rangle$	Function from worlds to functions from entities to sets of entities, that is, a <i>two-place first-order relation</i>
$\langle s, \langle \langle e, t \rangle, t \rangle \rangle$	Function from worlds to sets of sets of entities
$\langle \langle s, e \rangle, t \rangle$	Function from individual concepts to truth values, that is, a (characteristic function of a) set of individual concepts
$\langle s, \langle \langle s, e \rangle, t \rangle \rangle$	Function from worlds to sets of individual concepts, that is, a property of individual concepts
$\langle \langle s, t \rangle, t \rangle$	Function from propositions to truth values, that is a (characteristic function of a) set of propositions

als. We shall have more to say on the relationship between predicates and properties, and more generally on the relationship between reference and intension, once we have dealt with the models of the intensional theory of types. Table 5.1 contains more examples of the interpretations of expressions of types involving s (examples of types not involving s are to be found in table 4.2). The italicized terms in table 5.1 are expressions commonly used for the intensional entities in question.

A model M for an intensional type theoretical language L consists of a non-empty set D , its domain, a nonempty set W of possible worlds, and an interpretation function I . As in §3.3.4, we have thus chosen to deal with a single domain. This is not the only course open to us, since we might also introduce an accessibility relation between the worlds. We will take every world to be accessible from itself and from every other world, however, choosing R as the universal relation including every pair of worlds in W . Then the semantics may be defined avoiding all mention of R (as it indeed will be defined), the modal system thus obtained being $S5$ (see §2.3.2). But obviously R might just as well have been chosen differently.

Next, the interpretation function I assigns an interpretation to each constant in the type-theoretical language L . In extensional logical systems like the type theory dealt with in chapter 4, the interpretation of a constant of type a is always an element of D_a . In an intensional system like the present one, however, this will not do. We want the interpretations of expressions, for example, the truth value of a given formula, to be able to vary from world to world. Thus the references of the constants must be able to vary from world to world. To this end, the interpretation function I assigns to each constant a function which gives, for each world, the interpretation of that constant in that world. That is, if α is a constant of type a , then $I(\alpha) \in D_a^W$, i.e., $I(\alpha)$ is a function mapping possible worlds onto elements of D_a . $I(\alpha)(w)$ is thus an element of D_a , namely, the reference of α in w . Note how this differs from the way things were set up in modal predicate logic in §3.3. There it was only the

predicates that were allowed to vary their interpretations from world to world. Individual constants were treated as rigid designators, their references being the same in every world. Here that is not the case. Constants of type e are allowed to refer to different entities in different worlds. Of course, individual constants might have been treated as rigid designators simply by stipulating that for every constant α of type e , $I(\alpha)$ is a function which takes the same value in every world, that is, a *constant function*. Besides this way of interpreting the constants, we also make use of assignments g in order to interpret the variables. As usual, if v is a variable of type a , then $g(v)$ is an element of D_a .

Given this definition of models M and assignments g , we can define $\llbracket \alpha \rrbracket_{M,w,g}$ in the usual inductive manner. We refer to $\llbracket \alpha \rrbracket_{M,w,g}$ as the *extension* (the reference) of α in w , given M and g . The definition runs as follows:

Definition 4

- (i) If $\alpha \in \text{CON}_a^L$, then $\llbracket \alpha \rrbracket_{M,w,g} = I(\alpha)(w)$.
If $\alpha \in \text{VAR}_a$, then $\llbracket \alpha \rrbracket_{M,w,g} = g(\alpha)$.
- (ii) If $\alpha \in \text{WE}_{(a,b)}^L$ and $\beta \in \text{WE}_a^L$, then $\llbracket \alpha(\beta) \rrbracket_{M,w,g} = \llbracket \alpha \rrbracket_{M,w,g}(\llbracket \beta \rrbracket_{M,w,g})$.
- (iii) If $\phi, \psi \in \text{WE}_t^L$, then $\llbracket \neg \phi \rrbracket_{M,w,g} = 1$ iff $\llbracket \phi \rrbracket_{M,w,g} = 0$.
 $\llbracket \phi \wedge \psi \rrbracket_{M,w,g} = 1$ iff $\llbracket \phi \rrbracket_{M,w,g} = \llbracket \psi \rrbracket_{M,w,g} = 1$.
 $\llbracket \phi \vee \psi \rrbracket_{M,w,g} = 1$ iff $\llbracket \phi \rrbracket_{M,w,g} = 1$ or $\llbracket \psi \rrbracket_{M,w,g} = 1$.
 $\llbracket \phi \rightarrow \psi \rrbracket_{M,w,g} = 0$ iff $\llbracket \phi \rrbracket_{M,w,g} = 1$ and $\llbracket \psi \rrbracket_{M,w,g} = 0$.
 $\llbracket \phi \leftrightarrow \psi \rrbracket_{M,w,g} = 1$ iff $\llbracket \phi \rrbracket_{M,w,g} = \llbracket \psi \rrbracket_{M,w,g}$.
- (iv) If $\phi \in \text{WE}_t^L$ and $v \in \text{VAR}_a$, then
 $\llbracket \forall v \phi \rrbracket_{M,w,g} = 1$ iff for all $d \in D_a$: $\llbracket \phi \rrbracket_{M,w,g[v/d]} = 1$.
 $\llbracket \exists v \phi \rrbracket_{M,w,g} = 1$ iff for some $d \in D_a$: $\llbracket \phi \rrbracket_{M,w,g[v/d]} = 1$.
- (v) If $\alpha, \beta \in \text{WE}_a^L$, then $\llbracket \alpha = \beta \rrbracket_{M,w,g} = 1$ iff $\llbracket \alpha \rrbracket_{M,w,g} = \llbracket \beta \rrbracket_{M,w,g}$.
- (vi) If $\alpha \in \text{WE}_a^L$ and $v \in \text{VAR}_b$, then $\llbracket \lambda v \alpha \rrbracket_{M,w,g}$ is that function $h \in D_a^{D_b}$ such that for all $d \in D_b$: $h(d) = \llbracket \alpha \rrbracket_{M,w,g[v/d]}$.
- (vii) If $\phi \in \text{WE}_t^L$, then
 $\llbracket \Box \phi \rrbracket_{M,w,g} = 1$ iff for all $w' \in W$: $\llbracket \phi \rrbracket_{M,w',g} = 1$.
 $\llbracket \Diamond \phi \rrbracket_{M,w,g} = 1$ iff for some $w' \in W$: $\llbracket \phi \rrbracket_{M,w',g} = 1$.
- (viii) If $\alpha \in \text{WE}_a^L$, then $\llbracket \wedge \alpha \rrbracket_{M,w,g}$ is that function $h \in D_a^W$ such that for all $w' \in W$: $h(w') = \llbracket \alpha \rrbracket_{M,w',g}$.
- (ix) If $\alpha \in \text{WE}_{(s,a)}^L$, then $\llbracket \vee \alpha \rrbracket_{M,w,g} = \llbracket \alpha \rrbracket_{M,w,g}(w)$.

Just as in definition 4 of §4.2.3, we are now in a position to define the concepts of *truth relative to M*, *universal validity*, and *equivalence*. The same remarks apply here as there. What are new in comparison with the interpretation of the extensional theory of types are the clauses (i), (vii), (viii), and (ix). In (i), the extension of a constant α in a world w is defined as the result of applying the interpretation of α to this world w . Clause (vii) defines the truth value of formulas with modal operators: $\Box \phi$ is true in w just in case ϕ is true in all worlds, and $\Diamond \phi$ is true in w just in case ϕ is true in at least one world.

Clause (viii) defines the extensions of expressions of the form $\wedge\alpha$. Each of the expressions is of some intensional type $\langle s, a \rangle$, and its domain of interpretation is, according to definition 3, the set of functions D_a^w , which consists of all functions mapping worlds onto things of type a . In any given world w , $\wedge\alpha$ is interpreted as that function from worlds to extensions which, when applied to any world w' , takes the extension of α in w' as its value. An example will make this more tangible. If ϕ is a formula, then $\wedge\phi$ is an expression of type $\langle s, t \rangle$. The extension (or reference) of $\wedge\phi$ in w is that function from worlds to truth values which when applied to any given world takes the truth value of ϕ in that world as its value: $\llbracket \wedge\phi \rrbracket_{M,w,g}$ is that function $h \in \{0, 1\}^W$ such that: $h(w') = \llbracket \phi \rrbracket_{M,w',g}$, for all $w' \in W$. The expression $\wedge\phi$ is thus interpreted as the proposition expressed by ϕ . Note that the reference of $\wedge\phi$ is the same in every world w . This means that $\wedge\wedge\phi$ is a *constant* function from worlds to the proposition expressed by ϕ . Indeed, it is only because of this fact that we may speak of *the* proposition expressed by ϕ : the proposition expressed by ϕ does not vary from world to world.

Here is a second example. If M is a one-place first-order predicate constant, that is to say, it is of type $\langle e, t \rangle$, then $\wedge M$ is of type $\langle s, \langle e, t \rangle \rangle$. The extension of $\wedge M$ is that function from worlds to sets of entities which, when applied to a world w' , takes as its value the extension of M in w' : $\llbracket \wedge M \rrbracket_{M,w,g}$ is thus that $h \in D_{\langle e, t \rangle}^w$ such that for all $w' \in W$, $h(w') = \llbracket M \rrbracket_{M,w',g}$. The expression $\wedge M$ thus refers to the property expressed by the predicate letter M .

Clause (ix) defines the extensions of expressions of the form $\vee\alpha$. The expression α is then always of some intensional type $\langle s, a \rangle$, and its extension is always a function from worlds to D_a . The expression $\vee\alpha$ itself is of type a , and its extension is thus some element of D_a . In (ix) it is stipulated that the extension of $\vee\alpha$ in w is that element of D_a which may be obtained by applying the extension of α in w to w : $\llbracket \vee\alpha \rrbracket_{M,w,g} = \llbracket \alpha \rrbracket_{M,w,g}(w)$. Let us once again turn to an example: Let m be a constant of type $\langle s, e \rangle$. Its extension is a function from worlds to individuals. Let us just suppose that for a certain w $\llbracket m \rrbracket_{M,w,g}$ ($=I(m)(w)$ according to clause (i)) is that function which indicates the most powerful individual in each possible world; we shall call this the top-dog function. The extension of $\vee m$ in w is now obtained by applying $\llbracket m \rrbracket_{M,w,g}$ to w : $\llbracket \vee m \rrbracket_{M,w,g} = \llbracket m \rrbracket_{M,w,g}(w)$. The extension of $\vee m$ in w is the top dog of w , or the most powerful individual in w . Note that the extension of m may vary from world to world. We stipulated that m was to refer to the top-dog function in w , but that says nothing about the extension of m in any other world w' : there it might be any other individual concept, say the individual concept which indicates the richest individual in each world.

Here is a second example. The expression $\wedge M$ is of type $\langle s, \langle e, t \rangle \rangle$ and has as its reference, as we have seen, the property expressed by M . The expression $\vee \wedge M$ is of type $\langle e, t \rangle$. Its extension in any world w is obtained by applying $\wedge M$'s extension in w to w : $\llbracket \vee \wedge M \rrbracket_{M,w,g} = \llbracket \wedge M \rrbracket_{M,w,g}(w)$. The extension of $\wedge M$ is that function whose value in any world is the extension of M in that

world; thus $\llbracket \wedge M \rrbracket_{M,w,g}(w) = \llbracket M \rrbracket_{M,w,g}$. One difference between this example and the preceding one is that while the extension of m may depend on w , the extension of $\wedge M$ is independent of w . The statement that the property $\wedge M$ belongs to an individual j in a world w amounts to saying that j belongs to that set of entities which is the value of $\wedge M$ in w . Now $\vee \wedge M$'s reference in w is just that set, so that the above statement may be rendered as $\vee \wedge M(j)$. As we will see in §5.5, where we return to the interaction between \wedge and \vee , this is equivalent to the simpler formula $M(j)$.

Definition 4 defines the extensions of expressions, and we will now define their intensions. It should be clear from the foregoing that the intension of α can be defined in terms of its multiple reference, that is, the various denotations it has in different worlds. This means that α 's intension may be defined in terms of its extension. The intension of α in M relative to g , written $\text{Int}_{M,g}(\alpha)$, is then defined as follows:

Definition 5

If $\alpha \in WE_a^L$, then $\text{Int}_{M,g}(\alpha)$ is that $h \in D_a^W$ such that for all $w' \in W$: $h(w') = \llbracket \alpha \rrbracket_{M,w',g}$.

Definition 5 is such that the extension of α in a world w is the result of applying the intension of α to that world w . That is, $\text{Int}_{M,g}(\alpha)(w) = \llbracket \alpha \rrbracket_{M,w,g}$. According to clause (i) of definition 4, for a constant α we have $\llbracket \alpha \rrbracket_{M,w,g} = I(\alpha)(w)$. This means: $\text{Int}_{M,g}(\alpha) = I(\alpha)$. We see that constants are interpreted intensionally.

Comparing the definition of intensions with clause (viii) of definition 4, we see that the extension of $\wedge \alpha$ is just the intension of α . That is, $\llbracket \wedge \alpha \rrbracket_{M,w,g}$ and $\text{Int}_{M,g}(\alpha)$ have been defined in such a way that they are one and the same function. What this means is that given any expression α , the \wedge -operator enables us to form an expression $\wedge \alpha$ whose extension is α 's intension. This means that the \wedge -operator provides expressions referring to the various kinds of intensional objects which are the intensions of different expressions. Thus the requirement mentioned in § 5.2 has been met.

The following fact is worth mentioning. By definition, the familiar principles of extensionality do not apply in intensional logic. That is, in an intensional system we do not have:

$$(4) \quad \alpha = \beta \models \gamma = [\beta/\alpha]\gamma$$

(In the theory of types, placing the identity sign $=$ between any two expressions of the same type gives rise to a formula. Thus (4) expresses the principle of extensionality quite generally, covering substitution of formulas, predicates, terms and so on.) In intensional logic expressions with the same intension may be substituted for one another without changing the truth value of a formula. The intensional theory of types provides expressions which denote intensions, and it may happen that two such expressions denote the same in-

Table 5.2 Intensional Types and Expressions

Type	Variables	Constants	Extension	Intension
e	x, y, z	j, m	Entity	Function from worlds to entities, individual concept
t	p, q, r		Truth value	Function from worlds to truth values, proposition
$\langle s, e \rangle$	x, y, z	j, m	Function from worlds to entities, individual concept	Function from worlds to individual concepts
$\langle e, t \rangle$	X, Y, Z	M, N, P, W	(Characteristic function of) set of entities	Function from worlds to sets of entities, first-order property
$\langle s, \langle e, t \rangle \rangle$	X, Y, Z	M, N, P, W	Function from worlds to sets of entities, first-order property	Function from worlds to first-order properties
$\langle e, \langle e, t \rangle \rangle$		A, B, K	Function from entities to sets of entities, i.e., set of ordered pairs of entities	Function from worlds to sets of ordered pairs of entities, two-place first-order relation
$\langle s, \langle e, \langle e, t \rangle \rangle \rangle$		A, B, K	Function from worlds to sets of ordered pairs of entities, two-place first-order relation	Function from worlds to two-place first-order relations
$\langle \langle e, t \rangle, t \rangle$	\mathcal{X}, \mathcal{Y}	$\mathcal{C}, \mathcal{H}, \mathcal{M}$	(Characteristic function of) set of sets of entities	Function from worlds to sets of sets of entities, property of sets of entities
$\langle s, \langle \langle e, t \rangle, t \rangle \rangle$	S, T		Function from worlds to sets of sets of entities, property of sets of entities	Function from worlds to properties of sets of entities
$\langle \langle s, \langle e, t \rangle \rangle, t \rangle$	X, Y, Z	M, N, P	(Characteristic function of) set of first-order properties	Function from worlds to sets of first-order properties, property of first-order properties, second-order property
$\langle s, \langle \langle s, \langle e, t \rangle \rangle, t \rangle \rangle$	X, Y, Z	M, N, P	Function from worlds to sets of first-order properties, property of first-order properties, second-order property	Function from worlds to second-order properties

tension. If $\wedge\alpha$ and $\wedge\beta$ denote the same intension, then α and β may freely be substituted for one another. That is to say, the following theorem can be proved:

Theorem 1

$$\wedge\alpha = \wedge\beta \models \gamma = \{\beta/\alpha\}\gamma$$

Table 5.2 gives the extensions and intensions of various kinds of expressions by way of illustration. Note the following regularity: the extension of an expression of type $\langle s, a \rangle$ is the intension of some expression of type a . Note also the way the intensions and extensions of various expressions with and without \wedge and \vee are related.

Exercise 2*

Consider the following model **M**: $D = \{a, b, c, d\}$ and $W = \{w_1, w_2, w_3\}$. Let j and m be constants of type e . In w_1 and w_2 j denotes a , in w_3 it denotes b ; m denotes c in all w . M is a constant of type $\langle e, t \rangle$. In w_1 M is true of a and b , in w_2 of c and d , and in w_3 of no individual. M is a constant of type $\langle s, \langle e, t \rangle \rangle$. In w_1 and w_2 M denotes the property which M expresses, and in w_3 it refers to the property which holds of all entities.

(a) Write down the interpretation function I of **M**.

(b) Determine the following values:

- (i) $\llbracket j \rrbracket_{M, w_2, g}$
- (ii) $\llbracket \wedge j \rrbracket_{M, w_1, g}$
- (iii) $\llbracket \wedge j \rrbracket_{M, w_3, g}$
- (iv) $\llbracket M(j) \rrbracket_{M, w_2, g}$
- (v) $\llbracket \vee M \rrbracket_{M, w_3, g}$
- (vi) $\llbracket \vee M(j) \rrbracket_{M, w_1, g}$
- (vii) $\llbracket \vee M(j) \rrbracket_{M, w_2, g}$
- (viii) $\llbracket M = \wedge M \rrbracket_{M, w_1, g}$
- (ix) $\llbracket \vee M = M \rrbracket_{M, w_3, g}$

(c) Decide whether the following formulas are valid in **M**:

- (i) $\Diamond(M = \wedge M)$;
- (ii) $\Box(\vee M = M)$;
- (iii) $\exists x \Box(m = x)$

Exercise 3

Let x, y be variables of type $\langle s, e \rangle$, and j and m be constants of the same type. Further, let x be a variable of type e , and j a constant of that type. Determine which of the following formulas are valid:

- (i) $\forall x \forall y (x = y \rightarrow \Box(x = y))$
- (ii) $j = m \rightarrow \Box(j = m)$
- (iii) $\exists x \Box(x = j)$

- (iv) $\exists x \Box(x = \wedge j)$
- (v) $\exists x \Box(\wedge x = j)$
- (vi) $\exists x \Box(x = \vee j)$
- (vii) $\exists x \Box(x = \vee j) \rightarrow \exists x \Box(x = j)$

5.5 The Operators \wedge and \vee

As we have seen, $\wedge \alpha$ denotes the intension of α . That is, for any \mathbf{M} , w , and g we have:

$$(5) \quad \llbracket \wedge \alpha \rrbracket_{\mathbf{M}, w, g} = \text{Int}_{\mathbf{M}, g}(\alpha)$$

The denotation of $\vee \alpha$ in any world w is the extension of α in w applied to w : $\llbracket \vee \alpha \rrbracket_{\mathbf{M}, w, g} = \llbracket \alpha \rrbracket_{\mathbf{M}, w, g}(w)$. So now the question arises as to how the operators \wedge and \vee interact with one another.

First we turn to $\vee \wedge$. The expression $\vee \wedge \alpha$ denotes, in w , the result of applying the intension of α to w . This is because $\llbracket \vee \wedge \alpha \rrbracket_{\mathbf{M}, w, g} = \llbracket \wedge \alpha \rrbracket_{\mathbf{M}, w, g}(w)$, which, given (5) is just $\text{Int}_{\mathbf{M}, g}(\alpha)(w)$. Applying the intension of α to w , we obtain the extension of α in w . Thus $\text{Int}_{\mathbf{M}, g}(\alpha)(w) = \llbracket \alpha \rrbracket_{\mathbf{M}, w, g}$. We have thus demonstrated that $\llbracket \vee \wedge \alpha \rrbracket_{\mathbf{M}, w, g} = \llbracket \alpha \rrbracket_{\mathbf{M}, w, g}$. Since the above argument applies to any α , \mathbf{M} , w , and g , we have now proved the following theorem:

Theorem 2

$\vee \wedge \alpha$ is equivalent to α .

This means that α may always be written in place of $\vee \wedge \alpha$. The same does *not*, however, apply to $\wedge \vee \alpha$. This may be demonstrated by means of the following rather abstract example. Let \mathbf{M} be a model with two worlds, w_1 and w_2 . Consider the constant p of type $\langle s, t \rangle$. The extension of p is thus a proposition. We may assume the extension of p in w_1 to be that proposition k which has the truth value 1 when applied to w_1 and the truth value 0 when applied to w_2 . We may furthermore stipulate that the extension of p in w_2 is that proposition k' which has the truth value 0 when applied to w_1 and the truth value 1 when applied to w_2 . That is to say:

$$\begin{aligned} I(p)(w_1) &= k \\ k(w_1) &= 1 \\ k(w_2) &= 0 \\ I(p)(w_2) &= k' \\ k'(w_1) &= 0 \\ k'(w_2) &= 1 \end{aligned}$$

Now we have $\llbracket \wedge \vee p \rrbracket_{\mathbf{M}, w_1, g} \neq \llbracket p \rrbracket_{\mathbf{M}, w_1, g}$. This may be seen as follows (subscripts \mathbf{M} and g have been dropped for legibility's sake):

$$\begin{aligned} \llbracket \wedge \vee p \rrbracket_{w_1} &= \text{that function } h \in \{0, 1\}^W \text{ such that for all } w' \in W: \\ h(w') &= \llbracket \vee p \rrbracket_{w'} \end{aligned}$$

Exactly which proposition h is may be determined as follows:

$$\begin{aligned} h(w_1) &= \llbracket \vee p \rrbracket_{w_1} = \llbracket p \rrbracket_{w_1}(w_1) = I(p)(w_1)(w_1) = k(w_1) = 1; \\ \text{and} \\ h(w_2) &= \llbracket \vee p \rrbracket_{w_2} = \llbracket p \rrbracket_{w_2}(w_2) = I(p)(w_2)(w_2) = k'(w_2) = 1 \end{aligned}$$

The proposition h is thus that function from worlds to truth values which takes 1 as its value in each and every world. It is always true. But $\llbracket p \rrbracket_{w_1}$ is not this invariably true proposition; it is $I(p)(w_1) = k$. And k and h are different propositions. Thus $\llbracket \wedge \vee p \rrbracket_{w_1} \neq \llbracket p \rrbracket_{w_1}$, which serves as a counterexample to the equivalence of $\wedge \vee \alpha$ and α .

This rather abstract example can be made a little more tangible. Let us assume the sentence *Mary is coming* to be true in w_1 and false in w_2 . The proposition k may then be seen as the proposition expressed by this sentence; it is the proposition that *Mary is coming*. The proposition k' is expressed by the sentence *Mary is not coming*. It is the proposition that *Mary is not coming*. According to our stipulation above, the extension of the constant p in w_1 , the world in which she is coming, is the proposition that *Mary is coming*, while the extension of this constant in w_2 , the world in which she is not coming, is the proposition that *Mary is not coming*. This makes p a suitable representation of the expression *whether or not Mary is coming* as it appears in (6):

- (6) John knows whether or not Mary is coming.

For consider the following two valid arguments:

- (7) John knows whether or not Mary is coming.
Mary is coming.

John knows that Mary is coming.

- (8) John knows whether or not Mary is coming.
Mary is not coming.

John knows that Mary is not coming.

The validity of (7) and (8) shows that given that *Mary is coming*, the extension of *whether or not Mary is coming* is the proposition that *Mary is coming*, that is to say, k ; while given that she is not coming, the expression has as its extension the proposition that she is not coming, that is to say, k' . And p is just the expression which meets these requirements. The expression *whether or not Mary is coming* thus serves as a natural language example of an expression α for which $\wedge \vee \alpha \neq \alpha$.

There are, of course, expressions α for which $\wedge \vee \alpha$ and α are equivalent. The counterexamples which we have just seen exploit the fact that expressions may in general have different extensions in different worlds. And indeed, if α is an expression which allows no such variation in its extension, then $\wedge \vee \alpha$ and α turn out to be equivalent. One kind of expression whose extension may

not vary from world to world is the variable: its extension depends only on the assignment under consideration. Another example, as we have seen, is any expression of the form $\wedge\beta$: thus $\wedge\vee\wedge\beta$ is always equivalent to $\wedge\beta$ (a fact which also follows from theorem 2). More generally, we may define a class of expressions whose extension does not vary from world to world, the class of *intensionally closed expressions*:

Definition 6

ICE^L , the class of intensionally closed expressions in L , is the minimal subset of WE^L such that:

- (i) If $v \in \text{VAR}_a$, then $v \in \text{ICE}^L$.
- (ii) If $\alpha \in \text{WE}_a^L$, then $\wedge\alpha \in \text{ICE}^L$.
- (iii) If $\phi \in \text{WE}_t^L$, then $\Box\phi, \Diamond\phi \in \text{ICE}^L$.
- (iv) If α is constructed from elements from ICE^L using only connectives, quantifiers, and the λ -operator, then $\alpha \in \text{ICE}^L$.

It should be noted that where accessibility relations R other than the universal relation are allowed, the extensions of $\Box\phi$ and $\Diamond\phi$ may differ from world to world. In that case, the extensions of $\Box\phi$ and $\Diamond\phi$ will be invariant from world to world only if the extension of ϕ is. Definition 6 is then subject to the appropriate modifications.

We can now state the following theorem:

Theorem 3

If $\alpha \in \text{ICE}^L$, then $\llbracket \alpha \rrbracket_{M,w,g} = \llbracket \alpha \rrbracket_{M,w',g}$, for all M, w, w' .

We shall not prove this theorem here. As an immediate consequence of theorem 3 we have:

Theorem 4

If $\alpha \in \text{ICE}^L$, then $\wedge\vee\alpha$ is equivalent to α .

Finally, note that theorems 3 and 4 go in one direction only: membership of ICE^L in both cases is a sufficient condition, but not a necessary one. For example, the relevant properties hold for all valid formulas, too, but not all valid formulas are elements of ICE^L .

Exercise 4

Suppose $\llbracket q \rrbracket_{M,w_1,g}(w) = 1$ for all w , and $\llbracket q \rrbracket_{M,w_2,g}(w) = 0$ for all w . Is $\llbracket q \rrbracket_w$ a constant function? And are $\wedge\vee q$ and q equivalent?

Exercise 5*

Prove theorem 4 for the case that α is a variable (of some intensional type $\langle s, a \rangle$).

Exercise 6

Suppose we define an accessibility relation R on W as follows: for all w, w^1 : wRw_0 and if wRw^1 , then $w^1 = w_0$. Does this necessitate a modification of definition 6 in order for theorem 3 to remain valid? If so, what is this modification? If not, why not?

5.6 λ -Conversion

A final matter concerning the intensional theory of types which we shall now turn to is the determination of the conditions under which application of λ -conversion leads to an equivalent result. Recall that the corresponding condition in the case of extensional type theory was that no free variables become bound in the process of conversion (see theorem 1, §4.4.2). Clearly this condition remains in force. Over and above it, however, another condition has to be formulated for the intensional theory of types, a condition which has to do with the fact that we are now dealing with expressions whose extensions may vary from world to world.

The following informal example will serve as an illustration. Consider the expression $\lambda x \exists y \Box(x = y)$ (in which x and y are variables of type e). It expresses a property of entities, namely, the property that an entity d has if there is some entity which is necessarily, that is to say, in all possible worlds, identical to d . Now of course all entities have this property, since they are all necessarily identical to themselves. This means that the formula $\lambda x \exists y \Box(x = y)(j)$ (in which j is a constant of type e) is always true. Applying λ -conversion to it, we obtain the formula $\exists y \Box(j = y)$. This formula is true just in case there is an entity which, in any given possible world, is identical to whatever entity j has as its extension in that world. And that is so only if j is a rigid designator, in other words, if j denotes the same entity in every world. But this is not guaranteed: the extension of j may well vary from world to world in a model, which means that the truth of $\exists y \Box(j = y)$ is not guaranteed either. Thus this is a case where λ -conversion does not lead to an equivalent result.

We have just seen how problems can arise if an expression which is not intensionally closed is placed by λ -conversion within the scope of an intensional operator, such as \Box , \Diamond , or \wedge . (Treating individual constants as rigid designators would remove the above example, but other counterexamples involving, say, predicate constants could easily be generated.) Theorem 5 provides two conditions under which λ -conversion may freely be applied:

Theorem 5

$\lambda v \beta(\gamma)$ is equivalent to $[\gamma/v]\beta$ if

- (i) all free variables in γ are free for v in β ; and
- (ii) either $\gamma \in \text{ICE}^L$, or no free occurrence of v in β , lies within the scope of \Box , \Diamond , or \wedge .

Note that, again, the theorem states a sufficient condition only, for the very same reason that theorems 3 and 4 did. For reasons mentioned in §4.4, λ -abstraction has an important part to play if the intensional theory of types is to be applied as a formalism for the representation of natural language. The process of λ -conversion then enables us to reduce some long and complex formulas to relatively short and simple ones. Theorem 5 assures us that, for example, $\lambda x^\vee X(x)(j)$ and $\vee X(j)$ represent the same meaning.

Exercise 7*

Let j be a constant of type e ; j of type $\langle s, e \rangle$; M of type $\langle e, t \rangle$; M of type $\langle s, \langle e, t \rangle \rangle$, and B of type $\langle \langle s, e \rangle, \langle \langle s, e \rangle, t \rangle \rangle$. Let x and y be variables of type e ; x, y , and z of type $\langle s, e \rangle$, and X of type $\langle s, \langle e, t \rangle \rangle$. Reduce the following expressions as much as possible using theorems 2, 4, and 5:

- (i) $\lambda X(\vee X(j))(M)$
- (ii) $\lambda x \lambda X(\vee X(x))(j)(\wedge M)$
- (iii) $\lambda x \lambda X \Box(\vee X(x))(j)(\wedge M)$
- (iv) $\lambda X \lambda x \Box(\vee X(x))(\wedge M)(j)$
- (v) $\lambda x \wedge (M(x) \wedge \vee M(x))(y)$
- (vi) $\lambda x \lambda y \Box(B(x)(y))(\wedge j)(j)$
- (vii) $\lambda y(\lambda x \wedge \exists x(B(x)(y) \wedge \wedge x = y)(\wedge \vee z))(\wedge x)$

Exercise 8

Assume that descriptions are analyzed, not in the Russellian way, but by means of the \imath -operator (see vol. 1, §5.6). And suppose we want to express the de re statement that the unique individual that has the property F necessarily has the property G . What is wrong with the formula $\Box G(\imath x Fx)$? Try to give a better formalization, using both the λ -operator and the \imath -operator.

5.7 Temporal Operators

The intensional theory of types dealt with in the preceding pages is practically identical to Montague's system IL, which is applied in Montague grammar. The difference is that IL also includes temporal operators. In fact, IL is a straightforward extension of our formalism in which the temporal operators **P**, **F**, **H**, and **G** are added to the syntax. The semantics is adapted by adding to the models a set of temporal moments T linearly ordered by a relation $<$ (see §2.4). The interpretation domains come to depend on T . Only in clause (iv) of definition 3 (§5.4), however, does this make any real difference. This clause becomes:

$$(v) \quad D_{\langle s, a \rangle, D, W, T} = D_{a, D, W, T}^{W \times T}$$

$W \times T$ is the set of all ordered pairs $\langle w, t \rangle$, in which w is a world and t a moment in time. The contexts are thus no longer just possible worlds; they have become possible worlds at particular moments in time (see §2.5). Expressions of

an intensional type $\langle s, a \rangle$ thus come to denote functions from worlds at times to things of type a . To a constant the interpretation function I assigns a function which gives, for each world at a time, the denotation of that constant in that world at that time. If α is a constant of type a , then $I(\alpha) \in D_a^{W \times T}$.

The definition of extension now features pairs $\langle w, t \rangle$ instead of simple worlds w . The only real difference this makes is in clauses (vii), (viii), and (ix) of definition 4 in §5.4. The new versions of these clauses are as follows:

- (vii') If $\phi \in WE_t^L$, then $\llbracket \Box \phi \rrbracket_{M, \langle w, t \rangle, g} = 1$ iff for all $w' \in W$ and all $t' \in T$: $\llbracket \phi \rrbracket_{M, \langle w', t' \rangle, g} = 1$.
- (viii') If $\alpha \in WE_a^L$, then $\llbracket \wedge \alpha \rrbracket_{M, \langle w, t \rangle, g}$ is that function $h \in D_a^{W \times T}$ such that for all $w' \in W$ and all $t' \in T$: $h(\langle w', t' \rangle) = \llbracket \alpha \rrbracket_{M, \langle w', t' \rangle, g}$.
- (ix') If $\alpha \in WE_{\langle s, a \rangle}^L$, then $\llbracket \vee \alpha \rrbracket_{M, \langle w, t \rangle, g} = \llbracket \alpha \rrbracket_{M, \langle w, t \rangle, g}(\langle w, t \rangle)$.

Note that according to clause (vii'), \Box now means *necessarily at all times*. This interpretation is different from the one in §2.5, in which modality was relativized to a moment in time. As we then pointed out, there is a great deal of freedom to define these things just as one wishes. Definition 4 still needs to be supplemented with clauses for the temporal operators, but since these are completely straightforward, they are left out.

The definition of intension, finally, needs adaptation in the obvious manner: intensions have now become functions from worlds at times to extensions.

Other extensions besides this one are possible too. Some of them were discussed in §3.4.

5.8 Two-Sorted Type Theory

This section, which is not required for a full understanding of chapter 6, is concerned with the two-sorted theory of types. Just like the intensional theory of types, the two-sorted theory of types is an extension of the standard extensional theory of types dealt with in chapter 4. Here too, possible worlds will play an important part, but they will be introduced in a different manner. Here they are introduced as a second kind of entity which may be treated in the same manner as the 'normal' entities in the domain. (Thus, two-sorted type theory is a many-sorted logic, in the sense of volume 1, chapter 5.) In spite of the essential part played by possible worlds, two-sorted type theory, unlike intensional type theory, is an extensional system in the sense that it observes the extensionality principles familiar from predicate logic and standard type theory. The main reason for dealing with it here is that it can lead us to a better understanding of intensional type theory.

In intensional type theory s , unlike e and t , is not a basic type; it is purely a technical device enabling the construction of composite types of the form $\langle s, a \rangle$. In two-sorted type theory, on the other hand, s is treated as a basic type just like e and t . The set **T2** of types for two-sorted type theory is defined as follows:

Definition 7

T2 is the minimal set such that:

- (i) $e, t, s \in \mathbf{T2}$.
- (ii) If $a, b \in \mathbf{T2}$, then $\langle a, b \rangle \in \mathbf{T2}$.

The set **T** of types for intensional type theory is a proper subset of **T2**. Thus, for example, both s and $\langle e, s \rangle$ are contained in **T2** but not in **T**. Expressions of type s denote possible worlds of course. The domains of interpretation corresponding to the various types, given a domain D of entities and a set W of possible worlds, may be defined as follows:

Definition 8

- (i) $\mathbf{D}_{e,D,W} = D$
- (ii) $\mathbf{D}_{t,D,W} = \{0, 1\}$
- (iii) $\mathbf{D}_{s,D,W} = W$
- (iv) $\mathbf{D}_{\langle a,b \rangle,D,W} = \mathbf{D}_{b,D,W}^{\mathbf{D}_{a,D,W}}$

Note that the domain of interpretation corresponding to an intensional type $\langle s, a \rangle$ is not determined in a separate clause, as it is in intensional type theory (see definition 3), but falls under clause (iv). Such types are now generated by clause (ii) of definition 7, since s is now itself a type. And given that, according to clause (iii) in definition 8, $\mathbf{D}_s = W$, clause (iv) gives $\mathbf{D}_{\langle s,a \rangle} = \mathbf{D}_a^W$. The result, then, is just the same as in intensional type theory.

Now that s is just a normal type, we can have variables and constants of type s . Such expressions thus denote possible worlds. Also, it becomes possible to quantify and to abstract over possible worlds. The syntax of two-sorted type theory is quite analogous to that of extensional type theory. The only difference is that the vocabulary now contains variables and constants of type s , and of types constructed using s . As we shall see, the modal operators \Box and \Diamond and the operators \wedge and \vee now familiar from intensional type theory are quite dispensable in two-sorted type theory. The role they formerly played has been taken over by quantification and abstraction over, and application on, possible worlds.

In two-sorted type theory, as in extensional type theory, the extensions of expressions are relative to models and assignments. A model's interpretation function assigns to a constant of type a an element of \mathbf{D}_a , assignments do the same for variables, and the definition of the extension of an expression relative to a model and assignment is then just the same as the corresponding definition for extensional type theory (see definition 4 in §4.2.2). This diverges once again from the situation in intensional type theory. There the interpretation did not map a constant, say, of type a , onto an extension, that is, an element of \mathbf{D}_a , but onto an intension, that is, an element of $\mathbf{D}_{\langle s,a \rangle}$. The extension of an expression was relative to possible worlds, thus accounting for the context-dependent character of expressions. This is achieved in a different way in two-sorted type theory, namely, by equipping expressions whose extension

varies from world to world with a variable of type s . The interpretation of such a formula then becomes relative to the possible world assigned by the assignment to that variable. Context dependence is thus fixed into this kind of type theory via dependence on an assignment.

This may be illustrated by means of an example. One expression whose extension varies from context to context is the verb *to walk*. In intensional type theory, this verb was represented by means of a constant W of type $\langle e, t \rangle$. The interpretation function I then assigned to W an intension, that is, $I(W) \in D_{\langle s, \langle e, t \rangle \rangle}$. The extension of W in any given world w could be recovered from this intension as the value taken by the intension at w . In two-sorted type theory, *to walk* is rendered as a complex expression $W(a)$ in which W is a constant of type $\langle s, \langle e, t \rangle \rangle$ and a is a variable of type s . The result is of course that *to walk* once again is represented by an expression of type $\langle e, t \rangle$. The interpretation function maps W onto an extension, that is, an element in $D_{\langle s, \langle e, t \rangle \rangle}$. The extension of $W(a)$ is obtained by applying the extension of W to the extension of a . The latter is $g(a)$, the world that the assignment g assigns to a . The extension of this representation of *to walk*, $W(a)$, thus depends on what world g assigns to a . We see that the context dependence of *to walk* appears here as a dependence on assignments.

The relation between W and $W(a)$ may be made more precise by comparing with each other the models of intensional type theory and two-sorted type theory. Let M be a model for intensional type theory, and $M2$ a model for two-sorted type theory, both being based on the same underlying sets D and W . Thus W consists of D and W together with I_M , where I_M maps constants of type a onto elements of $D_{\langle s, a \rangle}$. And $M2$ consists of the same D and W together with I_{M2} , where I_{M2} maps constants of type a onto elements of D_a . Furthermore, let M be a constant in the intensional theory of types, its type being $\langle e, t \rangle$, and let W be a constant in two-sorted type theory, its type being $\langle s, \langle e, t \rangle \rangle$. Assume that $I_M(W) = I_{M2}(W)$. Now it follows that the extension of W relative to M , w , and g is just the extension of $W(a)$ relative to $M2$ and $g[a/w]$. Choose arbitrary w and g . $\llbracket W \rrbracket_{M, w, g} = I_M(W)(w)$ (see definition 4 in §5.4). Furthermore, we have $\llbracket W(a) \rrbracket_{M2, w, g} = I_{M2}(W)(g[a/w](a)) = I_{M2}(__)(w)$ (compare this with definition 4 in §4.2.2). According to our assumptions, $I_M(W) = I_{M2}(W)$, from which it immediately follows that $I_M(W)(w) = I_{M2}(W)(w)$. Finally, since w and g were chosen arbitrarily, we have $\llbracket W \rrbracket_{M, w, g} = \llbracket W(a) \rrbracket_{M2, g[a/w]}$.

It turns out that this relation between M and $M(a)$ may be generalized: intensional type-theoretical expressions can always be translated into two-sorted formulas which carry the same meaning. We will show this by defining a translation from expressions of intensional type theory to expressions of two-sorted type theory which preserves interpretation in the sense discussed above. Not surprisingly, constants c_a are always translated as new constants $c_{\langle s, a \rangle}$, which are then applied to a . The translation of a complex expression α is for the remainder built up just like α itself, except where \Box , \Diamond , \wedge , or \vee enter into α . Intuitively it should be clear that \Box corresponds to $\forall a$, \Diamond to $\exists a$, \wedge to λa , and \vee to application to a . \Box , \Diamond , and \wedge thus correspond to binding

free occurrences of a , while \vee introduces such an occurrence. We shall now give a precise, inductive definition of the translation, written as $\text{trans-}\alpha$:

Definition 9

- (i) $\text{trans-}c_a = c_{(s,a)}(a)$
 $\text{trans-}v_a = v_a$
- (ii) $\text{trans-}(\alpha(\beta)) = (\text{trans-}\alpha(\text{trans-}\beta))$
- (iii) $\text{trans-}\neg\phi = \neg\text{trans-}\phi$
 $\text{trans-}(\phi \wedge \psi) = (\text{trans-}\phi \wedge \text{trans-}\psi)$; likewise for $\vee, \rightarrow, \leftrightarrow$
- (iv) $\text{trans-}\forall v\phi = \forall v \text{trans-}\phi$; likewise for $\exists v\phi$
- (v) $\text{trans-}(\alpha = \beta) = (\text{trans-}\alpha = \text{trans-}\beta)$
- (vi) $\text{trans-}\lambda v\alpha = \lambda v \text{trans-}\alpha$
- (vii) $\text{trans-}\Box\phi = \forall a \text{trans-}\phi$
 $\text{trans-}\Diamond\phi = \exists a \text{trans-}\phi$
- (viii) $\text{trans-}\wedge\alpha = \lambda a \text{trans-}\alpha$
- (ix) $\text{trans-}\vee\alpha = (\text{trans-}\alpha(a))$

It is possible to convert any model \mathbf{M} for intensional type theory into a model $\mathbf{M2}$ for two-sorted type theory such that for all expressions α , for all $w \in W$ and for all assignments g , we have:

Theorem 6

$$\llbracket \text{trans-}\alpha \rrbracket_{\mathbf{M2}, g[a/w]} = \llbracket \alpha \rrbracket_{\mathbf{M}, w, g}$$

$\mathbf{M2}$ may be obtained from \mathbf{M} by taking the latter's domain D and set W of worlds, defining $I_{\mathbf{M2}}(c)$, for any constant c , as $I_{\mathbf{M}}(c)$. Now an inductive proof is needed in order to show that theorem 6 does indeed hold. We shall only deal with the nontrivial steps of the proof (assuming in the induction that theorem 6 holds for all $w' \in W$). The clauses have been numbered so as to correspond with those in definition 4.

- (i) $\llbracket \text{trans-}c \rrbracket_{\mathbf{M2}, g[a/w]} = \llbracket c(a) \rrbracket_{\mathbf{M2}, g[a/w]} = \llbracket c \rrbracket_{\mathbf{M2}, g}(w) = I_{\mathbf{M2}}(c)(w) = I_{\mathbf{M}}(c)(w) = \llbracket c \rrbracket_{\mathbf{M}, w, g}$.
- (vii) $\llbracket \text{trans-}\Box\phi \rrbracket_{\mathbf{M2}, g[a/w]} = 1 \Leftrightarrow (\text{translation}) \llbracket \forall a \text{trans-}\phi \rrbracket_{\mathbf{M2}, g[a/w]} = 1 \Leftrightarrow$
 $(\text{truth definition}) \text{ for all } w': \llbracket \text{trans-}\phi \rrbracket_{\mathbf{M2}, g[a/w']} = 1 \Leftrightarrow (\text{induction hypothesis}) \text{ for all } w': \llbracket \phi \rrbracket_{\mathbf{M}, w', g} = 1 \Leftrightarrow \llbracket \Box\phi \rrbracket_{\mathbf{M}, w, g} = 1$.
- (viii) $\llbracket \text{trans-}\wedge\alpha \rrbracket_{\mathbf{M2}, g[a/w]} = \llbracket \lambda a \text{trans-}\alpha \rrbracket_{\mathbf{M2}, g[a/w]}$, which is that function h such that for all w' : $h(w') = \llbracket \text{trans-}\alpha \rrbracket_{\mathbf{M2}, g[a/w']}$. By the induction hypothesis, this means that for all w' : $h(w') = \llbracket \alpha \rrbracket_{\mathbf{M}, w', g}$, which means that $h = \llbracket \wedge\alpha \rrbracket_{\mathbf{M}, w, g}$.
- (ix) $\llbracket \text{trans-}\vee\alpha \rrbracket_{\mathbf{M2}, g[a/w]} = \llbracket (\text{trans-}\alpha(a)) \rrbracket_{\mathbf{M2}, g[a/w]} = \llbracket \text{trans-}\alpha \rrbracket_{\mathbf{M2}, g[a/w]}(w) =$
 $(\text{induction hypothesis}) \llbracket \alpha \rrbracket_{\mathbf{M}, w, g}(w) = \llbracket \vee\alpha \rrbracket_{\mathbf{M}, w, g}$.

It is striking that the translation process makes use of just one variable a ranging over possible worlds. The expressive power of two-sorted type theory is clearly much greater than that of intensional type theory.

This translation from intensional into two-sorted type theory throws new light on the former. Thus, for example, the equivalence of $\vee^{\wedge}\alpha$ and α in intensional type theory reappears in two-sorted type theory simply as a case of valid λ -conversion. The formula corresponding to $\vee^{\wedge}\alpha$ is $\lambda a\alpha(a)$, and this, in view of the fact that theorem 1 of §4.4.2 also holds for two-sorted type theory, is equivalent to α .

It also becomes clearer why $\wedge^{\vee}\alpha$ is not always equivalent to α . The counterexample given in §5.5 may be reconstructed in this new setting in the following manner. Instead of the constant p which appeared there, we make use of the expression $\lambda w(q(w) = q(a))$, in which both w and a are variables of type s , while q is of type $\langle s, t \rangle$. If q is taken to represent the proposition *that Mary is coming*, then $q(w)$ is true in $g(w)$ if *Mary is coming* is true in $g(w)$, and $q(w)$ is false in $g(w)$ if *Mary is coming* is false in $g(w)$. The expression $\lambda w(q(w) = q(a))$ thus denotes the proposition that Mary is coming in those worlds in which it is true that she is coming, and the proposition that Mary is not coming in those worlds in which it is false that she is coming. That is, $\lambda w(q(w) = q(a))$ represents *whether or not Mary is coming*. Just as $\wedge^{\vee}p$ and p were not equivalent, we now have that $\lambda a(\lambda w(q(w) = q(a)))(a)$ and $\lambda w(q(w) = q(a))$ are nonequivalent. For $\lambda a(\lambda w(q(w) = q(a)))(a)$ reduces to $\lambda a(q(a) = q(a))$, and this expression refers to a proposition different from that referred to by $\lambda w(q(w) = q(a))$. Just like $\wedge^{\vee}p$ in §5.5, $\lambda a(q(a) = q(a))$ in fact refers to the proposition which is true in every world.

We noted in §5.5 that there are circumstances under which $\wedge^{\vee}\alpha$ is equivalent to α , among others, where α is intensionally closed. What is now the pendant in two-sorted type theory of intensionally closed expressions? Since context dependence is captured here by giving context-dependent expressions a variable a as an extra argument, intensional closure of an expression in intensional type theory amounts to the corresponding two-sorted expression's not containing any free occurrences of a . According to definition 6 in §5.5, $\Box\phi$, $\Diamond\phi$, and $\wedge\alpha$ are all examples of intensionally closed expressions. The corresponding two-sorted expressions $\forall a \text{ trans-}\phi$, $\exists a \text{ trans-}\phi$, and $\lambda a \text{ trans-}\alpha$ may be seen not to contain free occurrences of a .

It is now not difficult to see what condition (ii) in theorem 5 of §5.6 amounts to in this setting. This theorem, repeated below, concerns the admissibility of λ -conversion in intensional type theory:

Theorem 5

$\lambda v\beta(\gamma)$ is equivalent to $[\gamma/v]\beta$ if

- (i) all free variables in γ are free for v in β ; and
- (ii) either $\gamma \in \text{ICE}^{\perp}$, or no free occurrence of v in β lies within the scope of \Box , \Diamond , or \wedge .

In two-sorted type theory, condition (ii) amounts to the condition that if γ contains a free occurrence of a , then γ may not as a result of the substitution find

itself within the scope of $\forall a$, $\exists a$, or λa . But this is already required by condition (i). So in two-sorted type theory, as in normal extensional type theory, condition (i) suffices on its own.

So we see how certain matters which arose in intensional type theory now reappear, in this two-sorted setting, in a form already familiar from extensional type theory.

Two-sorted type theory enables us to refer to possible worlds and to quantify over them (and over moments in time if we choose to introduce them into the picture). Now one objection which may be raised against such a language is that it has more ontological commitments than a language with intensional operators. It seems to us that this objection is unfounded. A mere language doesn't have ontological commitments. What has ontological commitments is a language together with its semantics. And the semantics of a language with intensional operators, in the form of a truth definition which avails itself of possible worlds, plainly refers to possible worlds and quantifies over them every bit as much as the semantics of a two-sorted language. The ontological commitments are quite the same in both cases. Another reason for preferring a language with intensional operators over a two-sorted language has to do with their difference in expressive power. A language with intensional operators has the advantage that it may be given exactly as much expressive power as is needed for a particular application. With applications like the semantics of natural language in mind, it becomes an empirical question as to how expressive this power is. Do we or do we not, to take an example, need quantification over moments in time in order to render the temporal expressions of natural language in a satisfactory manner? These are complex but quite fascinating questions. In §2.4.3 we noted that in the course of time, temporal formalisms have been equipped with more and more operators in order to deal with temporal expressions and constructions, and that some have argued for languages which allow quantification over moments. An issue like this is difficult to settle; the question is not only whether the different theories capture all the phenomena. Their simplicity and elegance is also at stake.

6 Montague Grammar

6.1 Introduction

In this chapter we shall deal extensively with *Montague grammar*. Montague grammar, developed by the American logician Richard Montague in the beginning of the seventies, aims to define a model-theoretic semantics for natural language. The most common version of Montague grammar, which is the one introduced in this chapter, achieves this by relating in an explicit and systematic fashion the expressions of a natural language to those of an intensional, type-theoretical logical system in such a way that the interpretations of the latter may also serve as interpretations of the former.

Montague's model wasn't the only attempt made at that time to use the semantic methods of logic in the description of the meanings of natural language expressions. Others, e.g., Cresswell, Bartsch and Vennemann, and Lewis, made proposals that went in the same direction. A common denominator for such models is 'logical grammars'. In this chapter we will be concerned exclusively with Montague's model. For as it is, Montague grammar still serves as the standard model of a logical grammar. And the thorough introduction that follows will enable readers, if they desire, to master the particularities of other models relatively quickly. Also, the more recent developments in formal semantics which will be discussed in chapter 7 can be understood properly only against the background of Montague grammar. And the same holds for such current trends as situation semantics, which at least initially got its momentum by starting out as a full-blown attack on some of the essentials of Montague grammar.

The idea of the enterprise of a logical grammar for natural language is neither self-evident nor does it come out of the blue. In chapter 1 of volume 1, we discussed briefly the historical development of logic and linguistics and their interrelationship. The discussion tried to make it clear that the idea of a common goal certainly has its roots in history, but at the same time it is made feasible only by certain relatively recent developments in both disciplines. This chapter starts with a short discussion of three methodological presuppositions of the enterprise, three general principles which underlie most attempts at a systematic application of model-theoretic semantics to natural

language. These concern the principle of compositionality of meaning and its consequences for syntax, the problem of semantic closure, and the relationship between truth and meaning.

6.1.1 *Compositionality of Meaning and Syntax*

The principle of compositionality of meaning has important consequences for the relationship between syntax and semantics. Usually in a logical system the definition of the semantic interpretation of expressions closely follows the lead of their syntactic construction. The reason for this is that the semantics must specify the interpretation of an infinite number of expressions, but in a finite manner. The obvious way to proceed, then, is to let the definition of the semantics parallel the finite, recursive definition of the syntax. This method ensures that to every syntactic rule which allows us to construct a certain type of expression out of one or more simpler ones a semantic rule corresponds, which states how the interpretation of the newly formed expression is to be obtained from the interpretations of its component parts. Succinctly put, logical languages satisfy the following principle: the interpretation of a complex expression is a *function* of the interpretations of its parts. This is the principle of compositionality of meaning, also referred to as 'Frege's principle'.

The actual formulation of a logical system may not always carry its compositionality on its sleeve, but every such system actually conforms to, or can be reformulated so as to conform to, the principle of compositionality. For example, in volume 1, §2.7, an alternative definition of the system of propositional logic is given which is equivalent to the usual one but makes its compositionality explicit. Compare also the remarks made above, in §§4.3.4 and 4.4.3, about the possibility of introducing the existential and universal quantifiers in a categorematic rather than syncategorematic fashion. In fact compositionality is so basic a starting point for the logical way of doing semantics that in logic proper it almost always goes unnoticed.

If we consider natural language, however, the compositionality of meaning requires more attention, for the following reasons. It is evident that compositionality provides a finite method for the semantic interpretation of an infinite number of expressions of a given language. Given that a model specifies the interpretation of the basic components, the semantic rules which correspond to the syntactic rules uniquely determine the interpretation of every complex expression. But it should be noticed that in effect, compositionality puts heavy constraints on the syntax, the semantics, and their relation to each other. On the one hand, every syntactic rule should have a semantic interpretation; and on the other hand, every aspect of the semantics which is not related to the interpretation of basic expressions should be linked to a syntactic operation. In a logical system, we comply with these requirements simply by setting things up in accordance with them. But a natural language is not something we construct; it comes as given.

In fact, the assumption that natural language semantics can be defined compositionally is not uncontroversial. Tarski, one of the founders of modern model-theoretic semantics, didn't have much faith in the possibility of applying its methods to natural language. For this requires minimally that the syntax of a language be exactly specified in a rigorously formal way, and Tarski was of the opinion that for a natural language this is an unreachable goal. This view prevailed among philosophers and logicians for quite some time. It was not until the development of generative grammar in the late fifties and early sixties that a different attitude began to emerge. Because of the enormous impact of generative linguistics, the conviction grew that the required exact description of natural language syntax *can* be given and that the possibility of a logical semantics for natural language would become real.

Ironically, the idea of using the methods of logical semantics in the study of natural language didn't receive a very warm welcome among most linguists in the Chomskyan tradition, and still doesn't, officially at least, although some results and methods did catch on. On the other hand, people like Katz have argued for some version of compositionality along the following lines. A competent language user is capable of interpreting a theoretically infinite number of sentences. The language user's interpretation is based on his knowledge of the meaning of the finite number of lexical elements and his (implicit) knowledge of syntactic rules, also finite in number. The hypothesis of compositionality, requiring that semantic rules correspond to syntactic rules, seems to offer an explanation for this fact. (This line of argumentation can be found in various authors; cf., for example, Frege 1923. It should be noted that the argument does not prove that natural language is compositional; it only shows that there must be *some* effective means to compute meanings. And compositionality is only one of the candidates.)

There are probably two major reasons why generative grammarians are skeptical about the enterprise of logical grammar. One is their commitment to mentalism, a doctrine that doesn't seem to square with the truth-based approach to meaning that logical grammar starts from (see §6.1.3 below). The other is that logical grammar, with its principle of compositionality of meaning, goes straight against the autonomy of syntax so cherished in the generative tradition. For compositionality not only requires a well-defined syntax to base semantic interpretation on; it also puts some constraints on it. As we remarked above and will argue more fully below, in §6.2, it follows from compositionality that every nonlexical aspect of meaning must be syntax based. (Here it should be kept in mind that we are considering only a sentence grammar consisting of a syntax and a semantics, which has no recourse to discourse, intonation, extralinguistic context, and the like.)

And that means, at least in principle, that semantic considerations may influence the syntax, thus breaching the supposed autonomy of the latter.

Exactly what role the principle of compositionality plays in the overall layout of a logical grammar will be discussed in §6.2. In §6.5 we will discuss

in some detail its methodological status and its relationship with the contrast between logical form and grammatical form. For now it suffices to have pointed out the important role it plays in the logical grammar enterprise.

6.1.2 Object Language and Metalanguage: Semantic Closure

There is a more technical condition that a language must fulfill in order for it to be possible to define its semantics along the lines of a logical system. It concerns the relationship between the language we want to give a semantics *for*, which is called the *object language*, and the language we formulate the semantics *in*, which is called the *metalanguage*. These two languages may actually be different, as in the preceding chapters where we used English, the metalanguage, to formulate the semantics of several logical languages, the object languages. But it is possible for the object language to be part of the metalanguage. Thus, the semantics of a fragment of the English language can be stated in English. The terms 'object language' and 'metalanguage', then refer rather to different *functions* of language, which may be performed by different languages but also by one and the same language.

The question that now arises is whether it is possible for object language and metalanguage to be identical. At first sight, nothing seems to speak against the supposition that this can be the case, but closer scrutiny will show that this leads to some unexpected problems.

In describing the semantics of an object language, among other things we state in the metalanguage the truth conditions of sentences in the object language. Working in the familiar Tarski style, this means that in the metalanguage, names are available for object language sentences. Usually we use the sentences themselves, transforming them into names by putting them between quotation marks or writing them in italics. As a result, it is possible to define in the metalanguage a *truth predicate*. A truth predicate is a predicate of the metalanguage which holds of an object language sentence if and only if that sentence is true. That is, giving the truth conditions of object language sentences amounts to the same thing as specifying the extension of the truth predicate. All this is quite in order, except when object language and metalanguage are identical. For then the truth predicate would ipso facto be part of the object language as well, as would the names of object language sentences; this would give rise to *semantic paradoxes*. For instance, suppose that English is our object language and our metalanguage. Then it would contain its own truth predicate *is true* and would contain names for all its sentences; this would mean that we could formulate sentences such as:

- (1) Sentence (1) is not true.

This sentence yields a paradox. For (1) to be true, (1) should not be true, for that is what (1) asserts. But if (1) is not true, then the assertion expressed by (1) is true, and so (1) is true. We may conclude that we cannot formulate a

semantic theory for an object language in a metalanguage which is identical to that object language without getting into trouble. But the problem is more serious than that. For in a situation in which object language and metalanguage are different, the same unpleasant results may obtain. For instance, if we described the meaning of sentence (1) in language other than English, say in Dutch, the paradox would not disappear. It would just get a different wording. The trouble arises with any object language which allows self-reference, e.g., by containing names for its own sentences, and which contains its own truth predicate. Such a language is called *semantically closed*. It must be concluded that a consistent semantics, one which will not yield paradoxes such as those indicated above, can be defined only for those languages which are not semantically closed.

As example (1) shows, English is semantically closed, and this implies that it will not be possible to give a consistent semantic theory for it. Since this holds quite generally for any natural language, it seems that we must conclude that the enterprise of logical grammar is on the wrong track right from the very start. There are several ways to avoid this conclusion. Perhaps the most common one is to steer a safe course by formulating the semantics, not for the entire language, but only for those fragments of it which are not semantically closed. The loss of generality seems rather small. Of course if we take this approach, it again becomes possible to formulate the semantics in the language itself. In that case, the object language belongs to the metalanguage but is not identical with it, and no paradoxes ensue. This way to get around the problem is essentially Tarski's (1935, 1944). As a result, we get a hierarchy of ever more inclusive languages in which a language of level $n + 1$ functions as the metalanguage for the language of level n . From an empirical point of view, it is argued, the restriction this involves is not important, but there are several formal conditions which have to be met in a semantic theory set up along these lines.

Beside this rather formal approach, others have been proposed, which purport to take the paradoxes more seriously. Note that if we follow Tarski's approach, paradoxical sentences aren't assigned the meanings they intuitively seem to have, the ones that yield the paradoxes. If there were only such directly self-referential sentences as (1), this might seem a small price to pay, but semantic paradoxes are also generated by more natural examples. Compare the two statements in (2):

- (2) A: Everything B says about me is false.
 B: Everything A says about me is true.

(If we substituted, for instance, the names of two politicians running for the same office, we might get more natural examples.) Now, suppose that in fact A and B say only one thing about each other, that is, the statements in (2). Is what A says true? Then it should hold that B's statement that everything A says about B is true, is false. But since, by supposition, the only thing A says

about B is that whatever B says about A is false, that implies that A's statement is false. Similarly, A's statement is true if it is false. Paradox regained. Examples such as (2) can be multiplied and can be made more natural by describing the circumstances more fully. To many this has suggested that semantic paradoxes are a less marginal feature than is often thought, and that rather than avoiding them—which is what Tarski's approach essentially comes down to—they should be faced head-on. Various solutions have been proposed to the problem of setting up a semantic theory that does this, e.g., by Kripke, Herzberger, Gupta, and recently by Barwise and Etchemendy in the framework of situation semantics.

In conclusion, we can say that whatever position one wants to take here, the existence of semantic paradoxes does not preclude the possibility of a model-theoretic semantics for natural language. But of course we have to take measures. Either we must restrict ourselves in describing natural language, along the lines suggested by Tarski, or we must adapt one of the more direct, but also more complex, approaches developed by Kripke and others.

6.1.3 *Semantics and Truth Theory*

There seem to be no formal or methodological obstacles to an application of the methods of logical, model-theoretic semantics to natural language. However, it may be argued that although it is perhaps possible to describe the meanings of expressions of a natural language along the same lines as those of a logical one, it is by no means obvious that it is profitable to approach meaning in natural language this way. Here we touch upon, and only touch upon, a very fundamental issue.

As was explained in chapter 1, model-theoretic semantics can be viewed as a referential theory of meaning. Meaning is explained in terms of the relation of reference, or denotation, that holds between expressions and some independent set of entities. This holds for an intensional semantics, as well as for an extensional semantics: multiple reference, in terms of which intension is defined, is a relation of reference too, with an extra parameter. Hence, the notions of *reference* and *truth* are to be regarded as the key notions of model-theoretic semantics. And if we attempt to use the methods of the latter in our description of meaning in natural language, we assume, implicitly or explicitly, that at least a substantial part of the meaning of natural language expressions can be caught in terms of the notions of reference and truth. In other words, one of the starting points of logical grammar is the idea that a semantic theory for a natural language should at least contain a truth definition for that language.

Exactly at this point there is a marked opposition between the proponents of logical grammar and the adherents of generative grammar. Whereas the former view meaning essentially as a relation between the expressions of a language and something else, 'out there', that the expressions refer to, make

statements about, or ask questions about, the latter hold that meaning consists in the mental representation that accompanies a linguistic expression. The controversy may be very great, as between a logical grammarian who sees no role whatsoever for mental representations to play, and a mentalist who denies that language has any referring function. Or it may be less dramatic, since a logical grammarian may acknowledge the existence of mental representations, and the mentalist may admit that language somehow, indirectly, also relates to the nonmental. But in any case, the difference is a principled one. For although there might be agreement on other things, it represents a fundamental difference of opinion about what constitutes meaning: its relation to the 'world' or its relation to the 'mental'.

We cannot go into the arguments that have been proposed in its favor. We simply note that a long-standing tradition exists in philosophy and in logic which assumes that any adequate semantic theory should at least contain a theory of truth. About that there has been, and still is, agreement. And since the late sixties, several proposals have been made for applying the methods of model-theoretic semantics in the analysis of natural language. However, it should be remarked that here two main streams of thought have to be distinguished. All proposals take the formulation of truth conditions of sentences, or more generally, denotation conditions of expressions, to be an essential part of the theory, but they agree on neither the method to bring this about nor the question of whether a definition of truth conditions is sufficient in itself as an analysis of meaning in natural language. With respect to the latter question, we may distinguish between 'extensionalists' and 'intensionalists'. Such a noted extensionalist as Davidson holds that it is possible and necessary to do semantics for a natural language solely in terms of truth conditions. Davidson shares this view with Quine, who for philosophical and methodological reasons has always opposed the use of notions such as 'meaning', 'synonymy', and so on, even in logic. The position that any explanation of semantic facts using such notions would be an explanation of *obscurum per obscurior* has been put forward and defended forcefully by Quine in numerous places. According to Quine and Davidson, then, the use of theoretical notions like those of a possible world, intension, and hence of intensional semantics as such should be avoided, since no real insight into the meaning of natural language is to be expected from them. They are of the opinion that a semantic theory ought to be formulated in purely extensional terms.

The intensionalist stream of thought holds that such a position is inspired too much by purely philosophical motives, and that it pays too little attention to the requirements of an empirically adequate semantic theory of natural language. According to the intensionalists, the intensional character of natural language is obvious. Many expressions and constructions pose severe problems for a strictly extensional semantics. (Cf. the examples in the §§1.6 and 3.1.) An extensionalist semantics cannot cope with these difficulties. If our aim is an empirically adequate semantic theory for natural language rather

than a semantic theory that meets some independent philosophical constraints, the obvious way to proceed is to use an intensional semantics. This view has been put forward both theoretically and practically by several authors, notably by Montague, Lewis, and Cresswell. The models they have developed really aim to provide a framework for an empirically adequate semantic theory for natural language. For this reason, these models can truly be characterized as ‘logical grammars’.

This chapter is devoted to a detailed exposition of one of these logical grammars, and for reasons indicated in §6.1, the model we have chosen is the one developed by Richard Montague.

6.2 The Organization of a Montague Grammar

This section deals with the organization of Montague grammar. We adopt Montague’s model as proposed in “The Proper Treatment of Quantification in Ordinary English” (1973), the *PTQ model*. The PTQ model is one of the possible versions of Montague’s general theory of syntax and semantics which he formulated in his “Universal Grammar” (1970b). This general theory has led to several somewhat differently organized models; for instance, Montague’s approach in “English as a Formal Language” (1970a) differs in several respects from the PTQ model, and it is therefore strictly speaking misleading to present the PTQ model as *the* Montague grammar. On the other hand, the PTQ model is the best-known and most widely used implementation of Montague’s ideas, and it even counts as the paradigmatic formulation of logical grammar as such. So with these reservations in the back of our minds, we concentrate on the PTQ model in the remainder of this chapter.

Methodologically, Montague’s most important notion is this: the principal task of a linguistic theory must be the framing of a semantic theory. This presupposition has consequences for the organization of the grammar. Compositional semantics requires a syntactic theory. To put it differently: interpreting a complex expression compositionally is not merely interpreting the expression as such but interpreting it *given* a syntactic analysis. The syntactic analysis tells us what subexpressions the expression is composed of, what rules are used to form the expression, and in what order the expression is composed. And we need this information if we want to construct the meaning of the expression out of the meanings of its component parts, as compositionality requires. Semantic interpretation is done not on expressions as such but only on expressions given a syntactical analysis. In this sense a semantic theory presupposes a syntactic theory.

The usual model-theoretic interpretation of logical languages fulfills this requirement. Logical languages are characterized by the fact that expressions and their syntactic analysis cannot be distinguished: an expression shows its syntactic analysis in its structure. For each expression, there is exactly one construction tree that can be deduced from the expression unambiguously,

thanks to the use of brackets. Syntactically ambiguous sentences do occur in natural languages, where one expression can be the result of different syntactic processes and hence may have more than one syntactic analysis. Sometimes different syntactic analyses give rise to different meanings. As a result, we can determine the meaning of an expression only on the basis of a syntactic analysis of it.

The next three examples will illustrate this. Consider sentence (3):

- (3) John sees old men and women.

Sentence (3) has two readings: on one, John sees women and old men; on the other, John sees old men and old women. The constituent expression *old men and women* is the source of the ambiguity. The scope of the adjective *old* determines what set of persons the expression refers to. When the scope is just the noun *men*, the reference is as on the first reading, while the second reading of (3) is the one in which *old* applies to the complex expression *men and women*. The corresponding two different ways of constructing *old men and women* can be represented as follows:

- (4) a. [[old men] and women]
b. [old [men and women]]

The two meanings of (3) can be constructed using the syntactic analyses (4a) and (4b). Hence, the process of interpretation works on representations like (4a) and (4b), rather than on unstructured expressions like *old men and women*.

The ambiguity of (3) is *structural*: it resides in the two different structures that can be assigned to *old men and women*. But not all syntactic ambiguities can be reduced to such structural ambiguities. Consider sentences (5) and (6):

- (5) Everyone in this room speaks one language.
(6) John seeks a unicorn.

Each of these sentences has two readings as well. On one reading of (5), there is one language which everyone in this room speaks, for example, English, while some people may also speak other languages; the other reading has it that everyone speaks only one language, possibly every person a different one. A *de dicto/de re* ambiguity occurs in (6). The *de re* interpretation states that there is a unicorn which John seeks, but the *de dicto* interpretation does not imply the existence of a unicorn. Ambiguities such as these *cannot* be reduced to structural ambiguities, as was the case with (3). For a sentence such as (5) or (6) has only one constituent structure. Apparently for sentences such as these there are *different ways to derive* them, which all result in the same expression with the same structure but a different meaning. The two ways to derive (5) differ in the order in which the quantified terms *everyone in this room* and *one language* are introduced. If we introduce *one language*

first, we get the second reading; if we start with *everyone in this room* we get the first reading. Something similar happens in (6), as will be shown later on. The ambiguity of (5) or (6) is the result of the existence of more than one syntactic derivation; hence it is often referred to as a *derivational ambiguity*. The structural ambiguity of (3) can be regarded as a special kind of derivational ambiguity.

We may conclude that it is the derivation of an expression which determines its meaning. We will use the term *syntactic analysis* in this sense, referring to the derivational history of an expression, distinguishing it from the notion of a *structural analysis*.

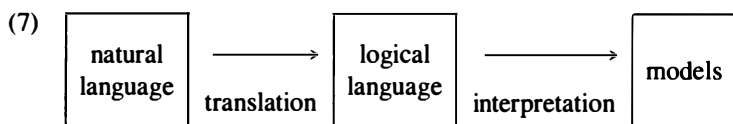
A compositional semantic theory not only presupposes a syntactic theory but also imposes certain conditions on it. Every nonlexical ambiguity, that is any ambiguity which cannot be reduced to the ambiguity of a lexical element, should correspond to a derivational ambiguity. Another condition requires that every syntactic operation be semantically interpretable; its semantic effect must be statable explicitly and in general terms.

All this means that the syntax of Montague grammar is not autonomous, or at least it need not be. Semantic considerations may be of paramount importance in the formulation of the syntax. A syntactic ambiguity might be built in, so to speak, merely for semantic reasons, and similarly, semantic considerations might force a choice between alternative syntactic analyses which cannot be based on purely syntactic considerations. It remains an open question whether this potential transgression of the autonomy of syntax by semantics will be encountered in reality, that is, in the actual description of some natural language.

Though it imposes conditions on the form and content of the syntax, the semantics enables us to explain certain semantic features without relying on syntactic considerations. Two sentences with the same meaning may differ considerably syntactically, even to the extent that from a syntactic point of view it is not at all plausible to relate them to each other. As an example, consider sentences in their active and passive forms. An explicit semantic theory enables us to account for the conformity in meaning without having to subscribe to the hypothesis that the sentences are equal at some syntactic level. It is sufficient that their syntactic analyses determine the same semantic interpretation. In this sense the nature and power of compositional semantics may make up for the transgression on the autonomy of syntax by enabling us to eliminate from the latter irrelevant, i.e., semantic, considerations.

The organization of the PTQ model is dictated not only by the above-mentioned motives and principles but also by its characteristic method of linking syntax and semantics. It is possible to define a model-theoretic semantics for a natural language *directly*. The syntactic analysis of an expression provides the basis for a direct interpretation in a model, in the same way as this is achieved for the expressions of logical languages. Montague adopted this method of direct interpretation in his "English as a Formal Language"

(1970). In the PTQ model he used the *indirect* method, however. In this model, expressions of a natural language are first translated into expressions of a logical language, as is shown in (7). The logical expressions are interpreted in the usual model-theoretic fashion. Hence, natural language expressions are interpreted indirectly through the interpretation of the logical expressions they are translated into.



As usual, the logical language is a unambiguous language; therefore the semantics assigns one meaning to each logical expression. If the semantic interpretation of a logical expression is to function indirectly as the semantic interpretation of a natural language expression, some conditions must be put on the process of translation. It is not possible simply to translate natural language expressions into logical expressions. If a natural language expression is ambiguous, we need to assign it more than one meaning. In the PTQ model, this implies that an ambiguous expression must be translated into distinct, nonequivalent, logical expressions. The principle of compositionality requires the correspondence of every nonlexical ambiguity to a derivational ambiguity, as we explained before.

Since the process of translation is nothing more than an indirect process of interpretation, it too should conform to the syntactic analysis. A translation is rendered, not for an expression as such, but only for an expression given a syntactic analysis. If there are different ways to analyze, that is, to derive an expression, there must be different translations, representing different meanings. (As we shall see later on, not every derivational ambiguity corresponds to a genuine semantic ambiguity, though the inverse is true.) So in the PTQ model, compositionality plays a role on two different levels. First there is the level of the interpretation of the logical language, which satisfies compositionality as usual. But compositionality is also involved at the level of the translation of the natural language into the logical language. This is necessary if we want to ensure the compositionality of the process of interpretation of the natural language: we must therefore organize the process of translation as a compositional process. All the lexical elements of the natural language are translated into logical language expressions and in this way are assigned unique meanings. Syntactic rules tell us how to construct new expressions out of existing ones. With every such rule a translation rule is associated which specifies the translation of the complex expression, given the translation of its component parts. Ultimately, every expression is constructed out of basic expressions by the application of a finite number of syntactic rules. The construction supplies the syntactic analysis, which determines a unique translation and thus a unique meaning. If an expression can be analyzed syntactically

in more than one way, then it has more than one translation and hence possibly more than one meaning.

The general structure of the PTQ model is thus defined. In the following sections we shall give an example of a grammar according to the PTQ model for a small fragment of English, covering roughly the same area as Montague himself covered in PTQ. This description aims at no more than a tangible illustration of the above-mentioned principles and notions, to enable the reader to understand Montague's work and the work of others in the tradition he founded. Some constructions will be examined more closely than others; we shall confine ourselves to a detailed account of those constructions and analyses which form the core of the PTQ fragment. Our exposition will proceed step by step. In order to facilitate exposition and, we hope, understanding, our fragment will initially be built up in a way which deviates in one important respect from PTQ itself. In a later stage, we will add the extra complications and thus obtain a fragment which is substantially that of PTQ. Compared with the richness and vastness of natural language itself, this fragment is of course very restricted. Nevertheless, it contains an analysis of several phenomena of which every adequate semantic theory should provide an account. The limited descriptive range of the fragment should not be taken as an indication of limited possibilities for the application of Montague grammar. A great variety of phenomena interesting both syntactically and semantically have been described and studied within the framework of Montague grammar. Some examples will be mentioned in §6.5.

6.3 A Montague Grammar for a Fragment of English

6.3.1 *Categories and Basic Expressions*

The PTQ model uses a categorial syntax to generate the expressions of a natural language. As we indicated in §4.3, a pure categorial syntax consists of four things: (i) an enumeration of the basic categories; (ii) a definition of the derived categories; (iii) a lexicon, i.e., a specification of the lexical elements of each category; and (iv) a specification of the behavior of the syntactic operation of concatenation. This form of categorial syntax is equivalent to a simple system of context-free rewrite rules, and we therefore know that it will have difficulties in coping with phenomena such as word order, deletion, discontinuous constituents, morphological features, and so on. Most of these will feature in the fragment to be treated. Over the years several proposals have been made to make up for the inadequacies of pure categorial syntax. An early suggestion, made by Lyons and Lewis, was to use a pure categorial syntax as a base component and to add a transformational component to deal with these phenomena. At present, other strategies are more popular. One of them is to introduce a certain systematic flexibility into the assignment of categories to expressions. This approach will be introduced in chapter 7.

In PTQ, Montague in effect overcame the shortcomings just referred to in a

rather crude and ad hoc way, viz., by simply allowing the use of all kinds of syntactic operations in the syntactic rules. In a pure, unidirectional categorial syntax, the only syntactic operation allowed in forming expressions is concatenation (either to the left or to the right). Consequently, such a syntax has just one syntactic rule. In PTQ, complex expressions can be obtained by other methods besides the simple concatenation of two expressions of suitable categories, and the number of syntactic rules grows accordingly. Also, the generative power of the syntax is increased.

The syntactic operations used in the syntactic rules of the PTQ fragment are a rather heterogeneous lot. We find the simple operations of (left and right) concatenation but also operations which introduce expressions syncategorematically, which change word order, which regulate the morphological form of expressions, and we even find operations that perform several of these tasks at the same time. From a modern linguistic point of view, the obvious thing to do is to look for restrictions on the admissible operations. Various proposals have been made in this direction.

Now let us define a categorial syntax for a small fragment of English, starting with a definition of the categories:

Definition 1

CAT, the set of categories, is the smallest set such that:

- (i) $S, CN, IV \in CAT$.
- (ii) If $A, B \in CAT$, then $A/B \in CAT$.

This syntax contains three basic categories: S , the category of sentences; CN , the category of common noun phrases; and IV , the category of intransitive verb phrases. The derived categories A/B are functor categories: an expression of category A/B takes an expression of category B as its argument to yield a new expression of category A . In fact, clause (ii) of definition 1 defines an infinite number of derived categories, but only a few of these will actually be used. For some derived categories special abbreviations will be introduced. Table 6.1 sums up the categories that will be used, giving a definition if a category name is an abbreviation of a derived category, and a more familiar linguistic characterization. It also gives the lexicon of our fragment; next to each of the categories the lexical elements of that category are listed. In the following sections, we will often designate the lexical elements (also called *basic expressions*) of a category A as B_A . According to table 6.1, then, B_{CN} is the set $\{man, woman, language, unicorn, elephant, queen, park\}$. Note that the number of basic expressions of category T is infinite. For every natural number n , $he_n \in B_T$. The function of these expressions, called *syntactic variables*, will become clear later on.

So far we have defined three of the four components of a categorial syntax: the basic categories, the derived categories, and the lexicon. What remains to be given is an enumeration of the syntactic rules that define how to create new

Table 6.1 Categories and Expressions

<i>Categorial definition</i>	<i>Description</i>	<i>Expressions</i>
S	Sentences	
CN	Common nouns	Man, woman, language, unicorn, elephant, queen, park
IV	Intransitive verb phrases	Smoke, sleep, walk, talk, stroll
T = S/IV	Terms	John, Mary, Bill, Elsie, he ₀ , he ₁ , . . .
TV = IV/T (= IV/(S/IV))	Transitive verb phrases	Love, kiss, know, seek, find, be
IV/S	Sentential complement verbs	Believe that, assert that
IV/IV	Infinitival complement verbs	Try to, wish to
CN/CN	Prenominal adjectives	Green, large, pink, square, imaginary
S/S	Sentence-modifying adverbs	Necessarily

expressions out of existing ones. The syntactic rules for the fragment will be the subject of the next sections. Their definitions will be given one by one: first a few to show how the resulting subfragment is translated; then we will gradually add more syntactic rules and translation rules.

6.3.2 Terms, Intransitive Verbs, Sentences

A syntactic rule should provide us with information on three things: (i) the categories of the expressions to which the rule can be applied, (ii) the category to which the new expression which results after application of the rule will belong, and (iii) the syntactic operation that should be applied to get the new expression. Together the syntactic rules supply us with a definition of the expressions of the language. Hence, we may look upon an enumeration of the syntactic rules as a definition which lays down for every category A the expressions that belong to it. In other words, if we denote the set of all expressions belonging to A as P_A , the set of syntactic rules can be considered to give a simultaneous recursive definition of P_A for all $A \in \text{CAT}$.

In this section we shall give rules for the generation of simple sentences such as:

- (8) John walks.
- (9) Every man sleeps.
- (10) The elephant smokes.

(11) A woman strolls.

(12) One unicorn talks.

Our first concern here is the analysis of terms, both proper names and quantified expressions.

The starting point of the entire syntax is the following *basic rule*:

S1: $B_A \subseteq P_A$, for every category A.

This rule states that every lexical element of a category A is an expression of category A: the basic expressions are included in the set of all expressions.

The next rule is a *rule of functional application*. A rule of functional application describes how expressions of a certain functor category, that is, a category of the form A/B, combine with expressions of category B to yield expressions of category A. Rule S2 states how terms and intransitive verb phrases combine to form sentences. The category of terms, T, is a functor category: T is defined as S/IV. The category of the intransitive verb phrases, IV, is a basic category, like the category of sentences S. The rule is as follows:

S2: If $\delta \in P_{IV}$ and $\alpha \in P_T$, then $F_I(\alpha, \delta) \in P_S$ and $F_I(\alpha, \delta) = \alpha \delta'$,
where δ' is the result of replacing the main verb in δ by its third-person singular present form.

For instance, the term *John* and the intransitive verb *walk* are combined by S2 to form the sentence *John walks*. For $John \in P_T$ (by rule S1, because $John \in B_T$) and $walk \in P_{IV}$ (analogously). If we apply F_I to *John* and *walk*, the result is a sentence: $F_I(John, walk) \in P_S$. The function of the syntactic operation F_I is twofold: it inflects the main verb in the IV, and it concatenates the T and the 'inflected' IV. In this example *walk* is the only verb, and therefore $F_I(John, walk) = John walks$. Some IVs contain more than one main verb: *stroll and talk*, *sleep or love Mary*. These IVs are formed by the rules for conjunction and disjunction, which will be dealt with in §6.3.10. Applying F_I to a T and such a coordinated IV results in the inflection of all the main verbs in the IV: $F_I(John, stroll and talk) = John strolls and talks$. By way of contrast, consider *try to talk*, which also contains two noninflected verbs, but of which only one is a main verb: $F_I(John, try to talk) = John tries to talk$.

The fragment does not contain plural terms; therefore F_I need do no more than inflect the main verb. And it need do so only for the third person, since first and second person pronouns do not occur in the fragment. Furthermore, F_I assumes a procedure for recognizing a main verb in an IV. PTQ does not supply such a procedure, though several have been proposed in later work on the PTQ model. Also, the definition of F_I assumes that the third-person singular present form is known for every verb. From the point of view of linguistic organization, it would be preferable for such morphological details to be treated in a separate morphological component instead of in the syntax. But

surely such a component is not essential for the PTQ model. Considering the restricted objective of PTQ, it is clear why Montague didn't bother with that type of detail. Showing how syntax and model-theoretic semantics can be systematically related was his main objective; syntactic and morphological details, however interesting in their own right, were put aside for the time being. In the following sections we will come across similar instances, but we will not discuss them explicitly; we will simply assume that certain procedures are available and not deny that certain functions of the syntactic operations might better be dealt with elsewhere in the grammar.

Rules S1 and S2 provide us with the procedure to form sentences with proper names (or syntactic variables) in subject position. In sentences like (9)–(12), the subject is a quantified term. To produce this kind of sentence, we need rules to form quantified terms of category T out of CNs, common nouns, by syncategorematically introducing the expressions *every*, *the*, *a(n)*, *one*:

S3: If $\zeta \in P_{CN}$, then $F_2(\zeta) \in P_T$ and $F_2(\zeta) = \textit{every } \zeta$.

S4: If $\zeta \in P_{CN}$, then $F_3(\zeta) \in P_T$ and $F_3(\zeta) = \textit{the } \zeta$.

S5: If $\zeta \in P_{CN}$, then $F_4(\zeta) \in P_T$ and $F_4(\zeta) = \textit{a(n)} \zeta$.

S6: If $\zeta \in P_{CN}$, then $F_5(\zeta) \in P_T$ and $F_5(\zeta) = \textit{one } \zeta$.

Rules S3–S6 give four different ways of producing quantified terms. The operations F_2 – F_5 each take a CN as argument and syncategorematically introduce a different determiner, combining it with the CN to yield a T. Note that these rules take elements of P_{CN} as input, not only elements of B_{CN} . So far we have only seen basic expressions of category CN, but CN also contains complex expressions: for instance, noun phrases consisting of an adjective and a common noun, like *large woman*, *pink unicorn*, or a noun with a restrictive relative clause like *man who sleeps*, *woman who seeks Bill*. In §6.3.11 we shall return to this subject.

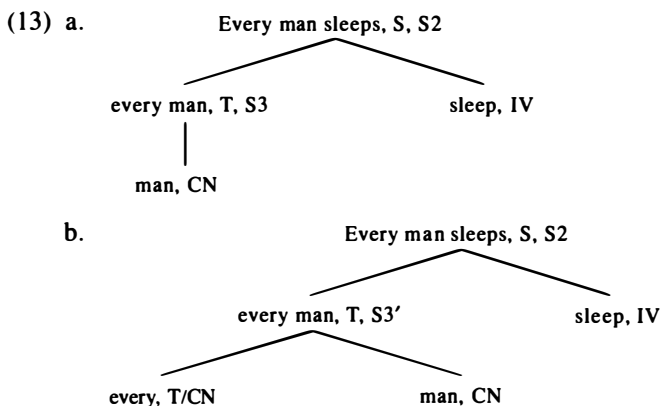
The method employed here, which introduces determiners syncategorematically, as quantifiers are introduced in logical languages, is the one Montague used in PTQ. This means that a separate rule, or separate operation, must be provided for every determiner. We might choose to treat determiners categorically. In that case, we would consider *every*, *the*, *a(n)*, *one* to be elements of category T/CN. One rule of functional application suffices to combine determiners and CNs to Ts:

S3': If $\sigma \in P_{T/CN}$ and $\zeta \in P_{CN}$, then $F'_2(\sigma, \zeta) \in P_T$, and $F'_2(\sigma, \zeta) = \sigma\zeta$.

(Rules and operations that are not part of the fragment but that are introduced to show alternative options are marked with a prime to distinguish them from their 'actual' counterparts.) The term *every man* is formed by applying F'_2 to the T/CN *every* and the CN *man*. The operation F'_2 concatenates them. If we restrict ourselves to simple determiners, there is no compelling reason to prefer one method to the other. But once we take complex determiners into con-

sideration, the situation changes. For instance, we may consider the possessive construction *John's*, as in *John's mother*, to be a determiner. A CN, *mother*, is combined with it to form a term, *John's mother*. For such determiners, syncategorematic introduction will no longer be feasible. Rather, *John's* must be regarded as a complex element of $P_{T/CN}$. And perhaps another reason to prefer the categorematic approach is provided by the fact that there is a (potentially) infinite number of numerals, which act as determiners.

Using the rules S1–S6, or S1, S2, and S3', we can construct the sentences (8)–(12). In Montague grammar the derivation of an expression is represented in an *analysis tree*. Figure (13) shows two analysis trees for sentence (9): analysis tree (a) shows the derivation using S3, and analysis tree (b), the derivation using S3'.



Every node of an analysis tree is labeled with an expression, its category, and the name of the rule used in its formation. (S1 is never mentioned. PTQ gives the number of the syntactic operation instead of the name of the syntactic rule, as we do here.)

Before adding more rules to the syntax, which will allow us to deal with more constructions and expressions, we shall first look at the organization of the process of translation using this very simple fragment. The translation of terms is especially important. Proper names and quantified terms are treated alike by the syntactic rules: they belong to the same syntactic category. As we will see, this has rather far-reaching implications for the translation of the fragment.

6.3.3 The Organization of the Translation Process

In the PTQ-model, English expressions are assigned a meaning via a translation into expressions of a logical language. This language is the language of intensional type theory, which was defined in chapter 5. In various places we

have argued that for an adequate representation of the meanings of natural language expressions, we need at least a language with a type structure and an intensional semantics (cf. §§3.1, 4.2.1, 5.2).

Before proceeding with the specification of the translation process, we first introduce some notational conventions. (cf. also table 5.2).

j, m, b, e are constants of type e .

x, y, z, x_0, \dots, x_n are variables of type e .

X, Y, Z, X_0, \dots, X_n are variables of type $\langle s, \langle e, t \rangle \rangle$.

X, Y, Z, X_0, \dots, X_n are variables of type $\langle s, \langle \langle s, \langle e, t \rangle \rangle, t \rangle \rangle$.

In the translation process, each expression of English is to be associated with a logical expression of a suitable type, i.e., of a type that fits the semantic import of its category. So first of all, we set up a systematic correspondence between the categories of our categorial syntax and the types of the intensional theory of types. Given this correspondence, the translation process should yield for each expression of some category a logical expression of the corresponding type. For lexical expressions the translation can be given in a finite list. For the (infinite number of) derived expressions, the translation process will follow the lead of the syntactic rules. Since, as we observed in §6.2, if we want the semantics of our English fragment to be compositional, the process of translation must be compositional as well. Hence, the translation of derived expressions will be given by specifying for each syntactic rule a translation rule which defines what, given the translations of the input expressions of the syntactic rule, the translation of its output will be.

Let us start by defining the correspondence between categories and types. One of the leading principles behind categorial syntax is that the syntactic category of an expression reflects its semantic function. In Montague grammar this idea is incorporated by means of the correspondence between categories and types. In type theory, the type of an expression matches its semantic function directly: the extension of an expression of type e is an individual, one of type t denotes a truth value, the interpretation of an expression of type $\langle a, b \rangle$ is a function assigning objects of type b to objects of type a , and so on. If we define a correspondence between categories and types, we get, indirectly, a relation between the syntactic categories of English expressions and their semantic functions. Hence, we define a function f mapping categories on types as follows:

Definition 2

f is a function from CAT to T such that:

- (i) $f(S) = t$
- (ii) $f(CN) = f(IV) = \langle e, t \rangle$
- (iii) $f(A/B) = \langle \langle s, f(B) \rangle, f(A) \rangle$

The category *S* of sentences corresponds to the type *t* of formulas. Both the category *IV* of intransitive verb phrases and the category *CN* of common nouns correspond to the type of one-place first-order predicates, $\langle e, t \rangle$. The last example illustrates that not every distinct category corresponds to a distinct type. Although two expressions which belong to the same syntactic category necessarily have the same semantic function, the inverse does not hold. For instance, *man* and *walk* have the same semantic function, but they belong to different categories. That is, the syntactic differences between them are regarded as having no bearing on their semantic function.

Clause (iii) of definition 2 defines the types corresponding to derived categories. Quite generally, a functor category *A/B* corresponds to a function from intensions of objects of type *f*(*B*) to objects of type *f*(*A*). That is, an expression of a functor category semantically operates on the intension of its argument. The reason for setting things up in this way is that some expressions create intensional contexts. (Cf. the elaborate though incomplete list in §3.1). For example, the transitive verb *seek* creates an intensional context, as is evident from the fact that (16) does not follow from (14) and (15). (In earlier examples we used the more common *look for*):

(14) John seeks the supreme commander of the U.S. armed forces.

(15) The president of the United States of America is the supreme commander of the U.S. armed forces.

(16) John seeks the president of the United States of America.

Two terms with the same extension but different intension, *the supreme commander of the U.S. armed forces* and *the president of the United States of America*, cannot be substituted *salva veritate* in the context *John seeks*. . . . In a Montague grammar this is accounted for by stipulating that semantically *seeks* operates on the intension of its object. Note that expressions with the same intension can be substituted for each other in an intensional context, as is shown by the fact that (18) follows from (17):

(17) John seeks Peter's barber.

(18) John seeks Peter's hairdresser.

In §6.3.5 we shall return to the subject of the representation of the intensional nature of some transitive verbs, and in §6.3.7, to the representation of extensional verbs. The fact that in every functor category expressions can be found that create an intensional context is the justification for associating every functor category with a type of the form $\langle \langle s, b \rangle, a \rangle$.

The second step in setting up the translation process consists in specifying the translation of the lexical elements of the fragment. Most elements will be associated with constants of the logical language, but others will be associated with complex logical expressions. To the latter group belong the elements of

B_T , the transitive verb *be*, and the sentence-modifying adverb *necessarily*. All other lexical elements are translated as constants of the appropriate type. We will make sure that no two lexical elements are mapped onto the same constant and will account for synonymy of lexical expressions by other means (see §6.3.6).

To increase the legibility of translations, it is common practice in Montague grammar to use constants which resemble the corresponding lexical elements. Of course they must be distinguishable: this book prints constants in small capital letters. For instance, the IV *walk* is associated with the constant WALK of type $\langle e, t \rangle$. Formally, the association of lexical elements with constants is performed by a function g , meeting the following requirements:

- (i) g is a function from B_A (with the exception of B_T , *be*, and *necessarily*) to $CON^L_{f(A)}$.
- (ii) If $\alpha \neq \beta$, then $g(\alpha) \neq g(\beta)$.

Clause (i) expresses that g assigns constants of type $f(A)$ to the basic expressions of a category A , with the exceptions stated, which get a separate treatment. For instance $g(man) = MAN$. Clause (ii) ensures that no two lexical elements are mapped onto the same constant. The function g is the foundation of the process of translation. The first translation rule, in fact, is formulated as follows:

T1(a): If α is in the domain of g , then α translates into $g(\alpha)$.

Except for the elements of B_T , *be*, and *necessarily*, the lexical elements of a category A are translated by T1 to constants of type $f(A)$. T1 will be extended with the translations of the other lexical elements. Then it functions analogously to its syntactic counterpart S1: it gets the translation process started.

The third step of the translation process consists in a definition of the translation rules corresponding to the syntactic rules. A syntactic rule operates on one or more expressions to produce a new one. The corresponding translation rule defines the translation of the new expression in terms of the translations of the expressions which are its component parts. Examples will be given in the following sections.

6.3.4 The Translation of Terms

This section deals mainly with the translation rules which correspond to the rules S3–S6, the syntactic rules that construct quantified terms out of CNs. The translation of terms is one of the most interesting features of PTQ.

Compositionality of semantics requires compositionality of translation, and consequently terms should get a separate translation. Some terms, like proper names and definite descriptions, may be regarded as referring to individuals: *John* refers to a certain individual John, *the smallest prime* to the number 2.

But quantified terms are different, and this complicates the story. For instance, consider the quantified term *every man*. This term cannot be regarded as referring to an individual, since there is no individual that is every man. In §4.4.3, we showed how a separate translation of quantified terms can be obtained. Consider for example:

(19) Every man sleeps.

Sentence (19) is interpreted as the assertion that the property of sleeping has the property of being true of every man. The quantified term *every man* is considered to be the second-order predicate that is true of a property of individuals if every individual that is a man has that property. Of course in the end this is just the more familiar first-order reading of (19), which reads that for every individual it is true that if it is a man, it sleeps. So semantically *every man* is considered to be a function applicable to a property. Note that this is in keeping with the fact that terms also behave syntactically as functors. Category T is defined as the functor category S/IV . Syntactically, the application of a term to an IV results in a sentence; semantically, the application of a term interpretation to a property, expressed by an IV, results in a truth value (of that sentence). The type which corresponds to the category T reflects this:

$$(20) f(T) = f(S/IV) = \langle \langle s, f(IV) \rangle, f(S) \rangle = \langle \langle s, \langle e, t \rangle \rangle, t \rangle$$

Expressions of this type refer to (characteristic functions of) sets of first-order properties (cf. table 5.2.). The term *every man* is translated into expression (21), which refers to the set of properties which are true of every man:

$$(21) \lambda X \forall x (\text{MAN}(x) \rightarrow \forall X(x))$$

In (21), X is a variable of type $\langle s, \langle e, t \rangle \rangle$ and x is a variable of type e . With respect to a value assignment, the variable X refers to a function from possible worlds to sets of entities, that is, X refers to a property. Applied to a possible world, a property yields the set of entities which have that property in that world. With respect to a value assignment, $\forall X$ refers in a possible world to the set of entities which in that world have the property to which X refers: $\forall X$ is of type $\langle e, t \rangle$ and $\forall X(x)$ is a formula which is true in a world if the individual referred to by x has the property referred to by X , i.e., belongs to the set of entities referred to by $\forall X$ in that world. The constant MAN , the translation of the CN *man*, is of type $f(\text{CN}) = \langle e, t \rangle$. The expression $\text{MAN}(x)$ is true in w iff x belongs to the set of men in w . Formula (22)

$$(22) \forall x (\text{MAN}(x) \rightarrow \forall X(x))$$

asserts, with respect to an assignment, that all individuals that are men have the property referred to by X . Applying λ -abstraction over the variable X results in (21), the translation of *every man*. This expression is of type $\langle \langle s, \langle e, t \rangle \rangle, t \rangle$. It refers in a world w to the set of all those properties such that

every man in w has those properties, which is exactly what we wanted for the translation of the quantified term *every man*. When we apply (21) to an expression of type $\langle s, \langle e, t \rangle \rangle$, a property-denoting expression, we get a formula which is true in w if the property in question is an element of the set of properties to which (21) refers in w . For example, the IV *sleep* translates into the constant SLEEP of type $f(IV) = \langle e, t \rangle$. The expression $\wedge \text{SLEEP}$ refers to the property expressed by SLEEP and is of type $\langle s, \langle e, t \rangle \rangle$. If we apply (21) to $\wedge \text{SLEEP}$, the result is formula (23):

$$(23) \quad \lambda X \forall x (\text{MAN}(x) \rightarrow \vee X(x)) (\wedge \text{SLEEP})$$

This formula is true in w iff sleep is one of the properties every man in w has. By means of λ -conversion (theorem 5 in §5.6), we can shorten (23) to (24), which in turn can be reduced to (25) by the theorem of $\vee \wedge$ -elimination (theorem 2 in §5.5):

$$(24) \quad \forall x (\text{MAN}(x) \rightarrow \vee \wedge \text{sleep}(x))$$

$$(25) \quad \forall x (\text{MAN}(x) \rightarrow \text{SLEEP}(x))$$

And this brings us back to the standard predicate-logical representation of sentence (19).

However, two points should be borne in mind. First, unlike in predicate logic, the translation of (19) is obtained compositionally. Secondly, the transition of (23) to (24), and of (24) to (25), is admissible precisely because (23) is logically equivalent to (24), and (24) to (25). As for being a representation of the meaning of (19), there is no difference whatsoever between (23), (24), and (25): they are all logically equivalent and hence represent the same meaning. We may regard (23), (24), and (25) as three different notations for one and the same semantic object. We convert (23) into (25) only because (25) is a shorter, more common notation.

The compositional translation of quantified terms is illustrated by the translation of the example *every man*. We will now give the rule of translation T3 corresponding to syntactic rule S3. Rule S3 takes an arbitrary CN ζ and yields the term *every* ζ ; T3 defines the translation of *every* ζ in terms of the translation of ζ . From now on we will abbreviate the phrase ' α translates into β ' as ' $\alpha \mapsto \beta$ '.

$$\text{T3: If } \zeta \in P_{\text{CN}} \text{ and } \zeta \mapsto \zeta', \text{ then } F_2(\zeta) \mapsto \lambda X \forall x (\zeta'(x) \rightarrow \vee X(x)).$$

Given that *man* $\mapsto \text{MAN}$ by rule T1(a), *every man* ($= F_2(\text{man})$) translates into (21) by rule T3.

The translation rules corresponding to S4–S6 follow the same pattern:

$$\text{T4: If } \zeta \in P_{\text{CN}} \text{ and } \zeta \mapsto \zeta', \text{ then } F_3 \mapsto \lambda X \exists x (\forall y (\zeta'(y) \leftrightarrow x = y) \wedge \vee X(x)).$$

T5: If $\zeta \in P_{CN}$ and $\zeta \mapsto \zeta'$, then $F_4 \mapsto$

$$\lambda X \exists x (\zeta'(x) \wedge \forall X(x)).$$

T6: If $\zeta \in P_{CN}$ and $\zeta \mapsto \zeta'$, then $F_5 \mapsto$

$$\lambda X \exists x \forall y ((\zeta'(y) \wedge \forall X(y)) \leftrightarrow x = y)$$

By rule T5, the quantified term *a woman* is translated into (26):

$$(26) \lambda X \exists x (\text{WOMAN}(x) \wedge \forall X(x))$$

In a world w , expression (26) refers to a set of properties. A property belongs to this set if there is a woman in w having this property. Different women have different properties. If Mary sleeps in w , then the property of sleeping belongs to the extension of (26) in w , and if Elsie is awake, so does the property of being awake. The set of properties referred to by *a woman* will contain mutually exclusive properties as soon as there is more than one woman.

T4 gives a Russellian analysis of definite descriptions like *the unicorn*, which translates as:

$$(27) \lambda X \exists x (\forall y (\text{UNICORN}(y) \leftrightarrow x = y) \wedge \forall X(x))$$

In a world w , (27) refers to the set of those properties which hold for the unique individual that is a unicorn in w . Existence and uniqueness are asserted in this analysis, not presupposed. (For a discussion of these matters, see §§5.2 and 5.5 in volume 1. The analysis given here is not the only one possible within the framework of Montague grammar, but it is the analysis employed in PTQ.) Given T6, the translation of *one unicorn* refers to the set of properties such that there is precisely one unicorn (not necessarily always the same one) that has that property.

The analysis of quantified terms outlined above affects the analysis of proper names. Proper names belong to the same syntactic category as quantified terms. Syntactically there is hardly any difference between the two. Their distribution, i.e., the position they may take in a sentence, is virtually equal; to regard them as belonging to different syntactic categories would be very unnatural, and moreover it would complicate the syntax enormously. Every rule applicable to both quantified terms and proper names would have to be doubled. Another reason for analyzing proper names in the same way as quantified terms is that on that level it becomes possible to give a straightforward interpretation of proper names as elements in a conjunction, such as in *John and Bill*, *Bill or a woman*, which cannot be obtained if we stick to an analysis which views them as individual denoting expressions (see §6.3.10). Enough reasons, therefore, to regard proper names and quantified terms as belonging to the same category. As a result, proper names will translate into expressions of type $f(T) = \langle \langle s, \langle e, t \rangle \rangle, t \rangle$ and not into constants of type e as in chapter 5. Now this looks more problematic than it actually is. For proper names can also be viewed as referring to sets of properties. Sentence (8), *John*

walks, can be interpreted as asserting that the property of walking belongs to the set of John's properties. Of course this is just another, more elaborate way of saying that John has the property of walking. Just as a proper name is analyzed syntactically as a functor which takes an IV to give an S, it can be viewed semantically as a function which, applied to a property (expressed by the IV), yields a truth value (of the S). The proper name *John* then translates as:

$$(28) \lambda X \vee X(j)$$

where X is a variable of type $\langle s, \langle e, t \rangle \rangle$ and j a constant of type e . The formula $\vee X(j)$ is true in w with respect to an assignment if the individual referred to by j in w belongs to the set of entities referred to by $\vee X$ in w , which is the set of entities that have the property to which X refers under that assignment. The type of (24) is $\langle \langle s, \langle e, t \rangle \rangle, t \rangle$, and its reference in w is the set of properties that are true of the individual j in w .

The translation of the proper name *John* is used in the translation of sentence (8) in the same way as quantified terms were used. The IV *walk* translates as the constant WALK of type $\langle e, t \rangle$, and $\wedge \text{WALK}$ refers to the property expressed by WALK. Applying (28) to $\wedge \text{WALK}$ yields formula (29):

$$(29) \lambda X \vee X(j)(\wedge \text{WALK})$$

It asserts that the property of walking belongs to the set of properties of John. By means of λ -conversion and $\vee \wedge$ -elimination, (29) is reduced to (30):

$$(30) \text{WALK}(j)$$

So in this case, too, we get the standard representation of first-order predicate logic.

This treatment of proper names is made possible by the fact that the set of properties of an individual and the individual itself are uniquely related. Two individuals are identical iff they have the same properties, i.e., iff their sets of properties are identical. Formula (31) is a formalization of this:

$$(31) \forall x \forall y (x = y \leftrightarrow \lambda X \vee X(x) = \lambda X \vee X(y))$$

Principle (31) is a valid principle of the intensional theory of types. Not every set of properties corresponds to an individual, of course. For example, (21) and (26) refer to sets of properties which, except in borderline cases, do not define an individual.

The translation of the basic expressions of category T, the proper names *John*, *Mary*, *Bill*, and the syntactic variables he_0, \dots, he_n is given in the following addition to rule T1:

$$\begin{aligned} \text{T1(b): } & \textit{John} \mapsto \lambda X \vee X(j) \\ & \textit{Mary} \mapsto \lambda X \vee X(m) \\ & \textit{Bill} \mapsto \lambda X \vee X(b) \\ & he_n \mapsto \lambda X \vee X(x_n) \end{aligned}$$

Syntactic variables, like proper names, are translated into expressions referring to sets of properties of individuals: with respect to an assignment g , he_n refers to the set of properties of $g(x_n)$, the individual denoted by x_n . We will come back to syntactic variables in §6.3.8.

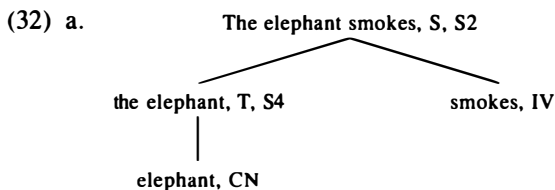
Now we have a uniform definition of the translation of all the terms in our fragment. What remains to be defined in order to complete the translation of the fragment of §6.3.2 is the translation of syntactic rule S2, which combines Ts and IVs to form sentences. In view of the foregoing, the translation rule seems obvious. A sentence of the form $T + IV$ asserts that the property expressed by the IV belongs to the set of properties referred to by the T. An IV δ translates into an expression δ' of type $\langle e, t \rangle$. The expression $\wedge \delta'$ of type $\langle s, \langle e, t \rangle \rangle$ refers to the intension of δ' , a function from possible worlds to sets of individuals, the property expressed by δ' . The desired result is obtained by applying the translation of the T to the intension of the translation of the IV. Translation rule T2, then, reads as follows:

T2: If $\delta \in P_{IV}$ and $\alpha \in P_T$ and $\delta \mapsto \delta'$ and $\alpha \mapsto \alpha'$, then $F_1(\alpha, \delta) \mapsto \alpha'(\wedge \delta')$.

We will see this pattern in every translation rule corresponding to a syntactic rule of functional application: a functor operates on the intension of its argument.

The type corresponding to a functor category A/B is $f(A/B) = \langle \langle s, f(B) \rangle, f(A) \rangle$. The type corresponding to the argument category is $f(B)$. If the category of δ is B , and $\delta \mapsto \delta'$, then $\wedge \delta'$ is of the type $\langle s, f(B) \rangle$ and an appropriate argument for the translation α' of an expression α of category A/B .

We do not translate expressions as such but expressions given a syntactic analysis, since, as we observed in §6.2, only this guarantees compositionality of translation, and hence of interpretation. For there are ambiguous expressions which should get more than one translation. So, the translation process operates on analysis trees, which represent the derivations of expressions. We start at the bottom of the tree, with the translation of the lexical elements, and climb up from node to node, applying at each step the translation rule corresponding to the syntactic rule that labels the node. Using this method, we construct a unique translation which gives the meaning that corresponds to the analysis encoded in the tree. The translation itself can also be represented in a tree structure. Figure (32) shows an analysis tree and the corresponding translation tree for sentence (10) *The elephant smokes*:



$T1(c')$: *every* $\mapsto \lambda Y \lambda X \forall x (\vee Y(x) \rightarrow \vee X(x))$
the $\mapsto \lambda Y \lambda X \exists x (\forall y (\vee Y(y) \leftrightarrow x = y) \wedge \vee X(x))$
a(n) $\mapsto \lambda Y \lambda X \exists x (\vee Y(x) \wedge \vee X(x))$
one $\mapsto \lambda Y \lambda X \exists x \forall y ((\vee Y(y) \wedge \vee X(y)) \leftrightarrow x = y)$

The translation of the rule of functional application S3', combining determiners and CNs to yield terms, is predictable:

$T3'$: If $\sigma \in P_{T/CN}$ and $\zeta \in P_{CN}$ and $\sigma \mapsto \sigma'$ and $\zeta \mapsto \zeta'$, then $F'_2(\sigma, \zeta) \mapsto \sigma'(\wedge \zeta')$.

In the categorematic approach, the translations of complex expressions containing terms are constructed analogously. The reader can check that both approaches yield the same results.

Exercise 1*

- Construct analysis and translation trees for the sentence *A woman strolls*, using the categorematic and syncategorematic definitions of determiners.
- Give a translation for the determiner *no*.

6.3.5 Transitive Verbs

In this section we add a syntactic rule, S7, to our fragment, which combines a transitive verb with a term to form a complex, nonlexical intransitive verb phrase. Such nonlexical IVs combine with terms to form sentences, just as lexical elements of that category do, by rule S2. This enables us to give an analysis of sentences with relational predicates which preserves the traditional subject-predicate analysis. For the new syntactic rule S7 there is a corresponding translation rule which defines the interpretation of complex IVs.

The treatment of transitive verbs will lead us into a second important aspect of the PTQ model: the analysis of intensional verbs and of scope ambiguities. We are concerned with finding a correct, compositional analysis of sentences like:

- (37) John kisses a unicorn.
- (38) John seeks a unicorn.
- (39) Every woman loves one man.

Various aspects of the analysis that PTQ offers will be discussed separately in subsequent sections. The translation of sentences in which the transitive verb *be* appears will not be treated until §6.3.9. There we shall demonstrate how its use both in identity statements such as (40) and in predicative assertions such as (41) can be accounted for in a uniform way.

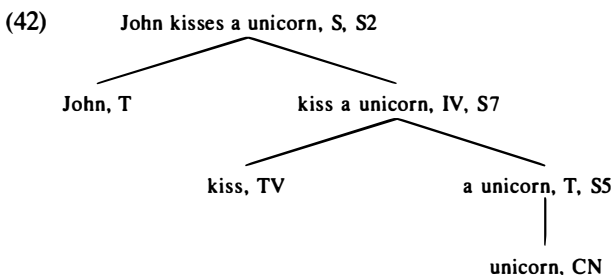
(40) John is Mary.

(41) John is a woman.

The syntactic rule S7 which combines TVs and Ts to IVs is a rule of functional application, because TV is defined as IV/T. Transitive verbs are therefore functors whose input is a term and whose output is an IV:

S7: If $\delta \in P_{TV}$ and $\alpha \in P_T$, then $F_6(\delta, \alpha) \in P_{IV}$, and $F_6(\delta, \alpha) = \delta\alpha'$, where α' is the accusative form of α if α is a syntactic variable; otherwise $\alpha' = \alpha$.

Note that syntactic rule S7 may have a morphological effect. If the term the rule operates on is a syntactic variable, its morphological form has to be adjusted. Figure (42) shows an analysis tree of (37).



This example illustrates that sentences with relational predicates are assigned the traditional subject-predicate structure.

Our fragment contains only basic expressions of category TV. If a three-place verb like *give (to)* were to be added, it would when combined with the indirect object *Mary* yield the complex expression *give to Mary* of category TV.

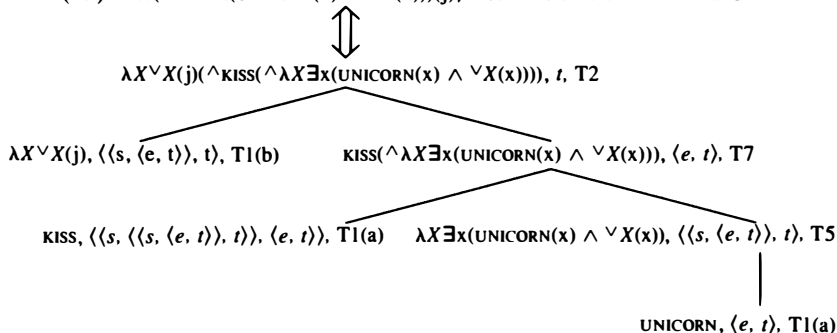
The translation rule corresponding to S7 follows the pattern of other rules of functional application. The type that is associated with the category TV is $f(TV) = f(IV/T) = \langle \langle s, f(T) \rangle, f(IV) \rangle = \langle \langle s, f(S/IV) \rangle, f(IV) \rangle = \langle \langle s, \langle \langle s, \langle e, t \rangle \rangle, t \rangle \rangle, \langle e, t \rangle \rangle$. Semantically a TV is a function which, when applied to a second-order property, yields a set of individuals. So its reference can be viewed as a relation between individuals and second-order properties. The latter are properties of first-order properties, i.e., functions from possible worlds to sets of properties of individuals. The reference of a term is a set of properties of individuals, so its intension is a second-order property. As usual the TV, being the functor, operates semantically on the intension of the T, its argument. There are a number of reasons to prefer this translation of TVs to a translation which treats them as relations between individuals, i.e., as expressions of type $\langle e, \langle e, t \rangle \rangle$. In the first place, it enables us to regard TVs quite generally as functors, taking direct object terms as their argument, whether these terms are proper names or quantified terms. Second, it makes it possible

to account for the intensional nature of verbs such as *seek*. We will go into this subject at greater length. But first we formulate the translation rule:

T7: If $\delta \in P_{TV}$ and $\alpha \in P_T$ and $\alpha \mapsto \alpha'$ and $\delta \mapsto \delta'$, then $F_6(\delta, \alpha) \mapsto \delta'(\wedge \alpha')$.

The effect of T7 is illustrated by the translation trees in figure (43), which display the translation of sentence (37), given the analysis tree in figure (42).

(43) $KISS(\wedge \lambda X \exists x(UNICORN(x) \wedge \vee X(x)))(j)$, λ -conversion and $\vee \wedge$ -elimination



The translation of the TV *kiss* is applied to the intension of the translation of the term *a unicorn* in accordance with rule T7. The result, $KISS(\wedge \lambda X \exists x(UNICORN(x) \wedge \vee X(x)))$ is of type $\langle e, t \rangle$. In a world, the denotation of this expression is a set of individuals, viz., those that kiss a unicorn in that world. The property expressed by *kiss a unicorn* is referred to as $\wedge KISS(\wedge \lambda X \exists x(UNICORN(x) \wedge \vee X(x)))$, which is of type $\langle s, \langle e, t \rangle \rangle$. Formula (37) asserts that this property of individuals is a property of John.

This assertion is expressed by the formula $\lambda X \vee X(j)(\wedge KISS(\wedge \lambda X \exists x(UNICORN(x) \wedge \vee X(x))))$. By means of λ -conversion and $\vee \wedge$ -elimination, this will reduce to $KISS(\wedge \lambda X \exists x(UNICORN(x) \wedge \vee X(x)))(j)$.

PTQ has a notation for two-place relations which is more reminiscent of that of standard predicate logic. We will adopt this convention too:

Notational convention 1

If γ is an expression of type $\langle a, \langle b, t \rangle \rangle$, α an expression of type a , and β an expression of type b , then we may write $\gamma(\beta, \alpha)$ for $(\gamma(\alpha))(\beta)$.

This notational convention, from now on referred to as NC1, tells us that we can treat functions from objects of type a to sets of objects of type b as relations between objects of type b and objects of type a . The translation of (37) can now also be written as follows:

(44) $KISS(j, \wedge \lambda X \exists x(UNICORN(x) \wedge \vee X(x)))$

Exercise 2*

Construct analysis and translation trees for the sentences *John loves Mary* and *Every woman loves one man*. Reduce the result of the translation, using λ -conversion, \vee / \wedge -elimination, and NC1.

Now up to what point are the results of the application of the rules adequate? We will try to answer this question by first considering another example: (45), the translation of sentence (38):

$$(45) \text{ SEEK}(j, \wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$$

This translation is obtained in exactly the same way as (44) is obtained for sentence (37), only a different transitive verb is used. Formula (45) asserts that the seek relation holds between the individual John and the second-order property of being a property of a unicorn. Can this be regarded as an adequate representation of the meaning of (38)? When answering this question, we must keep in mind what our aims are. We are concerned with finding a correct representation of the meaning of the sentences of our fragment. The meaning of a sentence appears, among other ways, in its logical relations to other sentences. We must therefore assign a logical interpretation to a sentence in such a way that the fact that it bears the relation of logical entailment to some sentences and not to others is accounted for. In the case of sentence (38), we may formulate the relevant semantic facts as follows: (46) does not follow from (38) (*John seeks a unicorn.*), (48) does not follow from (38) and (47), and (38) and (49) are not equivalent:

(46) Unicorns exist.

(47) Unicorns and centaurs do not exist.

(48) John seeks a centaur.

(49) John seeks every unicorn.

Obviously sentence (38) has other logical relations with sentences, but if we want to make certain that (45) is a correct representation of an important part of (38), we must be able to account for at least the entailment relations mentioned above. So we want to construe *seek* as a relation between an individual and such a semantic object that the facts just mentioned are accounted for. It is evident that we cannot consider *seek* to be a relation between two individuals. If we did, the meaning of (38) would be that John stands in the seek relation to a certain individual that is a unicorn, but this would imply that (46) follows from (38), which contradicts the facts. In the foregoing, we resolved to analyze terms quite generally as sets of properties of individuals, keeping in mind that it is not necessary for every such set to define a unique individual. For instance, as we observed in §6.3.4, there is no individual with all the properties in the set $\lambda X \exists x(\text{WOMAN}(x) \wedge \vee X(x))$ (unless there is only one woman). And every individual that is a man has more properties than those in the set

$\lambda X \forall x (\text{MAN}(x) \rightarrow \forall X(x))$ (again, unless there is only one man): for instance, take the property of being equal to John, which is a property of John, but not of Bill. An analysis which regards *seek* as a relation between an individual and a set of first-order properties is to be preferred to one which takes it as a relation between two individuals, since the former, but not the latter, will account for the fact that (46) does not follow from (38).

Note that the *seek* relation may hold between John and the set of first-order properties denoted by $\lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))$, even in those situations where there are no such things as unicorns. In those situations, the *seek* relation is true of the individual John and the empty set. If, in a certain world, the set of properties in question is empty because there are no unicorns, this means that in that world John will never find what he is seeking, but it does not imply that John cannot seek what does not actually exist. To put it in other words, (46) does not follow from (38), and the negation of (38) does not follow from (46). These results are quite satisfactory.

But if we look at (45) we see that *seek* is not regarded as a relation between an individual and a set of first-order properties but as a relation between an individual and a function from possible worlds to sets of first-order properties. *Seek* does not operate on the term itself but on its intension. Why is that? The reason is that *seek* does not allow substitution of merely extensionally equivalent expressions. With respect to (38) the reason for this is quite evident. Suppose (47) is true. In such a situation $\lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))$ and $\lambda X \exists x (\text{CENTAUR}(x) \wedge \forall X(x))$ would be extensionally equivalent, since both would denote the empty set. Still, (48) does not follow from (38), not even when (47) is true. But notice that although in this situation the terms *a unicorn* and *a centaur* have the same extension, they have different intensions. There are worlds where unicorns exist but centaurs do not, and vice versa. And in a world where both mythological creatures exist, their sets of properties would be different: a unicorn has one horn, a centaur has a human head. These semantic facts, as well as the fact that (38) and (49) are not equivalent, are neatly accounted for by taking *seek* to be a relation between an individual and a second-order property.

With regard to the semantic facts observed above, the PTQ analysis of transitive verbs and their direct objects is satisfactory. A second-order property, though perhaps not the first thing that comes to mind, is semantically an adequate second argument for the *seek* relation. Surely the analysis as developed so far leaves several aspects of the meaning of *seek* unaccounted for. To put it more precisely, so far we have specified only those aspects that make *seek* an intensional transitive verb, i.e., we have accounted for what all intensional transitive verbs have in common, but not for what differentiates one from the others. We will see in §6.3.1 how one more specific aspect of the meaning of *seek* is explained in the PTQ model.

Our analysis of transitive verbs, then, is satisfactory for intensional transitive verbs, but in fact it treats every transitive verb as if it were intensional.

If we return to translation (44) of sentence (37), we see why this is so. In (44) *kiss* is also taken to be a relation between an individual and a second-order property, implying that (37) does not entail the existence of unicorns. But this is not correct; one simply cannot kiss things that do not exist. So *kiss* does express a relation between individuals. Unlike the *seek* relation, this relation is extensional, and (52) follows from (50) and (51):

(50) John kisses the supreme commander of the U.S. armed forces.

(51) The president of the United States of America is the supreme commander of the U.S. armed forces.

(52) John kisses the president of the United States of America.

Although intensional and extensional verbs have different semantic properties, they belong to the same syntactic category. The analysis of intensional verbs seems to necessitate a similar analysis of extensional verbs. In §6.3.7 we shall see that we can account for the extensional nature of certain transitive verbs by putting further restrictions on their interpretation without losing the advantage gained by the analysis with respect to intensional verbs. Besides, this will provide a solution to the following problem: we want *seek* to express a relation between individuals in a case like:

(53) John seeks Mary.

According to (53) the *seek* relation holds between two individuals, namely, John and Mary, and in this case existential conclusions may be drawn, for (54) follows from (53):

(54) There is someone whom John is seeking.

A second problem is the explanation of scope ambiguities. Exercise 2 illustrates that deriving sentence (39), *Every woman loves one man*, with the syntactic means available up to now only gives the reading in which the scope of *every woman* is wider than the scope of *one man*. The other reading cannot be produced yet, which is unsatisfactory. Something similar holds for the representation of de dicto/de re ambiguity that sentences with intensional verbs give rise to. Until this point we have only been concerned with an adequate representation of the de dicto reading of a sentence such as *John seeks a unicorn*. But quite generally, sentences in which expressions occur which create an intensional context have besides a de dicto reading also a de re reading, as we saw in §3.1. The de re reading of (38) can be paraphrased as follows:

(55) There is something that is a unicorn that John is seeking.

Section 6.3.8 is dedicated to the treatment of these phenomena; it consists in the introduction into our grammar of another method of construction for sentences. But first we treat the problems mentioned before. Meaning postulates play a central part in this treatment, and therefore §6.3.6 will discuss the function of meaning postulates in general.

6.3.6 *The Function of Meaning Postulates*

In the PTQ model the semantic interpretation of the intensional theory of types serves as the indirect interpretation of a fragment of English. The interpretation determines what suitable models are for the intensional theory of types. But not every model that is adequate for an intensional theory of types will ipso facto be an adequate model for English. This is because such a model contains little information about the meanings of expressions of English. The meanings of the logical constants, i.e., the connectives, quantifiers, identity predicate, λ -operator, \wedge - and \vee -operators, and modal operators, are defined exactly, but for all other expressions only the general nature of their meaning is described. For instance, for the logical constant \neg the semantics not only states that its interpretation is a function from truth values to truth values but it further specifies what function this is: the one that maps 0 on 1 and 1 on 0. But the meaning of an expression which is not a logical constant is not fully specified in the same way. For such an expression the semantics only states what kind of interpretation it has, i.e., to what domain its interpretation belongs, but it does not specify which element of this domain this is. For example, the semantics determines that the constant j , of type e , is interpreted as a function from worlds to individuals, $I(j) \in D^W$, but the exact function from W to D is not given. Likewise, of the constants *BACHELOR* and *MARRIED* we learn that their meanings are functions from worlds to sets of individuals, but no more.

If we translate English into the language of the theory of types, we inherit this distinction. The meanings of some expressions are specified exactly. Those expressions are the ones which are translated into logical constants or in expressions in which only logical constants and bound variables occur. Examples in our fragment are the determiners. This is as it should be, for every semantic theory for the English language should contain at least a fixed interpretation of the logical constants. It is the only way to account for valid inferences in English that depend on the meanings of the English counterparts of the logical constants.

On the other hand, we also want our semantic theory to account for those inferences whose validity depends on something more than the interpretation of the logical constants. For example, we want to account for the fact that (57) follows from (56):

(56) John walks rapidly.

(57) John walks.

The validity of this inference may be considered to be due to a semantic property of the adverb *rapidly*. There is a large class of adverbs with the same property, but not every adverb has it: for example, (57) does not follow from (58):

(58) John walks often.

Another example of a valid derivation that depends on the semantic content of expressions which are not logical constants is provided by (59) and (60):

(59) John is a bachelor.

(60) John is not married.

As we pointed out above, the models of the intensional theory of types only give us the information that **BACHELOR** and **MARRIED** (the translations of the English expressions *bachelor* and *married*) are interpreted as first-order properties. But if we are to give an account of the relationship between (59) and (60), those constants should bear a specific relation to each other: in any model, the interpretation function should assign them interpretations such that the extension of **BACHELOR** and the extension of **MARRIED** are disjoint in every world. The semantics of the intensional theory of types allows both models in which this is the case and models in which it is not. For the intensional theory of types as a *logical theory* both kinds of models are admissible. But if we want to use the theory to give a *semantics for English*, clearly the latter should be excluded. We want the semantic theory to admit only those models in which **BACHELOR** and **MARRIED** are interpreted in such a way that they are related as indicated above.

The function of meaning postulates is to restrict the class of all models to a certain subclass. The subclass is to consist of those models in which some kind of semantic relation between (classes of) predicates is valid, certain subclasses of expressions have specific semantic properties, and so on. For the examples just mentioned, we would want only those models which validate the following formulas:

(61) $\forall x \forall X \Box ((\gamma(X))(x) \rightarrow \forall X(x))$, for $\gamma = \text{QUICKLY, SLOWLY, . . .}$
(but not: **OFTEN**, . . .)

(62) $\forall x \Box (\text{BACHELOR}(x) \rightarrow \neg \text{MARRIED}(x))$

Meaning postulate (61) tells us something about the interpretation of a subclass of adverbs; we have encountered it before, in §4.2.3. In every model in which (61) is valid, it holds that (57) follows from (56). Meaning postulate (62) accounts for the fact that (59) entails (60), because in every model in which (62) is valid, the extensions of **BACHELOR** and of **MARRIED** are disjoint in every world.

Meaning postulates are formulas from our logical language. We use them to impose restrictions on the models for the logical language, and hence indirectly for the natural language, by stipulating that we consider only those models in which they are valid. (Sometimes this type of semantic information can be fitted into a model in other ways too, without using a meaning postulate. See §6.3.9 for an example.)

Using meaning postulates we enter the twilight zone between *sentence semantics* and *word semantics*. Meaning postulates are used to capture part of

the lexical meanings of expressions. But there is a limit to the amount of information that we can build into our models by means of meaning postulates, or would want to. For instance, (62) states about the meaning of *bachelor* just that it is related in the way specified to the meaning of *married*. But it certainly does not give the full semantic content of the noun. Whether it is possible, and for what purposes it would be necessary, to represent the entire meaning of a lexical element is a hard question to answer. The most troublesome problem is how to make a distinction between entailments which depend on the meanings of expressions and those which are due to factual circumstances. We want the semantic theory to account for the former, not for the latter (which it probably could not anyway). This problem actually is the same one every lexicographer struggles with: to distinguish between the semantic and the factual information carried by a lexical element. There is no simple solution for this problem.

6.3.7 Meaning Postulates for the Fragment

Before formulating a meaning postulate for transitive verbs which will account for the extensionality of a verb like *kiss*, we will first formulate a meaning postulate that concerns proper names.

It was argued in §3.2 that proper names, as opposed to definite descriptions, are rigid designators. In every possible world they refer to the same individual. We have regarded all individual constants as rigid designators in modal predicate logic, but we did not in the intensional theory of types. The extension of a constant of type e may differ from world to world. Recall that the proper name *John* is translated as $\lambda X \vee X(j)$. In a world w this refers to the set of properties which (in that world w) are true of the individual which is the reference of j in that world. In different worlds, $\lambda X \vee X(j)$ may refer not just to different sets of properties but to different sets of properties of different individuals. Assuming that *John* is a rigid designator, this is not what we want: $\lambda X \vee X(j)$ should refer to (possibly different) sets of properties of one and the same individual, namely, John, in every world. By requiring the reference of the constant j to be the same individual in every world, we capture that the proper name *John* functions as a rigid designator. We do not impose this constraint on every constant of type e but only on those which are used in the translations of proper names. In the fragment, these are the constants j , m , b , and e . So we formulate the following meaning postulate:

MP1 $\exists x \Box (x = \alpha)$, where $\alpha = j, m, b$, or e

For each of the constants j , m , b , and e , MP1 asserts that there is an individual that is identical with the extension of the constant in every possible world, i.e., that it is a rigid designator. And if we allow only those models of the intensional theory of types in which MP1 is valid, we ensure that *John*, *Mary*, *Bill*, and *Elsie* are rigid designators.

An immediate result of MPI is that if two proper names refer to the same individual in some world, they do so in all worlds. In other words, (63) is a valid principle:

(63) $\alpha = \beta$ is equivalent to $\wedge \alpha = \wedge \beta$, where α, β is j, m, b , or e .

An equivalent principle (see the discussion in §3.3.2) is (64):

(64) $\alpha = \beta$ is equivalent to $\Box(\alpha = \beta)$, where α, β is j, m, b , or e .

Another consequence of MPI concerns the relation between expressions referring to individuals and the corresponding expressions referring to sets of properties of individuals. An individual is characterized by the set of its properties. This means that if two expressions α and β refer to the same individual in a world w , the two expressions $\lambda X \vee X(\alpha)$ and $\lambda X \vee X(\beta)$ refer to the same set of properties in that world. In other words, (65) is a valid principle:

(65) $\alpha = \beta$ is equivalent to $\lambda X \vee X(\alpha) = \lambda X \vee X(\beta)$, where α, β are expressions of type e .

(Compare principle (31) discussed in §6.3.4.) Now if in addition α and β are rigid designators, that is, α and β refer to the same individual in each and every world, then $\lambda X \vee X(\alpha)$ and $\lambda X \vee X(\beta)$ refer to the same set of properties in every world. Therefore principle (66) and its equivalent (67) are also valid:

(66) $\alpha = \beta$ is equivalent to $\wedge \lambda X \vee X(\alpha) = \wedge \lambda X \vee X(\beta)$, where α and β are rigid designators.

(67) $\alpha = \beta$ is equivalent to $\Box(\lambda X \vee X(\alpha) = \lambda X \vee X(\beta))$, where α and β are rigid designators.

Not only are the constants j, m, b , and e rigid designators, but variables are too. The extension of a variable in a world does not depend on the world but on the assignment. For nonrigid designators principle (65) is valid, but (66) and (67) are not. For example, if c and d both refer to the same individual in w_1 , but not in w_2 , then $\lambda X \vee X(c)$ and $\lambda X \vee X(d)$ refer to the same set of properties only in w_1 and not in w_2 . And thus $\wedge \lambda X \vee X(c) = \wedge \lambda X \vee X(d)$ does not hold.

So for rigid designators α , and only for rigid designators α , it holds that α and $\wedge \lambda X \vee X(\alpha)$ are entirely equivalent ways of identifying a certain individual. If α is a rigid designator, then any expression in which $\wedge \lambda X \vee X(\alpha)$ occurs can be converted to an equivalent expression in which α occurs, since both α and $\wedge \lambda X \vee X(\alpha)$ are necessarily tied to the same individual.

The second meaning postulate to be discussed here formulates a property of certain transitive verbs. We observed in §6.3.5 that some transitive verbs, such as *seek*, can be regarded as relations between individuals and second-order properties. Such an approach accounts for important semantic facts

about such intensional TVs, like the fact that the sentence *John seeks a unicorn* does not entail the existence of unicorns. On the other hand, this analysis fails for an extensional TV like *kiss*: we want the existence of unicorns to be implied by *John kisses a unicorn*. Hence, we must be able to regard these extensional TVs as relations between individuals. In the meaning postulate MP2, we state that the second-order properties which are related to individuals by the relation expressed by an extensional TV are determined by a relation between individuals:

$$\text{MP2 } \exists S \forall x \forall X \square (\delta(x, X) \leftrightarrow \forall X (\wedge \lambda y \vee S(x, y))), \text{ where } \delta = \text{LOVE, KISS, KNOW, or FIND.}$$

The variable S in MP2 is of type $\langle s, \langle e, \langle e, t \rangle \rangle \rangle$, a two-place first-order relation, and X is a variable of type $\langle s, \langle \langle s, \langle e, t \rangle \rangle, t \rangle \rangle$, a second-order property. MP2 expresses that for each δ for which it is defined, there is a relation S between individuals such that $\delta(x, X)$ is true iff $\lambda y \vee S(x, y)$, or in other words, the property of standing in the relation S to x belongs to the set of properties $\vee X$, viz., $\vee X(\lambda y \vee S(x, y))$. We will see that for every δ for which MP2 is defined, there is exactly one S that fulfills this condition.

First let us introduce some conventions. Whenever we say that the sentence ϕ is *universally valid* ($\models \phi$) we mean that $\llbracket \phi \rrbracket_{\mathbf{M}} = 1$ for every model \mathbf{M} in which the meaning postulates are true. Such models are called 'admissible'. Furthermore, we will say that formulas ϕ and ψ are *equivalent* if $\llbracket \phi \rrbracket_{\mathbf{M},g} = \llbracket \psi \rrbracket_{\mathbf{M},g}$ for every assignment g and admissible model \mathbf{M} ; and we say that ϕ and ψ are *equivalent* by MP2 if $\llbracket \phi \rrbracket_{\mathbf{M},g} = \llbracket \psi \rrbracket_{\mathbf{M},g}$ in all admissible models for every assignment g that fulfills MP2 with respect to S , i.e., which assigns a value to S so that $\llbracket \forall x \forall X \square (\delta(x, X) \leftrightarrow \vee X (\wedge \lambda y \vee S(x, y))) \rrbracket_{\mathbf{M},g} = 1$, where δ is as in MP2.

Let us look at an example. The translation of (68) (= (37)) is (69) (= (44))

(68) John kisses a unicorn.

(69) $\text{KISS}(j, \wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x)))$

By MP2 it now holds that there exists a relation S between individuals such that (69) is equivalent to (70):

(70) $\vee \wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x)) (\wedge \lambda y \vee S(j, y))$

In other words, there is a relation S such that (69) is equivalent to the assertion that it is true of the property of standing in the relation S to j , i.e., the property $\wedge \lambda y \vee S(j, y)$ that there is a unicorn which has that property. Formula (70) is reduced to (71) by λ -conversion and $\vee \wedge$ -elimination, and (71) in its turn is equivalent to (72):

(71) $\exists x (\text{UNICORN}(x) \wedge \vee \wedge \lambda y \vee S(j, y)(x))$

(72) $\exists x (\text{UNICORN}(x) \wedge \vee S(j, x))$

In other words, by MP2 there is a relation S between individuals such that (69) is equivalent to (72), the assertion that there is a unicorn that has the property of standing in the relation S to j .

The following notational convention, NC2, provides for each extensional TV an expression which will be seen to play the role of the relation S .

Notational Convention 2

If δ is an expression of type $\langle\langle s, \langle\langle s, \langle e, t \rangle \rangle, t \rangle \rangle, \langle e, t \rangle \rangle$, then we may write δ_* instead of $\lambda y \lambda x \delta(x, \wedge \lambda X \vee X(y))$.

The expression δ_* refers to the relation between individuals that holds between x and y iff the relation δ holds between x and the intension of the set of all properties of y , that is, $\wedge \lambda X \vee X(y)$. By this notational convention, $\forall x \forall y (\delta_*(x, y) \leftrightarrow \delta(x, \wedge \lambda X \vee X(y)))$ is universally valid, implying that we may write $\delta_*(x, y)$ for the assertion $\delta(x, \wedge \lambda X \vee X(y))$.

Not every assertion of the form $\delta(x, X)$ can be rewritten into an equivalent one with δ_* , since not every expression of type $\langle s, \langle\langle s, \langle e, t \rangle \rangle, t \rangle \rangle$ refers to a second-order property which is related to a specific individual. For instance, $\lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$ does not define a specific individual, because if there is more than one unicorn or if there are no unicorns, then $\lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$ does not refer to a set of properties which is the set of properties of one specific individual. Although NC2 is defined for all expressions of type $\langle\langle s, \langle\langle s, \langle e, t \rangle \rangle, t \rangle \rangle, \langle e, t \rangle \rangle$, the foregoing shows that NC2 is merely a notational convention and that it does not entail that δ_* can always replace S in MP2. The relation SEEK may be true of an individual, for instance j , and a second-order property, for instance, $\wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$, without it being the case that the relation SEEK_* is true of j and a certain individual in that situation. The intended connection between δ and δ_* holds only in those situations where δ is true of an individual x and a second-order property corresponding to a specific individual y , that is, the second-order property that refers in every world to the set of properties of y (in that world).

An instance of such a second-order property is the one referred to by $\wedge \lambda X \vee X(m)$. Thanks to MP1, m is a rigid designator, and therefore $\wedge \lambda X \vee X(m)$ refers to the function that for every possible world gives the set of properties of one and the same individual, Mary.

By NC2 we may write $\delta_*(x, m)$ for $\delta(x, \wedge \lambda X \vee X(m))$, and thus we have a relation between individuals whenever a proper name occurs as the second argument of an intensional verb like *seek*. That is, (74), the translation of (73) (= (53)) is equivalent to (75).

(73) John seeks Mary.

(74) $\text{SEEK}(j, \wedge \lambda X \vee X(m))$

(75) $\text{SEEK}_*(j, m)$

This does not do away with the intensionality of *seek*. The equivalence of (74) to (75) is simply due to the treatment of proper names and the notational convention. Furthermore, we get an account of the fact that (73) entails (76) (= (54)), because (75) entails (77):

(76) There is someone whom John is seeking.

(77) $\exists x \text{ SEEK}_*(j, x)$

For those δ 's for which MP2 postulates their extensionality, every assertion of the form $\delta(x, X)$ is equivalent to an assertion about individuals. For instance, as we saw above, by MP2, $\delta(j, \wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x)))$ is equivalent to the assertion that there is an individual x which is a unicorn and stands in the relation S to j . In the same manner, MP2 assures that $\delta(j, \wedge \lambda X \forall x (\text{UNICORN}(x) \rightarrow \vee X(x)))$ is equivalent to the assertion that for every individual x , if it is a unicorn, it stands in the relation S to j . The extension of S , its existence being guaranteed by MP2, is nothing else than δ_* . We can prove that MP2 and NC2 entail theorem 1:

Theorem 1

$\models \forall x \forall X \Box (\delta(x, X) \leftrightarrow \vee X(\wedge \lambda y \delta_*(x, y)))$, where $\delta = \text{LOVE, KISS, KNOW, or FIND}$.

Proof: First we demonstrate the equivalence of $\delta_*(x, y)$ and $\vee S(x, y)$ by MP2. According to NC2, $\delta_*(x, y)$ is equivalent to $\delta(x, \wedge \lambda X \vee X(y))$. By MP2, this is equivalent to $\vee \wedge \lambda X \vee X(y)(\wedge \lambda y \vee S(x, y))$, which by $\vee \wedge$ -elimination and λ -conversion is equivalent to $\vee \wedge \lambda y \vee S(x, y)(y)$ and, applying $\vee \wedge$ -elimination and λ -conversion once more, to $\vee S(x, y)$. Then, by replacing δ_* by $\vee S$ in MP2 we prove the theorem. \square

Theorem 1 allows us to substitute formulas $\vee X(\wedge \lambda y \delta_*(x, y))$ for all formulas $\delta(x, X)$ if δ is an extensional verb.

Let us take a look again at the translation of (68), *John kisses a unicorn*. We have already reduced (78) (= (44) and (69)), the direct outcome of the translation process, to (79) (= (72)):

(78) $\text{KISS}(j, \wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x)))$

(79) $\exists x (\text{UNICORN}(x) \wedge \vee S(j, x))$

It has been proven that $\vee S$ is equivalent to δ_* , and hence we may reduce (79) to (80) (= (66)):

(80) $\exists x (\text{UNICORN}(x) \wedge \text{KISS}_*(j, x))$

Formula (80) gives us the familiar first-order predicate-logical style representation of (68). It must be stressed that formulas (78), (80), and all the intermediate steps are equivalent by MP2, and that hence all represent the same meaning. The reduction of (78) to (80) is done merely for our convenience.

Note also that we can get the same result without the intermediate step (79), because by theorem 1, (78) is equivalent to (81):

$$(81) \quad \forall^{\wedge} \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)) (\wedge \lambda y \text{ KISS}_*(j, y))$$

Theorem 1 therefore states that the assertion that the relation *KISS* is true of *j* and the second-order property $\wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))$ is equivalent to the assertion that the property of standing in the relation *KISS* to *j* belongs to the set of properties of a unicorn. The reduction of (81) to (80) proceeds via the intermediate steps (82), (83), and (84):

$$(82) \quad \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)) (\wedge \lambda y \text{ KISS}_*(j, y)), \text{ by } \forall^{\wedge}\text{-elimination}$$

$$(83) \quad \exists x (\text{UNICORN}(x) \wedge \forall^{\wedge} \lambda y \text{ KISS}_*(j, y)(x)), \text{ by } \lambda\text{-conversion}$$

$$(84) \quad \exists x (\text{UNICORN}(x) \wedge \lambda y \text{ KISS}_*(j, y)(x)), \text{ by } \forall^{\wedge}\text{-elimination}$$

All this shows that restricting the models of the intensional theory of types by MP2 gives us a more satisfactory semantics for English expressions. Extensional verbs are interpreted as relations between individuals, while at the same time the representation of intensional verbs remains as it was defined in §6.3.5. The mixed method of meaning postulates and notation conventions was taken over from PTQ. Note, however, that it would also be possible to introduce theorem 1 as a meaning postulate instead of MP2. In §6.3.9, in the discussion on the transitive verb *be*, we will explore this other approach a little further.

Exercise 3*

Reduce the translations of the two sentences from exercise 2 using NC2, MP1, and MP2. For each step in the reduction, indicate what validates it.

6.3.8 Scope Ambiguities, *de Re* Readings, and Rules of Quantification

In §6.3.5, we mentioned two problems concerning the analysis of sentences with transitive verb phrases. In §6.3.7 we gave a solution for the first problem, of how to represent the extensional nature of certain TVs while keeping a satisfactory representation of the intensionality of others. The second problem is how to account for scope ambiguities and the representation of *de re* readings of sentences with intensional verbs. To illustrate the scope ambiguity problem, we look at sentence (85) (= 39)):

$$(85) \quad \text{Every woman loves one man.}$$

If we analyze this sentence using the rules we have defined so far, we get as its reduced translation (86):

$$(86) \quad \forall x (\text{WOMAN}(x) \rightarrow \exists y \forall z ((\text{MAN}(z) \wedge \text{LOVE}_*(x, z)) \leftrightarrow y = z))$$

This formula states that for each woman there is precisely one man whom she loves, possibly different men for different women. Now there is also a reading of (85) which may be paraphrased as (87):

(87) There is one man whom every woman loves.

On this reading, it is possible that some women love more than one man, but only one man is loved by every woman. Our semantic theory must give both interpretations of (85) and of similar ambiguous sentences. The ambiguity of (85) is due to the ambiguity of the scope of *every* and *one*.

The distinction between *de dicto* and *de re* readings can also be formulated in terms of relative scope of expressions. For instance, consider (88) (= (38)):

(88) John seeks a unicorn.

This sentence has two readings; one, the *de dicto* reading, results from application of the rules we have so far. The other, the *de re* reading, was paraphrased as (89) (= (55)):

(89) There is something that is a unicorn that John is seeking.

As opposed to the *de dicto* reading, the *de re* reading of (89) entails the existence of unicorns. As the wording of (89) suggests, *a unicorn* has wide scope over the intensional verb *seek* in the *de re* reading, while in the *de dicto* reading *a unicorn* occurs within the scope of *seek*. There are many ambiguities which are based on the relative scope of certain expressions, such as determiners, temporal expressions, modal expressions, intensional verbs, negation and so on. The principle of compositionality requires that every (nonlexical) semantic ambiguity correspond to a derivational ambiguity. Whenever a sentence has more than one meaning, there should be more than one way of constructing it. In the case of scope ambiguities, it seems obvious that the different syntactic constructions are a result of the order in which the scope-bearing elements are introduced. However, the syntactic rules which we have defined so far force a certain order. Applying a TV to a T and applying a T to the resulting IV is the only way to form a sentence like (85) or (88). The functor will always have wider scope than the argument. The subject, therefore, has widest scope, then comes the transitive verb, and the scope of the direct object is narrowest.

In order to represent scope ambiguities, among which we now include the *de dicto/de re* ambiguities just mentioned, we introduce a second method of sentence construction. The syntactic variables, the basic expressions of the form he_n , play a key role in this. They are of category T and therefore may occur in subject position or in object position with a transitive verb. Thus we get sentences like (90), (91), and (92):

(90) He_6 walks.

(91) Every woman loves him_2 .

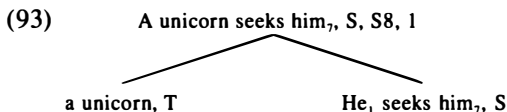
(92) He_1 seeks him_7 .

We will formulate a rule which enables us to form a new sentence out of a term and a sentence containing a syntactic variable by substituting the term for the syntactic variable. This rule is called a *rule of quantification* and it is formulated as follows:

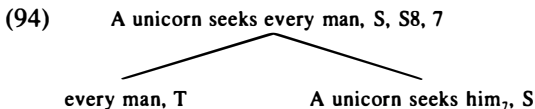
S8, n: If $\alpha \in P_T$ and $\phi \in P_S$, then $F_{7,n}(\alpha, \phi) \in P_S$ and $F_{7,n}(\alpha, \phi) = \phi'$, where ϕ' is the result of the following substitution in ϕ :

- (i) If α is not a syntactic variable he_k , then replace the first occurrence of he_n or him_n with α , and the other occurrences of he_n or him_n with appropriate anaphoric pronouns;
- (ii) if $\alpha = he_k$, then replace every occurrence of he_n with he_k and of him_n with him_k .

Rule S8, n is not one syntactic rule but a *rule schema*. It is an abbreviation for an infinite number of rules. For every number n we have a syntactic rule as an instantiation of the rule schema S8, n . The index n indicates the syntactic variable for which we substitute the term α . For example, if we want the term *a unicorn* substituted for he_1 in (92), we apply S8, 1 as shown in figure (93):



If we next want to substitute the term *every man* for him_7 , then we must use S8, 7:



If the same syntactic variable occurs in a sentence more than once, we replace its first occurrence with the term in question and the other occurrences with appropriate anaphoric pronouns. Thus, quantifying the term *a woman* into (95) results in (96):

(95) He_1 strolls and John loves him_1 .

(96) A woman strolls and John loves her.

(The construction of conjoined sentences like (95) and (96) will be treated in §6.3.10.) The formulation of the rule of quantification, essentially taken from PTQ, has some shortcomings. For instance, it does not provide us with reflexive pronouns, which are required in some cases: if we quantify *John* into (97), we get (98) and not (99) as we should:

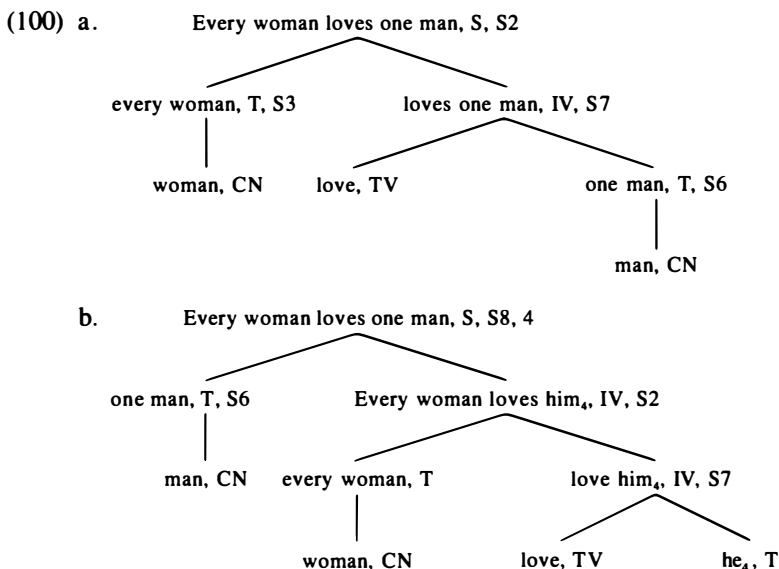
(97) He_1 loves him_1 .

(98) John loves him.

(99) John loves himself.

Another shortcoming is the fact that S8, n permits 'vacuous' quantification. S8, n can be applied to a term and a sentence in which no variable with index n occurs: the syntactic result will be well-formed, but an incorrect meaning will be assigned to it. Such problems have been dealt with in the literature, and we will not dwell upon them. Finally, note that S8, n by clause (ii) permits the substitution of variables for variables; this fact will not bear on the following.

The construction of a sentence by means of a rule of quantification will be called an *indirect (way of) construction*. The direct (way of) construction of sentence (85) and an indirect one are given in figure (100).



Now we must demonstrate that the indirect construction yields the reading of (85) which was lacking up to now, the one paraphrased in (87).

The translation rule corresponding to S8, n reads as follows:

T8, n: If $\alpha \in P_T$ and $\phi \in P_S$ and $\alpha \vdash \alpha'$ and $\phi \vdash \phi'$, then
 $F_{7,n}(\alpha, \phi) \vdash \alpha'(\wedge \lambda x_n \phi')$.

A simple example demonstrates how T8, n works. Suppose we have constructed sentence (101) by substituting *John* for he_3 in (102):

(101) John walks.

(102) he_3 walks.

By rule T1(b), described in §6.3.4, syntactic variables translate into expressions that refer to a set of first-order properties, just like any other term. The variable he_3 translates as $\lambda X \vee X(x_3)$. The translation of (102) is (103), by T2; (103) is reduced to (104) in the usual way.

(103) $\lambda X \vee X(x_3)(\wedge \text{WALK})$

(104) $\text{WALK}(x_3)$

Formula (104), with the free variable x_3 , is converted into a predicate by means of λ -abstraction: $\lambda x_3 \text{WALK}(x_3)$. Its intension:

(105) $\wedge \lambda x_3 \text{WALK}(x_3)$

refers to the property of walking. According to rule T8, 3, quantification of the term *John* into (102) semantically boils down to the assertion that the property which the sentence with the syntactic variable he_3 expresses is a property of John. Formula (105) refers to that property, and formula (106) ascribes it to John. By way of (107), formula (106) is reduced to (108):

(106) $\lambda X \vee X(j)(\wedge \lambda x_3 \text{WALK}(x_3))$

(107) $\vee \wedge \lambda x_3 \text{WALK}(x_3)(j)$

(108) $\text{WALK}(j)$

Formula (108) gives the meaning of (101), which in this case is the same for both the direct and the indirect way of construction.

This simple example illustrates concretely the semantic effect of the application of rule S8, n. Quite generally, the semantics of this process can be described as follows: substitution of a term α for a syntactic variable he_n in a sentence ϕ amounts to the assertion that the property which is expressed by the sentence ϕ with the free variable he_n belongs to the set of properties referred to by α . By abstracting over x_n and applying the \wedge -operator, we transform the translation ϕ' of ϕ into the expression $\wedge \lambda x_n \phi'$, which refers to the property in question. By applying α' , the translation of α , to $\wedge \lambda x_n \phi'$ we get the formula $\alpha'(\wedge \lambda x_n \phi')$, which expresses the intended assertion.

In the case of (101), the indirect construction yields the same result as the direct way. And rightly so, since (101) is not ambiguous. But in other cases, of course, the different construction methods yield different results. The translation of (85) by the indirect method is shown in (109), this time not in a tree structure but in a list:

(109)	1. $he_4 \mapsto \lambda X \vee X(x_4)$	T1b
	2. $love \mapsto \text{LOVE}$	T1a
	3. $F_6(love, he_4) \mapsto \text{LOVE}(\wedge \lambda X \vee X(x_4))$	T7
	4. $woman \mapsto \text{WOMAN}$	T1a

5. $F_2(\text{woman}) \mapsto \lambda X \forall x (\text{WOMAN}(x) \rightarrow \vee X(x))$ T3
6. $F_1(\text{every woman, love him}_4) \mapsto$
 $\lambda X \forall x (\text{WOMAN}(x) \rightarrow \vee X(x))$
 $(\wedge \text{LOVE}(\wedge \lambda X \vee X(x_4)))$ T2
7. $= \forall x (\text{WOMAN}(x) \rightarrow$
 $\vee \wedge \text{LOVE}(\wedge \lambda X \vee X(x_4))(x))$ λ -conv.
8. $= \forall x (\text{WOMAN})(x) \rightarrow \text{LOVE}(\wedge \lambda X \vee X(x_4))(x))$ $\vee \wedge$ -elim.
9. $= \forall x (\text{WOMAN}(x) \rightarrow \text{LOVE}(x, \wedge \lambda X \vee X(x_4)))$ NC1
10. $= \forall x (\text{WOMAN}(x) \rightarrow \text{LOVE}_*(x, x_4))$ NC2
11. $\text{man} \mapsto \text{MAN}$ T1a
12. $F_5(\text{MAN}) \mapsto \lambda X \exists z \forall y ((\text{MAN}(y) \wedge$
 $\vee X(y)) \leftrightarrow z = y)$ T6
13. $F_{7,4}(\text{one man, every woman loves him}_4) \mapsto$
 $\lambda X \exists z \forall y ((\text{MAN}(y) \wedge \vee X(y)) \leftrightarrow z = y)$
 $(\wedge \lambda x_4 \forall x (\text{WOMAN}(x) \rightarrow \text{LOVE}_*(x, x_4)))$ T8, 4
14. $= \exists z \forall y ((\text{MAN}(y) \wedge \vee \wedge \lambda x_4 \forall x (\text{WOMAN}(x) \rightarrow$
 $\text{LOVE}_*(x, x_4))(y)) \leftrightarrow z = y)$ λ -conv.
15. $= \exists z \forall y ((\text{MAN}(y) \wedge \lambda x_4 \forall x (\text{WOMAN}(x) \rightarrow$
 $\text{LOVE}_*(x, x_4))(y)) \leftrightarrow z = y)$ $\vee \wedge$ -elim.
16. $= \exists z \forall y ((\text{MAN}(y) \wedge \forall x (\text{WOMAN}(x) \rightarrow$
 $\text{LOVE}_*(x, y))) \leftrightarrow z = y)$ λ -conv.

A few remarks. Steps 7–10 show that we reduce subexpressions as much as possible during the translation process. It is of course not necessary to do this, but it does make things easier. The reduction of 9 to 10 is possible merely by NC2, because the variable, x_4 , is a rigid designator; MP2 is not needed here yet. We may choose a variable other than the one used in the description of the translation rule, as we did in step 12; again, this is not necessary, but it serves to avoid confusion. The general pattern of the translation of a sentence which is derived by means of a quantification rule is shown here in step 13. From the translation of *every woman loves him*₄, step 10, the expression $\wedge \lambda x_4 \forall x (\text{WOMAN}(x) \rightarrow \text{LOVE}_*(x, x_4))$ is formed, which refers to the property of being loved by every woman. The translation of the result of quantifying *one man* in *every woman loves him*₄, 13, states that this property belongs to the set of properties for which it is true that exactly one man has those properties, which is the same as the assertion that there is exactly one man that is loved by every woman, which is what 16 expresses in the simplest way.

Comparing translation 16 in (109) of the indirect construction of (85) to (86), the translation of the direct construction, we see that both readings of (85) are now adequately represented. The direct and indirect ways of construction are different derivations of one and the same sentence. However, no difference in constituent structure corresponds to it. Were we to represent constituent structure in the grammar, (85) would get the same structure, (110), in both cases:

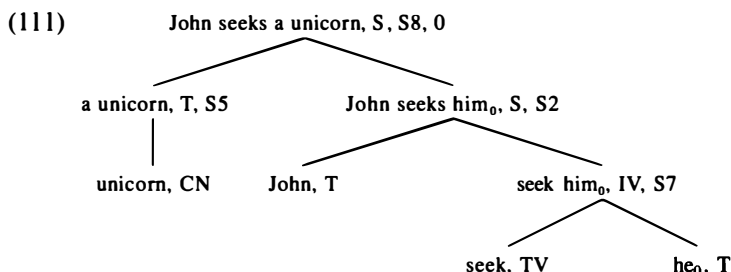
(110) $s[_T[\text{every}_{CN}[\text{woman}]]_{IV}[_{TV}[\text{loves}]]_T[\text{one}_{CN}[\text{man}]]]$

This example illustrates what was stated in §6.2: it is the derivational history of a sentence, and not its constituent structure, which determines its meaning.

The rule of quantification S8, n is an example of a syntactic rule that is primarily semantically motivated. Of course it also performs a certain syntactic function, generating anaphoric pronouns, but the main reason for its introduction is semantic, viz., to account for scope ambiguities. So here we see concretely how the semantic principle of compositionality affects the contents of the syntax.

Note that the quantification process provides us with an infinite number of ways to derive a sentence: we may use another syntactic variable every time. Since these variants do not lead to semantic differences (the resulting formulas, though different, are all equivalent), we will not take them into consideration.

The quantification rule also enables us to derive the de re reading of (88). Figure (111) shows the analysis tree for this reading.



The important steps in the translation of this tree are the following:

- (112)
1. $John\ seeks\ him_0 \mapsto SEEK(j, \wedge \lambda X \vee X(x_0))$
 2. $= SEEK_*(j, x_0)$ NC2
 3. $F_{7,0}(a\ unicorn, John\ seeks\ him_0) \mapsto$
 $\lambda X \exists x (UNICORN(x) \wedge \vee X(x))$
 $(\wedge \lambda x_0 SEEK_*(j, x_0))$ T8, 0
 4. $= \exists x (UNICORN(x) \wedge SEEK_*(j, x))$ λ -conv.,
 $\vee \wedge$ -elim.

The transition from 1 to 2 depends on the fact that x_0 is a rigid designator. The sentence in which it occurs, *John seeks him₀*, can be viewed as expressing a property of individuals, namely, the property of being sought by John. The translation of (88) by the indirect way of construction is 4 in (112), which asserts that there is something which is a unicorn and which is being sought by John, and this assertion expresses the de re reading of (88).

Exercise 4*

- (a) Construct three analysis trees for the sentence *Every man seeks a unicorn* which give rise to three logically distinct translations.

- (b) Show that the direct and the indirect constructions of the sentence *John kisses a unicorn* lead to equivalent results.

Some final remarks. The first one concerns the difference between the de dicto and de re readings of such sentences as (88). We have observed that the de dicto reading of (88) does not entail the existence of unicorns, whereas the de re reading does. The question now is whether its two translations, (113) (= (45)) and (114) express this difference.

$$(113) \text{ SEEK}(j, \wedge \lambda x \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$$

$$(114) \exists x(\text{UNICORN}(x) \wedge \text{SEEK}_*(j, x))$$

Actually, the answer must be no; this is not yet guaranteed. In the intensional theory of types we quantify over *possible* individuals, and thus (114) entails no more than that there is a possible individual that is a unicorn. And by the way, (116), the translation of (115), does not entail the existence of women:

$$(115) \text{ John kisses a woman.}$$

$$(116) \exists x(\text{WOMAN}(x) \wedge \text{KISS}_*(j, x))$$

The introduction of an existence predicate *E* provides us with several ways to cope with this problem. For example, the information that lexical elements relate to actually existing individuals can be added to their translation. *T*_{1a} would then be replaced by clauses like:

$$\text{woman} \mapsto \lambda x(\text{WOMAN}(x) \wedge E(x))$$

Another method is the introduction of a meaning postulate:

$$\forall x \Box(\delta(x) \rightarrow E(x)), \text{ for } \delta = \text{WOMAN, UNICORN, } \dots$$

This postulate imposes restrictions on the interpretation functions of models, like:

$$\text{For all } w: I(\text{UNICORN})(w) \subseteq I(E)(w)$$

Each solution claims, not that unicorns exist, but only that if there is something which is a unicorn, it actually exists. Now (114) does entail the existence of unicorns: if (114) is true, then there is an actually existing individual that is a unicorn. But this conclusion cannot be drawn from (113). The de re/de dicto distinction is thus explicated as a distinction between *existential* and *nonexistential* readings. Sometimes the distinction between (113) and (114) is described as the distinction between the *specific* and *nonspecific* readings of (88). When someone is seeking a unicorn, she might be looking for the unicorn which was presented to her on her birthday only the day before and which has run away. This corresponds to the specific reading of (88). On the other hand, she might be happy with any old unicorn that she can find: this would correspond to the nonspecific reading. Whether we may take (113) and (114) to represent the nonspecific and specific readings of (88), respectively,

is not clear. The problem here is that the intentions and convictions of the seeking person seem also to play a part in establishing the distinction between specific and nonspecific readings. For instance, suppose John believes that there are unicorns, and that one of these unicorns has a golden mane. John wants to find that unicorn. In that case John is looking for a specific unicorn. We cannot represent this situation with (113), since that is supposed to be the nonspecific reading of (88). On the other hand, (114) entails the existence of unicorns, and from the fact that John is looking for a specific unicorn we would not want to conclude that unicorns exist. Also, (114) states that there is an individual which is a unicorn and which John is seeking, but it does not imply that John knows that the individual he is looking for is a unicorn. Suppose that in fact he doesn't (he believes that it is a centaur he is after). In that case, we would not say that John has the intention of finding a specific unicorn. Even worse, (114) does not imply that John has any intention whatsoever of finding that specific individual which (114) says he is looking for. If there is exactly one unicorn and John sets out to find one, (114) is true even if John in fact believes there are many more and would be happy with any one of these. But we could hardly call that a situation in which John seeks a specific unicorn. Therefore, if we want formulas like (113) and (114) to represent the distinction between the specific and nonspecific readings of sentences like (88), preserving the representation of existential and nonexistential readings, it seems that much more is called for and that we must subject verbs like *seek*, which are not only intensional but also intentional, to closer scrutiny.

Our second remark concerns possible alternatives for this method of accounting for scope ambiguities. In considering a sentence like *Every man loves a woman*, one could be led to the idea that representation (117) is sufficient, since the other reading, (118), implies (117) and is thus only a special case of it:

$$(117) \forall x(\text{MAN}(x) \rightarrow \exists y(\text{WOMAN}(y) \wedge \text{LOVE}_*(x, y)))$$

$$(118) \exists y(\text{WOMAN}(y) \wedge \forall x(\text{MAN}(x) \rightarrow \text{LOVE}_*(x, y)))$$

The result of the direct construction of the sentence is (117), and one might conclude that the indirect construction, (118), is superfluous. If it were always true that the indirect method yields a special case of the situation described by the directly constructed sentence, much could be said for representing only the latter. However, this is not the case. First, there are sentences whose direct and indirect readings are logically independent. For instance, consider our example (85), *Every woman loves one man*. Its two readings (86) and (109) are logically independent; neither of them entails the other. The *de re/de dicto* ambiguity of sentences presents a case of logically independent readings too. Second, even if one reading implies the other, it is not always true that the direct construction yields the most comprehensive reading. Consider, for instance, sentences like *It is not the case that every woman loves a man* and *A*

man loves one woman. Here the direct method yields a special case of the reading constructed by the indirect method. An adequate representation of scope ambiguities, it seems, presupposes a syntax that provides several ways to construct a sentence which shows a scope ambiguity.

The quantification rule provides us with a method for representing scope ambiguities. In this section we have looked at the relative scope of quantified terms in subject and object position, and that of intensional verbs and quantified terms. But there are many other sources of scope ambiguity which can be handled in the same way. We shall come across some instances in §6.3.10 and 6.3.11.

6.3.9 The Transitive Verb *Be*

Until now the verb *be* has not been discussed. Recall that we introduced *be* as a basic expression of category TV in §6.3.1. In PTQ, the *be* of identity and the copula are regarded as one and the same transitive verb. This is quite remarkable, because logical tradition has it that it is necessary from a logical point of view to distinguish between *is* in identity assertions and *is* in predicative assertions. Compare, for example, (121), the standard logical translation of (119), and (122), the translation of (120):

(119) John is Mary.

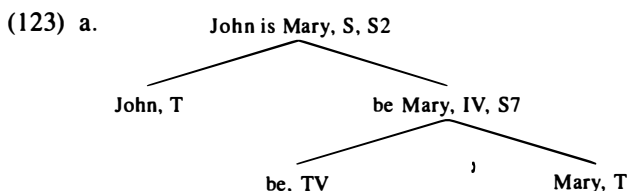
(120) John is a man.

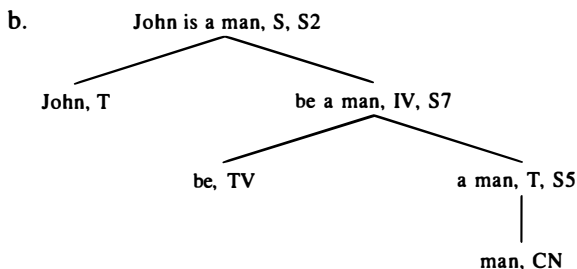
(121) $j = m$

(122) $\text{MAN}(j)$

Is as it appears in (119), turns up as the identity relation in (121), while the copula *is* in (120) seems to have disappeared in translation and instead has merged with the application of the predicate *MAN* to the constant *j*. So it seems that two different verbs need to be distinguished. However, as we shall see, it is possible to regard the occurrences of *is* in (119) and (120) as occurrences of the same verb, with one and the same meaning, and yet come up with representations of the respective sentences which are equivalent, and indeed reducible, to (121) and (122).

Since the verb *be* is regarded as an ordinary transitive verb, the syntactic derivation of both sentences is the same, as figure (123) illustrates.





As was remarked in §6.3.3, the verb *be* gets a special translation. This translation is the following:

$$T1(c): be \mapsto \lambda X \lambda x \lambda y (X(\wedge \lambda y (x = y)))$$

Since *be* is of category TV, its translation is of type $f(TV) = \langle \langle s, \langle \langle s, \langle e, t \rangle \rangle, t \rangle \rangle, \langle e, t \rangle \rangle$. Semantically it expresses a relation between individuals and second-order properties. T1c in fact defines what relation this is, namely, the relation which holds in a world between an individual x and a second-order property X iff the property of being equal to x , i.e., $\wedge \lambda y (x = y)$, belongs to the set of properties which in that world is the reference of X , i.e., iff $\vee X(\wedge \lambda y (x = y))$. At first sight this translation seems to render only the meaning of the *be* of identity, but as we will see, it gives the other meaning as well. First let us look at (124), the translation of (119):

- | | |
|--|----------------------|
| (124) 1. $MARY \mapsto \lambda X \vee X(m)$ | T1b |
| 2. $be \mapsto \lambda X \lambda x \lambda y (X(\wedge \lambda y (x = y)))$ | T1c |
| 3. $F_6(be, Mary) \mapsto \lambda X \lambda x \lambda y (X(\wedge \lambda y (x = y)))$
$(\wedge \lambda X \vee X(m))$ | T7 |
| 4. $= \lambda x \vee \wedge \lambda X \vee X(m) (\wedge \lambda y (x = y))$ | λ -conv. |
| 5. $= \lambda x \lambda X \vee X(m) (\wedge \lambda y (x = y))$ | $\vee \wedge$ -elim. |
| 6. $= \lambda x \vee \wedge \lambda y (x = y)(m)$ | λ -conv. |
| 7. $= \lambda x \lambda y (x = y)(m)$ | $\vee \wedge$ -elim. |
| 8. $= \lambda x (x = m)$ | λ -conv. |
| 9. $John \mapsto \lambda X \vee X(j)$ | T1b |
| 10. $F_1(John, be Mary) \mapsto \lambda X \vee X(j) (\wedge \lambda x (x = m))$ | T2 |
| 11. $= \vee \wedge \lambda x (x = m)(j)$ | λ -conv. |
| 12. $= \lambda x (x = m)(j)$ | $\vee \wedge$ -elim. |
| 13. $= (j = m)$ | λ -conv. |

The IV *be Mary* refers to the set of those entities x for which it is true that the property of being identical to x is a property of Mary, as formula 5 states; as is evident from the equivalent formula in 8, this is the set of entities identical to Mary, which is of course the singleton set with Mary as its sole element. The sentence as a whole asserts that the property of belonging to this set is a property of John. This assertion is equivalent to assertion 13, namely, that John is identical to Mary.

The translation of (120) is constructed in (125):

(125)	1. $man \mapsto MAN$	T1a
	2. $F_4(MAN) \mapsto \lambda X \exists z (MAN(z) \wedge \vee X(z))$	T5
	3. $be \mapsto \lambda X \lambda x \vee X(\wedge \lambda y (x = y))$	T1c
	4. $F_6(be, a\ man) \mapsto \lambda X \lambda x \vee X(\wedge \lambda y (x = y))$ $(\wedge \lambda X \exists z (MAN(z) \wedge \vee X(z)))$	T7
	5. $= \lambda x \vee \wedge \lambda X \exists z (MAN(z) \wedge \vee X(z))$ $(\wedge \lambda y (x = y))$	λ -conv.
	6. $= \lambda x \lambda X \exists z (MAN(z) \wedge \vee X(z))(\wedge \lambda y (x = y))$	$\vee \wedge$ -elim.
	7. $= \lambda x \exists z (MAN(z) \wedge \vee \wedge \lambda y (x = y)(z))$	λ -conv.
	8. $= \lambda x \exists z (MAN(z) \wedge \lambda y (x = y)(z))$	$\vee \wedge$ -elim.
	9. $= \lambda x \exists z (MAN(z) \wedge (x = z))$	λ -conv.
	10. $John \mapsto \lambda X \vee X(j)$	T1b
	11. $F_1(John, be\ a\ man) \mapsto \lambda X \vee X(j)$ $(\wedge \lambda x \exists z (MAN(z) \wedge (x = z)))$	T2
	12. $= \vee \wedge \lambda x \exists z (MAN(z) \wedge (x = z))(j)$	λ -conv.
	13. $= \lambda x \exists z (MAN(z) \wedge (x = z))(j)$	$\vee \wedge$ -elim.
	14. $= \exists z (MAN(z) \wedge (j = z))$	λ -conv.
	15. $= MAN(j)$	first-order logic

As is apparent from step 6, the IV *be a man* refers to the set of those entities x for which it is true that the property of being equal to x belongs to the set of those properties X such that there is man that has X . The equivalent formula in 9 makes it clear that this is the set of entities x such that x is identical with a man. According to 12, (120) asserts that the property of belonging to this set is a property of John; this assertion is reduced to 14, the assertion that there is an individual which is a man and which is identical to John. Of course this is equivalent to 15, the assertion that John is a man. Formulas 14 and 15 are equivalent in standard predicate logic.

In fact, there is a shorter route to this result. The translation of *be a man* is, after all, $\lambda x \exists z (MAN(z) \wedge (x = z))$, and it refers to the set of entities x such that there is a man that is identical to x . Of course this is nothing but the set of all men, i.e., the same set that our constant MAN refers to. So after line 9, we might have continued (125) as follows:

(126)	10. $= MAN$	type theory
	11. $F_1(John, be\ a\ man) \mapsto \lambda X \vee X(j)(\wedge MAN)$	T2
	12. $= MAN(j)$	λ -conv., $\vee \wedge$ -elim.

We conclude that the translation of *be* as defined in T1c gives the right semantics for its use both in predicative statements and in identity assertions.

As the above results illustrate, the verb *be* is treated as an extensional verb. Like any transitive verb, it is taken to express a relation between individuals

and second-order properties. But our translation T1c actually defines this relation. And it turns out that it is an extensional relation.

The same facts could also have been accounted for in a different manner. Suppose we had not given a separate translation for *be*, but that we had translated it into a constant BE like any other TV. In view of its extensionality, MP2 would have been applicable to this constant BE, too. Then (127) and (128) would have resulted as translations of (119) and (120):

$$(127) \text{ BE}_*(j, m)$$

$$(128) \exists x(\text{MAN}(x) \wedge \text{BE}_*(j, x))$$

But notice that these results only account for the extensionality of *be*; they do not represent its full meaning. For instance, according to (127), *John is Mary*, asserts that some extensional relation holds between John and Mary, but it is not expressed what relation this is, viz., the identity relation. In order to account for that we add another meaning postulate:

$$\text{MP3 } \forall x \forall y \square(\text{BE}_*(x, y) \leftrightarrow (x = y))$$

This meaning postulate defines BE_* as the identity relation. And using MP3, (127) and (128) can be reduced to (124) and (125).

This illustrates that giving the special translation T1c for *be* is not the only way to account for its meaning. We can get the same result by means of meaning postulates. We can also combine the effects of MP2 and MP3 in a single meaning postulate for BE:

$$\text{MP4 } \square(\text{BE} = \lambda X \lambda x \forall X(\wedge \lambda y(x = y)))$$

Here we have a single meaning postulate which expresses exactly what was formulated in a translation rule. This procedure can be applied in other cases as well. For instance, we could have translated the determiners into constants, EVERY, THE, and so on, explaining their relation to the logical quantifiers in a meaning postulate. (Of course this assumes that the determiners are introduced categorically.) This ability to switch between translation rules and meaning postulates shows that the purposes of the two are the same: to give a further specification, in some cases a full definition, of the meanings of lexical elements.

In fact, things work out the other way round as well. Instead of using a meaning postulate, we could express the extensionality of TVs by means of special translation rules. For example, instead of translating *kiss* as KISS and relating the latter to KISS_* by MP2, we can express the extensionality of *kiss* directly in its translation:

$$\text{T1: } (a') \text{ kiss} \mapsto \lambda X \lambda x \forall X(\wedge \lambda y(\text{KISS}_*(x, y)))$$

The similarity to the translation rule for *be* is obvious. But note that there is a difference: T1c not only expresses that *be* is extensional, i.e., expresses a relation between individuals, it also states what relation this is: the identity rela-

tion. T1a says that kissing is an extensional relation, but it does not define it any more precisely than that.

Exercise 5

Show how the special translation rule T1b for proper names can be replaced by a meaning postulate.

The next exercise is concerned with logical relationships between natural language sentences. For natural language sentences we define the relation *follow from* as follows: a sentence B follows from sentences A_1, \dots, A_n iff for every syntactic analysis of B and A_1, \dots, A_n , it holds that the translation of B in that analysis follows from the translations of A_1, \dots, A_n in that analysis. (Sometimes also use the notion *follow from on analysis* \dots , is used, which is a relativized version of this notion.)

Exercise 6*

Show that *John seeks Elsie* does not follow from *John seeks the queen* and *Elsie is the queen*.

6.3.10 Conjunction Rules, Disjunction Rules, and Negation Rules

We start with conjunction and disjunction. Consider the following sentences:

- (129) John sleeps and Mary strolls.
- (130) A man smokes and strolls.
- (131) Every man smokes or kisses a woman.
- (132) Every man loves Mary or Elsie.

Sentence (129) is a simple case of sentence conjunction, in contradistinction to (130)–(132), which are not reducible to conjoined sentences. For surely the corresponding sentences (133)–(135) mean something different:

- (133) A man smokes and a man strolls.
- (134) Every man smokes or every man kisses a woman.
- (135) Every man loves Mary or every man loves Elsie.

We shall account for these facts by introducing not only rules for conjunction and disjunction of sentences but also conjunction and disjunction rules for IVs and a disjunction rule for terms. (We will not introduce a conjunction rule for the latter, since this would involve us in an analysis of plurality, which is beyond the scope of this introduction.) The syntactic rules and corresponding translation rules are the following:

S9: If $\phi, \psi \in P_s$, then $F_8(\phi, \psi) \in P_s$ and $F_8(\phi, \psi) = \phi \text{ and } \psi$.

T9: If $\phi, \psi \in P_s$ and $\phi \mapsto \phi'$ and $\psi \mapsto \psi'$, then $F_8(\phi, \psi) \mapsto (\phi' \wedge \psi')$.

S10: If $\phi, \psi \in P_S$, then $F_9(\phi, \psi) \in P_S$ and $F_9(\phi, \psi) = \phi$ or ψ .

T10: If $\phi, \psi \in P_S$ and $\phi \mapsto \phi'$ and $\psi \mapsto \psi'$, then $F_9(\phi, \psi) \mapsto (\phi' \vee \psi')$.

S11: If $\gamma, \delta \in P_{IV}$, then $F_8(\gamma, \delta) \in P_{IV}$.

T11: If $\gamma, \delta \in P_{IV}$ and $\gamma \mapsto \gamma'$ and $\delta \mapsto \delta'$, then $F_8(\gamma, \delta) \mapsto \lambda x(\gamma'(x) \wedge \delta'(x))$.

S12: If $\gamma, \delta \in P_{IV}$, then $F_9(\gamma, \delta) \in P_{IV}$.

T12: If $\gamma, \delta \in P_{IV}$ and $\gamma \mapsto \gamma'$ and $\delta \mapsto \delta'$, then $F_9(\gamma, \delta) \mapsto \lambda x(\gamma'(x) \vee \delta'(x))$.

S13: If $\alpha, \beta \in P_T$, then $F_9(\alpha, \beta) \in P_T$.

T13: If $\alpha, \beta \in P_T$ and $\alpha \mapsto \alpha'$ and $\beta \mapsto \beta'$, then $F_9(\alpha, \beta) \mapsto \lambda X(\alpha'(X) \vee \beta'(X))$.

The syntactic operations F_8 and F_9 introduce *and* and *or* syncategorematically between sentences, between IVs and between terms. The syntactic rules are not in need of further explanation. The effect of translation rules T9 and T10 is also obvious. T11 and T12 make use of the λ -operator to define conjunction and disjunction of predicates in terms of the sentential connectives \wedge and \vee . This process was discussed in §4.4.1. Rule T13 gives the translation of a disjunction of terms similarly. For example, the translation of *Mary or Elsie* proceeds as follows. We apply the translation of *Mary* and the translation of *Elsie* to a variable X of type $\langle s, \langle e, t \rangle \rangle$. (As usual, we take a variable Y in the translation of *Mary* and *Elsie* to avoid confusion, but this is not strictly necessary.) We take the resulting formulas and form their disjunction $\lambda Y \vee Y(m)(X) \vee \lambda Y \vee Y(e)(X)$, which can be reduced to $\vee X(m) \vee \vee X(e)$. This formula is true if X is a property of Mary or a property of Elsie. Abstracting over X yields an expression that in a world w refers to the set of those properties that in w are a property of Mary or a property of Elsie (or of both): $\lambda X(\vee X(m) \vee \vee X(e))$.

In order to illustrate how rules S9–S13 work, (136) gives the most important steps of the translation of (130):

- (136) 1. $F_8(\text{smoke}, \text{stroll}) \mapsto \lambda x(\text{SMOKE}(x) \wedge \text{STROLL}(x))$ T11
2. $F_1(a \text{ man}, \text{smoke and stroll}) \mapsto \lambda x \exists y(\text{MAN}(y) \wedge \vee X(x))(\wedge \lambda x(\text{SMOKE}(x) \wedge \text{STROLL}(y)))$ T2
3. $= \exists y(\text{MAN}(y) \wedge \lambda x(\text{SMOKE}(x) \wedge \text{STROLL}(x))(y))$ λ -conv., $\vee \wedge$ -elim.
4. $= \exists y(\text{MAN}(y) \wedge \text{SMOKE}(y) \wedge \text{STROLL}(y))$ λ -conv.

If we compare (136) with the translation (137) of sentence (133), we see that (133) is indeed assigned a different meaning from (130):

- (137) $\exists x(\text{MAN}(x) \wedge \text{SMOKE}(x)) \wedge \exists x(\text{MAN}(x) \wedge \text{STROLL}(x))$

Sentence (135) is translated as (138):

- (138) $\forall x(\text{MAN}(x) \rightarrow \text{LOVE}(x, m)) \vee \forall x(\text{MAN}(x) \rightarrow \text{LOVE}(x, e))$

Sentence (132) has another meaning, as its step-by-step translation (139) shows:

- (139) 1. $F_9(\text{Mary}, \text{Elsie}) \mapsto \lambda X(\lambda Y \vee Y(m)(X) \wedge \lambda Y \vee Y(e)(X))$ T13
 2. $= \lambda X(\vee X(m) \vee \vee X(e))$ λ -conv.
 3. $F_6(\text{love}, \text{Mary or Elsie}) \mapsto \text{LOVE}(\wedge \lambda X(\vee X(m) \vee \vee X(e)))$ T7
 4. $F_1(\text{every man}, \text{love Mary or Elsie}) \mapsto \lambda Y \forall x(\text{MAN}(x) \rightarrow \vee Y(x)) (\wedge \text{LOVE}(\wedge \lambda X(\vee X(m) \vee \vee X(e))))$ T2
 5. $= \forall x(\text{MAN}(x) \rightarrow \text{LOVE}(x, \wedge \lambda X(\vee X(m) \vee \vee X(e))))$ λ -conv., $\vee \wedge$ -elim., NC1
 6. $= \forall x(\text{MAN}(x) \rightarrow \vee \lambda X(\vee X(m) \vee \vee X(e))(\wedge \lambda y \text{LOVE}_*(x, y)))$ MP2
 7. $= \forall x(\text{MAN}(x) \rightarrow (\text{LOVE}_*(x, m) \vee \text{LOVE}_*(x, e)))$ λ -conv., $\vee \wedge$ -elim.

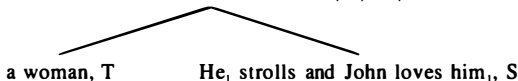
This gives the correct meaning of (132). MP2 is necessary in line 6, because NC2 does not suffice here: $\lambda X(\vee X(m) \vee \vee X(e))$ does not refer to one individual (assuming that m and e refer to different individuals), unlike $\lambda X \vee X(m)$ and $\lambda X \vee X(e)$.

Another thing that becomes apparent now is that if we quantify a term into a conjoined sentence in which the same syntactic variable occurs more than once, we get semantic coreference of the different occurrences. Consider (140) (= (96)):

(140) A woman strolls and John loves her.

In sentence (140), the anaphoric pronoun *her* refers back to *a woman*: it asserts that there is a woman that strolls and that is loved by John. Figure (141) shows the relevant part of the analysis of sentence (140). The relevant steps of the translation are given in (142).

(141) A woman strolls and John loves her, S, S8, 1



- (142) 1. $he_1 \text{ strolls and John loves him}_1 \mapsto \text{STROLL}(x_1) \wedge \text{LOVE}_*(j, x_1)$ T9
 2. $F_{7,1}(\text{a woman}, he_1 \text{ strolls and John loves him}_1) \mapsto \lambda X \exists x(\text{WOMAN}(x) \wedge \vee X(x)) (\wedge \lambda x_1(\text{STROLL}(x_1) \wedge \text{LOVE}_*(j, x_1)))$ T8, 1
 3. $= \exists x(\text{WOMAN}(x) \wedge \text{STROLL}(x) \wedge \text{LOVE}_*(j, x))$ λ -conv., $\vee \wedge$ -elim.

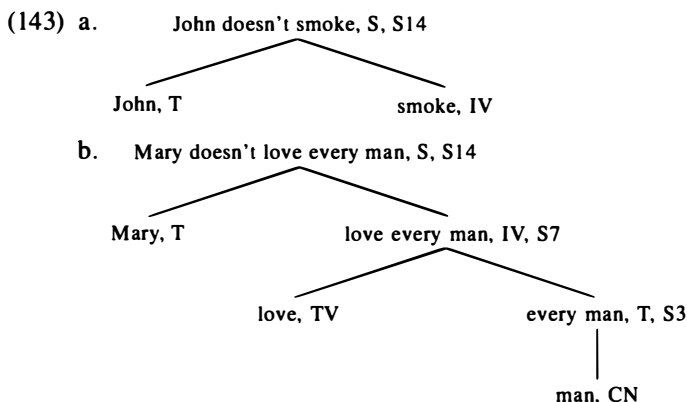
PTQ gives a separate rule to combine terms and IVs to form negated sentences:

S14: If $\alpha \in P_T$ and $\delta \in P_{IV}$, then $F_{10}(\alpha, \delta) \in P_S$ and $F_{10}(\alpha, \delta) = \alpha\delta'$, where δ' is the result of replacing the first verb with its negative third-person singular present form.

The corresponding translation rule reads as follows:

T14: If $\alpha \in P_T$ and $\delta \in P_{IV}$ and $\alpha \mapsto \alpha'$ and $\delta \mapsto \delta'$, then $F_{10}(\alpha, \delta) \mapsto \neg\alpha'(\wedge\delta')$.

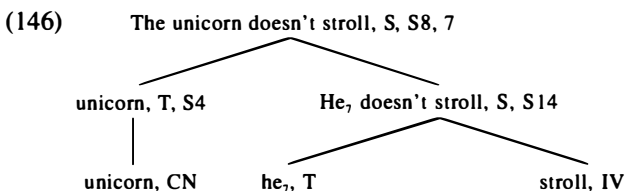
Figure (143) shows the analysis trees of two sentences formed by rule S14. Their translation is given in (144) and (145).



(144) $\neg \text{SMOKE}(j)$

(145) $\neg \forall x(\text{MAN}(x) \rightarrow \text{LOVE}_*(m, x))$

Using this construction, negation always gets widest scope. For sentences (144)–(145) this analysis is correct. Sometimes though, a subject term should have wider scope than the negation. This is achieved by quantifying the term into a sentence formed with S14. For instance, the derivation of the sentence *The unicorn does not stroll* is given in figure (146); the relevant steps of the translation are given in (147).



(147) 1. $F_{10}(\text{he}_7, \text{stroll}) \mapsto \neg \text{STROLL}(x_7)$ T14

2. $F_{7,7}(\text{the unicorn}, \text{he}_7 \text{ does not stroll}) \mapsto$
 $\lambda X \exists x (\forall y (\text{UNICORN}(y) \leftrightarrow x = y) \wedge$
 $\vee X(x)) (\wedge \lambda x_7 \neg \text{STROLL}(x_7))$ T8, 7

$$\begin{array}{ll}
 3. = \exists x(\forall y(\text{UNICORN}(y) \leftrightarrow x = y) \wedge & \\
 \neg \text{STROLL}(x)) & \lambda\text{-conv.}, \\
 & \vee \wedge\text{-elim.}
 \end{array}$$

Note that, given the way it is formulated above, the negation rule functions properly only if it is applied to an IV that contains just one main verb. It is possible, though, to formulate a rule for the more general case. In PTQ the rules for introducing tenses are similar to the rule that deals with negation.

Like their logical counterparts, natural language conjunction, disjunction, and negation (and tenses) are introduced syncategorematically. This is no more necessary here than it was for determiners. In §4.3 we discussed to some extent the categorematic introduction of conjunction and the like. A categorematic analysis along the lines sketched there can surely be implemented in PTQ.

Exercise 7*

Give a derivation tree and a translation for the sentence *John kisses Mary or the queen and loves her*.

6.3.11 Sentential and Infinitival Complements, Adjectives, Relative Clauses and Adverbs

In this section we shall briefly discuss some rules that produce sentences like (148)–(153):

- (148) John asserts that the queen strolls.
- (149) John tries to find a unicorn.
- (150) Elsie is an imaginary pink unicorn.
- (151) Mary loves a man who walks.
- (152) John walks slowly.
- (153) Necessarily, every man is a man.

For the construction of sentences like (148), it suffices to formulate the rule of functional application that defines how an expression of category IV/S combines with one of category S to form an IV:

$$\begin{array}{l}
 \text{S15: If } \delta \in P_{IV/S} \text{ and } \phi \in P_S, \text{ then } F_{11}(\delta, \phi) \in P_{IV} \text{ and } F_{11}(\delta, \phi) \\
 = \delta\phi.
 \end{array}$$

(Note that F_{11} is merely the operation of concatenation.) The translation rule corresponding to S15 follows the pattern of the other rules of functional application:

$$\begin{array}{l}
 \text{T15: If } \delta \in P_{IV/S} \text{ and } \phi \in P_S \text{ and } \delta \mapsto \delta' \text{ and } \phi \mapsto \phi', \text{ then } F_{11}(\delta, \phi) \\
 \mapsto \delta'(\wedge \phi').
 \end{array}$$

The translation of sentence (148) in the direct construction is the following:

$$(154) \text{ ASSERT}(j, \wedge \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \text{STROLL}(x)))$$

Semantically, the functor *assert that* operates on the intension of its argument. The type corresponding to category IV/S is $\langle\langle s, t \rangle \langle e, t \rangle\rangle$. So *assert that* is interpreted as a relation between an individual and a proposition; in (154) these are John and the proposition that the queen strolls. The expression $\wedge \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \text{STROLL}(x))$ refers to the proposition which is true in a world w if the individual that is the unique queen in w strolls in w . This proposition is the intension of the sentence *The queen strolls*. Translation (154) gives the de dicto reading of (148). On this reading, (156) does not follow from (148) and (155):

$$(155) \text{ Elsie is the queen.}$$

$$(156) \text{ John asserts that Elsie strolls.}$$

And indeed (154) and (157), the translations of (148) and (155), respectively, do not entail (158), the translation of (156):

$$(157) \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge e = x)$$

$$(158) \text{ ASSERT}(j, \wedge \text{STROLL}(e))$$

Besides the de dicto reading of (148), there is the de re reading, in paraphrase:

$$(159) \text{ Of the queen it is asserted by John that she strolls.}$$

A representation of the de re reading is obtained by quantifying the term *the queen* into a sentence with a syntactic variable, the sentence *John asserts that he₃ strolls*. This is the result:

$$(160) \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \text{ASSERT}(j, \wedge \text{STROLL}(x)))$$

This reading of (148) entails (158), given the additional premise (157).

These semantic results are satisfactory. A less attractive aspect of the analysis is the fact that *assert that* is considered a syntactic constituent, implying that *assert that Mary comes* is split into *assert that* and *Mary comes*. It would be more natural to analyze the expression as being composed of *assert* and *that Mary comes*. The latter expression occurs in other contexts as a separate, independent constituent: for instance, in *That Mary comes amazes John*. It is possible to analyze *that* as an expression which when combined with a sentence yields an expression which refers to a proposition. ASSERT, in that case, would take such a proposition as an argument. In the end the result will be the same: the representation of the meaning of such sentences as above.

For the construction of sentences with an infinitival complement, like (149), it is sufficient to add a rule of functional application:

$$S16: \text{ If } \gamma \in P_{IV/IV} \text{ and } \delta \in P_{IV}, \text{ then } F_{11}(\gamma, \delta) \in P_{IV}.$$

The translation rule corresponding to S16 is the following:

T16: If $\gamma \in P_{IV/IV}$ and $\delta \in P_{IV}$ and $\gamma \mapsto \gamma'$ and $\delta \mapsto \delta'$, then $F_{11}(\gamma, \delta) \mapsto \gamma'(\wedge \delta')$.

Sentence (149), *John tries to find a unicorn*, has two readings as well. The direct construction results in the following translation:

(163) $\text{TRY}(j, \wedge \text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))))$

Although the extensionality of *find* is not apparent in (161), it is guaranteed by MP2, and so we could leave it at that. We cannot apply theorem 1 at this point, since *FIND* does not have a subject. The expression $\text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$, however, is equivalent to (162), and now it is possible to apply theorem 1, resulting in (163), which is in turn reducible to (164):

(162) $\lambda y \text{FIND}(y, \wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$

(163) $\lambda y(\lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))(\wedge \lambda z \text{FIND}_*(y, z)))$

(164) $\lambda y \exists x(\text{UNICORN}(x) \wedge \text{FIND}_*(y, x))$

Formula (161) as a whole can then be reduced to (165):

(165) $\text{TRY}(j, \wedge \lambda y \exists x(\text{UNICORN}(x) \wedge \text{FIND}_*(y, x)))$

This formula renders the de dicto reading of (149), and it therefore does not entail the existence of unicorns. Again, the de re reading is obtained by means of quantification:

(166) $\exists x(\text{UNICORN}(x) \wedge \text{TRY}(j, \wedge \lambda y \text{FIND}_*(y, x)))$

PTQ introduces a meaning postulate to account for the relation between *seek* and *try to find*:

MP5 $\forall x \forall X \Box(\text{SEEK}(x, X) \leftrightarrow \text{TRY}(x, \wedge \text{FIND}(X)))$

By this meaning postulate, (165) is equivalent to (167)(= (113)), the representation of the de dicto reading of *John seeks a unicorn*, and (166) is equivalent to (168)(= (114)), the representation of the de re reading.

(167) $\text{SEEK}(j, \wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$

(168) $\exists x(\text{UNICORN}(x) \wedge \text{SEEK}_*(j, x))$

Note that PTQ considers *try* a relation between an individual and a property. Sentence (149) means that John stands in the relation of trying to the property of finding a unicorn. This does not assert, as it should, that John will not be happy unless he is the one that has this property. If we want to account for this, we must account for the several so-called control properties of verbs taking infinitival complements. The following examples may clarify this:

(169) John promises Mary to catch a unicorn.

(170) John asks Mary to catch a unicorn.

In (169), *John* is the subject of the verb *catch*; in (170) it is *Mary*. Meaning postulates can account for this type of difference. On the other hand, one could also analyze infinitival complements as expressions referring to propositions instead of properties, just like sentential complements.

Prenominal adjectives are expressions that combine with a noun to form a noun: they belong to category CN/CN. Sentence (150) illustrates how an adjective is applied to a CN which already consists of an adjective and a noun. The rule of functional application of a CN/CN to a CN and the corresponding translation rule are the following:

S17: If $\gamma \in P_{CN/CN}$ and $\zeta \in P_{CN}$, then $F_{11}(\gamma, \zeta) \in P_{CN}$.

T17: If $\gamma \in P_{CN/CN}$ and $\zeta \in P_{CN}$ and $\gamma \mapsto \gamma'$ and $\zeta \mapsto \zeta'$, then $F_{11}(\gamma, \zeta) \mapsto \gamma'(\wedge \zeta')$.

Semantically, adjectives are functions from first-order properties to sets of individuals: $f(CN/CN) = \langle \langle s, \langle e, t \rangle \rangle, \langle e, t \rangle \rangle$. Sentence (150) is translated by T17 as (171):

(171) $IMAGINARY(\wedge PINK(\wedge UNICORN))(e)$

As usual, adjectives, being functors, operate on the intensions of their arguments. The reason for this is that among the adjectives there are intensional ones. *Imaginary* is a case in point; note that (150) does not entail (172):

(172) Elsie is a pink unicorn.

What individuals are imaginary pink unicorns does not depend on what individuals are pink unicorns. So the extension of *pink unicorn* plays no part in establishing the extension of *imaginary pink unicorn*. Rather, it is the property of being a pink unicorn, i.e., the function which assigns to every possible world the set of pink unicorns in that world, which determines what is the set of imaginary pink unicorns in a given world w . To simplify, one could put it like this: something is an imaginary pink unicorn in w if it is a pink unicorn in a world w' that is epistemically accessible from w . Hence an adjective like *imaginary* is intensional; it requires the full intension of its argument and not just its extension. Other examples of intensional adjectives are *possible*, *former*, *future*, *alleged*, *supposed*.

There are extensional adjectives as well, of course; (173) follows from (172).

(173) Elsie is a unicorn.

The extension of *pink unicorn* depends on the extension of *unicorn*. We can express this aspect of the meaning of extensional adjectives in a meaning postulate:

MP6 $\forall X \forall x \Box (\gamma(X))(x) \rightarrow \forall X(x)$, where γ is PINK, GREEN, LARGE, SQUARE.

Further distinctions can be made within the class of extensional adjectives. For example, *pink* and *square* have the following property: whenever something is a pink (square) A, it is not only an A but it is also pink (square). Such adjectives are sometimes called 'intersective', since the set denoted by the combination of such an adjective and a CN can be looked upon as the intersection of the set denoted by the adjective and the set denoted by the CN. So-called relative adjectives are extensional but lack the property of intersectivity. From *Jumbo is a small elephant* we conclude that Jumbo is an elephant but not that he is small. These semantic properties of various classes of adjectives can be accounted for in meaning postulates as well.

Another way to form complex CNs presented in PTQ is to combine common nouns with restrictive relative clauses. The complex CN in sentence (151), *Mary loves a man who walks*, consists of a simple lexical CN, *man*, and a restrictive relative clause, *who walks*. This CN expresses a complex property: to be a man and to walk. PTQ forms such complex CNs by combining a CN with a sentence with a syntactic variable. In this case these are *man* and *he₀ walks*, respectively. The latter sentence expresses a property, since it translates into a formula that contains a free occurrence of a logical variable, viz., $\text{WALK}(x_0)$. The result of abstraction over x_0 is the predicate $\lambda x_0 \text{WALK}(x_0)$. Tying *man* to *he₀ walks* semantically comes down to the conjunction of the predicates MAN and $\lambda x_0 \text{WALK}(x_0)$: $\lambda x (\text{MAN}(x) \wedge \lambda x_0 \text{WALK}(x_0)(x))$, which reduces to $\lambda x (\text{MAN}(x) \wedge \text{WALK}(x))$. A shorter way to obtain this translation is to write directly $\lambda x_0 (\text{MAN}(x_0) \wedge \text{WALK}(x_0))$.

The formation of relative clauses is a rather complicated syntactic process, and we will not go into the formulation of the syntactic operation that it involves. The syntactic rule and the translation rule have the following general form:

S18, n: If $\zeta \in P_{\text{CN}}$ and $\phi \in P_s$, then $F_{12,n}(\zeta, \phi) \in P_{\text{CN}}$.

T18, n: If $\zeta \in P_{\text{CN}}$ and $\phi \in P_s$ and $\zeta \mapsto \zeta'$ and $\phi \mapsto \phi'$, then

$F_{12,n}(\zeta, \phi) \mapsto \lambda x_n (\zeta'(x_n) \wedge \phi')$.

These rules, like the rules of quantification, are rule schemata; for every n there is an actual rule. By way of illustration we give the relevant steps in the translation of (151):

- | | |
|---|------------------------|
| (174) 1. $F_{12,0}(\text{man}, \text{he}_0 \text{ walks}) \mapsto \lambda x_0 (\text{MAN}(x_0)$ | |
| $\wedge \text{WALK}(x_0))$ | T18, 0 |
| 2. $F_4(\text{man who walks}) \mapsto \lambda X \exists x (\lambda x_0 (\text{MAN}(x_0)$ | |
| $\wedge \text{WALK}(x_0))(x) \wedge \forall X(x))$ | T5 |
| 3. $= \lambda X \exists x (\text{MAN}(x) \wedge \text{WALK}(x) \wedge \forall X(x))$ | $\lambda\text{-conv.}$ |

The final translation of (151) is (175):¹

(175) $\exists x (\text{MAN}(x) \wedge \text{WALK}(x) \wedge \text{LOVE}_*(m, x))$

Predicate adverbs like *slowly* in (152) *John walks slowly* are expressions which yield an IV when applied to an IV. It seems obvious to consider them to be of category IV/IV, but we have already reserved this category for expressions like *try to* and *wish to*. Predicate adverbs cannot be taken as belonging to the same category, because their syntactic behavior is different. On the other hand, we do want them to operate on IVs. PTQ solves this dilemma by introducing a new kind of functional category A//B, besides A/B (and if necessary, even A///B, and so on). However, such distinct categories are mapped onto the same semantic type: $f(A//B) = f(A/B) = \langle \langle s, f(B) \rangle, f(A) \rangle$. So we distinguish *slowly* syntactically from *try to* by categorizing it as an IV//IV. But semantically there is no difference: both verbs of category IV/IV and predicate adverbs of category IV//IV are regarded semantically as functions from properties of individuals to sets of individuals.

The rule of functional application and the translation rule introducing predicate adverbs are:

S19: If $\gamma \in P_{IV//IV}$ and $\delta \in P_{IV}$, then $F_{13}(\gamma, \delta) \in P_{IV}$ and $F_{13}(\gamma, \delta) = \delta\gamma$.

T19: If $\gamma \in P_{IV//IV}$ and $\delta \in P_{IV}$ and $\gamma \mapsto \gamma'$ and $\delta \mapsto \delta'$, then $F_{13}(\gamma, \delta) \mapsto \gamma'(\wedge \delta')$.

The syntactic operation F_{13} concatenates the two arguments in the reverse order. The translation of (152) is (176):

(176) SLOWLY(\wedge WALK)(j)

Predicate adverbs translate into expressions of the same type as prenominal adjectives; since $f(IV) = f(CN)$, $f(IV//IV) = f(CN/CN)$. The extensionality of some expressions of this type is stated by MP6. MP6 holds not only for the translations of some prenominal adjectives but also for those of several predicate adverbs. The latter include *slowly*, but not *often*, since sentence (152) entails (178), but (177) does not:

(177) John walks often.

(178) John walks.

In addition to predicate adverbs, we have included a sentence-modifying adverb in the fragment. *Necessarily* in (153), *Necessarily, every man is a man*, is an expression which when applied to a sentence yields another sentence. The rule of functional application and the translation rule are:

S20: If $\gamma \in P_{S/S}$ and $\phi \in P_S$, then $F_{11}(\gamma, \phi) \in P_S$.

T20: If $\gamma \in P_{S/S}$ and $\phi \in P_S$ and $\gamma \mapsto \gamma'$ and $\phi \mapsto \phi'$, then $F_{11}(\gamma, \phi) \mapsto \gamma'(\wedge \phi')$.

The translation rule says that a sentence modifier operates semantically on the proposition expressed by the sentence that is its argument. It is a function, not from truth values to truth values, but from propositions to truth values. Surely

almost every sentence modifier is intensional. The truth value of *necessarily* ϕ in a world w depends not only on the truth value of ϕ in w ; the truth value of ϕ in other worlds plays a role too. The expression $\wedge\phi'$ refers to the intension of ϕ' , the function assigning to every possible world the truth value of ϕ' in that world. So semantically the sentence-modifying adverb *necessarily* is a function from propositions to truth values. Using the logical constant \Box we will define the exact function. *Necessarily* is translated as the constant NECESSARILY, of type $f(S/S) = \langle\langle s, t \rangle, t\rangle$ and the following meaning postulate is added (the variable p is of type $\langle s, t \rangle$):

$$\text{MP7 } \forall p \Box(\text{NECESSARILY}(p) \leftrightarrow \Box \vee p)$$

An alternative method would be to specify the relation between *necessarily* and \Box in the translation rule. This translation, which is the one we find in PTQ, is T1d:

$$\text{T1(d): } \downarrow \textit{necessarily} \mapsto \lambda p \Box \vee p$$

The translation of (153) in the direct way of construction is (179):

$$(179) \Box \forall x(\text{MAN}(x) \rightarrow \text{MAN}(x))$$

Of course the indirect construction of (153) would yield a different result.

We claimed that nearly all sentence modifiers are intensional. Semantically, an extensional sentence modifier is a function from truth values to truth values. There are exactly four such functions (see §4.3.4); one of them is negation. Sentence negation taken as a sentence modifier would be an extensional adverb, which can be defined in terms of \neg by means of a meaning postulate.

PTQ also contains rules for tenses, for prepositions, and for quantification of terms into expressions other than sentences. Tenses are treated by means of rules which are similar to rule S14 for negation. This rule is problematic in several respects, and the objections that can be raised against it hold for PTQ's rules for tenses as well. We will not go into this matter here.

Prepositions are treated as expressions of type $(IV//IV)/T$. They combine with a term to form a (complex) predicate adverb. Among the prepositions one may also find intensional and extensional expressions. Compare (180) and (181):

(180) John strolls in a garden.

(181) John talks about a unicorn.

The rules introducing and translating prepositions follow the familiar pattern of rules of functional application. The extensional nature of certain prepositions is accounted for in a meaning postulate. These are left for the reader as an exercise.

Beside the rule of quantification S8, n described in §6.3.8, which allows us to quantify terms into sentences, PTQ also introduces rules for the quantification of terms into IVs and CNs. The reason for having the last rule, allowing

quantification into CNs, is not clear. There are no known examples for which this rule is essential. The rule of IV quantification, on the other hand, is a necessary addition. For instance, consider sentence (182). Its de dicto reading can only be accounted for by means of such a rule:

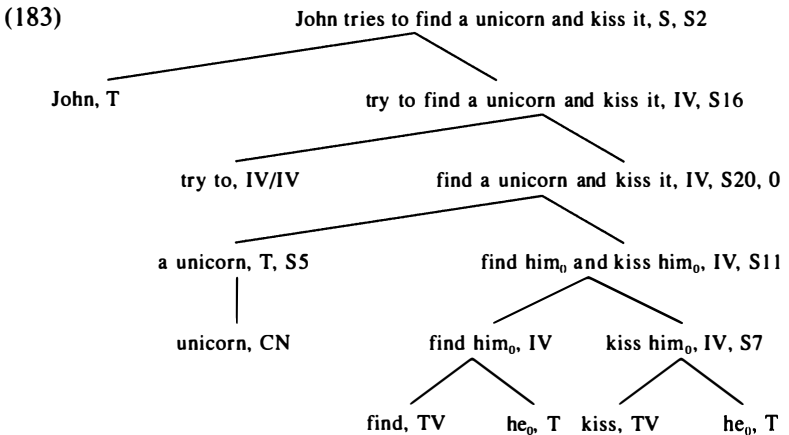
(182) John tries to find a unicorn and kiss it.

Quantification of the term *a unicorn* into the sentence *John tries to find him₀ and kiss him₀* yields the de re reading of (182), whereas the direct construction leaves the coreference of *it* and *a unicorn* unaccounted for. Hence we need a third way of construction. This is supplied by rule S20, n and the corresponding translation rule:

S20, n: If $\alpha \in P_T$ and $\delta \in P_{IV}$, then $F_{7,n}(\alpha, \delta) \in P_{IV}$.

T20, n: If $\alpha \in P_T$ and $\delta \in P_{IV}$ and $\alpha \mapsto \alpha'$ and $\delta \mapsto \delta'$, then $F_{7,n}(\alpha, \delta) \mapsto \lambda y \alpha' (\wedge \lambda x_n (\delta'(y)))$.

These rules are again rule schemata. The syntactic operation $F7, n$ is the same as the one used in rule of S-quantification S8, n. The derivation of (182) is represented in figure (183) and the relevant steps in its translation in (184):



- (184) 1. $F_{7,0}(a \text{ unicorn}, \text{find him}_0 \text{ and kiss him}_0) \mapsto$
 $\lambda y \lambda X \exists x (\text{UNICORN}(x) \wedge$
 $\vee X(x)) (\wedge \lambda x_0 (\lambda z (\text{FIND}_*(z, x_0) \wedge$
 $\text{KISS}_*(z, x_0))(y)))$ T20, 0
2. $= \lambda y \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$
 $(\wedge \lambda x_0 (\text{FIND}_*(y, x_0) \wedge \text{KISS}_*(y, x_0)))$ λ -conv.
3. $= \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x) \wedge$
 $\text{KISS}_*(y, x))$ λ -conv.,
 $\vee \wedge$ -elim.
and λ -conv.
4. $F_{11}(\text{try to}, \text{find a unicorn and kiss him})$
 $\mapsto \text{TRY}(\wedge \lambda y \exists x (\text{UNICORN}(x) \wedge$
 $\text{FIND}_*(y, x) \wedge \text{KISS}_*(y, x)))$ T16

5. $F_1(\text{John, try to find a unicorn and kiss him})$
 $\mapsto \text{TRY}(j, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x)$
 $\wedge \text{KISS}_*(y, x)))$ T2, NCI

This derivation results in the de dicto reading of (182) with the required co-reference of *a unicorn* and *it*.

Exercise 8*

- (a) Give an analysis tree and reduced translation for every reading of the sentence *John asserts that Elsie tries to find a unicorn*.
- (b) Formulate a syntactic rule and a translation rule for prepositions.
- (c) Give two analysis trees with two nonequivalent translations for *John walks in a garden*.
- (d) Formulate a meaning postulate for extensional prepositions such as *in* which accounts for the fact that (c) notwithstanding, *John walks in a garden* is not ambiguous.

6.4 Individual Concepts

6.4.1 Arguments for the Introduction of Individual Concepts

One aspect of the PTQ model that has not been discussed so far is Montague's use of individual concepts. In the fragment treated in §6.3, nouns and intransitive verbs are analyzed as properties of entities. In PTQ, however, CNs and IVs are analyzed as properties of individual concepts. They express not properties of entities but properties of functions from contexts to entities. Montague's argument for this approach is that it provides an explanation for the invalidity of inferences like the following:

- (185) The percentage of Dutchmen opposing nuclear energy is 38.
- (186) The percentage of Dutchmen opposing nuclear energy is rising.
- (187) 38 is rising.
- (188) The population of Amsterdam equals the population of Rotterdam.
- (189) The population of Amsterdam is declining.
- (190) The population of Rotterdam is declining.

These examples require that contexts be interpreted as points in time, or as worlds at points in time. The invalidity of (185)–(187) is indisputable: at any point in time, 38 is equal to 38; 38 can neither rise nor fall. The value of an individual concept can rise or fall. Sentence (185) states that 38 is at this point in time the value of the function that for any point in time gives a number which represents the percentage of Dutchmen opposing nuclear energy at that point in time. Sentence (186) states that this individual concept rises, which is

an assertion about a relationship between its values over a certain period. Rising, falling, and changing are characteristic properties of functions from points in time to numbers. Sentences (185) and (186) make different assertions about the individual concept *the percentage of Dutchmen opposing nuclear energy*: the first makes a statement about the value of the concept at this moment, and the second ascribes a property to the concept.

Nor is (188)–(190) a valid argument. From the fact that at this moment the two individual concepts *the population of Amsterdam* and *the population of Rotterdam* happen to have the same value, it does not follow that if the concept *the population of Amsterdam* declines, then so does the concept *the population of Rotterdam*. Other nouns referring to individual concepts are *price*, *temperature*, *traveling time*, and so on. It should be stressed that although such properties of individual concepts as rising, declining, and changing may be definable in terms of the values of the individual concepts at several points in time, they are not as a consequence properties of these values. Rising, for instance, is not a property of the value of an individual concept at a given point of time but a property of the concept itself.

Not everyone agrees that the introduction of individual concepts is the most appropriate way to account for the invalidity of these and similar inferences. One objection that is often heard is that numbers should not be considered basic entities, and hence, that *temperature*, *number*, and *percentage* are not functions from points in time to entities, i.e., they are not individual concepts. However, the present approach has certain distinct advantages. First, it explains the invalidity of these inferences. Second, if numbers are analyzed not as basic entities but as higher-order entities, a uniform treatment of nouns and verbs no longer seems feasible. Third, there are other sentences and expressions for which an analysis in terms of individual concepts seems particularly appropriate. Consider the following:

(191) The treasurer of the charity organization is the chairwoman of the Entertainment Committee.

(192) The treasurer of the charity organization resigns.

(193) The chairwoman of the Entertainment Committee resigns.

This inference is not valid either. The invalidity is explained by the assumption that (192) and (193) make assertions about individual concepts. Sentence (192) states that the individual who is now the value of the individual concept 'treasurer' will no longer be its value; in (193) the same is asserted of the concept 'chairwoman'. But from the fact that both individual concepts have the same value at this moment, which is expressed by the first premise, (191), and the fact that this individual will no longer be treasurer, which is what (192) states, it does not follow that this individual will no longer be chairwoman, as (193) would have it.

Another example with a 'functional' noun is (194):

- (194) The president is a Republican, but next year he will be a Democrat.

This sentence has two readings. The most improbable one claims that the same individual who is now president will remain in office next year but change his political color in the meantime. The reading on which it is asserted of the individual concept *the president* that its value is now some member of the Republican party but will be a different individual, a member of the Democratic party, next year, is surely more probable.

Not every sentence containing such a functional noun is an assertion about an individual concept. For example, in (195) there is talk of the action of an individual, and hence (196) follows from (191) and (195):

- (195) The treasurer of the charity organization is on the run.

- (196) The chairwoman of the Entertainment Committee is on the run.

Sentences (185), (188), and (191) are assertions about individuals as well.

An analysis using individual concepts seems to enable us to explain these and similar phenomena. The way PTQ accounts for the fact that some CNs and IVs express properties of individual concepts, while others express properties of entities, is interesting but hard to grasp. It is for this reason that the model was not presented in its final form in the preceding sections. In §6.4.2 we will summarize the changes which need be made to the model in order to allow for the introduction of individual concepts and indicate how the resulting model accounts for the invalidity of inferences such as those discussed in this section.

6.4.2 *Consequences of the Introduction of Individual Concepts*

Individual concepts are objects of type $\langle s, e \rangle$. CNs and IVs refer to sets of individual concepts and should be translated into logical expressions of type $\langle \langle s, e \rangle, t \rangle$. This makes it possible to treat CN and IV not as basic categories but as functor categories of type A/B and $A//B$, where $f(A) = t$ and $f(B) = e$, yielding $f(A/B) = f(A//B) = \langle \langle s, f(B) \rangle, f(A) \rangle = \langle \langle s, e \rangle, t \rangle$.

This means that A is S ; for B we choose to create E , a new basic category, and state that $f(E) = e$. This gives rise to the following definitions:

Definition 3

CAT, the set of categories, is the smallest set such that:

- (i) $S, E \in \text{CAT}$.
- (ii) If $A, B \in \text{CAT}$, then $A/B, A//B \in \text{CAT}$.

Definition 4

f is a function from CAT to T such that:

- (i) $f(S) = t, f(E) = e$.
- (ii) $f(A/B) = f(A//B) = \langle \langle s, f(B) \rangle, f(A) \rangle$.

Given these new categorial definitions of CN and IV, the definition of categories defined in terms of them (in our fragment these are all derived categories except S/S) change accordingly. Also, they are assigned another type as a result of the new type assignment of IV and CN. In Table 6.2 we give the new definitions and corresponding types for the most important categories. We also introduce several new lexical elements. Note that there are neither lexical elements nor derived expressions of category E. An expression of category E would refer directly to an entity: $f(E) = e$. The only type of expressions which could be considered to belong to this category are proper names, but as we argued in §6.3.4, there are good reasons for treating these on a par with quantified terms. Note that it is not strictly necessary to introduce this new category E; we could instead retain IV and CN as basic categories and change only the definition of f . Apart from the new definition of the categories and the introduction of additional lexical items, the syntax remains as it was. The changes in the process of translation are minor; the main difference concerns the types of the constants and variables used in the translation rules. Table 6.3 gives a description of three types which will be used regularly in the following sections:

The translation of those basic expressions which are not translated separately is altered only in the sense that the type of the constants related to them by g is modified. For example, MAN and NUMBER are now constants of type

Table 6.2 Categories and Expressions

Category	Definition	Type	New lexical elements
S		t	
E		e	
CN	S/E	$\langle \langle s, e \rangle, t \rangle$	Number, treasurer, chairwoman, temperature, percentage, population
IV	S//E	$\langle \langle s, e \rangle, t \rangle$	Rise, fall, change, resign
T	S/IV = S/(S//E)	$\langle \langle s, \langle \langle s, e \rangle, t \rangle \rangle, t \rangle$	38
TV	IV/T = IV/(S/IV) = (S//E)/ (S/(S//E))	$\langle \langle s, \langle \langle s, \langle \langle s, e \rangle, t \rangle \rangle, t \rangle \rangle, \langle \langle s, e \rangle, t \rangle \rangle$	

Table 6.3 Variables and Interpretations

Type	Variable	Description of extension
$\langle\langle s, e \rangle, t\rangle$	U, V	Set of individual concepts
$\langle s, \langle\langle s, e \rangle, t\rangle\rangle$	U, V	Property of individual concepts
$\langle s, \langle\langle s, \langle\langle s, e \rangle, t\rangle\rangle, t\rangle\rangle$	U, V	Second-order property of individual concepts

$f(\text{CN}) = (\text{S}/\text{E}) = \langle\langle s, e \rangle, t\rangle$. The translations of basic IVs, for example, *STROLL* and *RISE*, are also constants of this type. The translation rule T1 remains the same:

T1(a): If α is in the domain of g , then $\alpha \mapsto g(\alpha)$.

Terms are now translated into logical expressions of type $f(\text{T}) = \langle\langle s, \langle\langle s, e \rangle, t\rangle\rangle, t\rangle$. They refer to sets of properties of individual concepts. In addition to the usual proper names, we also have names of numbers as lexical elements, for instance 38. In the translation of this proper name we use 38 as a constant of type e . The translation of proper names and syntactic variables is now defined as follows:

T1(b): $\text{John} \mapsto \lambda U^\vee U(\wedge j)$
 $\text{Mary} \mapsto \lambda U^\vee U(\wedge m)$
 $\text{Elsie} \mapsto \lambda U^\vee U(\wedge e)$
 $38 \mapsto \lambda U^\vee U(\wedge 38)$
 $\text{he}_n \mapsto \lambda U^\vee U(x_n)$

The translation of *John*, $\lambda U^\vee U(\wedge j)$, refers in a world w to the set of properties of individual concepts that in w are the properties of the individual concept which is the reference of $\wedge j$ in w . The reference of $\wedge j$ in w is the function from worlds to individuals that assigns to every world w' the individual that in w' is the reference of j . As was remarked in §5.4, the reference of $\wedge \alpha$ is the same in every w , so $\wedge j$ refers always to the same function. We also retain MP1; hence j is a rigid designator, and $\wedge j$ refers to a *constant function*, a function that assigns the same value to every argument. So in every world, $\wedge j$ refers to the same constant function, viz., the function which in every world delivers the unique referent of j . An individual and the constant function from worlds to that individual are of course related one to one. Likewise, if α is a rigid designator, then α and $\wedge \alpha$ are uniquely related as well: they identify the same individual. We will extend MP1 to constants like 38, which figure in the translation of names of numbers; these constants are also considered rigid designators.

The translation of a quantified term also results in a logical expression which has as its extension a set of properties of individual concepts. The translation rules T3–T6 are formulated as follows:

T3: If $\zeta \in \text{P}_{\text{CN}}$ and $\zeta \mapsto \zeta'$, then $F_2 \mapsto \lambda U \forall x(\zeta'(x) \rightarrow \vee U(x))$.

- T4: If $\zeta \in P_{CN}$ and $\zeta \mapsto \zeta'$, then $F_3 \mapsto \lambda U \exists x (\forall y (\zeta'(y) \leftrightarrow x = y) \wedge \vee U(x))$.
 T5: If $\zeta \in P_{CN}$ and $\zeta \mapsto \zeta'$, then $F_4 \mapsto \lambda U \exists x (\zeta'(x) \wedge \vee U(x))$.
 T6: If $\zeta \in P_{CN}$ and $\zeta \mapsto \zeta'$, then $F_5 \mapsto \lambda U \exists x \forall y ((\zeta'(y) \wedge \vee U(y)) \leftrightarrow x = y)$.

For example, the term *the temperature* translates into (197) by T4:

$$(197) \lambda U \exists x (\forall y (\text{TEMPERATURE}(y) \leftrightarrow x = y) \wedge \vee U(x))$$

The reference of (197) in a world w is the set of properties of individual concepts of which it is true that they are properties of the unique individual concept that has the property *TEMPERATURE* in w .

The translation of the transitive verb *be* is modified as follows:

$$T1(c): be \mapsto \lambda U \lambda x \vee U (\wedge \lambda y (\vee x = \vee y))$$

As before, the translation accounts for the extensional nature of *be*. It expresses a relation between individual concepts and second-order properties of individual concepts which is defined in terms of the identity of individuals: $\vee x = \vee y$ is true in w with respect to an assignment g iff $g(x)(w) = g(y)(w)$, i.e., iff the value of $g(x)$ in w is identical to the value of $g(y)$ in w . The expression $\wedge \lambda y (\vee x = \vee y)$ refers to the property of individual concepts of having the same value as $g(x)$. So the interpretation of *be* is that relation which holds in a world w between an individual concept and a second-order property iff the property of having the same value as that individual concept is a property which belongs to the set of properties of individual concepts which is the value of that second-order property in w .

The translation of *necessarily*, the last basic expression which is translated separately, stays as it was, since the categorial definition of S/S has remained the same in this new approach.

No changes are necessary in the translation rules which correspond to the syntactic rules of functional application. However, in the translation rules of the quantification rules and the relative clause rule, we need to make a slight modification. In the old rules, we quantified over variables of type e , but now we quantify over variables of type $\langle s, e \rangle$. For instance, the new quantification rule T8, n is formulated as follows:

$$T8, n: \text{ If } \alpha \in P_T \text{ and } \phi \in P_S \text{ and } \alpha \mapsto \alpha' \text{ and } \phi \mapsto \phi', \text{ then } F_{7,n}(\alpha, \phi) \mapsto \alpha' (\wedge \lambda x_n \phi').$$

Similar adjustments have to be made in the rules of conjunction and disjunction.

6.4.3 Some Examples

By way of illustration we give the translations of two examples discussed in §6.4.1. First let us consider (185)–(187). For the sake of convenience, we

will represent the complex CN *percentage of Dutchmen opposing nuclear energy* as PERCENTAGE. The translation of (185)–(187) is:

$$(198) \exists x(\forall y(\text{PERCENTAGE}(y) \leftrightarrow x = y) \wedge \vee x = 38)$$

$$(199) \exists x(\forall y(\text{PERCENTAGE}(y) \leftrightarrow x = y) \wedge \text{RISE}(x))$$

$$(200) \text{RISE}(\wedge 38)$$

The relevant steps in the translation of (198) are:

$(201) \quad 1. F_6(\text{be}, 38) \mapsto \lambda U \lambda x \vee U(\wedge \lambda y(\vee x = \vee y))$ $(\wedge \lambda U \vee U(\wedge 38))$ $2. = \lambda x(\lambda U \vee U(\wedge 38)(\wedge \lambda y(\vee x = \vee y)))$ $3. = \lambda x(\lambda y(\vee x = \vee y)(\wedge 38))$ $4. = \lambda x(\vee x = \vee \wedge 38)$ $5. = \lambda x(\vee x = 38)$ $6. F_1(\text{the percentage}, \text{be } 38) \mapsto$ $\lambda U \exists x(\forall y(\text{PERCENTAGE}(y) \leftrightarrow x = y) \wedge$ $\vee U(x))(\wedge \lambda x(\vee x = 38))$ $7. = (198)$	<p style="text-align: right;">T7</p> <p style="text-align: right;">λ-conv., ∨^∧-elim.</p> <p style="text-align: right;">λ-conv., ∨^∧-elim.</p> <p style="text-align: right;">λ-conv. ∨^∧-elim.</p> <p style="text-align: right;">T2</p> <p style="text-align: right;">λ-conv., ∨^∧-elim.</p>
--	--

The IV *be* expresses the property of individual concepts of having the value 38. Sentence (185) asserts that the unique concept that is the percentage of Dutchmen that are against nuclear energy has that property, in other words, that the value of that concept is 38. It is evident that (200) does not follow from (198) and (199). The property of rising can be true of a concept at a given moment; the value of that concept can be 38 at that moment, but this does not entail that the concept $\wedge 38$ has the property of rising. On the contrary, the concept referred to by $\wedge 38$ is a constant function, and further specification of the property referred to by RISE will certainly require the concepts of which it is true to have different values at different moments. Assertion (200) will never be true then. But even without such specification of what it means for a concept to rise, a counterexample can easily be constructed.

The second example that we will discuss is (191)–(193). The CNs *treasurer of the charity organization* and *chairwoman of the Entertainment Committee* will be represented by the constants TREASURER and CHAIRWOMAN. The sentences (191)–(193) are translated to the following formulas:

$$(202) \exists x(\forall y(\text{TREASURER}(y) \leftrightarrow x = y) \wedge$$

$$\exists z(\forall y(\text{CHAIRWOMAN}(y) \leftrightarrow z = y) \wedge \vee x = \vee z))$$

$$(203) \exists x(\forall y(\text{TREASURER}(y) \leftrightarrow x = y) \wedge \text{RESIGN}(x))$$

$$(204) \exists x(\forall y(\text{CHAIRWOMAN}(y) \leftrightarrow x = y) \wedge \text{RESIGN}(x))$$

The relevant steps of the translation of (202) are:

- (205) 1. $F_6(\text{be}, \text{the treasurer}) \mapsto \lambda U \lambda x \forall U (\wedge \lambda y (\forall x = \forall y)) (\wedge \lambda U \exists z (\forall y (\text{TREASURER}(y) \leftrightarrow z = y) \wedge \forall U(z)))$ T7
2. $= \lambda x \lambda U \exists z (\forall y (\text{TREASURER}(y) \leftrightarrow z = y) \wedge \forall U(z)) (\wedge \lambda y (\forall x = \forall y))$ $\lambda\text{-conv.}, \forall\wedge\text{-elim.}$
3. $= \lambda x \exists z (\forall y (\text{TREASURER}(y) \leftrightarrow z = y) \wedge \lambda y (\forall x = \forall y)(z))$ $\lambda\text{-conv.}, \forall\wedge\text{-elim.}$
4. $= \lambda x \exists z (\forall y (\text{TREASURER}(y) \leftrightarrow z = y) \wedge \forall x = \forall z)$ $\lambda\text{-conv.}$

Formula (202) asserts that the unique concept that is the treasurer and the unique concept that is the chairwoman have the same value: both positions are filled by the same person. From the fact that both concepts have the same value at this moment and the fact that one concept has the property of resigning at this moment, however, it does not follow that the other concept also has that property. Again, a counterexample is easy to construct. Hence the argument is not valid.

The other examples discussed in §6.4.1 are explained in an analogous fashion. We conclude that as far as an explanation of these kinds of phenomena is concerned, the introduction of individual concepts gives quite satisfactory results. But of course we cannot simply leave it at that, since *every* expression is now considered to be a statement about individual concepts, and this certainly is too much of a good thing. Sentences like *John walks* and *Every man loves a woman* are, after all, assertions about individuals. And surely, as we have observed, not even every sentence containing a functional noun like *treasurer* is an assertion about an individual concept. We have accounted for the fact that the argument in (191)–(193) is not valid, but we have not yet shown that, nevertheless, (191) and (195) entail (196). Along with CNs like *percentage* and IVs like *rise*, which express properties of individual concepts, there are CNs like *unicorn* and IVs like *be on the run* that express properties of individuals. The situation we find ourselves in resembles the dilemma we encountered in the analysis of transitive verbs. Because there are intensional TVs, we analyzed all TVs as relations between individuals and second-order properties. To express the extensionality of some TVs, a meaning postulate was introduced. Here we will proceed along the same lines.

6.4.4 Meaning Postulates

PTQ formulates two separate meaning postulates for IVs and CNs. The postulate for IVs is the following:

MP8 $\exists X \forall x \Box (\delta(x) \leftrightarrow \forall X (\forall x))$, where δ is the translation of an IV other than *rise*, *fall*, *change*, *resign*.

This postulate guarantees that certain properties of individual concepts correspond to properties of individuals. For instance, MP8 implies that there is a property of individuals such that in every possible world, WALK is true of an individual concept iff that property is true of the individual that is the value of the individual concept of which WALK is true.

PTQ introduces a notational convention, similar to the one introduced earlier for transitive verbs:

Notational Convention 3

If δ is an expression of type $\langle\langle s, e \rangle, t\rangle$, then we may write δ_* instead of $\lambda x(\delta(\wedge x))$.

For instance, WALK_* is an expression of type $\langle e, t \rangle$ which in every world w refers to the set of entities of which it is true that the constant function of that entity belongs to the set of individual concepts that is the denotation of WALK in w .

Now let us take a look at the translation of the sentence *John walks*:

(206) $\text{WALK}(\wedge j)$

As we remarked above, it follows by MP1, which states that j is a rigid designator, that $\wedge j$ refers to a constant function. The constant function of an individual and the individual itself are uniquely correlated. Every constant function is related to one individual, and for every individual there is one constant function whose value is that individual. In this sense, (206) is an assertion about an individual as much as it is one about an individual concept. NC3 provides a more perspicuous notation. By NC3, (206) is written as:

(207) $\text{WALK}_*(j)$

This is the representation of the sentence *John walks* that we end up with. It is a formula which expresses the assertion that the individual John has the property of individuals WALK_* . (Note that WALK_* is of the same type as WALK in the fragment without individual concepts.) So the introduction of individual concepts does not make any difference as long as we are dealing with constant individual concepts.

For those IVs δ for which it is defined, MP8 guarantees that there exists a corresponding property of individuals. The latter property can be written as δ_* . For under the assumption of MP8, the following theorem is true for the δ for which MP8 is postulated:

Theorem 2

$$\forall x \Box (\delta(x) \leftrightarrow \delta_*(\vee x))$$

We will not give a proof of this theorem. The proof is analogous to the proof of theorem 1 in §6.3.7.

We could have formulated the meaning postulate differently, for example, like MP9 (for the same expressions δ):

MP9 $\forall x \Box(\delta(x) \leftrightarrow \exists x(x = \vee x \wedge \delta(\wedge x)))$

Formulated in this way, the meaning postulate expresses that for any individual concept x (constant or otherwise) in any world w , the assertion that x has the property δ is equivalent to the assertion that a certain constant concept has that property, viz., that concept which is the constant function which in every world w' has as its value the value of x in w . And assertions about constant concepts are in fact assertions about individuals, as was pointed out above.

PTQ gives another meaning postulate for CNs:

MP10 $\forall x \Box(\delta(x) \rightarrow \exists x(x = \wedge x))$, where δ is the translation of a CN other than *number*, *treasurer*, *chairwoman*, *price*, *temperature*, or *percentage*.

This meaning postulate asserts that every individual concept falling under a CN like *man*, *woman*, *unicorn* is a constant function. It is not clear why there are different meaning postulates for IVs and CNs. That there is a difference is established by the fact that in spite of PTQ's claim to the contrary, theorem 2 does not hold for the translations of CNs for which MP10 is defined. On the other hand, the following theorems do hold for these CNs:

Theorem 3

$\exists x(\delta(x) \wedge \vee U(x))$ is equivalent to $\exists x(\delta_*(x) \wedge \vee U(\wedge x))$.

Theorem 4

$\forall x(\delta(x) \rightarrow \vee U(x))$ is equivalent to $\forall x(\delta_*(x) \rightarrow \vee U(\wedge x))$.

Theorem 5

$\exists x(\forall y(\delta(y) \leftrightarrow x = y) \wedge \vee U(x))$ is equivalent to $\exists x(\forall y(\delta_*(y) \leftrightarrow x = y) \wedge \vee U(\wedge x))$.

Theorem 6

$\exists x \forall y((\delta(y) \wedge \vee U(y)) \leftrightarrow x = y)$ is equivalent to $\exists x \forall y((\delta_*(y) \wedge \vee U(\wedge y)) \leftrightarrow x = y)$.

We will not prove here that theorem 2 does not hold for CNs, nor will we prove theorems 3–6.

As in PTQ, the CNs treated in this fragment only occur in contexts where theorems 3–6 hold. Therefore the fact that theorem 2 does not hold for CNs will have no consequences.

Finally, we will give some instances of assertions about individual concepts which by theorems 2–6 are equivalent to assertions about individuals:

- (208) A unicorn strolls $\mapsto \exists x(\text{UNICORN}(x) \wedge \text{STROLL}(x))$
 $= \exists x(\text{UNICORN}(x) \wedge \text{STROLL}_*(\vee x))$, by theorem 2
 $= \exists x(\text{UNICORN}_*(x) \wedge \text{STROLL}_*(\vee \wedge x))$, by theorem 3
 $= \exists x(\text{UNICORN}_*(x) \wedge \text{STROLL}_*(x))$, by $\vee \wedge$ -elimination.

Another way of obtaining this result:

$$\begin{aligned}
 (209) \quad & \text{A unicorn strolls} \mapsto \exists x(\text{UNICORN}(x) \wedge \text{STROLL}(x)) \\
 & = \exists x(\text{UNICORN}_*(x) \wedge \text{STROLL}(\wedge x)), \text{ by theorem 3} \\
 & = \exists x(\text{UNICORN}_*(x) \wedge \text{STROLL}_*(x)), \text{ by NC3.}
 \end{aligned}$$

A second example is the (slightly simplified) representation of (195):

$$\begin{aligned}
 (210) \quad & \text{The treasurer of the charity organization is on the run} \mapsto \\
 & \exists x(\forall y(\text{TREASURER}(y) \leftrightarrow x = y) \wedge \text{ON THE RUN}(x)) = \\
 & \exists x(\forall y(\text{TREASURER}(y) \leftrightarrow x = y) \wedge \text{ON THE RUN}_*(\vee x)).
 \end{aligned}$$

This formula cannot be reduced any further, since MP10 is not defined for the CN *treasurer*: Formula (210) says that the individual that is the value of the concept *the treasurer* has the property of being on the run; sentence (196) gets a similar translation. It follows from (202), the translation of (191), and (210) that the chairwoman of the Entertainment Committee has bolted. Formula (202) asserts that the value of the concept treasurer and the concept chairwoman is the same individual. Formula (210) asserts that that individual has the property of being on the run. Thus, it is true of the individual who is the value of the concept chairwoman that it has this property.

A last remark concerns the reformulation of meaning postulate MP2, which accounts for the extensionality of certain TVs, and of NC2. The contents of MP2 are hardly changed: it now relates relations between individual concepts and second-order properties of individual concepts to relations between individuals. MP2, NC2, and theorem 1 are reformulated as follows:

$$\text{MP2 } \exists S \forall x \forall U \square(\delta(x, U) \leftrightarrow \vee U(\wedge \lambda y(\vee S(\vee x, \vee y)))), \text{ where } \delta \text{ is as formerly defined.}$$

NC2

If δ is an expression of type $\langle\langle s, \langle\langle s, \langle\langle s, e \rangle, t \rangle \rangle, t \rangle \rangle, \langle\langle s, e \rangle, t \rangle \rangle$, then we may write δ_* instead of $\lambda y \lambda x \delta(\wedge x, \wedge \lambda U \vee U(\wedge y))$.

Theorem 1

$$\forall x \forall U \square(\delta(x, U) \leftrightarrow \vee U(\wedge \lambda y(\vee \delta_*(\vee x, \vee y)))), \text{ where } \delta \text{ is as formerly defined.}$$

By means of the reformulated theorem 1, we can reduce (211) to (212), which is in turn reducible to (213) by theorem 3:

$$(211) \quad \text{KISS}(\wedge j, \wedge \lambda U \exists x(\text{UNICORN}(x) \wedge \vee U(x)))$$

$$(212) \quad \exists x(\text{UNICORN}(x) \wedge \text{KISS}_*(j, \vee x))$$

$$(213) \quad \exists x(\text{UNICORN}_*(x) \wedge \text{KISS}_*(j, x))$$

In this way, the introduction of individual concepts does not affect the results of the fragment without individual concepts; these carry over without restriction. But it enables us to account for the examples given in §6.4.1 and hence to formulate a more adequate semantics of English.

Exercise 9

Which of the following sentences can be reduced to formulas with the $*$ -notation? What does the reduction depend on?

- (i) John kisses Mary.
- (ii) John kisses a unicorn.
- (iii) John seeks Mary.
- (iv) John seeks a unicorn.

6.5 Compositionality, Logical Form, and Grammatical Form

In this section we will briefly take up an issue from §6.1.1: the methodological status of the principle of compositionality and its relationship with the contrast between logical form and grammatical form.

The term 'logical form' nowadays may mean two quite different (though perhaps historically not unrelated) things. It may refer to a notion widely used and explored in generative grammar since the seventies. In generative grammar it stands for a specific level of description in grammar, distinct from surface structure and underlying structure. Or the term may be used to denote a concept which is much older and which is of philosophical origin.

To begin with the latter, a distinction between the grammatical form of an expression and its logical form has been made in logic and philosophy at various times in history. Especially since the development of modern quantificational logic at the end of the nineteenth century, the idea that the 'observable' grammatical form of a sentence may mislead as to its real logical form has been formulated with vigor and conviction by such notables as Frege, Russell, and Wittgenstein. (See chapter 1 of volume 1 for more details.) And this 'Misleading Form Thesis' and the concomitant view of natural language as irregular, unsystematic, vague, and deficient has dominated philosophical and logical thinking about language well into the second half of the twentieth century.

This distinction between logical form and grammatical form is quite alien to logical grammar. The following quotation from Montague's 'Universal Grammar' (1970b) may serve as an illustration:

There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed, I consider it possible to comprehend the syntax and semantics of both kinds of languages within a single natural and mathematically precise theory. On this point I differ from a number of philosophers, but agree, I believe, with Chomsky and his associates. It is clear, however, that no adequate and comprehensive semantical theory has yet been constructed, and arguable that no comprehensive and semantically significant syntactical theory yet exists.

From this quotation it is also evident that Montague feels that his conviction that natural language can be given a rigorously formal description is shared by linguists in the Chomskyan tradition. However—and this brings us to the second meaning of the phrase ‘logical form’ and to the issue of compositionality—he is also convinced that work in the generative tradition offers no semantic theory, and that the syntactic theory it presents probably fails to provide an adequate basis for semantics.

As for the first claim, ‘semantics’ in Montague’s book means ‘truth-conditional semantics’, and indeed this kind of semantics has never been the concern of generative linguistics. To be sure, the level of logical form that figures in Chomsky-style grammar serves to account for some facts that a logical grammarian would classify as semantic and hence would want to account for too, but the way this is done in the two approaches greatly differs. And more importantly, the very idea of a model-theoretic semantics has been flatly rejected by some of the leading proponents of generative grammar as being irrelevant for natural language. We will not try to give a description of the exact nature of the generativist notion of logical form, nor of the role it plays in the grammar. It may suffice to note the following. The level of logical form is considered to be a level of description that is different from both the surface structure and the underlying (or ‘deep’) structure of an expression. It serves to account for various phenomena, such as coreference and scope, but not through a process of interpretation. In fact, it is not always clear whether logical form is part of semantics or of syntax. The logical form of an expression is only partly determined by its surface syntactic structure, the relation being one-many. The latter fact means that even if the level of logical form were something like a representation of a full-blown model-theoretic interpretation (which it isn’t), it still wouldn’t fit the most basic principle of logical grammar, that of the compositionality of meaning.

The disagreement between logical grammar and generative grammar is in fact nicely illustrated by the second claim attributed to Montague above, viz., that existing syntactic theories are not ‘semantically significant’. To put it rather bluntly, while the starting point for Montague is semantics, for Chomsky it is syntax: the former embraces the principle of compositionality of meaning, and the latter advocates the autonomy of syntax. As we remarked in §§6.11 and 6.2, these two principles may be in conflict with each other.

So two things may be noted. First, neither the traditional philosophical notion of logical form nor the modern one which is used in generative grammar seems to play a role in logical grammar. And second, the main watershed between generative grammar and logical grammar seems to be formed by the principle of compositionality. This raises two questions. If we want to identify a level of representation in logical grammar as a level of logical form which approximates the one used in generative grammar, what level is this? And what exactly is the status of the principle of compositionality and hence of the disagreement between generative grammar and logical grammar? Is it factual,

or is it rather methodological? In what follows we will try to answer these two questions briefly.

From now on we will use the phrase 'logical form' to mean the representation of an expression that determines its meaning. In the PTQ model there are three distinct levels of representation. We can distinguish between:

- (i) expressions
- (ii) analysis trees
- (iii) logical expressions

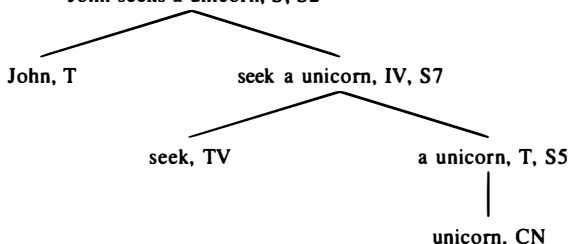
An expression is a string generated by the syntax, i.e., a sequence of symbols which the grammar declares is well-formed. The analysis tree encodes the derivational history of an expression. It contains the following information. It specifies which basic expressions were used and which rules were employed to form complex expressions. The logical expression is the result of the translation process applied to the analysis tree.

Now consider our by now worn-out example:

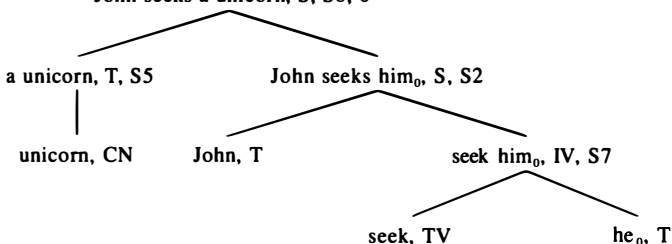
(214) John seeks a unicorn.

As a string generated by the syntax, i.e., at level (i), (214) represents itself. On the second level, there are two representations: the analysis trees corresponding to the direct and indirect constructions. These are repeated in figure (215).

(215) a. John seeks a unicorn, S, S2



b. John seeks a unicorn, S, S8, 0



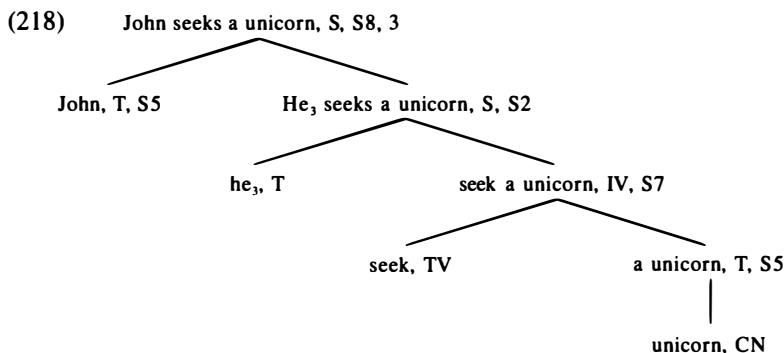
And on the third level we again find two representations: the two (reduced) translations (216) and (217) (disregarding individual concepts again):

(216) $\text{SEEK}(j, \wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$

(217) $\exists x(\text{UNICORN}(x) \wedge \text{SEEK}_*(j, x))$

With respect to the analysis trees and the logical expressions, it should be remarked that representations at these levels are not unique. For the translations we have stressed this several times: all the reduction and simplification steps that we apply are 'meaning preserving' in the sense that we always proceed from an expression to one that is logically equivalent with it. So it makes no sense, to speak, for example, of *the* logical expression which is the representation of the *de re* reading of (214). For obvious practical reasons, we will use (217) as representation in most cases, and in certain circumstances, e.g., when we want to apply a syntactic proof mechanism, it may even be necessary to do so. But as a representation of one of the meanings of (214), (217) has no privileged status: any of an infinite number of equivalent expressions would do.

The same holds for analysis trees. This is perhaps most clear in the case of (215b) where a different choice of syntactic variable would have resulted in a different analysis tree which would obviously determine the same meaning, since it would have resulted in a translation which is equivalent to (217). As for the direct construction exemplified in (215a), note that PTQ also allows us to use the quantifying mechanism in an 'empty' way, i.e., with no semantic effects. The analysis tree in figure (218) illustrates this.



This tree leads to a translation which is equivalent to (216) and hence also represents the *de dicto* meaning of (214).

Now what about logical form? Which of these representations determine the meaning of (214)? Since (214) is ambiguous, it must be assigned two distinct logical forms, and hence, (214) itself, i.e., the representation of our example as a string generated by the categorial syntax, cannot count as such. And this was not to be expected either, since we saw in §6.2 that compositionality implies that meaning is determined *given* an analysis. This leaves the level of analysis trees and that of translations to be considered. In fact, both may very well be considered as levels of logical form, as the following considerations may help to make clear.

Let us start with the translations. We have to bear in mind that it is not

proper to speak of a single logical expression as the representation; rather an entire equivalence class of expressions is to be counted as such. Two things should be noted. First of all, such a class determines, by definition, a unique meaning. But second, there isn't much that is *formal* about the resulting notion of logical form, for the logical equivalence of two expressions need not rest on any interesting resemblance between their forms.

As for the analysis trees, the important thing to note is that an analysis tree determines a unique translation. It has basic expressions as its leaves, and it has at each node a unique derived expression and an indication of the syntactic rule by means of which the latter was formed. The basic expressions have a unique translation, which is given in translation rule T1. And each derived expression has a unique direct translation, too, which is determined by the translation rule which corresponds to the syntactic rule. Given that logical expressions are unambiguous, it follows that an analysis tree determines a unique meaning. We noted above that, as was the case with translations, we must take an equivalence class of an analysis tree, rather than a single tree as the representation of the meaning of an expression. The relevant equivalence relation is that of generating an equivalent logical expression as a translation. Here, too, we note that there is no formal identity between the members of one and the same logical form, although it is to be expected that the formal resemblances are stronger here than in the case of logical expressions.

So we conclude that we have two candidates to serve as logical forms. Is there any reason to choose one rather than the other? As a matter of fact, compositionality again provides us with one. Note that it is the principle of compositionality of meaning which dictates that our grammar must contain the level of representation of the analysis trees. For compositionality states that the meaning of an expression is determined by, i.e., is a function of, the meanings of its parts. The meanings, it should be stressed once more, are the semantic objects in the model, i.e., the individuals, properties, propositions, second-order properties and so on that we associate with the expressions. The logical expressions serve to represent these but are not to be confused with them. The point is that this description of compositionality refers to an informal notion of the 'parts of' an expression. In a sense, the main point of doing syntax, that is, 'semantically significant' syntax, is to explicate this notion. A simple example of an ambiguous expression suffices to show that the parts cannot be identified with the lexical elements from which an expression is built. Example (214) has two different meanings, both of which are conveyed by the same set of basic expressions. And as was remarked in §6.2, the constituents of an expression are not the relevant objects either, since an expression may be ambiguous without having two distinct constituent structures. But compositionality simply requires that there *be* different 'parts' whenever there is nonlexical ambiguity, and if none of the known notions will do, the parts have to be 'invented'. In the PTQ framework this has led to the notion of derivations encoded in analysis trees, to the introduction of quantification rules, and so on. Other options are available, and other techniques have been

developed, but the important thing to note here is that *compositionality demands a disambiguated level of representation in the syntax*.

This establishes the necessity of (something like) the level of analysis trees, but it also shows the optional character of the level of logical expressions. For if the analysis trees determine meaning, the translations cannot add anything to it: they must be superfluous. And in fact they are. Given that an analysis tree determines a unique translation, and that logical expressions are unambiguous, it is always possible to bypass the translation level and interpret the analysis trees directly. Simply assign the semantic objects that are the interpretations of the translations of basic expressions to these expressions directly, and instead of the translation rules, use the semantic operations which correspond to these to operate on semantic objects. This is the method used by Montague in "English as a Formal Language" (1970a). So we conclude that if we should label one of the three levels of representation in the PTQ model as a level of logical form, the most reasonable choice is that of the analysis trees. They determine meaning uniquely, and they are a necessary ingredient of the grammar, in accordance with the principle of compositionality.

Let us now finally turn to our second question, which concerns the status of compositionality. Is compositionality to be regarded as a kind of empirical hypothesis, or is it rather a methodological principle? The considerations above suggest an answer. We have noted that the principle of compositionality requires a disambiguated level of representation in the syntax. When dealing with artificial languages, we simply set up the syntax in such a way that we comply with this requirement. For example, the expressions of logical languages have their derivational history encoded in their structure through the use of brackets or similar devices: each expression corresponds to a unique derivation tree and hence can be interpreted completely compositionally. For a natural language things are different. On the one hand, we have a notion of syntactic (constituent) structure which we may assume is motivated independently from semantic considerations. On the other hand, we are faced with the task of explicating the notion 'part of' that compositionality speaks of. Suppose we come across ambiguous expressions which are not 'structurally ambiguous', i.e., which cannot be broken up into constituents in two different ways. Then we may proceed in two ways. We may admit a level of syntactic representation other than that of constituent structure and so comply with compositionality. Or we may stipulate that the constituents are the relevant 'parts' and that hence natural language meaning cannot be described compositionally.

The important thing to note is that the first way is always open to us unless we decide in advance what can and cannot be part of our syntax. If we start out with the assumption that constituent structure and only constituent structure is what our syntax should represent, then we may indeed say that the hypothesis that the semantics of a natural language such as English is compositional is 'falsified' by the facts. But note that this initial assumption is not an empirical fact but rather a methodological decision. So it seems reasonable to conclude that whatever approach we take, compositionality is a method-

ological issue: we choose to describe the semantics of our language in a compositional fashion, or we decide not to, but in both cases what is at stake is a matter of methodology rather than of facts.

6.6 Concluding Remarks

The foregoing has been devoted to an in-depth introduction to the peculiarities of one particular model of logical grammar, Montague's PTQ. The reasons for choosing this model rather than another one were given in §6.1. Having mastered PTQ, the reader will find it a relatively easy task to get acquainted with other models and approaches which share its main background assumptions. See, for example, Bartsch and Vennemann 1972; Bartsch 1976b; Cresswell 1973, 1985; and Lewis 1972.

It should also be possible for such readers to find their way through the enormous amount of theoretically and empirically oriented literature that Montague's original papers have generated. It is quite impossible to give here a survey of the work done in this area, and we must content ourselves with pointing out a few particularly important individual contributions and collections.

Let us first of all mention some other introductory and exegetical literature. Dowty, Wall, and Peters 1981 provides an extensive introduction to PTQ in English; Link 1979 and Löbner 1976 provide introductions in German. Partee 1975 is a lengthy article which introduces PTQ to generative linguists. Thomason's introduction to Montague 1974 focuses more on the philosophical and logical aspects of Montague's work. Halvorsen and Ladusaw 1979 spells out and explains Montague's general semiotic theory, formulated in his "Universal grammar" (1970a).

A very thoroughgoing, mathematically oriented study of the content and role of the principle of compositionality, to which this introduction owes much, can be found in Janssen 1986. This also contains several contributions to empirical subjects (relative clauses, tense, and aspect), and an application of several of Montague's techniques to problems in the semantics of programming languages. A more empirically oriented study of compositionality is Partee 1984a. Compositionality was described above as the watershed between generative grammar and logical grammar. But these enterprises may also contribute to each other. Several authors have been concerned with this issue; see Partee 1973, 1979a; Cooper and Parsons 1976; Bach 1979b; McCloskey 1979. A Montague-style model-theoretic semantics is also employed in other models of grammar, notably in generalized phrase structure grammar: see Klein and Sag 1984; Gazdar, Klein, Pullum, and Sag 1985. One of the main drawbacks of Montague grammar from the point of view of generative grammar is the introduction of the level of analysis trees in the syntax. Cooper has developed an alternative to the quantifying mechanism of PTQ which depends heavily on the availability of analysis trees as a theoretical tool, while remaining compositional. See Cooper 1983 for a fully worked out analysis of his

'storage' mechanism. An application of these ideas can be found in Partee and Bach 1981. A different approach to the problem of representing scope ambiguities, which requires a certain relaxation of the compositionality requirement, is Hendriks 1988. The issue of the possibility of a 'monostratal grammar', i.e., a grammar with only one level of representation, is at the center of various developments, e.g., that of generalized phrase structure grammar (see above). See also Hausser 1984.

Anaphora are intensively studied in both generative and logical grammar. The analysis that PTQ offers, though adequate for a large class of cases, is in need of extension and refinement. Representative early treatments are Bartsch 1979; Cooper 1979; Partee 1979b; and Hausser 1979. Two special issues of *Linguistics and Philosophy*, 6(1): (1983) and 7(3): (1984), contain several interesting contributions; the one by Landman and Moerdijk deserves special mention in this context. Problems concerning anaphora and indefinite terms generated an entirely new theory in the early eighties: discourse representation theory, which will be discussed in some detail in §7.4. References to the literature in this area will be given there.

Let us very briefly mention some contributions on other empirical issues. Dowty's work on the analysis of aspectual verbs and tense can be found in Dowty 1979. See also Bennett 1977; Verkuyl 1989. Of related interest is Kamp 1980. Control verbs are treated in Bach 1979a and in Klein and Sag 1984. Adjectives and adverbials are analyzed in Bartsch 1976a; Cresswell 1976, 1979; Kamp 1975; Klein 1980. See also Stalnaker and Thomason 1973. Plurality is another issue that has been the subject of many studies. See especially Scha 1981; Verkuyl 1981; Link 1983; Landman 1988. Another topic is negation: see Jacobs 1982; Verkuyl 1987. An analysis of generics is presented in Carlson 1977, 1982. Questions are studied in Hamblin 1973; Karttunen 1977; Hausser and Zaefferer 1978; Belnap 1982; Groenendijk and Stokhof 1982, 1984, 1988; Hausser 1983; Engdahl 1985. The semantics of nominalized constructions has led to interesting proposals concerning the kind of semantic theory that is best suited to natural language; see Turner 1983; Chierchia 1982, 1984; Chierchia and Turner 1988.

The above list is far from complete. The reader may get a glimpse of the many other subjects treated in the Montague tradition by browsing through collections such as Davidson and Harman 1972; Keenan 1975; Partee 1976; Guenther and Rohrer 1978; Guenther and Schmidt 1979; Davis and Mithun 1979; Rohrer 1980; Groenendijk, Janssen, and Stokhof 1981, 1984; Bauerle, Egli, and von Stechow 1979; Bauerle, Schwarze, and von Stechow 1983; Landman and Veltman 1984; Groenendijk, de Jongh, and Stokhof 1987a, 1987b; Klein and van Benthem 1988; Groenendijk, Stokhof, and Veltman 1988.

Besides discourse representation theory, two other important developments in model-theoretic semantics for natural language will be introduced in chapter 7. These are the theory of generalized quantifiers and flexible categorial grammar. References to relevant literature will be given there.

7

Recent Developments

7.1 Introduction

This chapter will introduce three subjects which are currently at the center of interest in the field of logical semantics. All three build upon the framework of Montague grammar as described in chapter 6, yet they deviate from the course set out there in some fundamental respects. The three subjects are the theory of generalized quantifiers, flexible categorial grammar, and discourse representation theory.

The theory of generalized quantifiers may be viewed as a further development of Montague's analysis of quantifying expressions in PTQ, using the tools of abstract model theory. The theory has various objectives. Its aims are partly descriptive, and its nature is partly theoretical. The descriptive work involves a variety of topics, such as the internal semantic structure of terms, the distribution of negative polarity items, *there* insertion, and conjunction reduction. The more theoretical research focuses on restrictions on possible meanings of natural language terms, the expressive power of natural languages with regard to possible meanings, semantic universals, and so on. Key references are Barwise and Cooper 1981; van Benthem 1983a, 1984a, 1987; Keenan and Moss 1984; Keenan and Stavi 1986; Keenan 1987.

The framework of categorial grammar has evolved in recent years into a flexible tool which is better suited to capture various generalizations concerning natural language syntax and semantics. The main difference with the system of Montague's PTQ lies in making the relationship between expressions and categories more flexible: it is no longer assumed that an (unambiguous) expression belongs to only one category; various rules are postulated that allow us to change the category initially assigned to an expression by the lexicon into a well-defined set of other categories. In this way, various phenomena which categorial grammar in its original form could not deal with, such as discontinuous constituents, can be described adequately. Moreover, the category-changing component allows us to simplify some of the complexity of the category and type assignment of Montague's PTQ. Furthermore, the strict functional tie between syntactic categories and semantic types has been loosened. Important work in this field can be found in Partee and Rooth 1983; Zwarts 1986; van Benthem 1986; Moortgat 1988.

Discourse representation theory is in a sense the most antagonistic to the framework of Montague grammar. One of the motives for its development was to find alternatives for several central aspects of Montague grammar, and one of its objectives is to transcend the restriction of the latter to sentences, by moving on to extended discourses, or texts. Discourse representation theory was developed by Hans Kamp (1981a) and Irene Heim (1982, 1983), but similar ideas have been proposed in quite different frameworks (see, for instance, Karttunen 1976; Seuren 1985). Discourse representation theory has both descriptive and more theoretical aspects. On the descriptive side we find topics such as the distinction between referential and nonreferential terms, particularly in connection with anaphora (for instance the notorious 'donkey' sentences). Furthermore, the theory is tested with respect to the treatment of time and aspect (e.g., Partee 1984a; Kamp and Rohrer 1983) and of propositional attitudes (Asher 1986; Zeevat 1987). A more theoretical ambition is the possible synthesis of two views on meaning, the truth-conditional, model-theoretical conception and the procedural, representational viewpoint. Another important objective has already been mentioned: the extension of the domain of semantic theories from sentences to texts ('discourses').

7.2 The Theory of Generalized Quantifiers

7.2.1 *Principal Objectives*

An important aspect of a semantic theory like Montague grammar is the association of a certain type of semantic objects (truth values, properties, and so on) with a certain category of syntactic objects (sentences, common nouns, and so on). In general, no further constraints are imposed on the association than those which are required in order to account for our intuitions about semantic relations, such as entailment. Where necessary, restrictions are formulated by means of *meaning postulates*. Most of the meaning postulates apply either to individual expressions or to restricted classes of expressions. Their function is to isolate certain elements within the totality of semantic objects of a certain type as the possible meanings for a (class of) expression(s). But except for this kind of restriction, Montague grammar is concerned with the *entire class* of semantic objects of a type.

The theory of generalized quantifiers deals with the semantic objects which are the interpretations of terms: sets of properties. Within this theory, a main point of interest is the structure of these semantic objects: what formal properties do they have, what natural subclasses can be distinguished, and which of these can be considered to actually represent meanings of natural language terms? The investigation of such topics goes beyond the mere formulation of a relation between a syntactic category and a semantic type. We will give some examples.

One of the first lines of research tries to achieve a *classification* of generalized quantifiers in terms of their formal properties, attempting to give an

explanation of several linguistic phenomena. A simple instance of this is *there* insertion. Some terms may occur in context (1); others may not, as the difference between (2a) and (2b) will show (note that the sentences in (2b) are not interpreted as exclamations, in which case they would be correct):

- (1) There is/are . . . in the garden.
- (2) a. There is someone in the garden.
 There is no one in the garden.
 There are two unicorns in the garden.
- b. *There is/are everyone in the garden.
 *There is John in the garden.
 *There are the two unicorns in the garden.

The question which now arises is the following: are there properties of term meanings that distinguish the terms which can occur in the context of (1) from those which cannot? And do these properties explain, together with a semantic analysis of the phrase *There is/are . . .*, why these terms do or do not fit? For answers to these question, see, e.g., Barwise and Cooper 1981; Zwarts 1981; de Jong and Verkuyl 1984.

Another example of research in this direction concerns the distribution of expressions with 'negative polarity'. Compare (3) and (4):

- (3) John needn't go there.
 *John need go there.
- (4) Nobody saw anything.
 *Somebody saw anything.

Traditionally, the possibility of the occurrence of expressions with negative polarity like *need* and *any* has been connected with the occurrence of a negative element in the sentence (whence the name): negation in (3), *nobody* versus *somebody* in (4). The traditional explanation, however, is problematic for the interpretation of sentences like (5) and (6):

- (5) John needn't get more than a B.
- (6) *John needn't get less than a B.

The postulate of a separate abstract negative element in syntactic deep structure, which will be merged into the element it operates on at a later stage of the derivation, is not a very attractive solution. An explanation in terms of the semantic properties of this type of expression (*nobody* versus *somebody*, *more than n* versus *less than n*) seems preferable. The subject is discussed at length in Zwarts 1981, 1986.

A last example deals with the phenomenon of *conjunction reduction* (the name is adopted from transformational grammar). Compare (7) and (8):

- (7) John plays and John sings.

(8) John plays and sings.

The two sentences are equivalent. Transformational tradition once had it that (8) is derivable from (7) by the transformation of 'conjunction reduction'. But, compare (9) and (10):

(9) Nobody plays and nobody sings.

(10) Nobody plays and sings.

The two sentences are not equivalent: (9) implies (10) but not the other way around, and the proposed transformation should not be applicable in this case (assuming that transformations should be meaning preserving). This problem is difficult to cope with in a traditional transformational perspective. On the other hand, if we give up the assumption that identity of meaning should be accounted for in terms of the identity of syntactic (deep) structure, the situation changes. If we have an explicit semantics, like the one in Montague grammar, which enables us to account for semantic relations like synonymy and implication in terms of relations between semantic (model-theoretical) objects, and not in terms of relations between syntactic structures, the question has to be reformulated as: what properties of the type of semantic object associated with terms guarantee such relations of synonymy?

It should be clear from this short exposition that even the most empirically oriented results always have more theoretical or methodological implications. This will be discussed later on.

A second branch of research within the theory of generalized quantifiers consists in the search for *universals*, i.e., the formulation of significant universal regularities governing the semantic objects which are the meaning of terms.

Characteristically, Chomskyan linguistics looks for the grammatical principles that isolate the subclass of all possible human languages from the class of all possible languages. Such grammatical principles would form a universal grammar. (It seems obvious to associate such a grammar with universal principles of human thinking, and appealing to the rationalist tradition, this is what Chomsky did.) However, Montague proceeded from a different starting point, with a different objective: he wanted one uniform and mathematically exact framework which would contain both human, natural languages and formal languages. This was Montague's conception of a 'universal grammar' (see Montague 1974, chapter 4).

The theory of generalized quantifiers seeks to explore the interest of the Chomskyan tradition within the framework of model-theoretic semantics: the semantic domain of terms, the set of all sets of properties of individuals, is extremely 'big'. A priori, the assumption that all these potential meanings are suitable, i.e., actually express meanings of natural language terms, does not seem plausible; hence universally valid restrictions are to be formulated. The ensuing research on universal properties of meanings for natural language

terms has been done mostly by Barwise and Cooper (1981). Examples of this kind of semantic universals will be discussed in §7.2.4.

A third topic in the theory of generalized quantifiers is the search for *constraints*, formal properties which define certain independently interesting classes of determiners. This kind of research is closely linked with the preceding. For instance, van Benthem (1983a) raises the problem of what properties characterize the class of logical determiners (*all*, *some*, *no*, *not all*, i.e., the traditional Aristotelian square). Surely this class is interesting not only from a logical perspective but also from the point of view of natural language semantics. The question can also be put the other way around: given a certain (set of) global constraint(s), what is the class of natural language expressions which fulfill it/them? Some results in this field will be discussed in §§7.2.5 and 7.2.6.

Another research topic, which is also connected with the ones mentioned before, is the *expressive power* of natural languages. This research looks for constraints which could reduce the number of all potential semantic objects to an expressible number. The strategy most commonly followed assumes intuitively plausible constraints and then tries to prove that all meanings in such a constrained class can be actually expressed in natural language. The more independent the motivation of the constraints, the more the results of the strategy support the principle of expressibility of natural languages (and, if one wishes, of human thought). The research in this area will not be discussed in the following (but see, for instance, Keenan and Moss 1984; Keenan 1971).

7.2.2 NPs as Generalized Quantifiers in Montague Grammar

In this section we will briefly repeat the most important characteristics of Montague's treatment of terms and its application in the theory of generalized quantifiers.

To ensure that our exposition agrees with the literature on this subject, we adopt the common linguistic terminology from now on. NP denotes the class of expressions such as proper names, descriptions, and quantified terms. NP corresponds to Montague's T for Terms. VP denotes all verb phrases, both IVs and TVs, and N denotes all nouns, in Montague grammar called CN. DET is used to refer to the category of determiners (i.e., the articles and expressions such as *all*, *some*). Furthermore, we will use E, instead of D, to refer to the domain, reserving D for the interpretation of determiners.

Montague's analysis of NPs as it was described in chapter 6 depends on two principles: *uniformity* and *compositionality*.

The effect of uniformity is twofold. First, expressions exhibiting similar syntactic behavior, i.e., obeying the same distributional laws whenever this is syntactically determined, are regarded as belonging to the same syntactic category. For this reason, both proper names and descriptions on the one hand and quantified NPs on the other are classified as NPs, though their semantic be-

havior is different. Second, a syntactic category corresponds to one semantic type, that is, all expressions of a category have the same kind of meaning. (Recall that categorial grammar was originally conceived as a system of semantic categories.) In the case of NPs, this meant that the useful analysis of quantified NPs as sets of properties was extended to proper names. Compositionality implies that an NP has an independent meaning. NPs are independent syntactic units, and their meanings are the building blocks for the meanings of larger units. Compositionality 'naturally' leads to a semantic analysis of NPs as generalized quantifiers, hence as sets of properties (see §§4.4.3 and 6.3.4).

We may put it very briefly as follows. Consider sentence (11a), its syntactic structure (11b), and its translation (11c) in predicate logic:

- (11) a. Every man walks.
 b. $s_{[NP[every\ man]_{VP}[walks]]}$
 c. $\forall x(MAN(x) \rightarrow WALK(x))$

The meaning assigned to the expression *every man* in (11c) is not an independent one; compare (12):

- (12) a. Every man sleeps.
 b. $\forall x(MAN(x) \rightarrow SLEEP(x))$

In the comparison of the meanings of (11a) and (12a), intuitively there is something which is different in the two sentences (viz., *walk*, *sleep*) and something which is the same (viz., *every man does it*). The procedure is now to make a variable of the thing that is different and abstract over it, thus retaining the constant factor of the meaning. In the intensional semantics of PTQ, we get the following representation:

- (13) $\lambda X \forall x(MAN(x) \rightarrow \vee X(x))$, where X is of type $\langle s, \langle e, t \rangle \rangle$

Notwithstanding the analysis of PTQ discussed in chapter 6, which is heavily intensional, the theory of generalized quantifiers is extensional. First, it uses only extensional models $M = \langle E, \llbracket \cdot \rrbracket \rangle$, where E is a set of individuals and $\llbracket \cdot \rrbracket$ is an interpretation function assigning extensional interpretations to expressions (i.e., assigning to an expression that which is the extension of that expression in the intensional logic of PTQ). Second, the extensions are extensional: individuals instead of individual concepts, sets of individuals instead of properties, and so on.

In other words, (14), and not (13), is the representation of the meaning assigned to the NP *every man* by the theory of generalized quantifiers:

- (14) $\lambda X \forall x(MAN(x) \rightarrow X(x))$, where X is of type $\langle e, t \rangle$.

However, it is common to write these meanings directly in the metalanguage, using some set-theoretical notation without an intermediary logical language. That is, we get representations such as (15):

- (15) $\llbracket every\ man \rrbracket = \{X \subseteq E \mid \llbracket man \rrbracket \subseteq X\}$

The restriction to extensional models looks more severe than it actually is. In fact there are not many really intensional NPs or determiners.

7.2.3 Determiners: Two Perspectives

Within the theory of generalized quantifiers we can distinguish two different perspectives with respect to determiners: a *relational* perspective and a *functional* perspective. The latter is closely related to the traditional linguistic analysis of a sentence as being composed of a subject and a predicate, or in other words, of an NP and a VP. Consider the following simple sentence (16a) and its constituent structure (16b):

- (16) a. All men sleep.
 b. $s_{[NP[DET[all]]_N[men]]_{VP}[sleep]}$

The determiner *all* combines with the noun *men* to form the NP *all men*. In terms of categorial syntax, a determiner is of category NP/N. Semantically this implies that it is interpreted as a function: a function which assigns an NP interpretation, a set of sets of individuals, to an N interpretation, a set of individuals. This functional perspective is reflected in the analysis of NPs in the framework of Montague grammar (compare §6.3.2 and the representation of *every man*, (13) in §7.2.2).

A different way of looking at determiners is to regard them as relations. In this perspective, *all* in (16a) is an expression which relates an N to a VP to form an S. Both *men* and *sleep* are interpreted as sets of individuals, and *all* is regarded as a relation between sets, namely, that relation which holds between two sets X and Y iff $X \subseteq Y$. Read in this way, sentence (16a) asserts that the set of men is a subset of the set of sleeping individuals.

At first sight, the two perspectives seem to differ greatly. But if we take another look at the representation of NP meanings in a type-theoretical framework, and in particular at the type of this representation, we see that both perspectives amount to the same thing. The type of a determiner in the functional perspective is $\langle\langle e, t \rangle, \langle\langle e, t \rangle, t \rangle\rangle$. The corresponding semantic domain is formed by the set of functions from (the characteristic functions of) sets of individuals to (the characteristic functions of) sets of sets of individuals. In general it holds that a function from objects of type a to a set of objects of type b can be identified with a relation between objects of type a and objects of type b (as we observed in §4.2.3). Let $f \in (\{0, 1\}^{E_a})^{E_b}$. The corresponding relation $R_f \subseteq E_a \times E_b$ is defined as follows: $R_f = \{\langle d_a, d_b \rangle : d_a \in E_a \ \& \ d_b \in E_b \ \& \ (f(d_b))(d_a) = 1\}$. Conversely, if R is a relation, i.e., a subset of $E_a \times E_b$, the corresponding function $f_R \in (\{0, 1\}^{E_a})^{E_b}$ (is defined as follows: for all $d_b \in E_b$, $f_R(d_b)$ is that function $\in \{0, 1\}^{E_a}$ such that for all $d_b \in E_b$, $(f_R(d_a))(d_b) = 1$ iff $\langle d_a, d_b \rangle \in R$). Thus, these 'determiner functions' correspond uniquely to 'determiner relations' between sets of individuals.

In the theory of generalized quantifiers, the functional perspective is used in

the kind of research which focuses on properties of NPs. And the functional perspective provides the framework for actual description, as in Montague grammar. The relational perspective is often more comprehensible for an analysis of determiners as such. In the following, both will be discussed, beginning with the functional and using the relational perspective later on.

A final remark concerns a possible misunderstanding. In §4.4.3, we observed that determiners like *most* and *more than half* are essentially relational, as opposed to, for instance, *all* and *some*. In view of what was said earlier, this could lead to a misunderstanding, since we just stated that the functional and relational perspectives are interchangeable, that they come down to the same thing.

The point is that some determiners, such as the 'logical' determiners *all* and *some*, provide reasons for a third alternative analysis: we can regard them as second-order predicates, i.e., as expressions of type $\langle\langle e, t \rangle, t\rangle$. For instance, *all* asserts of a (complex) property that every object in the domain has that property. Extensionally it is defined as follows:

$$(17) \llbracket \text{all} \rrbracket = \{E\}$$

The same goes for *some*, which asserts of a property that it is not empty:

$$(18) \llbracket \text{some} \rrbracket = \{X \mid X \subseteq E \text{ \& } X \neq \emptyset\}$$

We may want to apply this interpretation to the analysis of sentences like (16a) and (19):

$$(19) \text{ Some women stroll.}$$

Then it is essential that we are provided with a way to represent such complex predicates as *being a strolling woman* (for instance, by means of λ -abstraction). It is clear, however, that a determiner like *most* does not permit an analysis as a second-order predicate; we cannot define its interpretation without making a separate reference to the interpretation of the noun. Sentence (19) can be paraphrased as

$$(20) \text{ The set of strolling women is not empty / contains something (i.e., is an element of (18)).}$$

But a similar paraphrase of (21) doesn't make sense, as (22) shows:

$$(21) \text{ Most men are asleep.}$$

$$(22) \text{ The set of sleeping men contains most things.}$$

And other possible reductions fare equally badly. In other words, the interpretation of *most* essentially refers both to the set which is the interpretation of the VP and to the set which is the interpretation of the noun: it expresses a relation between the two. In this sense, *most* is an essentially relational determiner. However, note that in this latter sense of 'relational', both the rela-

tional perspective and the functional perspective which were described above give a 'relational' interpretation of determiners.

7.2.4 Some Fundamental Properties of NPs and Quantifiers

Terminology; Examples; Undefined Interpretation

It is important to make a systematic distinction between NPs and their interpretations. An NP is a linguistic, syntactic object, a natural language expression. A quantifier is a semantic object, a set of sets. Models \mathbf{M} are ordered pairs $\langle E, \llbracket \cdot \rrbracket \rangle$, where E is a set of individuals, the domain of the model, and $\llbracket \cdot \rrbracket$ is an interpretation function assigning interpretations to natural language expressions.

Unlike the method used in Montague grammar, the theory of generalized quantifiers does not use an intermediary translation level in an interpreted logical language (like *IL* in *PTQ*; but Montague also used *direct* interpretation; cf. "English as a Formal Language" in Montague 1974). Interpretations are written directly in the metalanguage, which is English enriched with some set-theoretic and logical notation.

By way of example, we give the interpretations of some common NPs in table 7.1, using $\text{card}(X)$ to refer to the cardinality of X . Note that the interpretation of an NP depends on the model: in the interpretations the domain of \mathbf{M} , E , occurs as a parameter. In other words, what actual quantifier is the interpretation of an NP depends on the model. Of course we are chiefly interested in those properties of NP interpretations which they have regardless of the model. (In the following we will omit ' $X \subseteq E$ ' whenever this is not misleading.)

Before going on to discuss some examples of such properties, we must consider briefly what treatment we should give to 'presuppositional' NPs, such as definite descriptions. The interpretation of *the king of France* is the set of sets X such that the king of France belongs to X , *if* there is a unique king of France. But what will the interpretation be if there is no such individual? In principle, there are several options to choose from, and our choice will de-

Table 7.1 NP Interpretations

NP	Interpretation
All N	$\{X X \subseteq E \text{ \& } \llbracket N \rrbracket \cap X = \llbracket N \rrbracket\}$
An N	$\{X X \subseteq E \text{ \& } \llbracket N \rrbracket \cap X \neq \emptyset\}$
Not all N	$\{X X \subseteq E \text{ \& } \llbracket N \rrbracket \cap X \neq \llbracket N \rrbracket\}$
No N	$\{X X \subseteq E \text{ \& } \llbracket N \rrbracket \cap X = \emptyset\}$
Only N	$\{X X \subseteq E \text{ \& } \llbracket N \rrbracket \cap X = X\}$
Exactly 2 N	$\{X X \subseteq E \text{ \& } \text{card}(\llbracket N \rrbracket \cap X) = 2\}$
At most 2 N	$\{X X \subseteq E \text{ \& } \text{card}(\llbracket N \rrbracket \cap X) \leq 2\}$
At least 2 N	$\{X X \subseteq E \text{ \& } \text{card}(\llbracket N \rrbracket \cap X) \geq 2\}$

pend, among other things, on what we think is the status of sentences in which presuppositional expressions occur, in case the presuppositions are not satisfied. If we think that (23) is false under the present circumstances, we can choose (24) as interpretation of *the king of France*:

(23) The king of France is bald.

(24) $\llbracket \text{the king of France} \rrbracket = \{X \mid \text{card}(\llbracket \text{king of France} \rrbracket) = 1 \ \& \ \llbracket \text{king of France} \rrbracket \subseteq X\}$

If, on the other hand, we consider (23) to be *without* any truth value, it is better to give the NP in question an interpretation only in a subclass of models and to consider the condition $\text{card}(\llbracket N \rrbracket) = 1$ to be a necessary condition for the interpretation to be defined:

(25) $\llbracket \text{the king of France} \rrbracket = \{X \mid \llbracket \text{king of France} \rrbracket \subseteq X\}$, if $\text{card}(\llbracket \text{king of France} \rrbracket) = 1$; otherwise undefined.

In general, people working in the theory of generalized quantifiers take the second approach, usually without much discussion. As a consequence, the interpretation function $\llbracket \cdot \rrbracket$ is partial: some sentences have no truth value, and this raises the (open) question whether and how a truth value is assigned to complex sentences in which a sentence without a truth value occurs. (See volume 1 §5.5 for a general discussion and some bibliographical references on the subject.)

In spite of the unresolved issues with respect to the 'undefined interpretation' approach, it is the one we will take, since, as we have said, this is generally done in the theory of generalized quantifiers. Other examples of NPs getting only a conditional interpretation are (26) and (27):

(26) $\llbracket \text{the two } N \rrbracket = \{X \mid \llbracket N \rrbracket \subseteq X\}$, if $\text{card}(\llbracket N \rrbracket) = 2$; otherwise undefined.

(27) $\llbracket \text{some } N \rrbracket = \{X \mid \text{card}(\llbracket N \rrbracket \cap X) \geq 2\}$, if $\text{card}(\llbracket N \rrbracket) \geq 2$; otherwise undefined.

Note that *some* is interpreted as a plural and not as a singular determiner, the latter being the usual choice in logical systems.

Exercise 1*

Give the interpretations of the following NPs:

- (i) *John*
- (ii) *a few N*
- (iii) *not only N*
- (iv) *neither N*
- (v) *a finite number of N*

Monotonicity

One fundamental property of quantifiers and NPs recognized by the theory of generalized quantifiers is *upward monotonicity*. Consider examples 28–33. (Here we use ‘ \models ’ to denote entailment between natural language sentences.)

- (28) All men walked rapidly \models All men walked.
- (29) All women walked \models All women moved.
- (30) A man smoked a cigar \models A man smoked.
- (31) A child was dreaming \models A child was asleep.
- (32) Both boys were out on the street playing \models Both boys were playing.
- (33) More than half of the girls live in Hoorn \models More than half of the girls live in a town.

Clearly, these entailments hold. And that they do is due to the *meaning* of the NPs in question. Apparently *all* N, *an* N, *both* N, and *more than half of the* N have something in common in their interpretation which accounts for these inferences.

An NP is interpreted as a quantifier, i.e., as a set of sets. The sets which form a quantifier can be taken as a (partial) interpretation of predicates: a sentence of the form [NP VP] is true iff $\llbracket \text{VP} \rrbracket \in \llbracket \text{NP} \rrbracket$. Now if we take another look at examples (28)–(33), we see that the predicate in the premise is necessarily subordinate to the predicate in the conclusion: $\llbracket \text{walked rapidly} \rrbracket \subseteq \llbracket \text{walked} \rrbracket$, $\llbracket \text{was dreaming} \rrbracket \subseteq \llbracket \text{was asleep} \rrbracket$, $\llbracket \text{live in Hoorn} \rrbracket \subseteq \llbracket \text{live in a town} \rrbracket$, and so on, in every suitable \mathbf{M} .

Apparently the interpretation of the NPs in (28)–(33) is such that whenever there is a set that belongs to it, all ‘larger’ sets belong to it as well. They express what are called *upward monotonic* (also: ‘monotonically increasing’) quantifiers. Let Q be a quantifier in \mathbf{M} , i.e., a set of sets of individuals from $E_{\mathbf{M}}$: then this property can be defined as follows:

Definition 1

Q is *upward monotonic* in \mathbf{M} iff for all $X, Y \subseteq E$: if $X \in Q$ and $X \subseteq Y$ then $Y \in Q$.

(Of course the interpretation is determined by \mathbf{M} as far as the quantification over subsets of E is concerned.) An NP is called *upward monotonic* if in every model in which its interpretation is defined it expresses an upward monotonic quantifier:

Definition 2

NP is *upward monotonic* iff for all \mathbf{M} : if $\llbracket \text{NP} \rrbracket$ is defined in \mathbf{M} , then $\llbracket \text{NP} \rrbracket_{\mathbf{M}}$ is upward monotonic in \mathbf{M} .

Another way of phrasing ' α is upward monotonic' is ' α is closed under extension'. Definition 1 provides a *test* for upward monotonicity:

Upward monotonicity test 1

If $\llbracket VP_1 \rrbracket \subseteq \llbracket VP_2 \rrbracket$, then $NP \ VP_1 \models NP \ VP_2$.

With this test, it becomes clear why the NP *not one boy* is not upward monotonic:

(34) Not one boy walked rapidly $\not\models$ Not one boy walked.

Exercise 2*

(a) Which of the following NPs are upward monotonic?

- (i) *at least n N*
- (ii) *the n N*
- (iii) *few N*
- (iv) *at most n N*
- (v) *Mary*
- (vi) *half of the Ns*

(b) Let P be a predicate such that for all M : $\llbracket P \rrbracket = E_M$. Show that if NP is upward monotonic and $\llbracket NP \rrbracket \neq \emptyset$, then for all M : $\llbracket P \rrbracket \in \llbracket NP \rrbracket$. Show also that this property is not a *sufficient* condition for upward monotonicity.

An equivalent definition of upward monotonicity is the following:

Definition 3

Q is *upward monotonic* in M iff for all $X, Y \subseteq E$: if $X \cap Y \in Q$ then $X \in Q$ and $Y \in Q$.

This definition also gives rise to a test:

Upward monotonicity test 2

$NP \ VP_1$ and $VP_2 \models NP \ VP_1$ and $NP \ VP_2$.

The interpretation of a conjunction of two VPs is the intersection of the interpretations of the two conjoined VPs. The following two examples illustrate this test:

- (35) All girls were smoking and drinking \models All girls were smoking and all girls were drinking.
- (36) Not one boy was singing and dancing $\not\models$ Not one boy was singing and not one boy was dancing.

The set of upward monotonic NPs is closed under conjunction and disjunction. In other words, the conjunction or disjunction of two upward monotonic NPs is an upward monotonic NP. Compare:

(37) All boys and a girl walked rapidly \models All boys and a girl walked.

- (38) Two men or both women were dreaming \models Two men or both women were asleep.

Semantically the conjunction of two NPs is the intersection of their interpretations:

$$\llbracket \text{NP}_1 \text{ and NP}_2 \rrbracket = \llbracket \text{NP}_1 \rrbracket \cap \llbracket \text{NP}_2 \rrbracket$$

It is easily seen that upward monotonicity is preserved under intersection. Let Q_1 and Q_2 both be upward monotonic. Assume for some X and Y : $X \in Q_1 \cap Q_2$ and $X \subseteq Y$. Then $X \in Q_1$ and $X \in Q_2$, hence by upward monotonicity of Q_1 and Q_2 , $Y \in Q_1$ and $Y \in Q_2$, from which it follows that $Y \in Q_1 \cap Q_2$.

Exercise 3

Define disjunction of NPs and show that upward monotonicity is preserved under disjunction.

Exercise 4*

Show that definition 1 is equivalent to definition 3.

Downward monotonicity is a property of NPs and quantifiers, which in a way is the mirror image of upward monotonicity. An exact formulation of this will be given below. Consider examples (39)–(42):

- (39) No man walked \models No man walked rapidly.
 (40) Not every woman was asleep \models Not every woman was dreaming.
 (41) Less than half of the girls smoked \models Less than half of the girls smoked cigars.
 (42) Few boys were playing \models Few boys were playing out on the street.

These inferences are all valid. If we compare them with (28)–(33), we see that the entailment goes the other way around. Upward monotonicity accounts for inferences in which the predicate (the VP) in the conclusion contains the predicate in the premise. In the above examples, the predicate in the premise contains the predicate in the conclusion: $\llbracket \text{walked} \rrbracket \supseteq \llbracket \text{walked rapidly} \rrbracket$, $\llbracket \text{was asleep} \rrbracket \supseteq \llbracket \text{was dreaming} \rrbracket$, and so on, for all M . Apparently it is true of the NPs in (39)–(42) that whenever a set belongs to the interpretation of the NP, so do all of its subsets. These NPs are ‘closed under inclusion’, which is another way of phrasing their downward monotonicity. The definition is as follows. First for quantifiers:

Definition 4

Q is *downward monotonic* in M iff for all $X, Y \subseteq E$: if $X \in Q$ and $X \supseteq Y$ then $Y \in Q$.

As before, we call an NP downward monotonic if its interpretation, whenever defined, is a downward monotonic quantifier:

Definition 5

NP is *downward monotonic* iff for all \mathbf{M} : if $\llbracket \text{NP} \rrbracket$ is defined in \mathbf{M} , then $\llbracket \text{NP} \rrbracket_{\mathbf{M}}$ is downward monotonic in \mathbf{M} .

Downward monotonic quantifiers and NPs are also called ‘monotonically decreasing’. We have a test for downward monotonicity similar to the one for upward monotonicity:

Downward monotonicity test 1

If $\llbracket \text{VP}_1 \rrbracket \supseteq \llbracket \text{VP}_2 \rrbracket$, then $\text{NP VP}_1 \models \text{NP VP}_2$

This test shows that, for instance, *all men* is not downward monotonic:

(43) All men walked $\not\models$ All men walked rapidly.

Exercise 5

(a) Which of the following NPs are downward monotonic?

- (i) *many* N
- (ii) *half of the* Ns
- (iii) *not John*
- (iv) *at most n* Ns
- (v) *exactly n* Ns
- (vi) *neither* N

(b) Do the properties of upward and downward monotonicity exclude one another?

An equivalent definition for downward monotonicity in terms of union is the following:

Definition 6

Q is *downward monotonic* in \mathbf{M} iff for all $X, Y \subseteq E$: if $X \cup Y \in Q$ then $X \in Q$ and $Y \in Q$

The corresponding test is:

Downward monotonicity test 2

$\text{NP VP}_1 \text{ or VP}_2 \models \text{NP VP}_1 \text{ and NP VP}_2$.

Semantically, the disjunction of two VPs is interpreted as union, while conjunction is interpreted as intersection:

$$\llbracket \text{VP}_1 \text{ or VP}_2 \rrbracket = \llbracket \text{VP}_1 \rrbracket \cup \llbracket \text{VP}_2 \rrbracket$$

Examples that illustrate the test are:

(43) Neither girl was drinking or smoking \models Neither girl was drinking and neither girl was smoking.

- (44) All boys sing or dance \neq All boys sing and all boys dance.

The set of all downward monotonic NPs is closed under disjunction and conjunction, just like its upward monotonic counterpart. The interpretation of a disjunction of NPs is of course the union of the interpretations of the constituent elements:

$$\llbracket \text{NP}_1 \text{ or } \text{NP}_2 \rrbracket = \llbracket \text{NP}_1 \rrbracket \cup \llbracket \text{NP}_2 \rrbracket$$

That downward monotonicity is preserved under union can be argued as follows: Let Q_1 and Q_2 both be downward monotonic. Take arbitrary X and Y such that $X \cup Y \in Q_1 \cup Q_2$. Then it holds that $X \cup Y \in Q_1$ or $X \cup Y \in Q_2$, and hence $X, Y \in Q_1$ or $X, Y \in Q_2$ since Q_1 and Q_2 are downward monotonic. Therefore in either case $X \in Q_1 \cup Q_2$ and $Y \in Q_1 \cup Q_2$.

Upward and Downward Monotonicity and the Negation of Quantifiers

We have observed that there is a special relation between upward and downward monotonicity. They mirror each other: upward monotonicity involves closure under extension, and on the other hand, downward monotonicity involves closure under inclusion. And not only do their definitions show this relation, but the various examples illustrate it as well. See the following table:

Table 7.2 Monotonic NPs

<i>Upward monotonic</i>	<i>Downward monotonic</i>
All N	Not all N
Many N	Few N
At least n N	At most n N
More than half of the N	Less than half of the N
John	Not John

These examples clearly show that downward monotonic NPs are negated upward monotonic NPs, sometimes literally, on the surface (*John* vs. *not John*), sometimes implicitly (*many* vs. *few*).

Usually two types of negation of quantifiers are distinguished, namely, *external* and *internal*:

Definition 7

The *external negation* $\neg Q$ of Q in \mathbf{M} is $\{X \subseteq E \mid X \notin Q\}$.

Definition 8

The *internal negation* Q^\neg of Q in \mathbf{M} is $\{X \subseteq E \mid (E - X) \in Q\}$.

According to an alternative but equivalent definition, the external negation $\neg Q$ of Q is obtained by taking the complement of Q with regard to the power set of the domain E :

Definition 9

$$\neg Q = \text{pow}(E) - Q.$$

This quantifier contains exactly those subsets of E (those elements of $\text{pow}(E)$) that are not in Q .

The internal negation of Q , $Q\neg$, can also be obtained by taking for every element of Q its complement with regard to the domain E :

Definition 10

$$Q\neg = \{Y \subseteq E \mid \text{there is an } X \in Q: Y = E - X\}.$$

This quantifier contains exactly those subsets of E whose complements with regard to E are in Q ; hence definition 10 is equivalent to definition 8.

Some examples. Recall the interpretation of *all N*, *an N*, *not all N*, and *no N* given in table 7.1 above. The external negation of *all N* is *not all N*, which can be shown as follows:

$$\begin{aligned} (45) \quad \neg \llbracket \text{all N} \rrbracket &= \{X \mid X \notin \llbracket \text{all N} \rrbracket\} \text{ (using definition 7)} \\ &= \{X \mid \llbracket N \rrbracket \cap X \neq \llbracket N \rrbracket\} = \llbracket \text{not all N} \rrbracket \end{aligned}$$

Its internal negation is *no N*:

$$\begin{aligned} (46) \quad \llbracket \text{all N} \rrbracket \neg &= \{X \mid (E - X) \in \llbracket \text{all N} \rrbracket\} \text{ (definition 8)} \\ &= \{X \mid \llbracket N \rrbracket \cap (E - X) = \llbracket N \rrbracket\} = \{X \mid \llbracket N \rrbracket \cap X = \emptyset\} = \llbracket \text{no N} \rrbracket. \end{aligned}$$

The external negation of *an N* is *no N*:

$$\begin{aligned} (47) \quad \neg \llbracket \text{an N} \rrbracket &= \{X \mid X \notin \llbracket \text{an N} \rrbracket\} \text{ (definition 7)} = \{X \mid \llbracket N \rrbracket \cap X = \emptyset\} \\ &= \llbracket \text{no N} \rrbracket. \end{aligned}$$

And its internal negation is *not all N*:

$$\begin{aligned} (48) \quad \llbracket \text{an N} \rrbracket \neg &= \{X \mid (E - X) \in \llbracket \text{an N} \rrbracket\} \text{ (definition 8)} \\ &= \{X \mid \llbracket N \rrbracket \cap (E - X) \neq \emptyset\} = \{X \mid \llbracket N \rrbracket \cap X \neq \llbracket N \rrbracket\} \\ &= \llbracket \text{not all N} \rrbracket. \end{aligned}$$

Both external and internal negation 'reverse' the monotonicity of a quantifier:

Fact 1

If Q is upward monotonic, then $\neg Q$ and $Q\neg$ are downward monotonic.

Fact 2

If Q is downward monotonic, then $\neg Q$ and $Q\neg$ are upward monotonic.

By way of illustration, we prove fact 1. Assume Q is upward monotonic. Take arbitrary X, Y such that $X \in \neg Q$ and $Y \subseteq X$. Now it follows that $X \notin Q$ and therefore $Y \notin Q$. For assume that $Y \in Q$. Then, since $Y \subseteq X$ and Q is upward monotonic, $X \in Q$, in contradiction with the assumption. Hence $Y \in \neg Q$ and

thus $\neg Q$ is downward monotonic. Similarly, assume that Q is upward monotonic, $X \in Q$, and $Y \subseteq X$. Then it follows that $(E - X) \in Q$. Since $Y \subseteq X$, it holds that $(E - X) \subseteq (E - Y)$, and hence (by upward monotonicity of Q) that $(E - Y) \in Q$, which implies that $Y \in Q$. This means that Q is upward monotonic. The proof of fact 2 proceeds analogously.

Both types of negation undo themselves:

Fact 3

$$\neg\neg Q = Q = Q\neg\neg$$

Not generally valid, however, is $\neg Q\neg = Q$. What does hold is that $\neg Q\neg$ is equivalent to the *dual* of Q , written as Q^* , which is defined as follows:

Definition 11

The *dual* Q^* of Q in \mathbf{M} is $\{X \subseteq E | (E - X) \notin Q\}$.

Some calculations with definition 11 and the interpretations of *an N* and *all N* will show that these are each other's duals.

There are also quantifiers which are equivalent to their own dual, the *self-dual* quantifiers. Proper names, for instance, are self-dual:

$$[\text{John}] = [\text{John}]^*$$

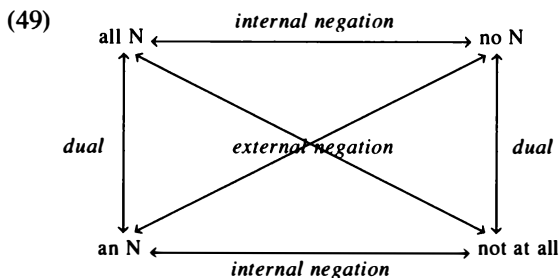
Since $Q^* = \neg Q\neg$, and both internal and external negation reverse the monotonicity of a quantifier, it follows that the dual of a quantifier has the same type of monotonicity as the quantifier itself:

Fact 4

If Q is upward (downward) monotonic, then Q^* is upward (downward) monotonic.

That this holds is easy to see: $Q^{**} = Q$.

Diagram (49) sums up the relationships between *all N*, *an N*, *not all N* and *no N*:



This is the 'Square of Opposition' of traditional logic.

Upward and downward monotonic NPs form two important classes of natural language NPs, which are related by negation. In general, most upward NPs

are unmarked, and most downward NPs are (implicitly or explicitly) negated upward NPs. But there is no reason for giving this negation a syntactic status, since the relation in question can be formulated in purely semantic terms.

The classification of NPs into upward and downward monotonic ones is not exhaustive; some NPs are neither. For instance, consider NPs of the form *exactly n N*:

(50) Exactly six boys walked rapidly \neq Exactly six boys walked.

(51) Exactly six boys were asleep \neq Exactly six boys were dreaming.

The first example shows that *exactly six boys* is not upward monotonic and the second that it is not downward monotonic either. Another instance of a non-monotonic NP is *a few N*:

(52) A few boys were dreaming \neq A few boys were asleep.

(53) A few boys walked \neq A few boys walked rapidly.

A final example is provided by NPs of the form *only NP*. Compare:

(54) Only John walked rapidly \neq Only John walked.

(55) Only John was asleep \neq Only John was dreaming.

(56) Only the men were dreaming \neq Only the men were asleep.

(57) Only the men walked \neq Only the men walked rapidly.

With regard to (55), it might be useful to keep in mind that *only John* means something like *John and nobody else*; a counterexample would be a situation in which John is the only one asleep and John does not dream.

Note that *only* as it occurs in (54)–(57) is not a determiner but an NP modifier. Further, it should be remarked that the determiner status of *few* and *many* is controversial. Some argue that these are adjectives. A similar analysis is proposed for numerals like *six* in *six girls*. In §7.2.5 we will take a closer look at this.

Exercise 6*

Is the interpretation of *exactly one boy* a nonmonotonic quantifier in every model?

Monotonicity and Semantic Universals

As a short detour we will take a brief look at two examples, both taken from Barwise and Cooper 1981, of attempts to apply the notions introduced above. Barwise and Cooper are interested in the formulation of *semantic universals*, properties of the meanings of natural language expressions, in our case NPs, which may be considered to hold for every natural language but which are not mere (logically or mathematically) necessary truths.

The first example of such a universal illustrates the importance of monotonicity, and it reads as follows:

Monotonicity constraint

In every natural language, noncompound NPs express monotonic quantifiers or conjunctions of monotonic quantifiers.

By 'noncompound' NPs we mean: proper names, NPs of the form 'simple determiner + N', and NPs such as *someone*, *everyone*, *nothing*. It should be clear that a constraint like this is not an a priori truth. There is no logical or mathematical law forbidding a simple NP to have the same meaning as the nonmonotonic (and compound) NP *an even number of men*. In other words, there is no logical reason why a natural language should not have a simple determiner with the meaning *an even number of*. If all languages satisfy the monotonicity constraint, and as far as we know they do (but see §7.2.5), it expresses a property of languages which on the one hand is not logically necessary but which on the other hand is universally valid for natural languages. And this is a significant contribution to the characterization of the notion of 'possible human language'.

The monotonicity constraint does not give any clue as to why this should be so. In general, this is true of all universals. A formulation of a universal property is one thing; the explanation for it is something else. In the case of the monotonicity constraint, Barwise and Cooper actually attempt such an explanation (which they do not undertake for other universals). The idea is that monotonic NPs are 'easier', i.e., that it is easier to verify or falsify sentences with monotonic NPs than sentences with nonmonotonic NPs (this topic will be discussed more generally in §7.2.5). Anyway, with or without an explanation, the simple *fact* (if indeed such it is) formulated by the monotonicity constraint is amazing. And the theory of generalized quantifiers provides us with the necessary tools for its formulation.

The second example concerns the relationship between monotonicity and NP conjunction. We have already observed that semantically, NP conjunction amounts to taking the intersection of the interpretations of the conjoined NPs. From this general perspective, there is no reason to doubt that every pair (i.e., every *n*-tuple) of NPs can be conjoined. However, there are apparently restrictions on NP conjunction in natural language. Compare the following examples:

- (58) a man and two women
all boys and Mary
Pete's father and many children

- (59) no man and few women
none of the girls and at most three boys
less than half of the children and not one adult

- (60) *a man and few women
 *John and no woman
 *two cellos and few violins

The NPs in the conjunctions in (58) are both upward monotonic, and those in the conjunctions in (59) are both downward monotonic. In (60) an attempt is made to conjoin NPs which have a different type of monotonicity, and the results are not well-formed. These examples and similar ones involving disjunction might lead to the conclusion that coordination by means of *and* and *or* is possible only when either both NPs are upward or both are downward monotonic. This restriction could be related to the fact that the properties of upward and downward monotonicity are preserved under intersection and union, whereas intersection and union of an upward and a downward monotonic quantifier normally do not result in a monotonic quantifier.

Exercise 7*

Show that the interpretation of *John and no woman* is not a monotonic quantifier.

It remains to be seen to what extent this last remark *explains* the restriction. As the monotonicity constraint indicates, natural language prefers monotonic NPs that are *simple*; but surely there are also nonmonotonic compound NPs. The fact that conjunction or disjunction of NPs with contrasting monotonicity results in a nonmonotonic quantifier provides no reason for the fact that these coordinated NPs are not well-formed.

Two more observations will suffice to show that the last word has not been said on the topic of monotonicity and coordination.

The first observation concerns the fact that a coordinated NP consisting of two NPs with contrasting monotonicity conjoined by *but* is actually well-formed:

- (61) many men but few women
 John but no woman
 many children but less than half of the adults

It even seems that *but* yields a well-formed NP only if it coordinates NPs with different monotonicity. Compare:

- (62) *some boys but two women
 *all boys but my sister
 *none of the girls but at most three boys

If we want to stick to the usual view, which goes back to Frege, that *but* is semantically (i.e., as far as truth conditions are concerned) equivalent to *and*, the examples in (62) cast doubt upon the explanation proposed above for the restriction on the coordination of NPs by means of *and*.

The second observation is directly concerned with the proposed restriction

itself. Not only should its explanation be regarded with suspicion, but the following examples raise doubts with respect to the phenomenon as such:

- (63) at most six girls and at least four boys
 Pete's father and a few women
 none of the boys and exactly one girl
 a few men and an even number of women

These are all well-formed NPs, but none consists of two NPs with equal monotonicity. In the first example, an upward monotonic NP and a downward monotonic NP are conjoined; in the second, an upward monotonic NP and a nonmonotonic NP; in the third, a downward monotonic NP and a nonmonotonic NP; and in the fourth, finally, two nonmonotonic NPs. In all cases, the result is a well-formed, nonmonotonic NP.

This phenomenon illustrates that the universal proposed by Barwise and Cooper, which claims that coordination of NPs is restricted to NPs with equal monotonicity, does not hold and has to be replaced by a more refined analysis. But it is still remarkable that the theory of generalized quantifiers allows us to formulate such falsifiable hypotheses.

Persistence and antipersistence

The properties of persistence and antipersistence that will be discussed now are, as will become apparent, closely related to the properties of upward and downward monotonicity. The main difference lies in the fact that persistence and antipersistence are properties, not of entire NPs but of determiners. This implies that we must look at things from a relational perspective (see §7.2.3). A determiner DET will be considered as an expression which takes an N and a VP to form an S. Semantically, a determiner interpretation D will be treated as a relation between sets. (From now on, we will be content to state definitions, facts, etc., for the semantic objects only, trusting that the reader will be able to furnish the corresponding definitions for the syntactic expressions.)

Definition 12

A determiner D is *persistent* iff for all X, Y, Z: if $D(X, Z)$ and $X \subseteq Y$ then $D(Y, Z)$.

Persistence is a property which relates to the first argument of a determiner relation. If we view determiners as linguistic expressions, it is a property adhering to the N to which the determiner is applied. This becomes clear when we convert the above definition to a test:

Persistence test

If $\llbracket N_1 \rrbracket \subseteq \llbracket N_2 \rrbracket$, then $\text{DET } N_1 \text{ VP} \models \text{DET } N_2 \text{ VP}$

A determiner is persistent if it is closed under the extension of its first argument, the N to which it is applied. A few examples:

- (64) Some men walked; men are human beings \models Some human beings walked.
- (65) At least four girls were smoking; girls are women \models At least four women were smoking.
- (66) All boys drink; boys are men $\not\models$ All men drink.

Some, *at least n* are persistent determiners; *all* is not a persistent determiner.

The mirror image of persistence is antipersistence:

Definition 13

D is *antipersistent* iff for all X, Y, Z : if $D(X, Z)$ and $Y \subseteq X$ then $D(Y, Z)$.

An antipersistent determiner is a determiner whose first argument is closed under inclusion. The corresponding test is of course the following:

Antipersistence test

If $\llbracket N_2 \rrbracket \subseteq \llbracket N_1 \rrbracket$ then $\text{DET } N_1 \text{ VP} \models \text{DET } N_2 \text{ VP}$

The following examples show that *all*, *no*, *at most n* are antipersistent determiners:

- (67) All children walked; toddlers are children \models All toddlers walked.
- (68) No woman was smoking; girls are women \models No girl was smoking.
- (69) At most three Englishmen agreed; Londoners are Englishmen \models At most three Londoners agreed.

Many determiners, including lexically simple ones, are neither persistent nor antipersistent: *many*, *few*, *the n*, *both*, *exactly n*, *more than half of*, *less than half of*. In this respect, the pair persistence/antipersistence differs remarkably from the pair upward/downward monotonicity.

Persistent and antipersistent determiners are linked by negation. Both external and internal negation of a determiner transform a persistent determiner into an antipersistent one, and vice versa; and hence, both external and internal negation of a nonpersistent determiner (i.e., a determiner which is neither persistent nor antipersistent) yield a nonpersistent determiner.

Exercise 8

Define internal and external negation of determiners and prove the assertions made above.

We have observed, and the definitions have illustrated that persistence and monotonicity are closely related. Persistence is upward monotonicity of the first argument of a determiner relation, and antipersistence is downward mono-

tonicity of the first argument. Analogously, the (upward or downward) monotonicity of an NP is (upward or downward) monotonicity of the second argument of the determiner in that NP. The usual terminology is: (upward or downward) ‘left-monotonicity’ and (upward or downward) ‘right-monotonicity’. Often the following notation is used: \uparrow mon, for upward left-monotonicity, mon \downarrow , for downward right-monotonicity and so on. In all, there are four possible combinations of these monotonicity properties of determiners. Examples of expressions which exhibit the four possible combinations are given in (70):

(70)

\uparrow mon	\uparrow	some	at least n	infinitely many
\downarrow mon	\uparrow	all		
\downarrow mon	\downarrow	no	at most n	a finite number of
\uparrow mon	\downarrow	not all		

The examples in the first and third rows may be regarded as climbing, from left to right, from the simple to the more general case. This is reflected in their linguistic form: the expressions in the first column are simple (except for *not all*), and those in the second and third columns are complex. Note also that in the first column we actually have the traditional determiners of the Aristotelian square of opposition. These observations illustrate once more that monotonicity and persistence are fundamental notions in the semantics of quantifying expressions. (Later, we will see what combination of properties will yield exactly the logical square).

A final observation about the central role of monotonicity is taken from van Benthem 1984b. It concerns the quantitative ‘ease’ with which a sentence of the form DET(A, B) can be falsified or verified. Consider table 7.3, which indicates, for a few examples of determiners, the number of elements that have to be checked in order to verify or falsify an assertion of the form DET(A, B). For instance, let $\llbracket \text{girl} \rrbracket$ contain six elements (i.e., $n = 6$). Then the assertion *Some girls are dancing* needs only one dancing girl in order to be verified, but for it to be falsified we must check all six elements in $\llbracket \text{girl} \rrbracket$. As another example, consider the sentence *At least three boys are smoking*. Suppose there are ten boys (i.e., $n = 10$). The assertion will be verified if we can find three smoking boys ($k = 3$). It will be falsified if we conclude that eight boys are not smoking, that is, $n - (k - 1) = 10 - (3 - 1) = 8$.

The number $n + 1$ turns out to be a provable minimum: for any determiner,

Table 7.3 Count Complexity of NPs

Determiner	Verification	Falsification	Total
All	n	1	$n + 1$
Some	1	n	$n + 1$
At least k	k	$n - (k - 1)$	$n + 1$
At most k	$n - k$	$k + 1$	$n + 1$
Exactly k	n	$k + 1$	$n + (k + 1)$

the sum of the numbers of verifying and falsifying instances is larger than or equal to $n + 1$. Also, it can be proved that all determiners which have 'minimal count complexity', i.e., for which this number is equal to $n + 1$, are (upward or downward) monotonic. This may serve to emphasize again the fundamental role of monotonicity with respect to natural language determiners and quantifiers.

7.2.5 Global Constraints

Introduction

We have mentioned several objectives of the theory of generalized quantifiers, including the search for constraints on determiners and quantifiers and the search for characterizations of specific classes of such expressions, which might be interesting for other, unrelated reasons. In the following we will discuss some global constraints that have been proposed in the literature and outline how they can be used to characterize the class of logical determiners (*all*, *some*, *not all*, *no*). Our exposition of this subject will be based mainly on the work of van Benthem (1983a, 1984a).

The search for global constraints does not remain restricted to just determiners and quantifiers. For it holds quite generally for all but a few simple types that the set of all semantic objects of a given type is 'too large' in the sense that natural language expresses only (a sometimes very small) part of it. This is particularly striking if we count only the lexically realized expressions but it still holds if we also take into account complex expressions. Besides, in almost every type there are rather 'wild' specimens, ill-behaved semantic objects which would never be classified as meanings of natural language expressions. This is also a reason for finding out whether there are perhaps global constraints on the entire class of objects of a certain type that help to reduce them to a smaller, preferably 'well-behaved' subset.

The hunt for global constraints is therefore not restricted to determiners and quantifiers. In fact, the meaning postulates familiar from Montague grammar can be interpreted as kinds of global constraints too (at least those which apply to a class of expressions).

Conservativity

Our perspective continues to be relational. The first global constraint we consider has to do with a property of determiners which is called *conservativity*:

Definition 14

D is *conservative* iff for all X, Y : $D(X, Y)$ iff $D(X, X \cap Y)$.

Definition 14 states that in order to verify or falsify an assertion of the form $\text{DET}(A, B)$, it is sufficient to look at the interpretation of A and the intersection of the interpretations of A and B . In other words, only what is in

$([A] - [B]) \cup ([A] \cap [B])$ is relevant; the contents of $[B] - [A]$ are not relevant, and neither is what falls outside $[A]$ and $[B]$, i.e., $E - ([A] \cup [B])$.

A related notion is the 'live on' property of quantifiers defined by Barwise and Cooper:

Definition 15

Q lives on X iff for all Y : $Y \in Q$ iff $X \cap Y \in Q$.

A quantifier can live on several sets; *every man*, for instance, lives on E (the domain) and on $[\text{man}]$. Intuitively, the latter is an interesting set: it is the natural restriction on the determiner *every* in the NP *every man*. According to Barwise and Cooper, the central part played by restriction on quantification in natural language ('all quantification in natural language is restricted quantification') is expressed in the following universal:

In every natural language, *simple* determiners together with an N yield an NP which lives on $[N]$.

According to this universal, it would hold for all simple determiners that $\text{DET}(A)$ lives on $[A]$, i.e., that $[\text{DET}]$ is a conservative determiner. Note that this universal is a strong constraint on the kind of relation that a simple determiner may express. There are certainly nonconservative determiners, and there is no logical reason why these could not be lexicalized in a language by means of a simple expression.

Apparent exceptions are *only* and *many*. As usual, a test can be derived from definition 14. (Here, we use ' \Leftrightarrow ' to denote mutual entailment between natural language sentences. I.e., ' $A \Leftrightarrow B$ ' means ' $A \models B$ and $B \models A$ '.)

Conservativity test

$\text{DET } N \text{ VP} \Leftrightarrow \text{DET } N \text{ are } N \text{ that VP}$

Compare the following examples:

(71) All boys walked \Leftrightarrow All boys are boys that walked.

(72) Some girls are dancing \Leftrightarrow Some girls are girls that are dancing.

(73) Only men smoke cigars \nLeftrightarrow Only men are men that smoke cigars.

It is quite clear that *only* is not a conservative determiner: *only men* does not live on $[\text{man}]$. The obvious way to save the universal is to regard *only* not as a determiner but as an NP modifier (i.e., as an expression of type NP/NP). Constructions like those in (74) seem to support this view; hence the analysis of *only men* in (75), in which Δ represents a morphologically 'null' plural determiner, seems reasonable. (Compare the discussion in the section on monotonicity in §7.2.4.)

- (74) only John
 only the neighbor
 only a few girls

- (75) $NP[NP/NP[only]NP[DET[\Delta]N[men]]]$

This takes care of *only*, but what about *many*?

Exercise 9

Give an interpretation of *many* which is not conservative.

Although exercise 9 has a solution, there might be reasons for doubting the determiner status of *many*. For *many*, just as for *few*, an analysis as an adjective (N/N) seems plausible. First, both may occur in prenominal position, preceded by a determiner (*the many mistakes*, *the few results*). Second, both have comparative and superlative forms (*more/fewer*, *most/fewest*). Third, they may occur as predicates (*The boys are many/few*). Finally, they may occur in constructions like *as ADJ as* (compare: *as many as/as big as*).

Whether we deal with these two exceptions in this or some other way, it seems safe to say that the universal proposed by Barwise and Cooper holds, at least for English.

But note that if we impose conservativity as such as a global constraint on the meanings of natural language determiners, what we get is considerably stronger. Conservativity as a global constraint amounts to the assertion that *all* determiners, both simple and complex, are conservative. Counterexamples to this much stronger claim occur in the form of 'intensional determiners'. Assuming that new determiners can be formed by adjectival modification of existing ones, the following are examples of determiners:

- (76) all
 all red
- (77) some
 some living
- (78) all
 all supposed

A little reflection shows that not only the simple determiners in (76) and (77) are conservative, but also the complex, restricted ones: adjectival restriction by means of intersective, extensional adjectives preserves conservativity. But (79) shows that this is not true of restriction by means of intensional adjectives as in (78):

- (79) All supposed women are men \Leftrightarrow All supposed women are women that are men.

These cases are usually excluded by means of the following reasoning: the

theory of generalized quantifiers was extensional anyway, and if we want to intensionalize it, we will have to modify the notion of conservativity.

At first sight this may seem a rather strange reaction, the more so since there is such an obvious alternative: (80) is not to be regarded as the constituent structure of the NP in (79), but (81):

(80) $NP[DET[DET[all]_{DET/DET}[supposed]]_N[women]]$

(81) $NP[DET[all]]_N[N/N[supposed]]_N[women]]$

Then we would have, instead of (79), the unobjectionable (82):

(82) All supposed women are men \Leftrightarrow All supposed women are supposed women who are men.

Of course, if we choose to analyze the NP in (79) along the lines of (81), we must do so generally, i.e., not only in the case of intensional adjectives, but also in the case of extensional ones. In other words, we conjecture that there is simply no such thing as adjectival restriction of determiners; adjectives are noun modifiers of category N/N.

But there are also good reasons for choosing not to save the universality of conservativity in this way. For if we disregard the possibility of adjectival restriction with intensional adjectives, we can prove the following (cf. Keenan and Stavi 1986, which contains a thorough discussion of conservativity):

Fact 5

The class of all conservative determiners is exactly the class of determiners generated by (i) *all* and *some*, (ii) Boolean combinations, and (iii) extensional adjectival restriction.

Exercise 10*

Show that Boolean operations and extensional restriction preserve conservativity.

Because of fact 5, it is attractive to retain adjectival restriction of determiners as a syntactic process: for then natural language can be said to be 'expressively complete' vis-à-vis possible determiner denotations.

A final potential counterexample against conservativity as a universal property of all natural language determiners is *all and only*. Compare:

(83) All and only boys skate \Leftrightarrow All and only boys are boys that skate.

Note that the right side sentence of (83) is equivalent to (84):

(84) All boys are boys that skate and only boys are boys that skate.

The second conjunct of (84) is a tautology, and therefore (84) is equivalent to (85):

(85) All boys are boys that skate.

And surely the meaning of (85) is different from the meaning of the left side sentence of (83). This counterexample can be disputed, however: its validity depends on whether *all and only*, as it occurs in *all and only boys*, has the status of a determiner. Earlier we observed that there might be reasons for giving *only* the status of NP modifier. However, *all and only* does not fit into that analysis easily. In general, it is assumed that coordination is not (very) 'cross-categorical': coordination is possible only between expressions which belong to the same (main) category. This would seem to argue for the determiner status of *only* after all, at least as it occurs in this construction.

We have paid a good deal of attention to conservativity, because it is the most important, most powerful global constraint proposed in the literature.

Variety

A simple, intuitively plausible global constraint on determiners requires them to have the property of *variety*:

Definition 15

D shows *variety* iff there are X, Y such that $D(X, Y)$ and there are X, Y such that $\neg D(X, Y)$.

Imposing the constraint that determiners must have this property excludes 'uninteresting' determiners which are either always or never true: only contingent relations are under consideration.

All simple determiners have this property. Apparent exceptions are *at least n* in a model with a domain of cardinality $< n$. But in such a case, we decided earlier, the interpretation of the determiner is undefined (see §7.2.4). If we keep this condition in mind ("in every model where DET is defined . . ."), all simple determiners have the property of variety.

Determiners which do not show variety are Boolean combinations of certain determiners, such as *one or no* (which holds of every pair $\langle X, Y \rangle$), or *at least four and at most three* (which holds of no pair $\langle X, Y \rangle$). The existence of this type of determiner sheds a different light on the status of variety as a global constraint. Surely determiners such as these are not very useful: in this sense they are not 'meaningful' expressions. On the other hand, they exist, and they have a meaning. Therefore we cannot regard variety as a constraint which excludes only determiner relations which are 'unnatural' in the sense that they are not expressed in natural language.

Continuity

In the section on applications, §7.2.4, we discussed the monotonicity constraint, which states that all simple natural language NPs express monotonic quantifiers or conjunctions of them.

An argument that not all simple determiners are monotonic can be derived from an example like *one* (the numeral), meaning *exactly one* (the other meaning, *at least one*, is $\text{mon } \uparrow$). This is not a monotonic determiner, as the following examples show

(86) One boy was dreaming \neq One boy was asleep.

(87) One boy was asleep \neq One boy was dreaming.

It should be added that the status of *one* as a determiner is not uncontroversial. On the other hand, if one accepts it as a determiner, monotonicity as such may not be a global constraint. In that case, it seems necessary to formulate a weaker property which characterizes exactly those determiners which are either monotonic or a conjunction of monotonic determiners. This weaker property is continuity:

Definition 16

D is *continuous* iff for all X, Y_1, Y_2, Y : if $D(X, Y_1)$ and $D(X, Y_2)$ and $Y_1 \subseteq Y \subseteq Y_2$ then $D(X, Y)$.

In Thijsse 1983, where continuity is proposed as the relevant property, we find the following fact:

Fact 6

The set of all continuous determiners is exactly the set of all monotonic determiners and their conjunctions.

Exercise 11*

Show that *all but one* can be taken as a conjunction of monotonic determiners and that *an even number of* cannot.

Extension

The last global constraint proposed and defended in the literature that we want to consider here concerns a form of context independence:

Definition 17

D has *extension* iff for all X, Y, E, E' : if $E \subseteq E'$ and $D(X, Y)$ in E then $D(X, Y)$ in E' .

Determiners which have extension are context independent, in the sense that extension of the number of elements in the domain does not make any difference to their interpretation. These are determiners which do not refer to the cardinality of the domain. An instance of an interpretation of a determiner which does not have extension is the following:

$$[[\text{many}]] = \{ \langle X, Y \rangle \mid \frac{\text{card}(X \cap Y)}{\text{card}(X)} > \frac{\text{card}(Y)}{\text{card}(E)} \}$$

In this interpretation, *many* means approximately 'relatively many', 'many in comparison with the entire domain'. This interpretation is essentially context dependent: if the cardinality of the domain increases or decreases, the determiner may hold of pairs other than those it held for before.

7.2.6 Logical Determiners

In this last section we will briefly discuss the set of constraints on determiners that yields exactly the class of logical determiners (*all*, *some*, *no*, *not all*). Our exposition is founded on van Benthem 1983a, 1984a.

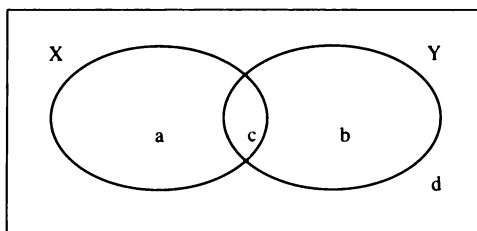
Unlike those discussed above, these constraints are not (all) global constraints, which express intuitive and universal properties of the meanings of natural language determiners. Rather, we are concerned here with principles which in a sense characterize the contents of these logical determiners, which explicate what it is to be a *logical* determiner. Of course the constraints discussed above still play a part; they stake out the field.

The Tree of Numbers

In this section, we introduce a constraint on determiners which, in combination with conservativity and extension, makes it possible to represent determiners which satisfy these constraints in a very simple and transparent way. This method of representation, in the form of patterns which determiners assign to a 'tree of numbers', gives a clear insight into the character of the various determiners and enables us to specify more exactly what distinguishes a logical determiner from a nonlogical determiner.

The constraint in question concerns the *quantitative* character of certain determiners. Consider (88):

(88)



$$\begin{aligned} a &= \text{card}(X - Y) \\ b &= \text{card}(Y - X) \\ c &= \text{card}(X \cap Y) \\ d &= \text{card}(E - (X \cup Y)) \end{aligned}$$

With the help of the numbers, *a*, *b*, *c*, and *d*, thus defined, we can define the property of quantitativeness. A quantitative determiner is a determiner which is not sensitive to the properties of, and the relations between, the elements in the domain and in the sets which it relates:

Table 7.4 Quantitative Determiners

Determiner	Definition
All	$a = 0$
Some	$c \neq 0$
No	$c = 0$
Most	$c > a$
Many	$\frac{c}{c + a} > \frac{c + b}{c + a + b + d}$

The interpretation of *Many* is the context-dependent interpretation given in §7.2.5.

Definition 18

D is *quantitative* iff for all X, Y: D(X, Y) depends only on a, b, c, and d

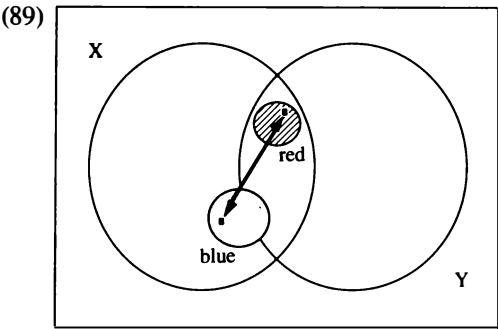
Thus, quantitative determiners can be defined in terms of a, b, c, and d only. Some examples are given in table 7.4.

Not all determiners are quantitative. Some depend on more than just the number of elements concerned: for instance they may be sensitive to the properties of (some of) these elements or to the relationships between them. Two instances of nonquantitative determiners which have been much discussed are adjectivally restricted determiners, such as *all red*, and possessive determiners, such as *Mary's*. The following equivalent definition of quantitativity might be useful for settling whether determiners are quantitative or not:

Definition 19

D is *quantitative* iff for every permutation π of E: D(X, Y) iff D($\pi(X)$, $\pi(Y)$)

A permutation of E does not affect the relevant numbers a, b, c, and d, but it may affect the properties of some of the elements in the domain and their relations with other elements. A quantitative determiner is insensitive to such permutations, but *all red*, for instance, is not:



In this model **M** it is true that *all red*(X, Y), and it is not true that *all blue*(X, Y), before the permutation π . Now if we substitute an element in the red X's

for an element of the blue X's, quantitatively nothing has changed; π results in a model M' in which a , b , c , and d have the same value as in M . But in M' it is no longer true that *all red*(X , Y): the red X's are no longer only in the shaded part (minus the blue intruder), but there is also a red element outside Y . So the determiner *all red* is sensitive to things other than just the number of elements; their nature (in this case their being red or not) is relevant too. This is of course exactly what adjectival restriction with *red* of the simple determiner *all* was supposed to achieve.

A similar example shows that a possessive determiner like *Mary's* is not quantitative either. In this case, the relation of possession is relevant to the validity of assertions of the form *Mary's*(X , Y). This relation is not preserved under permutations of E .

Cases like this can be covered by a somewhat more subtle notion, called 'quality'. A domain has a certain structure, and we may restrict ourselves to those permutations which preserve that structure. Determiners like *all red* and *Mary's* supply the information about the relevant structure themselves: *all red* is qualitative with respect to π such that $x \in \llbracket \text{red} \rrbracket \Leftrightarrow \pi(x) \in \llbracket \text{red} \rrbracket$, and so on.

Quantitativity is very powerful notion, and combined with conservativity and extension it yields the following result. If a determiner is quantitative, we know that only a , b , c , and d are relevant for its interpretation. If the determiner satisfies extension as well, d is no longer relevant. If it is also conservative, b does not play a part either. In other words, the interpretation of all determiners which are quantitative, conservative, and context independent with respect to E can be formulated solely in terms of the numbers a and c . Their meaning is fully specified by just stating what they yield for every pair of numbers (a , c): true or false. In other words, a quantitative, conservative determiner that satisfies extension can be regarded as an assignment of + or - to all pairs of numbers (a , c). The resulting representation uses the 'tree of numbers':

card(X) = 0				0, 0
= 1		1, 0		0, 1
= 2	2, 0		1, 1	0, 2
= 3	3, 0	2, 1	1, 2	0, 3
and so on.				

If X has no element, then neither have $X - Y$ and $X \cap Y$. Hence $a = c = 0:0, 0$. If the cardinality of X is 1, there are two possibilities for a and c : the one element belongs to $X - Y$, and therefore not to $X \cap Y$: 1, 0, or vice versa: 0, 1. If X has two members we get three distinct possibilities for a and c : both elements belong to $X - Y$, and hence $X \cap Y$ is empty: 2, 0; there is one ele-

ment in $X - Y$ and one in $X \cap Y$: 1, 1; and both elements may belong to $X \cap Y$, while $X - Y$ is empty: 0, 2. In this way the entire tree is constructed.

In terms of the tree of numbers we can characterize the meaning of a determiner which is quantitative, conservative, and which satisfies extension simply by stating to what *part* of the tree it will assign a plus, i.e., for what pairs of numbers (a, c) it is true. For instance, *all* is true on the right branch of the tree, *no* is true on the left side; *the* is true in (0, 1) and nowhere else; and so on.

We can also characterize in terms of the tree the various monotonicity concepts: a determiner is \uparrow mon if on the horizontal line to the right of any plus there are only pluses. Analogously, a determiner is \downarrow mon if this is true of the line to the left of any plus. For a determiner which is \uparrow mon, it holds that if it is plus on a certain point, it is plus on the downward triangle of which that point is the apex; the same for \downarrow mon, but with an upward triangle.

For many purposes, the tree of numbers is a useful tool, and it is often used in the literature.

Characterization of the Logical Determiners

What additional properties distinguish the logical determiners within the class of determiners which satisfy quantitativeness, conservativeness, and extension? We now present one analysis to this effect, which uncovers some semantic notions which may also be of independent interest.

The first two properties we have already discussed: they are continuity and variety. Logical determiners are \uparrow mon or \downarrow mon and therefore continuous; they also satisfy variety.

The two additional properties that we need both concern a kind of regularity in the behavior of determiners and a relative independence of specific numbers. These two properties appear to be fundamental characteristics of logical determiners (and of logical concepts in general).

The first of these two properties is the following. (Here, ' $D(a, c)$ ', etc. means that D assigns a plus to the pair a, c ; and ' $\neg D(a, c)$ ' etc., means that D assigns a minus to the pair a, c .)

Definition 20

D has the *plus property* iff: If $D(a, c)$, then $D(a + 1, c)$ or $D(a, c + 1)$; if $\neg D(a, c)$, then $\neg D(a + 1, c)$ or $\neg D(a, c + 1)$.

This definition of the 'plus' property states that D has no 'dead ends': if it assigns a certain truth value, then it must be possible to preserve this truth value if we add an element to X . In particular, the specific number of elements in X does not influence the behavior of D , whereas it is essential for a determiner like *the* n .

The second additional property is rather complicated to formulate exactly and generally (and there are several alternative formulations around). This

property, called *uniformity* or *homogeneity*, makes the truth value pattern of D 'smooth' by precluding those determiners that show different 'jump patterns' in the tree. For instance, if a determiner is to be uniform and shows truth value pattern *a* somewhere in the tree, it is not allowed to have another pattern, say *b*, somewhere else in the tree:

$$\begin{array}{cc} \text{a.} & + & & \text{b.} & + \\ & - & + & & + & - \end{array}$$

This informal characterization is used in the following definition:

Definition 21

D is uniform iff D shows only one truth value pattern.

If, as is done in van Benthem 1987, we consider quantifiers to be 'semantic automata' which calculate a truth value when given a pair of sets as input, uniformity amounts merely to a determiner always embodying the same procedure.

Given these two additional properties, it becomes clear how the logical determiners result.

Consider the apex of the tree of numbers, where $\text{card}(X) = 0$ or 1:

$$\begin{array}{cc} \text{card}(X) = 0 & 0, 0 \\ & = 1 \quad 1, 0 \quad 0, 1 \end{array}$$

A determiner prints a pattern of pluses and/or minuses on it. There are eight possible patterns:

$$\begin{array}{cccc} 1. & + & 2. & + \\ & + & & + & 3. & + & 4. & + \\ & + & & - & & - & + & - \\ 5. & - & 6. & - & 7. & - & 8. & - \\ & + & & + & & - & + & - \\ & + & & - & & - & + & - \end{array}$$

The property of variety precludes 1 and 8: the second row should have both plus and minus occurring. The plus property eliminates 4 and 5: the plus and minus, respectively, in the first row should re-occur in the second row. Four patterns, i.e., four determiners, remain:

$$\begin{array}{cc} 2. & + & 3. & + \\ & + & - & - & + \\ 6. & - & 7. & - \\ & + & - & - & + \end{array}$$

And in fact, these are the patterns made by *no*, *all*, *not all*, and *some*, respectively. To show that they are, we must show that these patterns persist in the entire tree in the appropriate fashion. That they do can be seen by reflecting on uniformity and continuity. Uniformity assures us that we get the same pattern

everywhere in the tree. This implies, for instance, that the first four horizontal lines of the tree for *all* look like this:

$$\begin{array}{ccccccc}
 \text{card}(X) = 0 & & & & & & + \\
 = 1 & & & - & & & + \\
 = 2 & & & & - & & + \\
 = 3 & \dots & & & & - & +
 \end{array}$$

The pattern of *all*, given above as 3, is: 'under a plus we find to the left a minus and to the right a plus'. This gives pluses on the right branch, with minuses immediately to the left. The other positions in the horizontal lines are determined by continuity: it says that there can only be minuses further to the left (for between two pluses, continuity allows only plus):

$$\begin{array}{ccccccc}
 \text{card}(X) = 0 & & & & & & + \\
 = 1 & & & - & & & + \\
 = 2 & & - & & - & & + \\
 = 3 & - & & - & & - & +
 \end{array}$$

So because of uniformity and continuity, the pattern 3 made on the apex of the tree of numbers can be expanded only into the pattern made by the determiner *all*.

Let us consider another example, *some*. The tree for $\text{card}(X) = 0, 1, 2$, and 3 looks like this:

$$\begin{array}{ccccccc}
 \text{card}(X) = 0 & & & & & & - \\
 = 1 & & & - & & & + \\
 = 2 & & - & & & & + \\
 = 3 & - & & + & & \dots &
 \end{array}$$

In view of uniformity, the pattern 7 is unique, and it fills the left branch of the tree and the immediately adjoining positions. The remainder is again completed by continuity and consists only of pluses:

$$\begin{array}{ccccccc}
 \text{card}(X) = 0 & & & & & & - \\
 = 1 & & & - & & & + \\
 = 2 & & - & & + & & + \\
 = 3 & - & & + & & + & +
 \end{array}$$

(Note that having a third row of the form $- + -$ would lead to a fourth row violating continuity.) By the same kind of reasoning, it can be shown that the patterns 2 and 6 can be expanded only into those of *no* and *not all*, respectively.

This of course is only a sketch of a proof. The literature mentioned above should be consulted for more details. There one can also find results concerning the effect of weakening or deleting some of the properties involved.

Exercise 12

What class of determiners results if we leave out the property of continuity?

7.2.7 Further Developments

It must be emphasized that the survey in the preceding sections gives only a first, superficial view of the field and of its main concepts and principles.

For instance, much important empirical research has been done which we have not mentioned. Ter Meulen 1983; van Benthem and ter Meulen 1984; Groenendijk, de Jongh, and Stokhof 1987b; and Gärdenfors 1988 are collections in which to look for such work. And we have paid no attention to the investigation of determiners by means of concepts from the theory of relations (see Zwarts 1983; and especially van Benthem 1984a, which studies the connection between the latter approach and the one by means of global constraints). And the research on questions of 'expressibility' has not been dealt with either (see, e.g., Keenan and Moss 1984; Thijsse 1984; Keenan 1987).

Another omission is that no attention has been paid to the conditions which have to be satisfied in order to incorporate the theory of generalized quantifiers into a grammar. It is obvious that certain conditions must be imposed on the semantic component of such a grammar, but the syntactic component will also have to satisfy some requirements. For some discussion on this subject, see Zwarts 1986; van Benthem 1986. The work of Keenan and Faltz (1985) should be mentioned in this context as well, since it tries to transfer the concept of Boolean structure, as we observed in the domain of NP interpretations, to other components of the grammar.

7.3 Flexible Categorical Grammar and Type Theory

In recent years, there have been some interesting developments in research on categorical grammar, partly inspired by developments in Montague grammar (see chapter 6). We shall discuss some aspects of this progress, because it involves some further links with type theory.

7.3.1 Category Change

Several of the objections against classical categorical syntax mentioned in chapter 4 concern the rigidity of the assignment of categories to expressions. Natural language is rather flexible in its behavior in categorical combinations. For instance, the negation *not*, which is usually classified as *s/s*, occurs not only as sentence negation (*It is not the case that Archibald cries*), but also as

predicate negation (*Most babies don't cry*), NP negation (*Not every baby cries*), adverbial negation (*not unkindly*), and so on. As far back as 1972, Peter Geach proposed to account for this phenomenon by introducing category change rules which operate on the basic category assigned to an expression and then produce successive further admissible categories.

In order to avoid an excess of slashes, we will use another notation than we did before:

(a, b): 'from category a to category b'

This notation is to be interpreted nondirectionally, i.e., it carries no information about which side a functor takes its argument from. If required, the directional aspect can be introduced at a later stage. (Note, however, that there are also more principled linguistic arguments for a nondirectional approach; cf. Hoeksema 1984.)

In this notation, Geach's category-change rule reads like this:

If an expression has category (a, b), then it also has category ((c, a), (c, b)), for all categories c.

For instance, sentence negation (s, s) may also occur as predicate negation ((n, s), (n, s)), or when Geach's rule is used repeatedly, as ((n, (n, s)), (n, (n, s))) for negation of a transitive verb. Another application of this mechanism concerns the categorial analysis of transitive verbs that take complex NPs in direct object position. An expression like *sings every ballad* yields the following categories:

sings	every ballad
(n, (n, s))	((n, s), s)

These categories cannot be combined by functional application to the desired final category, that is, (n, s). The Geach rule provides an instant solution: ((n, s), s) is changed into ((n, (n, s)), (n, s)), and now functional application suffices to yield the desired result.

Another, in fact equivalent, way of describing what happens here is as an increase in the possibilities of categorial combination. Besides the method of functional application (consisting of two rules, given our present nondirectional view):

$$a + (a, b) \Rightarrow b \text{ ('a combined with (a, b) yields b')}$$

$$(a, b) + a \Rightarrow b$$

we also admit *functional composition*:

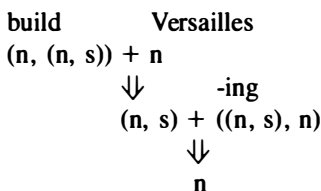
$$(a, b) + (b, c) \Rightarrow (a, c) \text{ ('(a, b) combined with (b, c) yields (a, c)')}$$

$$(b, c) + (a, b) \Rightarrow (a, c)$$

To see that this amounts to the same thing, note that with $a = n$, $b = (n, s)$,

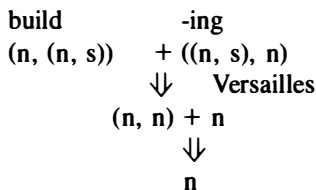
and $c = s$, the above verb phrase derivation becomes an instance of the first composition rule.

Many linguists have (re-)discovered the Geach rule as a descriptive tool. We will mention another example, this time of a morphological character, which is featured in Moortgat 1988 and Hoeksema 1984. Verbs can be nominalized, as in *Plumbing is a profitable activity*. It seems natural to categorize this nominalization as $((n, s), n)$, in other words: a property becomes an object. But this gives rise to a problem with an expression like *building Versailles*, where the nominalized verb *building* takes a direct object. One way of explaining this would be the following analysis.



Here we first combine *build* as a transitive verb with its direct object *Versailles*, nominalizing the result by combining it with the particle ending *-ing*. The problem is, of course, to get the right morphological form, i.e., to get the particle on the verb.

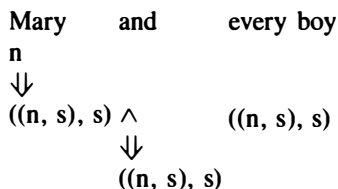
The following analysis, which uses the Geach rule, would therefore be more natural from a morphological point of view:



Several other type change rules have been proposed in recent years. One example is the 'Montague rule':

from category a to category $((a, b), b)$, for every category b

This principle accounts for phenomena of coordination, as in



Compare the treatment of terms and proper names in §6.3.4. Other forms of flexibility in categorial grammar can be found, for instance, in Partee and Rooth 1983; van Eyck 1985; Groenendijk and Stokhof 1984, 1988a.

7.3.2 A Logical Point of View

Not every transition between categories should be considered a well-motivated type change rule. As a matter of fact, the examples given above show a clearly defined pattern. This was observed early by Lambek (1958), in which an analogy was drawn with logical *implications*. In many respects, a functional type (a, b) behaves as an implication $a \rightarrow b$. This analogy provides an explanation of the above type changes in terms of logical entailments between implicational formulas:

$$a \rightarrow b \models (c \rightarrow a) \rightarrow (c \rightarrow b) \text{ (Geach rule)}$$

$$a \models (a \rightarrow b) \rightarrow b \text{ (Montague rule)}$$

This test fits other kinds of flexibility as well. For instance, Partee and Rooth use 'argument lowering':

$$(((a, b), b), c) \Rightarrow (a, c)$$

which is also valid as an implicational law:

$$((a \rightarrow b) \rightarrow b) \rightarrow c \models a \rightarrow c$$

These valid transitions can be described by an *implicational logic*, as Lambek did. And for this purpose, natural deduction, as presented in volume 1, chapter 4, turns out to be quite useful. The 'Lambek calculus' can be described as an *intuitionistic* implicational logic, with several additional restrictions on the 'bookkeeping' rules for assumptions which are used in derivations.

Example. A derivation of the Geach rule

1.	$a \rightarrow b$	assumption
2.	$c \rightarrow a$	assumption
3.	c	assumption
4.	a	$E \rightarrow (2, 3)$
5.	b	$E \rightarrow (4, 1)$
6.	$c \rightarrow b$	$I \rightarrow$
7.	$((c \rightarrow a) \rightarrow (c \rightarrow b))$	$I \rightarrow$

Not all implication laws from the system in volume 1 are admissible here. For example, $b \rightarrow a$ is not derivable from a . For the Lambek calculus only allows the withdrawal of actually *used* assumptions. And indeed, in natural language, transitions like $s \Rightarrow (n, s)$ ('a sentence becomes an intransitive verb') do not seem to occur.

Even so, there exist various defensible logical *options* for a reasonable category change system. For instance, the strictest variant does not validate:

$$a \rightarrow (a \rightarrow b) \vdash a \rightarrow b$$

The reason for this is that *multiple* use of the same assumption (which would be needed in this derivation) is not allowed in the Lambek calculus either. One might argue that this pattern does occur occasionally in natural language, for example, a transition like the following:

$$\begin{array}{ccc} \text{wash} & \Rightarrow & \text{wash oneself} \\ (n, (n, s)) & & (n, s) \end{array}$$

But there seems to be no general linguistic license to drop arguments in this way.

The general picture then becomes this. Below the intuitionistic or even minimal conditional logic of volume 1, there lies a spectrum of weaker implicational logics which can serve as ‘categorical engines’ for category change. One interesting system of this kind is the Lambek calculus, which allows the withdrawal of only *one* occurrence of an assumption in the introduction rule for implication. That is, it is a logic of *occurrences* of premises. But for certain applications, it is also wise to study stronger logics which allow multiple use of assumptions.

A more systematic way of viewing such options for category change is related to the following rather obvious question. On the analysis just presented, category transitions taken as implications show a nice syntactic, proof-theoretic pattern; but what is their semantic *meaning*?

This question is easy to answer in specific cases. For instance, the Geach rule is attractive precisely because of the underlying natural ‘recipe’ for converting a meaning in category (a, b) to one in category $((c, a), (c, b))$:

$$\text{from } M_{(a,b)} \text{ to } \lambda y_{(c,a)} \lambda z_c [M_{(a,b)}(y_{(c,a)}(z_c))]$$

Note how the λ -operator introduced in chapter 4 plays a key role here.

Analogously, here is the recipe for the Montague rule:

$$\text{from } M_a \text{ to } \lambda y_{(a,b)} [y_{(a,b)}(M_a)]$$

Evaluation of entire expressions then proceeds as follows, by an interplay of category change and ordinary functional application:

$$\begin{array}{ll} \text{Mary sings} & \text{every ballad} \\ n & (n, (n, s)) \\ A_n & B_{(n,(n,s))} \end{array} \quad \begin{array}{l} ((n, s), s) \\ C_{((n,s),s)} \\ \Downarrow \\ C_{((n,(n,s)),(n,s))} \\ \lambda x_{(n,(n,s))} \lambda y_n [C_{((n,s),s)}(x_{(n,(n,s))}(y_n))] \end{array}$$

$$\begin{array}{l}
+ \\
\Downarrow \\
(n, s) \\
\lambda x_{(n,(n,s))} \lambda y_n [C_{((n,s),s)}(x_{(n,(n,s))}(y_n))](B_{(n,(n,s))}) \\
\text{which reduces by } \lambda\text{-conversion to:} \\
\lambda y_n [C_{((n,s),s)}(B_{(n,(n,s))}(y_n))] \\
+ \\
\Downarrow \\
s \\
\lambda y_n [C_{((n,s),s)}(B_{(n,(n,s))}(y_n))](A_n) \\
\text{which reduces to the intended reading: } C_{((n,s),s)}(B_{(n,(n,s))}(A_n))
\end{array}$$

We could impose the general restriction on ‘reasonable’ category change rules that they must have such a type-theoretic explanation. Now some general logical observations can be made on this subject (see van Benthem 1986, chap. 7). All derivations in the Lambek calculus can be systematically provided with such type-theoretic terms: the corresponding category transitions are correct in the sense just stated. There are also converse results, indicating how type-theoretic terms can be effectively associated with implicational derivations. The general situation here is diverse. Different implicational calculi turn out to correspond to different *fragments* of a type-theoretic language with λ -operators—with the full intuitionistic calculus corresponding to the complete language, but the Lambek calculus using only part of it. Moreover, there has been further research into additional constraints on ‘natural’ category changes, that is, ‘ λ -recipes’, which are meant to ensure that the ‘ λ -transforms’ of the original denotations still retain roughly the same semantic behavior as the originals.

Thus, it appears that there is a family of flexible, category-changing categorial grammars. Current research concentrates on the technical properties of such grammars. One central issue here is the recognizing power of such grammars (see vol. 1, chap. 8). Probably categorial grammars based on the original Lambek rules recognize only context-free languages, but this question has not yet been settled. Stronger calculi, curiously enough, may lose recognizing power, even to the extent that only regular languages can be recognized.

A second series of questions concerns the semantic properties of the λ -terms which are assigned to expressions by our flexible grammars. For instance, it has been proved that no matter how many readings of an expression are produced by derivations in the Lambek calculus, the related ‘meaning recipes’ yield only a *finite* number of logically nonequivalent meanings. In this sense, the richness of category changes is kept within reasonable bounds.

Furthermore, in this setting various *generalizations* of important semantic properties, such as monotonicity (see §7.2.) are being investigated. Not only NPs and determiners can be monotonic but also, e.g., adjectives, adverbial

expressions, and prepositions. This is an instance of a current tendency to formulate semantic observations made in special categories (quantifiers, verbs, adverbs) quite generally. For instance, the central notion of conservativity encountered with determiners turns out to be an instance of a general *restricting* behavior of common noun phrases across whole expressions. Witness such patterns as:

All A fear a B \Leftrightarrow All A (fear \cap (A \times B)) a B.

No A gave every B a C \Leftrightarrow No A (gave \cap (A \times B \times C)) every B a C.

In fact we can systematically *derive* such more complex forms of conservativity through the λ -recipes accompanying the categorial derivation of such sentences. Thus, we increase our insight into the categorial system of natural languages as such.

7.3.3 Further Developments

At the moment, several extensions and variants of the approach explained above are being investigated. We have mentioned the discussion about the precise nature of the link between semantic type change and syntactic category change. Other topics concern various extensions of the type-theoretic approach formulated so far. One such extension concerns the following logical strengthening. Standard categorial grammar produces meanings which can be described by using only *function application*. The category changes introduced in §7.3.1 and §7.3.2 also give rise to λ -*abstraction*. The next step could be the admission of logical *identity* between type-theoretic terms. A linguistic example of this feature is the following (not uncontroversial) German example:

Der Heinrich.

The determiner *der* is of category $((n, s), ((n, s), s))$, and the proper name *Heinrich* is of category n . If an NP, i.e., $((n, s), s)$ is to be the result of the application of one to the other, the category of *Heinrich* should be converted to (n, s) . A possible recipe here would use identity:

$n \Rightarrow (n, s)$

from A_n to $\lambda y_n [A_n = y_n]$ (the property of being A_n)

Another extension leads to the addition of *intensional* types (see §5.6), since category changes also occur in intensional contexts. So far no definite system has been proposed for this purpose. Finally we mention a possibly more surprising issue. It is one thing to propose a general mechanism for natural language, such as category change. But we cannot leave matters at that, because this mechanism will display *interactive* behavior with other important features of natural language. For instance, if we can make logical deductions from a

certain expression, what happens if this expression is exposed to a category change? Will conclusions be preserved, and if so, in what form? In fact, there is a current line of research concerning inference and category change in combined calculi. Other types of interaction are still awaiting investigation.

Further literature on these subjects may be found in Oehrle, Bach, and Wheeler 1988; Buszkowski, Marciszewski, and van Benthem 1988; Klein and van Benthem 1988.

7.4 Discourse Representation Theory

7.4.1 Introduction

Discourse representation theory is a semantic theory for natural language which was developed in the early eighties by Hans Kamp (1981a). Many ideas incorporated in discourse representation theory were present in seminal form in earlier work of a number of other authors. And around the same time, similar proposals were developed independently by among others Irene Heim (1982, 1983) and Pieter Seuren (1985).

One of the characteristics of discourse representation theory, as the name suggests, is that it focuses on the semantic interpretation of *discourses*, i.e., on coherent sequences of sentences, also called 'texts', instead of on isolated sentences, as in Montague grammar. In discourse representation theory, henceforth DRT, the primary semantic (and syntactic) unit is not the sentence but the discourse (or the text).

Another characteristic of DRT is that it regards semantic interpretation not as a *direct* relation between expressions and (a model of) reality; instead, an intermediate level of semantic representation is postulated where the information conveyed by a discourse is stored. This characteristic, too, is reflected in the name of the theory.

Unlike the intermediate level in Montague grammar, where syntactic structures are translated into expressions of the system of intensional logic IL (see §6.2), the corresponding level of discourse representation in DRT is considered to be an essential component of the grammar. It is assumed that it is not possible to do without this level of analysis, whereas the level of translation in Montague grammar is there for convenience only, being eliminable because of the compositionality of the translation and interpretation processes. So the representationalism of DRT makes it a noncompositional semantic theory.

The idea is that a discourse representation reflects the information conveyed by a discourse. As such, it may be regarded as a partial description of reality. Certainly a text never gives information on everything which is true in some reality (fictional or otherwise); it describes at most only part of it. The meaning of an expression will be regarded primarily as the contribution of that expression to the discourse representation of the greater whole in which it occurs. This concept of meaning differs from the familiar concept of the inter-

pretation of an expression in a model. In a model, an expression is interpreted within a complete picture of reality.

This more familiar level of semantic interpretation, however, is also present in DRT, in the form of the definition of the truth of a discourse. The truth of a discourse in a model is defined in terms of whether the partial information represented by its discourse representation can be embedded in a complete model.

There are various motives for the development of DRT. First of all, general theoretical and methodological issues are at stake. DRT is claimed to bridge the gap between the (psycho-) linguistic view of meaning, in which syntactic structures are related to mental representations, and the logico-semantic view, in which syntactic structures are related to (a model of) reality. In this respect it is said that DRT reconciles the *declarative* or static view of meaning with the *procedural* or dynamic view. The procedural view, which is dominant in cognitive science, holds that the meaning of an expression is to be regarded as an instruction to the hearer to 'construct' (part of) a representation. The static view is usually held by logicians and philosophers of language; it connects meaning to truth conditions, or more generally, denotation conditions.

Of course the motives behind the development of DRT are not just methodological. DRT also aims to give an account of empirical issues which other semantic theories, e.g., Montague grammar, can not cope with. An important cluster of such phenomena concerns the interpretation of pronouns, and in particular, the anaphoric relations between pronouns and indefinite terms, both within and across sentence boundaries. DRT provides a solution to several problems in this field. Other areas to which DRT is applied include the interpretation of tense and aspect, in particular the role they play in establishing the coherence of texts, and the analysis of belief sentences and other reports of propositional attitude.

In this introduction we will concentrate on some central problems from the first cluster of phenomena. We do so for expository reasons, since here the contrast between DRT and Montague grammar can be displayed most clearly. Some distortion may result from this approach. Certainly there should be no suggestion that the application of DRT to other empirical phenomena is less important. The reader is referred to the works of Heim, Kamp, and Seuren and to the literature mentioned in §7.4.6.

We will introduce DRT from the point of view of Montague grammar. In §7.4.2 we present certain problems with anaphoric relations and indefinite terms that arise in Montague grammar. Next, we sketch the solution that DRT offers to these problems. In §7.4.3 we give an informal introduction to DRT. In §7.4.4 we give definitions of the syntax and semantics of the formal language used in DRT to represent the information conveyed by a discourse.

As we indicated above, the intermediate level of discourse representation that DRT postulates is at odds with the methodological principle of composi-

tionality, which occupies such a central position in Montague grammar. In §7.4.5 we address this issue, and we argue that contrary to what is often suggested, representationalism is *not* essential to DRT, in the sense that its explanatory power does not presuppose it. The empirical success of DRT rests rather upon its procedural, dynamic view of meaning.

7.4.2 Some Problems with Anaphoric Relations and Indefinite Terms

In Montague grammar as it was presented in chapter 6, anaphoric pronouns, i.e., pronouns which are interpreted as 'referring back' to the denotation of a term, are analyzed systematically as bound variables. The following examples illustrate this:

(90) John loves Mary and he kisses her.

(91) Every woman loves a man who admires her.

In sentence (90), *he* is understood as referring back to John, and *her* as referring back to Mary; in (91), the pronoun *her* is bound by the quantified term *every woman*. Of course, there are also readings of (90) and (91) in which the referents of the pronouns are determined otherwise, for instance, by pointing to persons present in the context of the utterance; but this deictic use of pronouns will not concern us here, and in the following we will systematically ignore this possibility. The intended readings of (90) and (91) are obtained in Montague grammar by means of the quantification rules (see §6.3.8).

For instance, sentence (90) is derived from a sentential structure in which two different syntactic variables occur, *he₁ loves him₂ and he₁ kisses him₂*, into which the terms *John* and *Mary* are successively introduced by means of the quantification rule S8,n. The effect of this rule is that the term in question is substituted for the first occurrence of the relevant syntactic variable, and that eventually suitable pronouns, i.e., pronouns which agree in gender, number, and case with the term in question, take the place of the other occurrences. By the way, it may be observed that the use of a type of pseudo-pronoun as a syntactic variable may lead to confusion: the syntactic variables themselves are not pronouns but can be replaced by them if necessary. We could also have used *x* and *y* instead of *he₁* and *he₂*.

Semantically the process of quantification corresponds to the following: a syntactic variable is translated into a logical variable. Next, the formula which is the translation of the sentence in which the syntactic variable occurs is turned into an expression which refers to a property by λ -abstraction over the logical variable. Terms are translated into expressions referring to sets of properties, and the translation of the final sentence, with the term quantified in, is the result of applying the translation of the term to the intension of the λ -expression. The result is a formula asserting that the property expressed by the λ -expression belongs to the set of properties which is the denotation of the

translation of the term. If this is a quantified term, like *every woman* or *a man*, then by λ -conversion the quantifier occurring therein binds the occurrences of the free variable in the original sentence. In the translation of proper names, all occurrences are replaced by the constant which occurs in the translation of the proper name. In this sense, anaphoric pronouns are regarded as bound variables in Montague grammar.

What then are the difficulties arising from this treatment of anaphoric pronouns, for which DRT attempts to give a solution? In this section we will confine ourselves to the discussion of three examples which, though apparently simple, illustrate the major problems. Of course, there are more phenomena related to terms and anaphoric relations. For a thorough and extensive overview we refer to Heim (1982, chap. 1).

The first topic centers on the treatment of anaphoric relations across sentence boundaries. Montague grammar cannot effectively treat this type of anaphoric relation. Consider example (92):

(92) A man walks in the park. He whistles.

On the reading we are concerned with here, the pronoun *he* in the second sentence is bound by the term *a man* in the first sentence. In other words, this sequence of sentences is given the same meaning as the single sentence (93):

(93) A man walks in the park and he whistles.

Deriving (93) with the intended reading in Montague grammar is easy. The process of quantifying in, sketched in the preceding paragraphs and discussed extensively in §6.3.8, enables us to derive (93) with reduced translation (94) (we ignore the internal structure of *walks in the park* and translate it in a single predicate constant WALK IN THE PARK):

(94) $\exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x) \wedge \text{WHISTLE}(x))$

This formula not only expresses the correct meaning of (93) but also gives the meaning of (92). At a first glance then, extending Montague grammar in order to get a satisfactory treatment of examples like (92) seems a simple matter. We can introduce a syntactic operation of 'sentence sequencing', which is interpreted semantically as conjunction, and apply the quantification rule to sequences of sentences as well. If we start with the sentences *He₁ walks in the park* and *He₁ whistles*, we form from these *He₁ walks in the park. He₁ whistles*, and into this structure we quantify the term *a man*. The (reduced) result is (94).

But there is a problem. The discourse (92) may be continued with sentences in which the pronoun *he* occurs again, with the intention of referring back to *a man*:

(95) A man walks in the park. He whistles. Apparently he is in a good mood.

If we derive the first two sentences of (95), i.e., (92), in the manner described above, it will not be possible to add the third sentence to it in such a way that the occurrence of *he* is bound by *a man*. In general—and therefore also in the case of (92)—this way of accounting for anaphoric reference across sentence boundaries presupposes that the entire text is generated first, with syntactic variables at the appropriate places, after which the introduction of the required terms and their dependent anaphoric pronouns can take place by means of the quantifying-in process.

But from the semantic perspective, doing it this way implies that the interpretation of a term and of the related anaphoric pronouns can take place only when we are sure that the discourse or text will not be continued but is closed. And this implies that the process of interpretation does not proceed step by step, even though intuitively that is how we perceive the process. When we read or hear a text, we analyze and interpret the first sentence, then the second sentence, and so on. In other words, interpretation is an incremental process. The interpretation of earlier sentences will influence the interpretation of later ones, and this presupposes that the interpretations of earlier sentences are available later on. The Montague grammar method of dealing with anaphoric relations by means of quantification rules, on the other hand, does this no justice. Whenever an anaphoric relation crosses a sentence boundary, the interpretation of the first sentence cannot be determined until the entire discourse is completed, that is, until the entire text is available. In other words, a text can be interpreted only holistically, not incrementally.

One could regard this as a somewhat counterintuitive result, as one of those inevitable instances of the theoretical explanation and the pretheoretical intuition diverging. But a second example will show that the problem is deeper. Consider the following variation of (92):

(96) Exactly one boy walks in the park. He whistles.

If we derive (96) in the same way as (92), by quantifying in the term *exactly one boy* in the open sequence of sentences *He₀ walks in the park. He₀ whistles*, (97) results as its (reduced) translation:

$$(97) \exists x \forall y ((\text{BOY}(y) \wedge \text{WALK IN THE PARK}(y) \wedge \text{WHISTLE}(y)) \leftrightarrow x = y)$$

But (97) does not represent the meaning of (96). This formula expresses that there is exactly one individual that has the properties of being a boy, of walking in the park, and of whistling; in other words, there is exactly one boy who walks in the park and whistles. But it does not exclude other boys walking in the park. The meaning of (96), on the other hand, is that there is exactly one boy walking in the park, and that this boy whistles. Therefore it is (98) and not (97), which gives the correct representation of the meaning of (96):

$$(98) \exists x(\forall y((\text{BOY}(y) \wedge \text{WALK IN THE PARK}(y)) \leftrightarrow x = y) \wedge \text{WHISTLE}(x))$$

This observation shows that the problem with extending Montague grammar in the way sketched above is not just that it accounts for anaphoric relations across sentence boundaries in an unintuitive way; it also makes wrong predictions.

The difference between examples (92) and (96) also makes it intuitively clear why this method will be incorrect in general. The underlying idea is to regard a discourse or a text as a description of a complex property subsequently ascribed to the term in question. In example (92), this results in the property 'walking in the park and whistling' being applied to a man, which is correct. On the other hand, the result of the operation in the case of (96) is the property 'walking in the park and whistling' being applied to exactly one boy, which is not the meaning of (96). While (92) and (93) are equivalent to (99), (96) is not equivalent to (100):

(99) A boy walks in the park and whistles.

(100) Exactly one boy walks in the park and whistles.

In the sequence of sentences (96), it is first asserted that there is exactly one boy who is walking in the park, and next it is asserted of this boy that he whistles. This could also be described as follows: the first sentence *introduces* an individual, the unique boy walking in the park, and the second sentence gives a further description of this individual: he whistles. This suggests a quite general way of dealing with the continuation of a text. For example, sequence (96) can be continued as in (101):

(101) Exactly one boy walks in the park. He whistles. He has blue eyes.

As we observed above, the quantifying-in approach runs into difficulties here. However, if we follow the suggestion made above, it seems that there is a way to avoid these difficulties. As the text proceeds, we dress up the introduced individual with more properties. Of course, we can deal with an example like (92) in the same way: the first sentence introduces a (not necessarily unique) individual that is a man and walks in the park. The succeeding sentences ascribe more properties to the individual: he whistles, he is in a good mood, and so on.

Such individuals, first introduced and then further described in a discourse or a text, are sometimes called *discourse referents*. They are stand-ins for the individuals to which a discourse or text refers. We observed earlier that a given discourse almost always gives only a partial description of a certain domain. Whether the discourse is true or not in some given model depends on whether a correspondence can be established between the discourse referents

introduced by the discourse and real individuals in the domain of the model in such a way that all assertions made in the discourse about the discourse referents are true. Sometimes there is exactly one way to get such a correspondence, and sometimes there is more than one.

These informal observations about how to interpret a discourse lie at the bottom of the DRT approach. But before we go into details, we want to discuss yet a third phenomenon, another one that Montague grammar cannot cope with and for which DRT proposes a solution.

The problem in the examples discussed above concerns sequences of sentences where a pronoun in a sentence is anaphorically related to an indefinite term in a preceding sentence. The third phenomenon also concerns indefinite terms and anaphoric pronouns, but this time within sentence boundaries. Consider the following examples:

(102) If John owns a donkey, he beats it.

(103) Every farmer who owns a donkey beats it.

In sentence (102), we find an indefinite term in the antecedent of an implication and a pronoun in the consequent; in (103) we find an indefinite term in a relative clause modifying a universally quantified term, and a pronoun in the main clause. Both sentences illustrate the same problem, which is known in the literature as the problem of the *donkey sentences*. The problem is the following.

A correct semantic analysis of the examples (102) and (103) should render the following (reduced) translations:

(104) $\forall x((\text{DONKEY}(x) \wedge \text{OWN}(\text{JOHN}, x)) \rightarrow \text{BEAT}(\text{JOHN}, x))$

(105) $\forall x\forall y((\text{FARMER}(x) \wedge \text{DONKEY}(y) \wedge \text{OWN}(x, y)) \rightarrow \text{BEAT}(x, y))$

The problem, of course, is not what the meanings of the sentences (102) and (103) are or how they should be represented; the first-order predicate-logical formulas (104) and (105) express their meanings adequately. The heart of the problem, as in the examples discussed above, is how to obtain the representations (104) and (105).

Let us take a closer look at example (102). We notice immediately that the indefinite term *a donkey* reappears in (104) not as an existential quantifier but as a universal one, and that it has scope over the entire implication. In view of the meaning of (102), this is correct; the question now is how to obtain this meaning in a compositional way. It seems reasonable to assume that the term *a donkey* as it occurs in sentences such as (102) and (103) is assigned its usual meaning, represented in IL by the familiar expression $\lambda X\exists x[\text{DONKEY}(x) \wedge \forall Y(X(Y))]$. But if we want to derive (102) in such a way that the pronoun *it* in the consequent of the implication is bound by the term *a donkey* in the antecedent, we get into difficulties. The only way to get this binding would be to

quantify the term *a donkey* into the sentence *If John owns him₀, then he beats him₀*. The result of this operation is (106):

$$(106) \exists x(\text{DONKEY}(x) \wedge (\text{OWN}(\text{JOHN}, x) \rightarrow \text{BEAT}(\text{JOHN}, x)))$$

But this formula does not express the meaning of (102). And the only alternative that Montague grammar offers is the direct introduction of the term, i.e., a derivation without any quantifying in. The result of this is (107):

$$(107) \exists x[\text{DONKEY}(x) \wedge \text{OWN}(\text{JOHN}, x)] \rightarrow \text{BEAT}(\text{JOHN}, x)$$

Here, the occurrence of x in the consequent is not bound by the existential quantifier in the antecedent, and therefore the anaphoric relation between *a donkey* and *it* is not accounted for; (107) is not equivalent to the correct translation (104). In general, a formula of the form $\exists x\phi \rightarrow \psi$ is equivalent to $\forall x(\phi \rightarrow \psi)$ only if ψ does not contain free occurrences of x . The consequent of (107) does contain a free occurrence of x , and hence (107) is not equivalent to (104).

We meet the same kind of problems, of course, if we try to get (105) as a translation of (103) in a compositional manner.

These examples are similar to the examples discussed earlier as far as the anaphoric relation is concerned. If there is no anaphoric pronoun, as in (108), then we can manage with the standard representation of *a donkey*, and the result is the adequate translation (109):

$$(108) \text{If John owns a donkey, then Jack is jealous.}$$

$$(109) \exists x(\text{DONKEY}(x) \wedge \text{OWN}(\text{JOHN}, x)) \rightarrow \text{JEALOUS}(\text{JACK})$$

The specific problem raised by the donkey sentences is finding a semantic way of dealing with these indefinite terms which accounts for the fact that in one construction their import is existential and in another construction it is universal. Section 7.4.3 will show that the analysis using discourse referents provides an adequate solution for the problem.

Exercise 13

Show that sentence (103) cannot be dealt with successfully in the Montague fragment of chapter 6.

7.4.3 An Informal Introduction to DRT

In section 7.4.2 we discussed several phenomena concerning indefinite terms and anaphoric pronouns that cannot be solved within the framework of Montague grammar as it was presented in chapter 6. One of the empirical claims of DRT is that it offers a semantic framework in which a uniform and elegant description of these facts can be given. This section will introduce this framework by showing how the problems are dealt with.

There is no really definitive formulation of DRT that is strictly adhered to. The definitions given below differ in several respects from the original version of DRT that is given in Kamp 1981a. However, we trust that readers will be able to explore the DRT literature after acquainting themselves with the version presented here.

Like Montague grammar, DRT offers a semantic interpretation of (a fragment of) natural language. The first difference lies in the fact that Montague grammar is a *sentence* grammar, while DRT in principle attempts to interpret *sentence sequences*. For the kind of examples that we discuss here, this difference turns out to be not that important, since all instances of sentence sequences that we will encounter can be paraphrased as sentence conjunctions. The claim is, however, that the way these simple sequences are dealt with provides a fruitful perspective for the treatment of more complicated cases.

DRT offers a semantic interpretation, and hence we are justified in expecting a *syntax* at the starting point of the analysis and a *model* at its finish. We will not go into the syntax of the fragment treated here. We will merely assume that a syntax assigning simple constituent structures to the sentences of the fragment is available. We further assume that the fragment contains: extensional intransitive and transitive verb phrases; common noun phrases; proper names; singular personal pronouns; existentially and universally quantified terms; restrictive relative clauses; and the sentential operations of negation, disjunction, implication, and sentence sequencing. Simple extensional first-order predicate-logical models can serve as models for the fragment.

A characteristic property of DRT is that, given a syntactic structure, (sequences of) sentences are provided with a *representation*. One of the mechanisms of DRT is a set of rules converting syntactic structures into *discourse representation structures*, DRSSs. These rules are called *DRS construction rules*. DRSSs themselves are expressions of a somewhat unorthodox formal language. We will introduce two kinds of notation for DRSSs, both of which can be found in the DRT literature: pictorial and linear notations. In this section we give an informal sketch of the DRS construction process and of the interpretation of the resulting DRSSs, using the pictorial notation. In §7.4.4 we turn to a formal definition of a syntax and semantics of DRSSs, using the linear notation.

As a first example of the construction of a DRS for a natural language sentence, consider (110):

(110) John loves a girl who admires him.

We assume that a constituent structure of sentence (110) is given. The first step in the DRS construction is to put the sentence in a box:

(111)

John loves a girl who admires him.

The second step in the construction leads to the following box:

(112)

<p>x</p> <p>John = x</p> <p>x loves a girl who admires him.</p>
--

In the transition from (111) to (112), three things have happened: (i) a variable x is introduced into the box, called a *reference marker* in DRT, which plays the role of what we called a discourse referent in §7.4.2; (ii) the term which is the subject of the sentence, the proper name *John*, is replaced in the sentence by the reference marker x ; (iii) an identity assertion $John = x$ is added. Application of the rule connected with proper names always makes these three things happen. The rule is applied whenever we come across something of the form $_s[_{NP}[proper\ name]_{IV}[\dots]]$ (though not exclusively in that case).

Proceeding with the construction, the third step leads to the following box:

(113)

<p>x y</p> <p>John = x</p> <p>x loves y</p> <p>girl(y)</p> <p>y admires him</p>

In the third step, a new reference marker y is introduced. Markers are introduced not only for proper names but also for indefinite terms, like *a girl*, or in this case, *a girl who admires him*. In what remained of the original sentence in box (112), this term is replaced by the newly introduced marker y , and the result is the formula $x\ loves\ y$. Finally, the formula $girl(y)$ is added, and simultaneously a reconstruction is applied to the relative clause. In essence, the latter amounts to the relative pronoun *who* being replaced by the marker y . If the original sentence had been *John loves a girl*, we would end up with box (113) minus the last line, and the construction of the DRS would be complete.

Only one thing remains to be done with our sentence: the anaphoric pronoun *him* must be taken care of. The relevant DRS construction rule substitutes a suitable previously introduced marker for the pronoun. Since *him* is masculine, the only suitable marker is the x which was introduced by *John*. In other instances there might be more than one candidate, and hence more than one DRS would be possible. In the present case, however, the result is not ambiguous:

(114)

x	y
John = x	
x loves y	
girl(y)	
y admires x	

Thus the final stage of the DRS construction process is reached. No further DRS construction rules are applicable to box (114). The result is a box containing two types of things: (i) a set of reference markers: $\{x, y\}$; (ii) a set of formulas: $\{\text{John} = x, x \text{ loves } y, \text{girl}(y), y \text{ admires } x\}$. The formulas which occur in a DRS are called *conditions*. In our example, all conditions are atomic formulas. Such simple DRSs, consisting of a set of reference markers and a set of atomic conditions, form the basic building blocks of DRSs. Later we will see that DRSs may also contain complex conditions.

If sentence (109) were to be continued with sentence (115):

(115) She loves him too.

then this sentence would be added to box (114), and the construction process would continue. The final result would then look like this:

(116)

x	y
John = x	
x loves y	
girl(y)	
y admires x	
y loves x	

Boxes like these are meant to represent the meanings of (sequences of) sentences. So let us now turn to the interpretation of DRSs.

We have observed that a DRS is considered to be a partial description of (a model of) reality. To put it somewhat differently, we may regard a DRS as a partial model of reality. In (116) and (114), this is a model with a domain containing two individuals, whose properties are (partly) specified by the formulas in (116) and (114). The idea now is that a DRS can be called *true* in relation to an ordinary, total model M if the partial model corresponding to the DRS can be taken to be a part of, i.e., can be embedded in, M .

The interpretation of a DRS proceeds as follows. A model M specifies a domain D and an interpretation function I . I interprets proper names, common

nouns, and verbs in the same way that individual constants and predicates are interpreted in predicate logic.

We define the notion of a *verifying embedding* of a DRS into a model M . Such a verifying embedding is a function f which assigns elements of D to the reference markers in the DRS in such a way that all conditions in the DRS come out true in M .

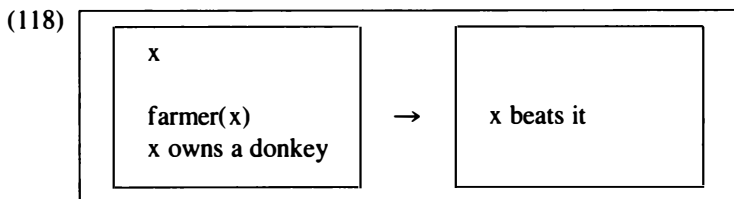
In terms of this notion of a verifying embedding, the notion of *truth* of a DRS in a model M is defined. A DRS is true in M iff there is at least one verifying embedding for that DRS in M .

For instance, DRS (116), the DRS of the sequence of sentences (109) and (115), is true iff there is a verifying embedding f assigning individuals from D to the reference markers x and y in such a way that $f(x) = \text{John}$, and $f(y)$ is a girl loved by John who in turn loves and admires John. In other words, the truth conditions of DRS (116) have the same effect as if the indefinite term *a girl who admires him* is analyzed as an existentially quantified term with wide scope over the conjunction of (109) and (115), and the pronouns are analyzed as bound variables. But no existential quantification is used to obtain this result. The effect of existential quantification is the result of the truth conditions for DRSs, which require the existence of *at least one* verifying embedding of the DRS into the model.

But we saw above that indefinite terms sometimes correspond to universal quantification. How does DRT manage to account for that? DRT's treatment of the donkey sentence, (103), repeated below as (117), illustrates this:

(117) Every farmer who owns a donkey beats it.

The first step in the construction of a DRS for (117) is again to put the entire sentence in a box. The sentence has the form $s_{NP}[every_{CN}[\alpha]]_{IV}[\beta]$, i.e., its subject is a universal term. The second step in the construction is the application of the DRS construction rule for universal NPs. This rule introduces something new, namely, an implication relation \rightarrow between DRSs. After these first two steps, the result looks like this:



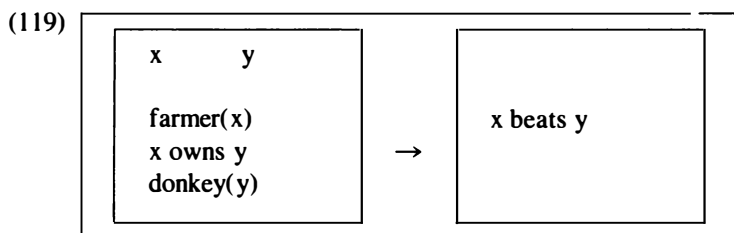
The DRS (118) consists of three boxes. The outer box, where the original sentence was placed, is called the main DRS. The implication relation \rightarrow between the two sub-DRSs turns them into a complex condition, and this complex condition is placed inside the main DRS.

The process also introduces a relation of *subordination* between DRSs. The two sub-DRSs related by \rightarrow are subordinated to the main DRS; and the one on the right of \rightarrow is subordinated to the one on the left.

In the left box a reference marker x is introduced. In the right box, a formula appears which results from replacing the universal NP in the sentence with the introduced marker x . The formulas in the left box correspond to the CN and its relative clause, which are treated in the same way as was illustrated in example (110) with an indefinite term.

The two sub-DRSs in (118) are subject to further DRS construction rules. So within the left box and the right box, we continue the DRS construction process. In the left box, the indefinite term *a donkey* occurs, and we apply the construction rule for indefinite terms as discussed in the analysis of sentence (110). This means that a new marker y is introduced into the left box, together with a formula asserting that y is a donkey; and the object NP *a donkey* is replaced by the newly introduced marker y in the phrase *x owns a donkey*.

Finally, in the right box, the pronoun *it* should be taken care of. In the previous example, there was only one box, and there we described the construction rule for pronouns as follows: *substitute a suitable introduced reference marker for the pronoun*. But this rule should be extended by use of the relation of subordination between DRSs introduced above: *substitute for the pronoun a suitable reference marker introduced in one of the boxes to which the box in which the pronoun occurs is subordinated*. In this case, this can only be the box to the left of the implication sign, since no marker has been introduced into the main DRS, and only y is suitable. The final result of the construction process is the following DRS:



Now let us take a look at the interpretation of this new type of DRS. The main DRS of (119) contains no reference markers and only one complex condition. So the definition of the notion of a verifying embedding introduced above requires for this DRS that the condition which consists of the two sub-DRSs joined by the implication sign be true. The latter is defined to be the case if *every* verifying embedding of the antecedent DRS gives rise to a verifying embedding for the consequent DRS. For (119), this implies that every assignment f that assigns a farmer to x and a donkey to y must verify that the farmer beats the donkey.

In other words, the truth conditions of the discourse representation (119) of

sentence (117) are exactly the same as those of its translation (105) in ordinary predicate logic, discussed in §7.4.2, with wide scope universal quantification over x and y . But this time we did not meet the problem mentioned there. In the DRS construction we treated the indefinite term *a donkey* in (117) in the same way as we treated the indefinite term in example (110). But the interpretation of DRSs ensures that in the latter case it gets the force of existential quantification, whereas in the former it acquires universal strength because it occurs inside the antecedent of a conditional DRS.

We end this section by noting that the conditional sentence (120) results in exactly the same DRS as the donkey sentence (117) we just dealt with:

(120) If a farmer owns a donkey, he beats it.

The DRS construction associated with a conditional sentence consists of introducing two sub-DRSs connected by \rightarrow in the main DRS. In the antecedent DRS we continue with the reconstruction of the antecedent of the sentence, in this case *a farmer owns a donkey*; in the consequent DRS we continue with the consequent of the sentence, in this case *he beats it*.

Exercise 14*

Construct DRSs for the following (sequences of) sentences:

- (a) A boy loves every girl.
- (b) Every boy loves every girl.
- (c) If John loves Mary, then she loves him. If she hates him, he hates her.

Exercise 15*

Formulate a DRS construction rule for subject NPs with the determiner *exactly one* and use it to construct a DRS for example (96), discussed in §7.4.2:

Exactly one boy walks in the park. He whistles.

7.4.4 Formal Definitions

In this section we give formal definitions of the syntax and semantics of DRSs in a linear, set-theoretical notation.

In the vocabulary of the DRS language we find individual constants and reference markers (together forming the class of terms), n -place predicate constants, identity, negation, disjunction, and implication. (So identifying reference markers with variables, the vocabulary of the DRS language forms a real subset of that of first-order predicate logic.)

As was indicated in §7.4.3, a DRS may be viewed as a pair $\langle V, C \rangle$, with V a (finite and possibly empty) set of reference markers, and C a (finite and possibly empty) set of conditions. The latter may be either atomic or complex. (Strictly speaking, C is not a set of formulas but rather a set of occurrences of formulas, or a bag or multiset of formulas. In the remainder we will ignore this technical detail.) Complex conditions are formed from DRSs, so the defi-

nitions of DRSs and of conditions have to go hand-in-hand. We use the lower-case Greek characters ϕ and ψ as metavariables ranging over conditions, and the uppercase Greek characters Φ and Ψ as metavariables ranging over DRSs.

Definition 1

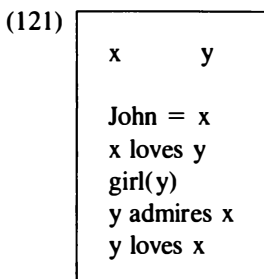
- (i) If P is a n -place predicate constant and t_1, \dots, t_n are terms, then $P(t_1, \dots, t_n)$ is a condition;
- (ii) If t and t' are terms, then $t = t'$ is a condition;
- (iii) If Φ is a DRS, then $\neg\Phi$ is a condition;
- (iv) If Φ and Ψ are DRSs, then $(\Phi \rightarrow \Psi)$ is a condition;
- (v) If Φ and Ψ are DRSs, then $(\Phi \vee \Psi)$ is a condition;
- (vi) If x_1, \dots, x_n are reference markers ($n \geq 0$), and ϕ_1, \dots, ϕ_m are conditions ($m \geq 0$), then $\langle\{x_1, \dots, x_n\}, \{\phi_1, \dots, \phi_m\}\rangle$ is a DRS;
- (vii) Nothing is a DRS or a condition except on the basis of (i)–(vi).

By means of clauses (i) and (ii), atomic conditions can be formed which differ in no respect from the atomic formulas in predicate logic. Clauses (iii)–(v) form negations, implications, and disjunctions. While in predicate logic these operations turn formulas into more complex formulas, here they turn DRSs into complex conditions. It is only by means of clause (vi) that DRSs can be formed, which are represented in a set-theoretical notation. This notation makes it possible to apply set-theoretical operations to DRSs. In fact, the operations on boxes which are used in the DRSs construction rules, such as adding reference markers and conditions to DRSs, can be viewed as such set-theoretical operations.

The set of reference markers in a DRS fulfills the role of a quantification mechanism. Free occurrences of reference markers in the (atomic or complex) conditions of the DRS are bound by it. The binding force of sets of reference markers is more powerful than that of the quantifiers in predicate logic. Quantifiers can only bind variables within their scope. If we identify the scope of a set of markers V in a DRS $\langle V, C \rangle$ with the conditions in C , then the set V can bind markers outside its scope. This happens in case $\langle V, C \rangle$ is the antecedent of a conditional $\langle V, C \rangle \rightarrow \langle V', C' \rangle$. In case a marker $x \in V$ has a free occurrence in the consequent $\langle V', C' \rangle$, that occurrence is bound by the set V in the antecedent. This more global notion of variable binding is an essential feature of DRT; it lies at the heart of its treatment of the donkey sentences, where an indefinite term within the antecedent of an implicational structure can be anaphorically linked to a pronoun outside its scope in the consequent.

In the DRS language defined above, this more relaxed notion of binding is restricted to implications. In a disjunction, it is not possible for the set of markers of one of the disjuncts to bind markers in the other disjunct. Similarly, a set of markers under the scope of negation has no binding force outside the negation. Of course the binding properties of DRSs discussed here informally are effected by their semantics, to which we turn below.

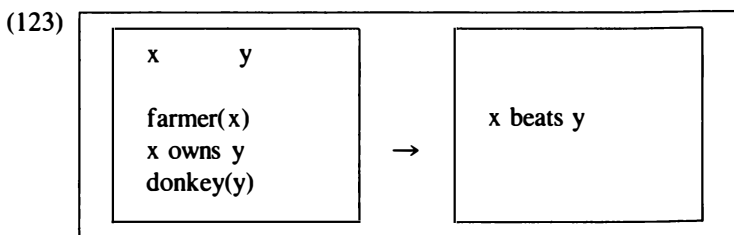
By way of illustration, we present the DRSs of the two examples discussed in §7.4.3 in the linear notation of definition 1. In the pictorial box notation, the sequence of sentences *John loves a girl that admires him. She loves him (too)* was represented as (121) (= (116)):



In the formalism of definition 1, (121) corresponds to the following DRS:

$$(122) \langle \{x, y\}, \{\text{JOHN} = x, \text{LOVES}(x, y), \text{GIRL}(y), \text{ADMIRE}(y, x), \text{LOVES}(y, x)\} \rangle$$

Our second example is the DRS of the donkey sentence *Every farmer who owns a donkey beats it* (= (119)):



The main DRS (123) consists of an empty set of reference markers and a set of conditions with one element: a complex condition which has the form of an implication. The box to the left of the implication sign, the antecedent DRS, is written as (124) in the linear notation of definition 1:

$$(124) \langle \{x, y\}, \{\text{FARMER}(x), \text{DONKEY}(y), \text{OWN}(x, y)\} \rangle$$

The box to the right of the implication sign, the consequent DRS, is now written as (125):

$$(125) \langle \emptyset, \{\text{BEAT}(x, y)\} \rangle$$

Together they form the complex condition (126):

$$(126) \langle \langle \{x, y\}, \{\text{FARMER}(x), \text{DONKEY}(y), \text{OWN}(x, y)\} \rangle \rightarrow \langle \emptyset, \{\text{BEAT}(x, y)\} \rangle \rangle$$

DRS (123) as a whole then corresponds to (127):

$$(127) \langle \emptyset, \{ \langle \{x, y\}, \{FARMER(x), DONKEY(y), OWN(x, y)\} \rangle \rightarrow \langle \emptyset, \{BEAT(x, y)\} \rangle \} \rangle$$

For reasons of readability, we will sometimes write $\langle \emptyset, \{\phi_1, \dots, \phi_n\} \rangle$ as $\{\phi_1, \dots, \phi_n\}$ and $\{\phi\}$ as ϕ , and leave off the outer brackets. With these conventions, (127) can then be written as:

$$(128) \langle \{x, y\}, \{FARMER(x), DONKEY(y), OWN(x, y)\} \rangle \rightarrow BEAT(x, y).$$

Exercise 16*

Write the DRSs for the (sequences of) sentences (a)–(c) of exercise 14, and the DRS for (d), which was the subject of exercise 15, in the linear notion of definition 1:

- (a) A boy loves every girl.
- (b) Every boy loves every girl.
- (c) If John loves Mary, then she loves him. If she hates him, he hates her.
- (d) Exactly one boy walks in the park. He whistles.

We now turn to the semantic interpretation of DRSs. We interpret DRSs in a model, just like formulas of an ordinary logical language. For the DRS language treated here, extensional first-order models are adequate. Thus, a model $\mathbf{M} = \langle D, I \rangle$ consists of a domain D and an interpretation function I . I interprets the individual constants and predicate constants in the usual way.

Since the syntactic definition defines both conditions and DRSs, the semantic definition also has to state the interpretations for the two kinds of expressions. Two notions are defined simultaneously:

- $\models_{\mathbf{M},g} \phi$ Condition ϕ is true in model \mathbf{M} with respect to assignment g .
- $h \models_{\mathbf{M},g} \Phi$ Assignment h is a verifying embedding for DRS Φ in model \mathbf{M} with respect to assignment g .

In terms of these two notions, the interpretation of conditions and DRSs is given by a simultaneous recursive definition. We use the following notational conventions:

$$\begin{aligned} \llbracket t \rrbracket_{\mathbf{M},g} &= I_{\mathbf{M}}(t) \text{ if } t \text{ is an individual constant;} \\ &= g(t) \text{ if } t \text{ is a variable.} \end{aligned}$$

$$h[x_1, \dots, x_n]g: \text{ Assignment } h \text{ differs at most from assignment } g \text{ in the values it assigns to the reference markers } x_1, \dots, x_n. \text{ (For } n = 0, \text{ this amounts to } h = g.)$$

Definition 2:

- (i) $\models_{\mathbf{M},g} P(t_1, \dots, t_n)$ iff $\langle \llbracket t_1 \rrbracket_{\mathbf{M},g}, \dots, \llbracket t_n \rrbracket_{\mathbf{M},g} \rangle \in I_{\mathbf{M}}(P)$;
- (ii) $\models_{\mathbf{M},g} t = t'$ iff $\llbracket t \rrbracket_{\mathbf{M},g} = \llbracket t' \rrbracket_{\mathbf{M},g}$;
- (iii) $\models_{\mathbf{M},g} \neg \Phi$ iff for no h : $h \models_{\mathbf{M},g} \Phi$;

- (iv) $\models_{\mathbf{M},g}(\Phi \rightarrow \Psi)$ iff for all h : if $h \models_{\mathbf{M},g} \Phi$, then there is a k such that $k \models_{\mathbf{M},h} \Psi$;
- (v) $\models_{\mathbf{M},g}(\Phi \vee \Psi)$ iff there is some h such that $h \models_{\mathbf{M},h} \Phi$ or there is some h such that $h \models_{\mathbf{M},g} \Psi$;
- (vi) $h \models_{\mathbf{M},g} \langle \{x_1, \dots, x_n\}, \{\phi_1, \dots, \phi_m\} \rangle$ iff $h[x_1, \dots, x_n]g$ & $\models_{\mathbf{M},h} \phi_1$ & \dots & $\models_{\mathbf{M},h} \phi_m$.

In terms of the notion of a verifying embedding for a DRS, we define the notion of truth of a DRS:

Definition 3:

A DRS Φ is true in model \mathbf{M} with respect to an assignment g , $\models_{\mathbf{M},g} \Phi$, iff there is an assignment h such that $h \models_{\mathbf{M},g} \Phi$.

Note that the notion of truth of a DRS implicitly plays a role in definition 2 in clauses (iii)–(v). Using the notion of truth of a DRS defined in definition 3, these clauses can be written more economically as:

- (iii') $\models_{\mathbf{M},g} \neg \Phi$ iff $\not\models_{\mathbf{M},g} \Phi$;
- (iv') $\models_{\mathbf{M},g}(\Phi \rightarrow \Psi)$ iff for all h : if $h \models_{\mathbf{M},g} \Phi$, then $\models_{\mathbf{M},h} \Psi$;
- (v') $\models_{\mathbf{M},g}(\Phi \vee \Psi)$ iff $\models_{\mathbf{M},g} \Phi$ or $\models_{\mathbf{M},g} \Psi$.

In words: a condition $\neg \Phi$ is true with respect to an assignment g iff the DRS Φ is false with respect to g ; a condition $\Phi \rightarrow \Psi$ is true with respect to g iff Ψ is true with respect to every assignment h which is a verifying embedding for Φ with respect to g ; a condition $\Phi \vee \Psi$ is true with respect to g iff Φ is true with respect to g or Ψ is true with respect to g .

The truth definition also makes it clear that the set of reference markers in a DRS fulfills the role of a quantification mechanism. A DRS $\langle \{x_1, \dots, x_n\}, \{\phi_1, \dots, \phi_m\} \rangle$ is true with respect to g iff there is some assignment h such that h differs at most from g with respect to the values it assigns to x_1, \dots, x_n and ϕ_1, \dots, ϕ_m are true with respect to h . This means that a DRS such as $\langle \{x, y\}, \{P(x), R(x, y)\} \rangle$ gets the same truth conditions as the predicate-logical formula $\exists x \exists y (P(x) \wedge R(x, y))$.

In case the set of reference markers of a DRS is empty, its truth conditions coincide with the truth conditions of its set of conditions. For example, according to clause (vi) of definition 2, h is a verifying embedding for $\langle \emptyset, \{Qx\} \rangle$ with respect to g iff h is an assignment which differs in no respect from g , i.e., iff $h = g$, and Qx is true with respect to g . In case Qx is false with respect to g , $\langle \emptyset, \{Qx\} \rangle$ has no verifying embedding with respect to g , and in case Qx is true with respect to g , g itself is the only verifying embedding for $\langle \emptyset, \{Qx\} \rangle$ with respect to g . According to definition 3, this means that the DRS $\langle \emptyset, \{Qx\} \rangle$ is true with respect to an assignment g iff the condition Qx is true with respect to g . In general it holds that $\models_{\mathbf{M},g} \langle \emptyset, \{\phi_1, \dots, \phi_m\} \rangle$ iff $\models_{\mathbf{M},g} \phi_1$ & \dots & $\models_{\mathbf{M},g} \phi_m$.

To illustrate the working of definitions 2 and 3, we consider again the DRSs

(122) of the sequence of sentences *John loves a girl that admires him. She loves him (too)*, and the DRS (127) of the donkey sentence *Every farmer who owns a donkey beats it*. First the interpretation of (129) (= (122)):

$$(129) \langle \{x, y\}, \{\text{JOHN} = x, \text{LOVE}(x, y), \text{GIRL}(y), \text{ADMIRE}(y, x), \text{LOVE}(y, x)\} \rangle$$

According to definition 3, DRS (129) is true (reference to an assignment can be omitted since (129) does not contain free occurrences of variables) iff there is some assignment h which is a verifying embedding for (129). According to clause (vi) of definition 2, this is the case iff there is some assignment h such that $\models_h \text{JOHN} = x$, $\models_h \text{LOVE}(x, y)$, $\models_h \text{GIRL}(y)$, $\models_h \text{ADMIRE}(y, x)$, and $\models_h \text{LOVE}(y, x)$. And according to clauses (i) and (ii), this is the case iff there is some assignment h such that $h(x) = I(\text{JOHN})$, $\langle h(x), h(y) \rangle \in I(\text{LOVE})$, $h(y) \in I(\text{GIRL})$, $\langle h(y), h(x) \rangle \in I(\text{ADMIRE})$ and $\langle h(y), h(x) \rangle \in I(\text{LOVE})$. In other words, DRS (129) is true under exactly the same circumstances as the predicate-logical formula (130), which is in turn equivalent to (131).

$$(130) \exists x \exists y (\text{JOHN} = x \wedge \text{GIRL}(y) \wedge \text{LOVE}(x, y) \wedge \text{ADMIRE}(y, x) \wedge \text{LOVE}(y, x))$$

$$(131) \exists y [\text{GIRL}(y) \wedge \text{LOVE}(j, y) \wedge \text{ADMIRE}(y, j) \wedge \text{LOVE}(y, j)]$$

As our second example, we consider the interpretation of the DRS (132) (= (127)) corresponding to donkey sentence (103):

$$(132) \langle \emptyset, \{ \langle \{x, y\}, \{\text{FARMER}(x), \text{DONKEY}(y), \text{OWN}(x, y)\} \rangle \rightarrow \langle \emptyset, \{\text{BEAT}(x, y)\} \rangle \} \rangle$$

Since the set of markers of the main DRS of (127) is empty, its truth conditions coincide with those of its only condition, (133) (= (126)):

$$(133) \langle \{x, y\}, \{\text{FARMER}(x), \text{DONKEY}(y), \text{OWN}(x, y)\} \rangle \rightarrow \langle \emptyset, \{\text{BEAT}(x, y)\} \rangle$$

According to clause (iv) of definition 2, (133) is true iff the consequent DRS $\langle \emptyset, \{\text{BEAT}(x, y)\} \rangle$ is true with respect to all assignments h that are verifying embeddings for the antecedent DRS $\langle \{x, y\}, \{\text{FARMER}(x), \text{DONKEY}(y), \text{OWN}(x, y)\} \rangle$. The truth of the consequent DRS $\langle \emptyset, \{\text{BEAT}(x, y)\} \rangle$, which has an empty set of markers, amounts to the truth of its condition $\text{BEAT}(x, y)$. This means that the truth conditions for (133), and hence for (132), amount to the following: for every assignment h , if $h(x) \in I(\text{FARMER})$, $h(y) \in I(\text{DONKEY})$, and $\langle h(x), h(y) \rangle \in I(\text{OWN})$, then $\langle h(x), h(y) \rangle \in I(\text{BEAT})$. In other words, the truth conditions of DRS (132) for the donkey sentence (117) are the same as those for its translation (134) (= (105)) in predicate logic.

$$(134) \forall x \forall y [[\text{FARMER}(x) \wedge \text{DONKEY}(y) \wedge \text{OWN}(x, y)] \rightarrow \text{BEAT}(x, y)]$$

It is DRT's interpretation of implication which takes care of the fact that a set of markers in the antecedent can bind occurrences of variables in the consequent and which lends it its universal quantificational force.

Exercise 17

Determine the truth conditions of the DRSs of the (sequences of) sentences (a)–(c), and (d), from exercise 16, by applying definitions 2 and 3:

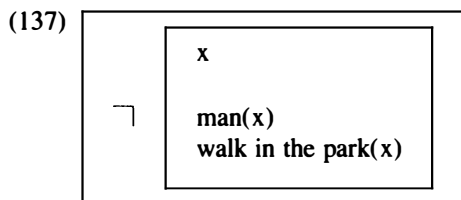
- (a) A boy loves every girl.
- (b) Every boy loves every girl.
- (c) If John loves Mary, then she loves him. If she hates him, he hates her.
- (d) Exactly one boy walks in the park. He whistles.

Until now we have paid little or no attention to negation and disjunction of DRSs. We end this section with a few remarks about these operations. First we look at negation. Consider the following two sequences of sentences:

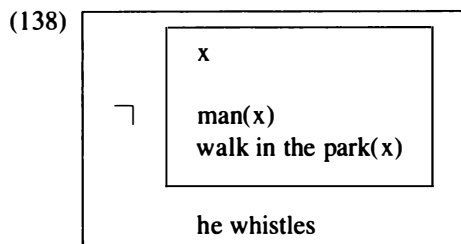
(135) It is not the case that a man walks in the park. He whistles.

(136) No man walks in the park. He whistles.

In both cases we observe that the pronoun in the second sentence cannot be anaphorically linked to the terms *a man* or *no man* in the first sentence. This fact is taken care of in DRT. Using the box notation once more, the DRS construction of the first sentence would lead to DRS (137):



This DRS has an empty set of reference markers and contains a single condition, the negation of the DRS that corresponds to *a man walks in the park*. If we add the second sentence to this DRS, we can get no further than the following:

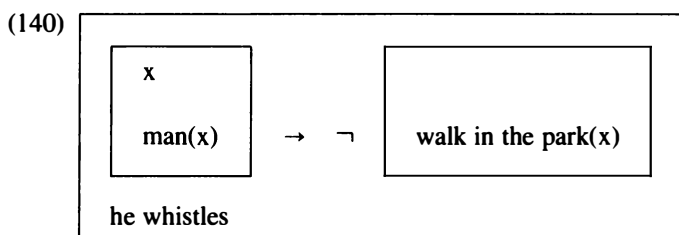


We cannot resolve the pronoun. In the main DRS, no reference markers have been introduced. The set of markers inside the negation is not accessible. We can also look at this in the following way. Take the linear representation (139) of DRS (138), where we replace the pronoun by the marker x :

$$(139) \langle \emptyset, \{\neg\langle\{x\}, \{\text{MAN}(x), \text{WALK IN THE PARK}(x)\}\rangle, \text{WHISTLE}(x)\}\rangle$$

The occurrence of the marker x in the condition $\text{WHISTLE}(x)$ is a free occurrence. The set of markers of the DRS to which it belongs is empty, and it is not in the consequent of an implication. The example is an instance of the general fact that terms in a negated sentence cannot enter into anaphoric relations with pronouns in subsequent sentences.

In the second example, (136), we get an equivalent result. In the box notation, we would get the following DRS as its representation:



Once again, the pronoun cannot be resolved. And correspondingly, in linear representation (141), we find an occurrence of the marker x in the condition $\text{WHISTLE}(x)$ which is not bound:

$$(141) \langle \emptyset, \{(\langle\{x\}, \{\text{MAN}(x)\}\rangle \rightarrow \neg\langle\emptyset, \{\text{WALK IN THE PARK}(x)\}\rangle), \text{WHISTLE}(x)\}\rangle$$

Under this representation it is not negation as such that blocks anaphoric relations in (136) but rather the fact that terms within a conditional sentence cannot enter into anaphoric relations with pronouns in subsequent sentences of the discourse. This means that in case of discourses (142)–(144), too, DRT correctly predicts that pronouns in the second sentence cannot be anaphorically related to terms in the first sentence:

(142) Every man walks in the park. He whistles.

(143) If a farmer owns a donkey, he beats it. He hates it.

(144) No man walks in the park. He whistles.

However, sometimes structurally similar discourses are OK:

(145) Every player chooses a pawn. He puts it on square one.

(146) If a client enters, you treat him politely. You offer him a cup of coffee and you ask him to wait.

In (145) and (146), the first sentence corresponds to a DRS that contains a condition that has the form of an implication. Correct representations of the meanings of (145) and (146) would result if the condition that corresponds to the second sentence were included *in the consequent* of the implication that corresponds to the first sentence. So it is not impossible to construct a DRS for these discourses, but the DRS construction process for the second sentences has to be different from the normal procedure. The normal procedure would enter the second sentences in the main DRS, whereas in these cases we would enter them in the consequent DRS in the representation of the first sentence. Further, one would have to explain why this procedure is not allowed for (142) or for many other cases.

Similar observations can be made concerning double negation. In some cases, though not in all, we get anaphoric relations:

- (147) It is not the case that John doesn't own a car. It is red and it is parked in front of his house.

DRT can not easily explain this; negation blocks anaphoric relations, and this feature of negation is not annihilated by double negation.

Similar limitations in accounting for anaphoric relations can be observed with disjunctions. The interpretation of disjunction as it is given in definition 2 prohibits anaphoric relations between a pronoun in the second disjunct and a term in the first. Hence, sentences like (148) and (149) cannot be accounted for by simply taking the disjunction of the two constituting sentences:

- (148) Either there is no bathroom here, or it is in a funny place.

- (149) Either John doesn't own a donkey, or he beats it.

This fact is surprising, since these disjunctions seem to be simple variations of ordinary donkey sentences; donkey disjunction (149) is equivalent to our earlier example (102). But there is no easy and straightforward way to improve the DRS language and its interpretation to get better results for these problematic examples.

Exercise 18*

Give DRSs for examples (143)–(146) which correctly present their meaning.

7.4.5 DRT and Compositionality

One of the starting points of model-theoretic semantics is that meaning resides in truth conditions. The notion of truth of DRSs defined in §7.4.4 provides truth conditions for DRSs, and in an indirect way, via their DRS reconstruction, truth conditions for natural language sentences and discourses.

Consider the following pair of examples:

- (150) A man walks in the park.

(151) Not every man does not walk in the park.

The DRSs corresponding to (150) and (151) are (152) and (153), respectively:

(152) $\langle \{x\}, \{\text{MAN}(x), \text{WALK IN THE PARK}(x)\} \rangle$

(153) $\langle \emptyset, \{\neg(\langle \{x\}, \{\text{MAN}(x)\} \rightarrow \neg\langle \emptyset, \{\text{WALK IN THE PARK}(x)\} \rangle)\} \rangle$

Using our abbreviation conventions, the latter can be written as:

(154) $\neg(\langle \{x\}, \text{MAN}(x) \rangle \rightarrow \neg \text{WALK IN THE PARK}(x))$

Some calculation will show that (152) and (153) do indeed have the same truth conditions in DRT, just like the corresponding formulas in predicate logic. Hence, if we identify logical meaning with truth conditions, we should conclude that (150) and (151) have the same logical meaning.

On the other hand, let us consider what happens if we follow each of the sentences (150) and (151) with the sentence *He whistles*. We then have the two following discourses (the first of which we met before; (155) = (92)):

(155) A man walks in the park. He whistles.

(156) Not every man does not walk in the park. He whistles.

Clearly, there is a difference now. Only in case of (155) can we interpret the pronoun in the second sentence as anaphorically linked to a term in the first sentence. This fact is mirrored in the two DRSs (157) and (158) for discourses (155) and (156):

(157) $\langle \{x\}, \{\text{MAN}(x), \text{WALK IN THE PARK}(x), \text{WHISTLE}(x)\} \rangle$

(158) $\langle \emptyset, \{\neg(\langle \{x\}, \{\text{MAN}(x)\} \rightarrow \neg\langle \emptyset, \{\text{WALK IN THE PARK}(x)\} \rangle), \text{WHISTLE}(x)\} \rangle$

The latter can again be abbreviated:

(159) $\{\neg(\langle \{x\}, \{\text{MAN}(x)\} \rightarrow \neg \text{WALK IN THE PARK}(x)), \text{WHISTLE}(x)\}$

While DRS (158) has an empty set of markers, the set of markers of (157) is the nonempty set $\{x\}$. It is precisely this difference which accounts for the fact that in (155) the pronoun in the second sentence can be anaphorically linked to the indefinite term in the first sentence, while such an anaphoric link is not possible in (156). Since in (158) the set of markers $\{x\}$ is inside a condition in its set of conditions, it is unable to bind the variable x in another condition, in this case $\text{WHISTLE}(x)$, in that set.

So despite the fact that (150) and (151) have the same truth conditions, i.e., have the same logical meaning, the difference between (155) and (156) shows that they have a different role in discourse, a different 'discourse meaning'. Hence, to be able to account for this difference, it seems to be essential that sentences (150) and (151) correspond to different DRSs, with different discourse properties.

From this it is a small step to draw the conclusion that the level of discourse *representation* is an essential level in semantics. If the two DRSs (152) and (153) which correspond to the sentences (150) and (151) are different but their logical meaning is the same, it is only their difference in *form*, their difference *as representations*, that can account for their difference in discourse behavior. Associating DRSs to sequences of sentences by means of the DRS construction rules is then an essential element of semantic interpretation which cannot be eliminated.

This conclusion is at odds with the principle of compositionality, the leading principle of Montague grammar. The situation we are confronted with is the following: DRT offers a noncompositional semantic theory which is able to account for certain empirical phenomena which, as we demonstrated in §7.4.2, cannot be accounted for in the compositional semantics offered by Montague grammar. This seems to suggest rather strongly that compositionality has been refuted by the facts.

But how can that be, since in chapter 6 we stressed the fact that compositionality is a methodological principle rather than an empirical hypothesis?

It is the principle of compositionality itself that points the way out of this dilemma. Consider again the two discourses (155) and (156) discussed above. The two discourses differ in meaning. They are both simple sequences of two sentences, and the second sentence is the same in both. Hence, since the meanings of the two discourses differ, compositionality dictates that the first sentences of (155) and (156) differ in meaning. But didn't we see that they have the same truth conditions? Then compositionality shows that their meaning does not reside in their truth conditions.

What notion of meaning can provide a means to make a difference between the two discourses?

In a sense, the required notion of meaning is already implicit in definition 2, in particular, in clause (vi), where the interpretation of DRSs is defined. For note that the basic recursive notion in the semantics of DRSs is that of an assignment h being a truthful embedding for a DRS with respect to an assignment g .

The notion of the truth of a DRS is not the basic recursive semantic notion in DRT; it is a *derived* semantic notion. In definition 3, the truth of a DRS is defined in terms of its embedding conditions. Truth is a global notion here; it is not the notion that oils the wheels of the definition of interpretation. For example, the truth of a condition of the form $\Phi \rightarrow \Psi$ is defined in terms of the verifying embeddings of Φ and Ψ and not in terms of their truth conditions.

Thus, what DRT—properly, i.e., compositionally, interpreted—really shows is that the meaning of a sentence or discourse cannot be identified with its truth conditions but rather resides in the embedding conditions of the DRS into which it translates.

And indeed two DRSs may have the same truth conditions even when their embedding conditions differ. For example, the two DRSs (152) and (153),

which correspond to the opening sentences of discourses (155) and (156) and which have the same truth conditions, differ not only in form but also in their embedding conditions.

So there is no need at all to draw the conclusion that we need a level of representation as an essential level of interpretation. There is no *empirical* reason to abandon compositionality. It is compositionality that leads the way to the conclusion that what we really need is a richer notion of meaning than that of standard semantics.

What all this adds up to is that in principle, nothing stands in the way of a unification of DRT and Montague grammar into one grand compositional theory of discourse meaning, as long as we properly interpret what is going on in DRT. We cannot discuss all the details of such a unification here. For one thing, DRT is a first-order extensional theory, while in Montague Grammar we use a higher-order intensional semantics. Again, this choice of logical framework is dictated by compositionality. So our unifying theory must extend discourse interpretation to such a higher order and intensional logic. Doing this is beyond the scope of this introduction. We will concentrate on the first-order case, just to show the way.

We will do so by comparing, for a few simple examples, the semantic representations in the DRS language with translations in first-order predicate logic. In this comparison, we concentrate on the question of to what extent these two ways of representation can be obtained by means of compositional processes. The answer to this question will be that sometimes DRSs can be obtained in a 'more compositional' way than the corresponding translations in predicate logic, but that in other cases, the DRS construction also leaves something to be desired. We will indicate a way to overcome this lack of compositionality in DRT, which will amount to returning to the language of first-order predicate logic, interpreting it, however, in a different way.

First we consider the simple donkey sentence (160):

(160) If a man walks in the park, he whistles.

Its translation (161) in predicate logic and the corresponding DRS (162) differ essentially in structure:

(161) $\forall x((\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x)) \rightarrow \text{WHISTLE}(x))$

(162) $\langle \emptyset, \{ \{ \{ x \}, \{ \text{MAN}(x), \text{WALK IN THE PARK}(x) \} \} \} \rightarrow \langle \emptyset, \{ \text{WHISTLE}(x) \} \} \rangle$

From a compositional point of view, DRS (162) is a far better representation of sentence (160) than formula (161). The two sentences *a man walks in the park* and *he whistles*, from which sentence (160) is built up, can be recovered in DRS (162) as the sub-DRSs $\langle \{x\}, \{ \text{MAN}(x), \text{WALK IN THE PARK}(x) \} \rangle$ and $\langle \emptyset, \{ \text{WHISTLE}(x) \} \rangle$. This is not the case for the predicate-logical translation (161). In fact, a compositional translation of sentence (160) in predicate logic would lead to formula (163):

$$(163) \exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x)) \rightarrow \text{WHISTLE}(x)$$

But of course (163) is not an appropriate translation of (160). The variable in the consequent is not bound by the existential quantifier in the antecedent; (163) is not equivalent to the correct but noncompositional translation (161). We have seen that the interpretation of the relation of implication in DRT ensures that the set of markers in the antecedent has binding force over variables in the consequent.

As a second example, let us consider once again the simple sequence of two sentences (164) (= (155) = (92)):

$$(164) \text{A man walks in the park. He whistles.}$$

This time, translation (165) (= (94)) in predicate logic and the corresponding DRS (166) (= (157)) have essentially the same structure:

$$(165) \exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x) \wedge \text{WHISTLE}(x))$$

$$(166) \langle \{x\}, \{\text{MAN}(x), \text{WALK IN THE PARK}(x), \text{WHISTLE}(x)\} \rangle$$

Unlike the previous example, the two sentences *a man walks in the park* and *he whistles* from which sentence (164) is built up cannot be recovered in DRS (166) as sub-DRSs. For this to be the case, we would need an operation on DRSs, say \wedge , that would turn two DRSs into a new one. If such an operation were available, sentence (164) could be represented more compositionally as (167):

$$(167) \langle \{x\}, \{\text{MAN}(x), \text{WALK IN THE PARK}(x)\} \rangle \wedge \langle \emptyset, \{\text{WHISTLE}(x)\} \rangle$$

In fact, the structure of (167) is like what the translation of (164) would be in predicate logic if we constructed it compositionally:

$$(168) \exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x)) \wedge \text{WHISTLE}(x)$$

But again, this formula does not give a correct representation of the meaning of (164).

To make sense of the conjunction of the DRSs in (167), we must add (169) as a clause to definition 1:

$$(169) \text{If } \Phi \text{ and } \Psi \text{ are DRSs, then } (\Phi \wedge \Psi) \text{ is a DRS.}$$

And we must add to definition 2 a clause presenting its interpretation.

To find the interpretation of the operator \wedge , we first look again at the interpretation of DRSs. As we have seen, definition 2 defines the interpretation of DRSs in terms of the relational notion $h \models_{\mathbf{M}, g}$: 'h is a verifying embedding of DRS Φ with respect to g'. What this amounts to is that we can take the interpretation of a DRS to be a *relation between assignments of values to reference markers*. In ordinary predicate logic, where we define the notion of a formula as true with respect to an assignment, we can view the meaning of a formula as a *set of assignments*: the assignments under which the formula is true. Similarly, the meaning of a DRS can be taken as a *set of ordered pairs of*

assignments. For example, the set of pairs of assignments which is the interpretation of the simple DRS $\langle\{x\}, \{Fx\}\rangle$ can be written as:

$$(170) \{ \langle g, h \rangle \mid h[x]g \text{ \& } h(x) \in I(F) \}$$

Note that the order of the pairs in (170) is the reverse of that in the notion $h \models_{M,g}$. The reason for this is that it makes sense to look at such pairs in input-output terms. With respect to an input assignment g , the output of the procedure of interpreting $\langle\{x\}, \{Fx\}\rangle$ is those assignments h which differ from g at most in that they assign an object to x such that the object belongs to the interpretation of the predicate F .

From this perspective, the task of finding an interpretation of $\Phi \wedge \Psi$ amounts to specifying the input-output relation of $\Phi \wedge \Psi$ in terms of the input-output relations associated with Φ and with Ψ . A candidate for the interpretation of \wedge that suggests itself almost immediately is the following: If h is to be a possible output for $\Phi \wedge \Psi$ with respect to g as input, then there has to be some k such that k is a possible output for Φ with respect to g as input, and h is a possible output for Ψ with respect to k as input. In terms of the notion $h \models_{M,g}$ of definition 2, this amounts to adding the following clause:

$$(171) \begin{array}{l} h \models_{M,g} \Phi \wedge \Psi \text{ iff there is some } k \text{ such that } k \models_{M,g} \Phi \text{ and} \\ h \models_{M,k} \Psi. \end{array}$$

For example (164) this means that we can represent it by means of the conjunction of DRSs (167). And under clause (171) for the interpretation of sequencing, the interpretation that (167) gets is the same as that of the original DRS (157). So we get the DRT way of interpreting sequences of sentences with an indefinite term in the first sentence and a pronoun in the second, but we can now represent it by means of a DRS in which the compounding sentences of (164) are represented by sub-DRSs in (167).

As our last example in discussing compositionality, we turn again to donkey sentence (172) (= (103)):

(172) Every farmer who owns a donkey beats it.

Sentence (172) gets the same translation (174 (= 105)) in predicate logic as (173) (= (120)), and both are represented by the same DRS (175) (= (127)):

(173) If a farmer owns a donkey, he beats it.

(174) $\forall x \forall y ((\text{FARMER}(x) \wedge \text{DONKEY}(y) \wedge \text{OWN}(x, y)) \rightarrow \text{BEAT}(x, y))$

(175) $\langle \emptyset, \{ \langle \{x, y\}, \{ \text{FARMER}(x), \text{DONKEY}(y), \text{OWN}(x, y) \} \rangle \rightarrow \langle \emptyset, \{ \text{BEAT}(x, y) \} \rangle \} \rangle$

As compared to the previous example, we find in this case a more dramatic breach of compositionality in the translation in predicate logic. Sentence (172) contains an indefinite term *a donkey* which normally translates as an existentially quantified phrase. In translation (174), however, we are forced to

associate it with universal quantification. Further, this quantifier has to be given wide scope over the implication as a whole, whereas the indefinite term *a donkey* occurs inside the relative clause which is part of the subject term of (172) and hence from a compositional point of view should belong inside the antecedent.

But similarly, no subexpression can be found in DRS (175) which corresponds to the common phrase *farmer who owns a donkey* which forms a constituent in sentence (172). And the same holds true for the intransitive verb phrase *owns a donkey* in sentence (173). A DRS that would correspond to the latter is (176):

$$(176) \langle \{y\}, \{\text{DONKEY}(y), \text{OWN}(x, y)\} \rangle$$

To get from here to *farmer who owns a donkey*, we can use our new operator \wedge :

$$(177) \langle \emptyset, \{\text{FARMER}(x)\} \rangle \wedge \langle \{y\}, \{\text{DONKEY}(y), \text{OWN}(x, y)\} \rangle$$

The verb phrase *beats it* can be associated with DRS (178):

$$(178) \langle \emptyset, \{\text{BEATS}(x, y)\} \rangle$$

What we need now is an operation to combine DRS (177), which corresponds to the common noun phrase of sentence (172), with DRS (178), which corresponds to its verb phrase. This operation has to turn these two DRSs into an implication which has (177) as its antecedent and (178) as its consequent. Moreover, the reference marker x should get bound, i.e., the antecedent should in addition contain a set of markers $\{x\}$. DRS (179) would do, in the sense that it would get the right interpretation; it has the same embedding conditions as the original DRS (133):

$$(179) (\langle \{x\}, \{\text{FARMER}(x)\} \rangle \wedge \langle \{y\}, \{\text{DONKEY}(y), \text{OWN}(x, y)\} \rangle) \rightarrow \langle \emptyset, \{\text{BEAT}(x, y)\} \rangle$$

But this result can only be obtained by replacing the empty set in the first conjunct of (177) with the set $\{x\}$. This is a kind of syntactic maneuvering which is, to say the least, hard to interpret semantically. It is not the kind of move that is allowed in a compositional framework. It forces you to break into a structure that has already been built up. It would be preferable if we could simply prepose $\{x\}$ to the antecedent:

$$(180) \{x\}(\langle \emptyset, \{\text{FARMER}(x)\} \rangle \wedge \langle \{y\}, \{\text{DONKEY}(y), \text{OWN}(x, y)\} \rangle) \rightarrow \langle \emptyset, \{\text{BEATS}(x, y)\} \rangle$$

But this is not in accordance with clause (vi) of definition 1 of the syntax of DRSs and conditions. According to that clause, the set of markers $\{x\}$ should combine with a set of conditions and not with a DRS.

In fact, the moment we can connect a set of markers with a DRS to form a new and more complex DRS, we can define an *iterative* notion of quantification, connecting a singleton set $\{x\}$ to a DRS, instead of the noniterative no-

tion of definition 1, which in one step combines a set $\{x_1, \dots, x_n\}$, for $n \geq 0$, to (a set of) conditions. The latter was needed in definition 1 because adding a set of markers changes syntactic status: it turns conditions into DRSs. An iterative notion of DRS quantification applies a singleton set $\{x\}$ to a DRS Φ , which results in a DRS $\{x\}\Phi$, to which another singleton set, say $\{y\}$, can in turn be preposed, resulting in $\{y\}\{x\}\Phi$. Of course, if quantification is defined in this way, we can just as well return to the familiar quantifier $\exists x$ and do away with sets of markers.

It then makes sense to drop the syntactic notion of a condition altogether. (Or alternatively, to add an operation which turns a condition into a DRS; but we will choose the former option here.) Then we no longer need sets of conditions; we replace them with conjunctions of DRSs, using the notion of conjunction already defined.

What all this amounts to is that the syntax of DRSs can be made identical to the syntax of ordinary first-order predicate logic. With one exception so far, we still lack the universal quantifier; but we shall see that it can be introduced and defined in terms of the existential quantifier and negation in the usual way.

To get the effects of DRT, we only have to adapt the semantics. Instead of defining the notion of 'formula ϕ is true with respect to assignment g ', we define the notion 'assignment h is a verifying embedding for a formula ϕ with respect to assignment g '. The system that thus results is called *dynamic predicate logic*. (See Groenendijk and Stokhof 1988a, 1990 for a more detailed introduction.)

We then arrive at the following definition for the semantic interpretation of dynamic predicate logic (DPL).

Definition 4:

- (i) $h \models_{M,g} P(t_1, \dots, t_n)$ iff $h = g$ and $\langle [t_1]_{M,h}, \dots, [t_n]_{M,h} \rangle \in F_M(P)$
- (ii) $h \models_{M,g} t = t'$ iff $h = g$ and $[t]_{M,h} = [t']_{M,h}$;
- (iii) $h \models_{M,g} \neg \phi$ iff $h = g$ and there is no k such that $k \models_{M,h} \phi$;
- (iv) $h \models_{M,g} (\phi \wedge \psi)$ iff there is a k : $k \models_{M,g} \phi$ and $h \models_{M,k} \psi$;
- (v) $h \models_{M,g} (\phi \rightarrow \psi)$ iff $h = g$ and for all k : if $k \models_{M,h} \phi$, then there is a j such that $j \models_{M,k} \psi$;
- (vi) $h \models_{M,g} (\phi \vee \psi)$ iff $h = g$ and there is ak : $k \models_{M,h} \phi$ or $k \models_{M,h} \psi$;
- (vii) $h \models_{M,g} \exists x \phi$ iff there is a k such that $k[x]g$ and $h \models_{M,k} \phi$;
- (viii) $h \models_{M,g} \forall x \phi$ iff $h = g$ and for all k : if $k[x]h$, then there is a j such that $j \models_{M,k} \phi$.

Clauses (i)–(iii) and (v)–(vi) are essentially the same as in definition 2. Clause (iv) introduces DRS conjunction as it was discussed above. In (vii) we find the iterative notion of DRS quantification. Clause (viii) introduces a new feature, universal quantification. As we already indicated, $\forall x \phi$ can be defined in the usual way as $\neg \exists x \neg \phi$. As the reader can verify, the following holds:

$$(181) \text{ For all } M, g, \text{ and } h: h \models_{M,g} \forall x \phi \text{ iff } h \models_{M,g} \neg \exists x \neg \phi.$$

The formulas $\forall x\phi$ and $\neg\exists x\neg\phi$ are semantically equivalent in the strong sense that they have the same embedding conditions. It should be noted, though, that the existential quantifier cannot be defined in terms of negation and the universal quantifier: $\exists x\phi$ and $\neg\forall x\neg\phi$ are not equivalent in the strong sense that they have the same embedding conditions. For example, the formulas $\exists xFx$ and $\neg\forall x\neg Fx$ are assigned the following embedding conditions by definition 4:

$$(182) h \models_{\mathbf{M},g} \exists xFx \text{ iff } h[x]g \text{ and } h(x) \in I(F).$$

$$(183) h \models_{\mathbf{M},g} \neg\forall x\neg Fx \text{ iff } h = g \text{ and for some } k \text{ such that } k[x]h: k(x) \in I(F).$$

In both cases it is required that some individual belong to the interpretation of F . But in (182), output assignments h will possibly differ from g in that they assign such an individual to x , while in (183) input and output assignments are required to be the same.

This makes all the difference if we add a new conjunct which contains the same variable x again, say Gx , to each of the two formulas. The resulting formulas: $\exists xFx \wedge Gx$ and $\neg\forall x\neg Fx \wedge Gx$, get the following embedding conditions:

$$(184) h \models_{\mathbf{M},g} \exists xFx \wedge Gx \text{ iff } h[x]g \text{ and } h(x) \in I(F) \text{ and } h(x) \in I(G).$$

$$(185) h \models_{\mathbf{M},g} \neg\forall x\neg Fx \wedge Gx \text{ iff } h = g \text{ and for some } k \text{ such that } k[x]h: k(x) \in I(F), \text{ and } h(x) \in I(G).$$

In case of (184), we find that the variable x in the second conjunct is still bound by the existential quantifier in the first conjunct. That is, we find that $\exists xFx \wedge Gx$ has the same embedding conditions as $\exists x(Fx \wedge Gx)$. In fact, the following holds in DPL:

$$(186) \text{ For all } \mathbf{M}, g, \text{ and } h: h \models_{\mathbf{M},g} \exists x\phi \wedge \psi \text{ iff } h \models_{\mathbf{M},g} \exists x(\phi \wedge \psi).$$

In (185), however, the variable in the second conjunct Gx is not bound by the universal quantifier. Condition (185) requires that $h(x)$, which must be the same as $g(x)$, is an element of $I(G)$.

Of course these facts are in accordance with our discussion earlier in this section of the difference between the two discourses (187) (= (164)) and (188) (= (156)):

(187) A man walks in the park. He whistles.

(188) Not every man does not walk in the park. He whistles.

These two discourses translate into the following two DPL formulas:

$$(189) \exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x)) \wedge \text{WHISTLE}(x)$$

$$(190) \neg\forall x(\text{MAN}(x) \rightarrow \neg\text{WALK IN THE PARK}(x)) \wedge \text{WHISTLE}(x)$$

As we have just seen, the first conjuncts of (189) and (190) are not equivalent in DPL, precisely because the existential quantifier in (189) has binding force over the second conjunct of (189) as well. In DPL, (189) is equivalent to (191):

$$(191) \exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x) \wedge \text{WHISTLE}(x))$$

This does not hold for (190).

Of course the truth conditions for the first conjuncts of (189) and (190) are the same, even though their full meaning is different. The truth definition remains essentially the same as it was in DRT:

Definition 5:

A formula ϕ is true in a model \mathbf{M} with respect to an assignment g , $\models_{\mathbf{M},g}\phi$, iff there is an assignment h such that $h \models_{\mathbf{M},g}\phi$.

Returning to our two simple examples again, given the embedding conditions (182) of $\exists xFx$ and (183) of $\neg\forall x\neg Fx$, it can easily be checked that their truth conditions are the same according to definition 5. In fact, it holds quite generally that:

$$(192) \text{ For all } \mathbf{M} \text{ and } g: \models_{\mathbf{M},g}\exists x\phi \text{ iff } \models_{\mathbf{M},g}\neg\forall x\neg\phi.$$

Another important fact about DPL is related to the interpretation of donkey sentences. Formulas (196), (197), and (198) are the DPL translations of donkey sentences (193), (194), and (195), respectively (= (160), (172), (173), respectively):

(193) If a man walks in the park, he whistles.

(194) Every farmer who owns a donkey beats it.

(195) If a farmer owns a donkey, he beats it.

$$(196) \exists x(\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x)) \rightarrow \text{WHISTLE}(x)$$

$$(197) \forall x((\text{FARMER}(x) \wedge \exists y(\text{DONKEY}(y) \wedge \text{OWN}(x, y))) \rightarrow \text{BEAT}(x, y))$$

$$(198) \exists x(\text{FARMER}(x) \wedge \exists y(\text{DONKEY}(y) \wedge \text{OWN}(x, y))) \rightarrow \text{BEAT}(x, y)$$

Of course (196)–(198) would not do as translations in ordinary predicate logic. The essential fact which determines their correctness as translations in DPL is best illustrated by the simplest of the three examples. Formula (196) is an implication with an existentially quantified antecedent and a consequent in which a ‘free’ variable occurs. In fact, the variable isn’t free at all in DPL; it is bound by the existential quantifier in the antecedent. Since we have a universal quantifier at our disposal, this fact can be stated in the following way. In DPL formula (196) has exactly the same interpretation as (199), which is the usual translation of donkey sentence (193) in predicate logic:

$$(199) \forall x((\text{MAN}(x) \wedge \text{WALK IN THE PARK}(x)) \rightarrow \text{WHISTLE}(x))$$

Quite generally, we have the following equivalence:

$$(200) \text{ For all } \mathbf{M}, g, \text{ and } h: h \models_{\mathbf{M},g} \exists x \phi \rightarrow \psi \text{ iff } h \models_{\mathbf{M},g} \forall x (\phi \rightarrow \psi).$$

Unlike in predicate logic, the equivalence of $\exists x \phi \rightarrow \psi$ and $\forall x (\phi \rightarrow \psi)$ holds irrespective of whether or not ψ contains free occurrences of x . Of course according to (200), formulas (197) and (198) are equivalent as well, and further, they are equivalent to (201), which is their correct translation in ordinary predicate logic:

$$(201) \forall x \forall y ((\text{FARMER}(x) \wedge \text{DONKEY}(y) \wedge \text{OWN}(x, y)) \rightarrow \text{BEAT}(x, y))$$

We end this section with the remark that apart from offering a more orthodox logical system which provides a better tool for compositional semantic analysis of natural language discourses, DPL and DRT have the same empirical impact. In particular, the problematic cases of anaphoric relations discussed at the end of §7.4.4 are equally problematic in the DPL framework. One needs an essentially richer dynamic semantics, a more dynamic semantics, to handle these problematic phenomena. (See Groenendijk and Stokhof 1988a.)

Exercise 19*

Consider again example (96), which was discussed in §7.4.2 and is the subject of exercise 15:

Exactly one boy walks in the park. He whistles.

Give a correct translation of (96) in (i) predicate logic; (ii) in the DRS language of definition 1; (iii) in dynamic predicate logic. Compare (i)–(iii) with respect to their ‘degree of compositionality’.

Exercise 20*

Consider the following alternative for the semantic interpretation of disjunction:

$$h \models_{\mathbf{M},g} \phi \vee \psi \text{ iff } h \models_{\mathbf{M},g} \phi \text{ or } h \models_{\mathbf{M},g} \psi$$

- (i) Discuss the differences between this interpretation of disjunction and its interpretation as given in clause (vi) of definition 4.
- (ii) Is it possible to account for examples (148) and (149) of §7.4.4 under this interpretation of disjunction?
- (iii) Try to find a typical example of a sequence of sentences which exhibits the kind of anaphoric relation that can be accounted for on the basis of this alternative interpretation of disjunction.

Exercise 21*

In ordinary predicate logic, it is possible to start from a minimal set of connectives and quantifiers and to define the others in terms of them. For example, \wedge , \vee , and \exists can be defined in terms of \neg , \forall , and \rightarrow . Determine such a minimal set for dynamic predicate logic.

Exercise 22*

Several notions of entailment are feasible for dynamic predicate logic (and DRT). Consider the following three alternatives:

- (a) $\phi \models_a \psi$ iff for all \mathbf{M}, g and h : if $h \models_{\mathbf{M}, g} \phi$ then $h \models_{\mathbf{M}, g} \psi$.
 - (b) $\phi \models_b \psi$ iff for all \mathbf{M}, g : if $\models_{\mathbf{M}, g} \phi$ then $\models_{\mathbf{M}, g} \psi$.
 - (c) $\phi \models_c \psi$ iff for all \mathbf{M}, g and h : if $h \models_{\mathbf{M}, g} \phi$ then $\models_{\mathbf{M}, h} \psi$.
- (i) Determine for which of these three notions of entailment it holds that:
 $\exists x Fx \models Fx$; $\exists x Fx \models \exists y Fy$.
 - (ii) Determine for which of the three notions it holds that:
 $\phi \models \psi$ iff $\models \phi \rightarrow \psi$.
 - (iii) In ordinary predicate logic, the entailment relation is reflexive and transitive. Does this hold for the three notions defined above? If not, present a counterexample.

7.4.6 Conclusion

The main conclusions to be drawn from the foregoing discussion are the following. First of all, it has been amply shown that the change of DRT from a static and sentential toward a dynamic and discourse semantics is a very successful move. Second, we have seen that some of the distinctive features of DRT, the unorthodox DRS language and in particular its postulate of an intermediate level of semantic representation between natural language syntax and semantic interpretation, are not necessary ingredients for its empirical success. Instead we can use the language of first-order predicate logic interpreted in the usual compositional manner, if only we use a richer dynamic notion of meaning.

Dynamic predicate logic is not simply a more orthodox notational variant of DRT; it enables us, to the extent that this is possible in a first-order language, to get straightforward compositional translations of the sentences that exemplify the empirical import of DRT.

Since DPL is an orthodox compositional logical framework, it does not seem to be a major task to transfer its dynamic interpretation to higher-order, intensional languages such as the one used in Montague grammar, thus aiming at a unification of discourse representation theory and Montague grammar.

And Montague grammar may also benefit from such an undertaking. It was shown in §7.4.2. that the mechanism of the quantification rules is not adequate to deal with the anaphoric relations to which DRT addresses itself. In fact, the DPL translations we get for donkey sentences and the DPL treatment of other anaphoric relations inside and outside sentence boundaries strongly suggest that we do not need the quantification mechanism to account for anaphoric relations. The translations DPL offers look like the ones we get on a direct construction, where the proper bindings are established by the dynamics of the interpretation mechanism instead of by the mechanism of quantifica-

tion rules. This robs the latter of one of their two functions: accounting for anaphoric relations. The sole function of the quantification rules that remains is accounting for scope ambiguities. (But as we indicated in §6.6, alternatives have been developed that deal with scope ambiguities in a different way as well.)

Let us finally remind the reader that the kinds of phenomena that have been studied in the DRT framework are not limited to what has been the main focus here, viz., anaphoric relations. A first example of another field of application is studies of tense and aspect. In Hinrichs 1986, Kamp 1981a, Kamp and Rohrer 1983, and Partee 1984a, it is argued that the role of tense and aspect in discourse is an important aspect of the semantics of tense and aspect, and that the dynamic approach to meaning sheds a new light on their analysis.

Another phenomenon that has been approached within the DRT framework is the semantic analysis of belief sentences and other propositional attitude reports. In Asher 1986, 1987 and Zeevat 1987, it is argued that the representational philosophy behind DRT provides a better means to deal with many of the long-standing problems in this field. So, this is an area where the representationalism of DRT is considered to contribute positively to our understanding of the phenomena in question.

Besides tense and aspect and propositional attitudes, other phenomena have also been addressed in DRT, such as the discourse impact of many other terms and determiners besides the small group discussed above and more complex anaphoric relations in intensional contexts. (See Kadmon 1987; Roberts 1987, 1989; van Eyck 1985.)

Solutions to Selected Exercises

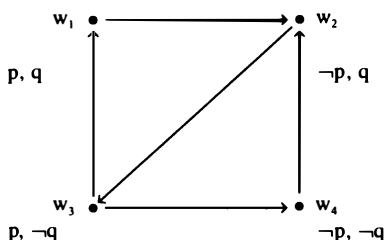
Chapter 2

Exercise 1

- (a) $\Diamond \neg p \wedge \neg \Box \neg p$ key: p : you understand me.
 (b) $\Diamond p \rightarrow \Box \Diamond p$ key: p : it is raining
 (c) $\Diamond(\Diamond p \rightarrow p)$ key: p : it is raining
 (d) $\Diamond \Box p \rightarrow \Box p$ key: p : it is raining
 (e) $\Diamond p \wedge \Diamond \Box p$ key: p : it is raining (*this = it is raining*)
 or: $\Diamond p \wedge \Diamond \Box \Diamond p$ key: p : it is raining (*this = it may be raining*)

Exercise 2

- (a) (i) in w_1 : $V_{w_1}(\Box p) = 0$ because $V_{w_2}(p) = 0$ and $w_1 R w_2$. So it follows that $V_{w_1}(\Box p \rightarrow \Box \Box p) = 1$; in w_2 : $V_{w_2}(\Box p) = 1$, for $V_{w_1}(p) = 1$ and only w_1 is accessible from w_2 , and $V_{w_2}(\Box \Box p) = 0$ because $V_{w_1}(\Box p) = 0$ and $w_2 R w_1$. So $V_{w_2}(\Box p \rightarrow \Box \Box p) = 0$. Since $V_{w_2}(\Box p \rightarrow \Box \Box p) = 0$, $\Box p \rightarrow \Box \Box p$ does not hold in \mathbf{M} .
 (ii) in w_1 : $V_{w_1}(\neg \Box p) = 1$ because $V_{w_1}(\Box p) = 0$; in w_2 : $V_{w_2}(\neg \Box p) = 0$ because $V_{w_2}(\Box p) = 1$. Thus, in model \mathbf{M} , $\neg \Box p$ is not valid.
 (iii) $V_{w_1}(\Diamond p) = 1$ because $V_{w_1}(p) = 1$ and $w_1 R w_1$, and $V_{w_2}(\Diamond p) = 1$ because $V_{w_1}(p) = 1$ holds and $w_2 R w_1$. This means that both $V_{w_1}(\Box \Diamond p) = 1$ and $V_{w_2}(\Box \Diamond p) = 1$ obtain, and hence both $V_{w_1}(p \rightarrow \Box \Diamond p) = 1$ and $V_{w_2}(p \rightarrow \Box \Diamond p) = 1$. So, in \mathbf{M} , $p \rightarrow \Box \Diamond p$ is valid.
- (b) (i) See the figure:



- (ii) 1. $V_{w_2}(q) = 1$ implies that $V_{w_1}(\Box q) = 1$, because only w_2 is accessible from w_1 .

2. $V_{w_3}(\neg(p \rightarrow q)) = 1$ implies $V_{w_2}(\Box \neg(p \rightarrow q)) = 1$ because only w_3 is accessible from w_2 .
 3. $V_{w_1}(p \wedge q) = 1$ implies that $V_{w_1}((p \wedge q) \vee (\neg p \wedge \neg q)) = 1$, and $V_{w_4}(\neg p \wedge \neg q) = 1$ implies that $V_{w_4}((p \wedge q) \vee (\neg p \wedge \neg q)) = 1$. So $V_{w_3}(\Box((p \wedge q) \vee (\neg p \wedge \neg q))) = 1$, because only w_1 and w_4 are accessible from w_3 .
 4. $V_{w_3}(p) = 1$, so $V_{w_2}(\Box p) = 1$, because only w_3 is accessible from w_2 , so $V_{w_1}(\Diamond \Box p) = 1$, because $w_1 R w_2$.
 5. $V_{w_2}(p) = 0$, so $V_{w_1}(\Diamond p) = 0$, because only w_2 is accessible from w_1 , so $V_{w_1}(\Diamond p \wedge \Diamond q) = 0$.
- (iii) 1. $V_w(\Box p) = 1$ iff $w = w_2$. Therefore $V_w(\Diamond \Box p) = 1$ iff $w = w_1$ or $w = w_4$, because w_2 is accessible only from w_1 and w_4 . From this it follows that $V_w(\Diamond \Diamond \Box p) = 1$ iff $w = w_3$, because both w_1 and w_4 are accessible only from w_3 . So $V_{w_2}(\Diamond \Box p \vee \Diamond \Diamond \Box p) = 0$ and $\Diamond \Box p \vee \Diamond \Diamond \Box p$ is not valid on \mathbf{M} .
2. $V_w(\Box p) = 1$ iff $w = w_2$. Also: $V_{w_2}(\neg p) = 1$. So for all $w \in W$, $V_w(\Box p \rightarrow \neg p) = 1$. In \mathbf{M} , $\Box p \rightarrow \neg p$ is valid.
 3. $V_{w_1}(p) = 1$ and $V_{w_1}(\Diamond p) = 0$, since $V_{w_2}(p) = 0$ and only w_2 is accessible from w_1 . It follows that $V_{w_1}(p \rightarrow \Diamond p) = 0$ and hence that $V_{w_1}((p \rightarrow \Diamond p) \wedge (q \rightarrow \Diamond q)) = 0$. In \mathbf{M} , $(p \rightarrow \Diamond p) \wedge (q \rightarrow \Diamond q)$ is not valid.
 4. The formula $p \vee \neg p$ is a tautology. So, $V_w(\Diamond(p \vee \neg p)) = 1$ for all $w \in W$ for which there is a $w' \in W$ such that $w R w'$. On the other hand, $V_{w_1}(\Box(p \vee \neg q)) = 0$, for $V_{w_2}(p \vee \neg q) = 0$ and $w_1 R w_2$. From this it follows that $V_{w_1}(\Diamond(p \vee \neg p) \rightarrow \Box(p \vee \neg q)) = 0$. So $\Diamond(p \vee \neg p) \rightarrow \Box(p \vee \neg q)$ is not valid in \mathbf{M} .
- (iv) 1. Suppose $V'_w(\Box p) = 1$ for some $w \in W$ and an arbitrary V' on the frame. This means that $V'_{w'}(p) = 1$ for all w' which are accessible from w . As there accessible worlds for any w , there will be such a w' . Hence $V'_{w'}(\Diamond p) = 1$. So $\Box p \rightarrow \Diamond p$ is valid on the frame.
2. Suppose $V'_w(\Diamond \Diamond \Box p) = 1$ for some $w \in W$ and an arbitrary V' on the frame. Then there are a w' and a w'' such that $w R w'$, $w' R w''$, and $V'_{w'}(\Box p) = 1$. To prove the validity of $\Diamond \Diamond \Box p \rightarrow p$, it is sufficient to show that $w'' R w$ must hold. (If this is true, $V'_w(p) = 1$ follows from $V'_{w'}(\Box p) = 1$). This is indeed the case:
 - if $w = w_1$, then it must be the case that $w' = w_2$ and $w'' = w_3$;
 - if $w = w_2$, then it must be the case that $w' = w_3$ and $w'' = w_4$ or $w'' = w_1$;
 - if $w = w_3$, then it must be the case that $w' = w_1$ or $w' = w_4$ and $w'' = w_2$;
 - if $w = w_4$, then it must be the case that $w' = w_2$ and $w'' = w_3$.
 In all these cases $w'' R w$ holds, so $\Diamond \Diamond \Box p \rightarrow p$ is valid in the frame.

Exercise 3

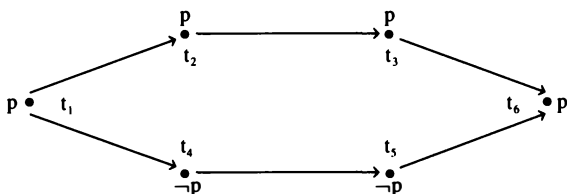
- (a) (i) Suppose \mathbf{M} is a model with an underlying symmetric frame. Suppose now that $V_w(\Diamond\Box\phi) = 1$. It must be shown that $V_w(\phi) = 1$. From $V_w(\Diamond\Box\phi) = 1$ it follows that for some w' such that wRw' : $V_{w'}(\Box\phi) = 1$. From the symmetry of R it follows that $w'Rw$. This fact, together with the fact that $V_{w'}(\Box\phi) = 1$, implies that $V_w(\phi) = 1$.
- (ii) Suppose now that the relation R on a frame is not symmetric. Then there are worlds w_1 and w_2 such that w_1Rw_2 , whereas w_2Rw_1 does not obtain. Now on that frame we define a model by stipulating that $V_w(p) = 1$ if and only if w_2Rw . Then $V_{w_2}(\Box p) = 1$ and so $V_{w_1}(\Diamond\Box p) = 1$. However, $V_{w_1}(p) = 0$, and so it follows that $V_{w_1}(\Diamond\Box p \rightarrow p) = 0$. In this model, $\Diamond\Box p \rightarrow p$ is not valid.
- (b) We only give the outcomes. On frames, $\Diamond\Diamond\Diamond\phi \rightarrow \phi$ corresponds to $\forall w\forall v_1\forall v_2((Rwv_1 \wedge Rv_1v_2) \rightarrow v_2 = w)$; $\Diamond\Diamond\Diamond\Diamond\phi \rightarrow \phi$ corresponds to $\forall w\forall v_1\forall v_2\forall v_3((Rwv_1 \wedge Rv_1v_2 \wedge Rv_2v_3) \rightarrow v_3 = w)$. Then the generalization to the case with arbitrary sequences $\Diamond_1, \dots, \Diamond_n$ will be obvious.

Exercise 7

- | | |
|--|--|
| (a) $p \wedge F\neg p$ | key: p : You are young. |
| (b) $p \wedge Gp$ | key: p : I am faithful to you. |
| (c) $Pp \wedge Pq$ | key: p : John reads <i>War and Peace</i> .
q : Charles reads <i>War and Peace</i> . |
| (d) $P(p \wedge Pq)$ | key: p : Mary enters in.
q : John puts the whiskey bottle
in the refrigerator. |
| (e) $P(p \wedge Fq)$ | key: as in (d) |
| (f) $(Fp \vee \neg Fp) \wedge (Fp \rightarrow HFp)$ | key: p : A sea battle is fought. |
| (g) $Gq \rightarrow Gp$ (or: $Gq \leftrightarrow Gp$) | key: p : You are with me.
q : I am really happy. |

Exercise 8

- (a) (i) See the figure:



- (ii) 1. Since $V_t(\neg p) = 1$ only in case $t = t_4$ and $t = t_5$, we need to determine $V_t(FGp)$ for these values of t only. Now $V_{t_5}(Gp) = 1$ because $V_{t_6}(p) = 1$ and t_6 is the only t later than t_5 . Also $V_{t_4}(Gp) = 1$

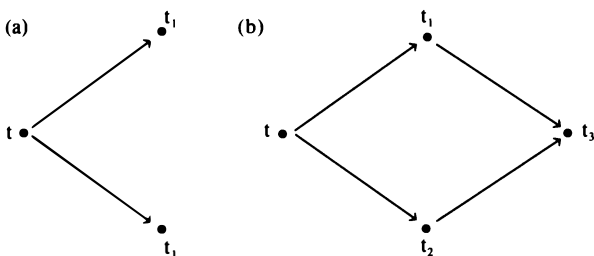
since there are no t such that t_6Rt . So $V_{t_4}(FGp) = 1$ and $V_{t_5}(FGp) = 1$ and $\neg p \rightarrow FGp$ is valid in the model.

2. $V_{t_4}(F\neg p) = 1$ since $V_{t_5}(p) = 0$ and t_4Rt_5 , but $V_{t_4}(FF\neg p) = 0$ because $\neg p$ is not true at t_6 . Hence $F\neg p \rightarrow FF\neg p$ is not valid in the model.

3. $V_{t_6}(P\neg p \rightarrow \neg p) = 0$, since $V_{t_6}(P\neg p) = 1$ and $V_{t_6}(\neg p) = 0$. It follows that, for example, $V_{t_5}(G(P\neg p \rightarrow \neg p)) = 0$ because t_5Rt_6 , and so $G(P\neg p \rightarrow \neg p)$ is not valid on the model.

4. $V_{t_6}(p \wedge Gp) = 1$, whereas $V_{t_6}(Hp) = 0$ because t_5Rt_6 and $V_{t_5}(p) = 0$. It follows that $(p \wedge Gp) \rightarrow Hp$ is not valid in the model.

(b) (i) Validity of $FG\phi \rightarrow GF\phi$ in a frame F means with respect to the time axis T of F that if two points of time are later than a given point in time, there will always be a point in time which is later than the two points; i.e., if tRt_1 and tRt_2 , then there is a t_3 in T such that t_1Rt_3 and t_2Rt_3 . In other words, a configuration like figure a. can always be extended to a configuration as sketched in figure b.



Assume first that this property holds for F and that $V_t(FG\phi) = 1$ for some $t \in T$ in a model on F . Then there is a $t_1 \in T$ such that tRt_1 , and for every t' with t_1Rt' it is the case that $V_{t'}(\phi) = 1$. Now take an arbitrary t_2 with tRt_2 . As assumed, there is a t_3 with t_1Rt_3 and t_2Rt_3 . Since t_1Rt_3 , $V_{t_3}(\phi) = 1$, and because t_2Rt_3 obtains, $V_{t_2}(F\phi) = 1$ holds. Since t_2 was arbitrarily chosen with the property tRt_2 , it follows that $V_t(GF\phi) = 1$. This means that if the given relational property holds for F , $FG\phi \rightarrow GF\phi$ is valid on every model in F .

Suppose now that the property does not hold for F . Then there are points of time $t, t_1, t_2 \in T$ of F such that tRt_1 and tRt_2 , whereas there is no t_3 such that both t_1Rt_3 and t_2Rt_3 . Now define a model on F by stipulating that $V_{t'}(p) = 1$ iff t_1Rt' . Then $V_{t_1}(Gp) = 1$ and so $V_t(FGp) = 1$. On the other hand, $V_{t_2}(Fp) = 0$ because there is no t' with t_2Rt' such that $V_{t'}(p) = 1$. So $V_t(GFp) = 0$. Hence $FGp \rightarrow GFp$ is not valid in this model, and hence it also fails in F .

(ii) Validity of $G(\phi \wedge \neg\phi) \vee FG(\phi \wedge \neg\phi)$ in a frame F means that on the time axis of the frame every point in time is followed by a final point (except for the latter point itself). We pointed out in the text that t is a final point in time iff $V_t(G(\phi \wedge \neg\phi)) = 1$. This implies

that $V_i(\mathbf{FG}(\phi \wedge \neg\phi)) = 1$ iff t is followed by a final point and $V_i(\mathbf{G}(\phi \wedge \neg\phi) \vee \mathbf{FG}(\phi \wedge \neg\phi)) = 1$ iff t is a final point or is followed by one.

- (iii) Validity of $\mathbf{PP}\phi \rightarrow \mathbf{P}\phi$ in a frame means that the accessibility relation R of the frame is transitive. To see this, suppose first that R is transitive and that $V_i(\mathbf{PP}\phi) = 1$. Then there will be a t' and a t'' with $t''Rt'$ and $t'Rt$ such that $V_{t'}(\phi) = 1$. Since R is transitive, $t''Rt$ also obtains and hence $V_i(\mathbf{P}\phi) = 1$. So for every model in a frame with a transitive accessibility relation, $\mathbf{PP}\phi \rightarrow \mathbf{P}\phi$ is valid.

Suppose now, on the other hand, that R is not transitive. In that case there are t'' , t' , and t with $t''Rt'$ and $t'Rt$, while $t''Rt$ does not hold. Now define a model by stipulating that $V_{t'}(p) = 1$; but V makes p false everywhere else. Then $V_i(\mathbf{PP}p) = 1$ indeed holds, whereas $V_i(\mathbf{P}p) = 0$. So $\mathbf{PP}p \rightarrow \mathbf{P}p$ is not valid in this model.

Chapter 3

Exercise 1

Keys and domains will be left implicit in the following solution. Furthermore, we abbreviate $\mathbf{H}\phi \wedge \phi \wedge \mathbf{G}\phi$ as $\mathbf{A}\phi$ (*always* ϕ) and $\mathbf{P}\phi \vee \phi \vee \mathbf{F}\phi$ as $\mathbf{E}\phi$ (*sometimes* ϕ). It will be clear that $\mathbf{A}\phi$ is equivalent to $\neg\mathbf{E}\neg\phi$, and $\mathbf{E}\phi$ with $\neg\mathbf{A}\neg\phi$.

- (a) $\Diamond\mathbf{FW}1w$
 (b) $\Diamond\forall x(\mathbf{H}x \rightarrow \mathbf{H}\exists y\mathbf{A}xy)$
 (c) $\Diamond\exists y\forall x(\mathbf{H}x \rightarrow \mathbf{H}\mathbf{A}xy)$
 (d) $\forall x\forall y(x \neq y \rightarrow \Diamond\mathbf{S}xy) \rightarrow \forall x\Diamond\forall y(x \neq y \rightarrow \mathbf{S}xy)$
 (e) $\forall x(\exists y\Diamond\mathbf{A}\mathbf{F}xy \wedge \forall y\Diamond\mathbf{E}\mathbf{F}xy \wedge \neg\Diamond\forall y\mathbf{A}\mathbf{F}xy)$

(You is interpreted here as *everyone*. This is just one possible reading.)

- (f) $\exists x(\forall y(\mathbf{P}y \leftrightarrow y = x) \wedge \mathbf{G}\mathbf{D}x)$

(The present president will always be a democrat)

and: $\mathbf{G}\exists x(\forall y(\mathbf{P}y \leftrightarrow y = x) \wedge \mathbf{D}x)$

(Whoever will be the president, he or she will be a democrat.)

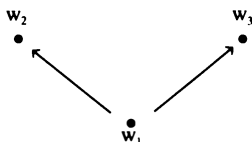
- (g) $\forall x(\mathbf{S}x \rightarrow \mathbf{B}(x, \exists y(\mathbf{M}y \wedge \mathbf{W}yt)))$ (de dicto)

$\exists y(\mathbf{M}y \wedge \forall x(\mathbf{S}x \rightarrow \mathbf{B}(x, \mathbf{W}yt)))$ (de re)

$\forall x(\mathbf{S}x \rightarrow \exists y(\mathbf{M}y \wedge \mathbf{B}(x, \mathbf{W}yt)))$ (de re, but for every schoolboy possibly a different one)

Exercise 4

- (a) We construct a counterexample to $\forall x\Diamond\phi \rightarrow \Diamond\forall x\phi$ in the figure.



$D_{w_1} = D_{w_2} = D_{w_3} = \{a, b\}$; $I_{w_1}(A) = \emptyset$, $I_{w_2}(A) = \{a\}$, $I_{w_3}(A) = \{b\}$. It is clear that for arbitrary g , $V_{M, w_2, g[x/a]}(Ax) = 1$ and hence $V_{M, w_1, g[x/a]}(\Diamond Ax) = 1$. Also $V_{M, w_3, g[x/b]}(Ax) = 1$ and consequently $V_{M, w_1, g[x/b]}(\Diamond Ax) = 1$. So it follows that $V_{M, w_1, g}(\forall x \Diamond Ax) = 1$. On the other hand, since both $V_{M, w_2}(\forall x Ax) = 0$ and $V_{M, w_3}(\forall x Ax) = 0$, $V_{M, w_1}(\Diamond \forall x Ax) = 0$. So $\forall x \Diamond Ax \rightarrow \Diamond \forall x Ax$ is not valid in M .

- (b) Suppose that M satisfies the assumption of decreasing domains and that $V_{M, w}(\Diamond \exists x \phi) = 1$. In order to show that (64) is valid in M we must now show that $V_{M, w}(\exists x \Diamond \phi) = 1$. Take an arbitrary g . The fact that $V_{M, w, g}(\Diamond \exists x \phi) = 1$ means that for some w' with wRw' , it is the case that $V_{M, w', g}(\exists x \phi) = 1$, and so $V_{M, w', g[x/d]}(\phi) = 1$ is the case for some $d \in D_{w'}$. Since M satisfies the assumption of decreasing domains, it is also true that $d \in D_w$. This means that $V_{M, w, g[x/d]}(\Diamond \phi) = 1$, and hence $V_{M, w, g}(\exists x \Diamond \phi) = 1$.

If M does not satisfy the assumption of decreasing domains, then there are worlds w and w' where wRw' , with an entity d such that $d \in D_{w'}$ and $d \notin D_w$. Now set $g(y) = d$. Then $V_{M, w', g[x/d]}(x = y) = 1$, and hence $V_{M, w', g}(\exists x (x = y)) = 1$ and $V_{M, w, g}(\Diamond \exists x (x = y)) = 1$. On the other hand, for $d' \in D_w$ (which is always distinct from the above d) and wRw' , the value $V_{M, w', g[x/d']}(x = y)$ is always 0 or undefined. So $V_{M, w, g[x/d']}(\Diamond (x = y)) = 0$. From this it follows that $V_{M, w, g}(\exists x \Diamond (x = y)) = 0$, and hence $V_{M, w, g}(\Diamond \exists x (x = y) \rightarrow \exists x \Diamond (x = y)) = 0$.

Exercise 5

- (a) We define a model M as follows: $W = \{w_1, w_2\}$; $R = \{\langle w_1, w_2 \rangle\}$; $D = \{a, b\}$; $I_{w_1}(E) = \{a\}$; $I_{w_2}(E) = \{a, b\}$; $I_{w_1}(A) = I_{w_2}(A) = \{a\}$. We shall show that $V_{M, w_1}(\forall x (Ex \rightarrow \Box Ax) \rightarrow \Box \forall x (Ex \rightarrow Ax)) = 0$. For an arbitrary assignment g , $V_{M, w_1, g[x/b]}(Ex) = 0$ holds, and so $V_{M, w_1, g[x/b]}(Ex \rightarrow \Box Ax) = 1$. Also $V_{M, w_2, g[x/a]}(Ax) = 1$ obtains, and so $V_{M, w_1, g[x/a]}(\Box Ax) = 1$ and $V_{M, w_1, g[x/a]}(Ex \rightarrow \Box Ax) = 1$. Therefore $V_{M, w_1, g}(\forall x (Ex \rightarrow \Box Ax)) = 1$. On the other hand, $V_{M, w_2, g}(\forall x (Ex \rightarrow Ax)) = 0$, whence $V_{M, w_1, g}(\Box \forall x (Ex \rightarrow Ax)) = 0$, which is what we wanted.
- (b) In a model M , (78) holds iff existence in M is increasing, that is, if wRw' implies that $I_w(E) \subseteq I_{w'}(E)$. Suppose first that the latter condition is met, and let $V_{M, w, g}(Ex) = 1$. This means that $g(x) \in I_w(E)$. Because of $I_w(E) \subseteq I_{w'}(E)$, $g(x) \in I_{w'}(E)$ will hold for all w' with wRw' , i.e., $V_{M, w', g}(Ex) = 1$. This proves that $V_{M, w, g}(\Box Ex) = 1$. Since g was chosen arbitrarily, it follows that $V_{M, w}(\forall x (Ex \rightarrow \Box Ex)) = 1$. Now assume that $V_{M, w}(\forall x (Ex \rightarrow \Box Ex)) = 1$ and let wRw' , $d \in I_w(E)$. It suffices to show that $d \in I_{w'}(E)$ holds as well. Now $V_{M, w, g[x/d]}(Ex \rightarrow \Box Ex) = 1$. From $d \in I_w(E)$ it also follows that $V_{M, w, g[x/d]}(Ex) = 1$, and thus $V_{M, w, g[x/d]}(\Box Ex) = 1$. But this means that $V_{M, w', g[x/d]}(Ex) = 1$ and so $d \in I_{w'}(E)$.

Chapter 4

Exercise 1

- (a) (i) no; (ii) no; (iii) yes, type $\langle e, t \rangle$; (iv) yes, type t ; (v) yes, type t ; (vi) no; (vii) yes, type t ; (viii) yes, type t
- (b) (i) $c_a(M)$ must be of type $\langle e, t \rangle$, since j is of type e , so $a = \langle \langle e, t \rangle, \langle e, t \rangle \rangle$.
- (ii) $M(j)$ is of type t , so $a = \langle t, t \rangle$.
- (iii) $S(M)$ is of type $\langle e, t \rangle$, so a is e .
- (iv) Again, $a = e$, so $b = \langle t, t \rangle$.
- (v) Because c_a is applicable to S , $a = \langle \langle \langle e, t \rangle, \langle e, t \rangle \rangle, a' \rangle$, for some type a' ; since $c_a(S)$, being of type a' , is applicable to M , $a' = \langle \langle e, t \rangle, a'' \rangle$ for some type a'' ; then $b = \langle a'', t \rangle$ for the same type a'' . So here we don't have a unique solution.

Exercise 2

j : John; h : Harry; m : Mary; s : the salami; c : the couch; M : sleep; S : slice; R : sit; T : soundly (type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$); C : carefully (type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$); c_1 : presumably (type $\langle t, t \rangle$); c_2 : do (type $\langle \langle e, t \rangle, \langle e, t \rangle \rangle$); c_3 : wrong, (in (d), type $\langle \langle \langle e, t \rangle, \langle e, t \rangle \rangle, \langle \langle e, t \rangle, \langle e, t \rangle \rangle \rangle$); c_4 : on (type $\langle e, \langle \langle e, t \rangle, \langle e, t \rangle \rangle \rangle$); c_5 : between (and), (type $\langle e, \langle e, \langle \langle e, t \rangle, \langle e, t \rangle \rangle \rangle \rangle$); W : wrong (in (e), type $\langle \langle e, t \rangle, t \rangle$).

Translations:

- (a) $(T(M))(j)$
- (b) $c_1(M(j))$
- (c) $(C(S(s)))(h)$
- (d) $\forall x(\neg \exists Y(c_2(Y)(x)) \rightarrow \neg \exists Y(c_3(c_2)(Y)(x)))$
- (e) $\forall x(\neg \exists Y(c_2(Y)(x)) \rightarrow \neg \exists Y(c_2(Y)(x) \wedge W(Y)))$
- (f) $((c_4(c))(R))(m)$
- (g) $((c_5(j)(h))(R))(m)$
- (h) $((c_5(j)(h))((c_4(c))(R)))(m)$.

Exercise 4

- (a) $I(e_1) = P_1$; $I(e_2) = P_2$; $I(e_3) = P_3$; $I(M)(P_1) = 1$; $I(M)(P_2) = 1$; $I(M)(P_3) = 0$; $I(A)(P_1)$ is the function f_1 such that $f_1(P_1) = 0$, $f_1(P_2) = f_1(P_3) = 1$. $I(A)(P_2)$ is the function f_2 such that $f_2(P_1) = f_2(P_2) = f_2(P_3) = 0$. $I(A)(P_3)$ is the function f_3 such that $f_3(P_1) = 0$ and $f_3(P_2) = f_3(P_3) = 1$. $I(\mathcal{E})(f_x) = 1$ iff $X \neq \emptyset$, $X \neq \{P_3\}$. $I(T)(f_x)(y) = 1$ iff $(f_x)(y) = 0$.
- (b) (i) This formula expresses that there is a point with arrows to two points, one of which is encircled, while the other one is not encircled. This holds for P_1 , so the sentence is true. The interpretation can be worked out with the help of definition 4, as follows:
Set $g(x) = P_1$; $g(y) = P_2$; $g(z) = P_3$. By this choice, $I(A)(g(y))(g(x)) = 1$ and $I(A)(g(z))(g(x)) = 1$, so $\llbracket A(y)(x) \rrbracket_{M,g} = \llbracket A(z)(x) \rrbracket_{M,g} = 1$.

Further, $I(M)(g(y)) = 1$, so $\llbracket M(y) \rrbracket_{\mathbf{M},g} = 1$ and $I(M)(g(z)) = 0$, so $\llbracket M(z) \rrbracket_{\mathbf{M},g} = 0$ and $\llbracket \neg M(z) \rrbracket_{\mathbf{M},g} = 1$. From this and the preceding facts it follows that $\llbracket A(y)(x) \wedge M(y) \wedge (A(z)(x) \wedge \neg M(z)) \rrbracket_{\mathbf{M},g} = 1$, and thus it follows that $\llbracket \exists x \exists y \exists z (A(y)(x) \wedge M(y) \wedge A(z)(x) \wedge \neg M(z)) \rrbracket_{\mathbf{M},g} = 1$.

- (ii) This formula expresses that an arrow is going from a point to itself if and only if it is not encircled. This is true because there is an arrow from P_3 to P_3 itself, and P_3 is the only point that is not encircled.

This can be worked out as follows: $\llbracket A(x)(x) \rrbracket_{\mathbf{M},g[x/P_1]} = \llbracket A(x)(x) \rrbracket_{\mathbf{M},g[x/P_2]} = 0$ and $\llbracket A(x)(x) \rrbracket_{\mathbf{M},g[x/P_3]} = 1$. Furthermore: $\llbracket \neg(M(x)) \rrbracket_{\mathbf{M},g[x/P_1]} = \llbracket \neg(M(x)) \rrbracket_{\mathbf{M},g[x/P_2]} = 0$ whereas $\llbracket \neg(M(x)) \rrbracket_{\mathbf{M},g[x/P_3]} = 1$. Hence, $\llbracket A(x)(x) \leftrightarrow \neg M(x) \rrbracket_{\mathbf{M},g[x/d]} = 1$ for all $d \in D$, which means that $\llbracket \forall x (A(x)(x) \leftrightarrow \neg M(x)) \rrbracket_{\mathbf{M},g} = 1$.

- (iii) This formula expresses that from every point with an arrow going to the point itself there is also an arrow going to an encircled point. This is true: P_3 is the only point with an arrow going to itself, and from P_3 there is an arrow going to P_2 , which is an encircled point.

This can be worked out as follows: we have $\llbracket A(x)(x) \rrbracket_{\mathbf{M},g[x/P_1]} = 0$ and $\llbracket A(x)(x) \rrbracket_{\mathbf{M},g[x/P_2]} = 0$, whereas $\llbracket A(y)(x) \rrbracket_{\mathbf{M},g[x/P_3][y/P_2]} = 1$ and $\llbracket M(y) \rrbracket_{\mathbf{M},g[x/P_3][y/P_2]} = 1$. From the latter two results it follows that $\llbracket \exists y (A(y)(x) \wedge M(y)) \rrbracket_{\mathbf{M},g[x/P_3]} = 1$, and also that $\llbracket A(x)(x) \rightarrow \exists y (A(y)(x) \wedge M(y)) \rrbracket_{\mathbf{M},g[x/P_3]} = 1$. From the first two results we obtain $\llbracket A(x)(x) \rightarrow \exists y (A(y)(x) \wedge M(y)) \rrbracket_{\mathbf{M},g[x/P_i]} = 1$, for $i = 1, 2$. Hence $\llbracket \forall x (A(x)(x) \rightarrow \exists y (A(y)(x) \wedge M(y))) \rrbracket_{\mathbf{M},g} = 1$.

- (iv) This formula means that every subset of the domain contains an element. This is not true: the empty set contains no element. More formally, take $g(X)$ to be the characteristic function of \emptyset , i.e., that function on the domain which assigns the value 0 to all three points. Then $\llbracket X(x) \rrbracket_{\mathbf{M},g[x/P_1]} = g(X)(g[x/P_1](x)) = g(X)(P_1) = 0$, and also $\llbracket X(x) \rrbracket_{\mathbf{M},g[x/P_2]} = \llbracket X(x) \rrbracket_{\mathbf{M},g[x/P_3]} = 0$. This implies that $\llbracket \exists x X(x) \rrbracket_{\mathbf{M},g} = 0$, and so $\llbracket \forall X \exists x X(x) \rrbracket_{\mathbf{M}} = 0$.

- (v) This formula expresses that if a set of points does not contain encircled elements, it either contains an element with an arrow pointing to itself or it is the empty set. This is true: a set without encircled elements can only be $\{P_3\}$ or \emptyset . This can be worked out as follows: it suffices—on the assumption that $\llbracket \forall y (M(y) \rightarrow \neg X(y)) \rrbracket_{\mathbf{M},g[X/C]} = 1$, for some arbitrary C in the domain—to show that $\llbracket \exists y (X(y) \wedge A(y)(y)) \vee \neg \exists y X(y) \rrbracket_{\mathbf{M},g[X/C]} = 1$.

First we show that C does not contain P_1 or P_2 , that is, that $\llbracket X(y) \rrbracket_{\mathbf{M},g[X/C][y/P_1]} = \llbracket X(y) \rrbracket_{\mathbf{M},g[X/C][y/P_2]} = 0$. For instance, suppose that $\llbracket X(y) \rrbracket_{\mathbf{M},g[X/C][y/P_1]} = 1$; then it would follow that $\llbracket \neg X(y) \rrbracket_{\mathbf{M},g[X/C][y/P_1]} = 0$, and consequently, given our assumption, $\llbracket M(y) \rrbracket_{\mathbf{M},g[X/C][y/P_1]} = 0$, and so $I(M)(g[X/C][y/P_1](y)) = I(M)(P_1) =$

0 which contradicts our assumptions. Now there are two possibilities: (a) $\llbracket X(y) \rrbracket_{\mathbf{M}, g[X/C][y/P_3]} = 0$; or (b) $\llbracket X(y) \rrbracket_{\mathbf{M}, g[X/C][y/P_3]} = 1$. In (a), $\llbracket \exists y X(y) \rrbracket_{\mathbf{M}, g[X/C]} = 0$, and so $\llbracket \neg \exists y X(y) \rrbracket_{\mathbf{M}, g[X/C]} = 1$. Because, in (b), $\llbracket A(y)(y) \rrbracket_{\mathbf{M}, g[X/C][y/P_3]} = 1$, and also $\llbracket X(y) \wedge A(y)(y) \rrbracket_{\mathbf{M}, g[X/C][y/P_3]} = 1$; so $\llbracket \exists y (X(y) \wedge A(y)(y)) \rrbracket_{\mathbf{M}, g[X/C]} = 1$. In both cases, $\llbracket \exists y (X(y) \wedge A(y)(y)) \vee \neg \exists y X(y) \rrbracket_{\mathbf{M}, g[X/C]} = 1$.

- (vi) This formula asserts that there is a set such that both the set and its complement contain encircled points. This is true. Take $\{P_1\}$, for example. Both $\{P_1\}$ and $\{P_2, P_3\}$ contain an encircled point. This is worked out as: $I(\mathcal{E})(f_{\{P_1\}}) = I(\mathcal{E})(f_{\{P_2, P_3\}}) = 1$; $I(T)(f_{\{P_1\}}) = f_{\{P_2, P_3\}}$. Set $g(X) = f_{\{P_1\}}$; then both $\llbracket \mathcal{E}(X) \rrbracket_{\mathbf{M}, g} = 1$ and $\llbracket T(X) \rrbracket_{\mathbf{M}, g} = f_{\{P_2, P_3\}}$ hold, and so $\llbracket \mathcal{E}(T(X)) \rrbracket_{\mathbf{M}, g} = 1$. With this $\llbracket \mathcal{E}(X) \wedge \mathcal{E}(T(X)) \rrbracket_{\mathbf{M}, g} = 1$ and finally $\llbracket \exists X (\mathcal{E}(X) \wedge \mathcal{E}(T(X))) \rrbracket_{\mathbf{M}, g} = 1$.

Exercise 5

- (a) The basic expressions obtain the following categories:

CN: *man, horse*

T: *John, Peter*

T/CN: *the, a(n)*

CN/CN: *green, big, honest*

T\S: *walk, swears*

(T\S)/T: *eats, makes, curses*

- (b) (i) predicate modifiers like *softly, quickly, . . .*, e.g., *swears softly*
 (ii) prepositions like *on, above, over, . . .*, e.g., *on the horse*
 (iii) possessive 's: e.g., *John's horse*
 (iv) the copula *is*: e.g., *is honest*. (There is a problem here: expressions like *is honest* and *is green* become of category T\S, so it should be possible to combine them with predicate modifiers, but this is not the case. In Montague grammar this problem is solved by 'duplication' of categories (see §6.2.11).)

The complete context-free grammar is now:

$S \Rightarrow NP VP$	$N \Rightarrow \textit{man, horse}$
$VP \Rightarrow VP Adv$	$Adv \Rightarrow \textit{softly, quickly}$
$VP \Rightarrow \begin{cases} V_{intr} \\ V_{tr} NP \\ V_{cop} Adj \end{cases}$	$Adv \Rightarrow P NP$
	$V_{intr} \Rightarrow \textit{walks, swears}$
	$V_{trans} \Rightarrow \textit{eats, makes, curses}$
$NP \Rightarrow \begin{cases} PropN \\ Det N \end{cases}$	$V_{cop} \Rightarrow \textit{is}$
	$Adj \Rightarrow \textit{green, big, honest}$
$PropN \Rightarrow \textit{John, Peter}$	$Det \Rightarrow \textit{the, a(n)}$
$N \Rightarrow Adj N$	$Det \Rightarrow NP Poss$
$P \Rightarrow \textit{on, above, over}$	$Poss \Rightarrow \textit{'s}$

Exercise 8

- | | |
|--|------------------------------------|
| (i) no | (viii) yes, $\langle e, t \rangle$ |
| (ii) yes, t | (ix) no |
| (iii) yes, $\langle e, t \rangle$ | (x) yes, t |
| (iv) yes, $\langle e, t \rangle$ | (xi) yes, t |
| (v) yes, t | (xii) no |
| (vi) yes, $\langle e, \langle \langle e, t \rangle, t \rangle \rangle$ | (xiii) yes, t |
| (vii) no | (xiv) yes, t |

Exercise 9

The key used in the following solution is:

- | | |
|--|-------------------------------------|
| E: wash | \mathcal{D} : healthy |
| C: put | \mathcal{M}_1 : lead to checkmate |
| B: restore | \mathcal{M}_2 : possible |
| F: have forgotten | \mathcal{H}_1 : good |
| G_1 : have known | \mathcal{H}_2 : bad |
| G_2 : know now | L_1 : grow |
| c_1 : always, type $\langle t, t \rangle$ | L_2 : glow |
| c_2 : again, type $\langle t, t \rangle$ | M: human |
| T: properly, type $\langle \langle e, \langle e, t \rangle \rangle, \langle e, \langle e, t \rangle \rangle \rangle$ | R_1 : perfect |
| S_1 : forwards | R_2 : be (exist) |
| S_2 : backwards | q: queen |
| L: love | m: Mary |
| \mathcal{H} : important | |

The translations are:

- (a) $\mathcal{H}(\lambda x((T(E))(x)(x)))$
 (b) $\mathcal{D}(\lambda x \exists y(L(y)(x)))$
 (c) $\lambda U \lambda y \lambda x((S_1(U))(y)(x) \vee (S_2(U))(y)(x))$ (compare answer (d))
 (d) $\mathcal{M}_1(\lambda x((S_1(C))(q)(x) \vee (S_2(C))(q)(x)))$
 (e) $\lambda x(L_1(x) \wedge L_2(x) \wedge \forall y(M(y) \rightarrow c_1(c_2(B(y)(x))))$, taken as the set of objects satisfying the description (there is also a noun phrase reading, which is omitted here)
 (f) $\lambda x \exists y(F(y)(x)) = \lambda x \exists y(G_1(y)(x) \wedge \neg G_2(y)(x))$
 (g) $\neg \mathcal{M}_2(\lambda x(c_1(x = x)))$
 (h) $\lambda x R_1(x) = \lambda x \forall X(\mathcal{H}_1(X) \rightarrow X(x))$
 (i) $\lambda x \forall X(\mathcal{H}_2(X) \rightarrow (X(x) \leftrightarrow X(m)))$
 (j) $\lambda y(\lambda x(R_2(x))(y) \vee \neg(\lambda x(R_2(x))(y)))$ i.e.: $\lambda y(R_2(y) \vee \neg R_2(y))$

Here translations arise in general from a rendering of general categorial forms, together with a spelling out of logical constants wherever possible.

Exercise 10

(i) $M(j)$; (ii) $M(j)$; (iii) $M(j)$; (iv) $\forall y(A(j)(y))$; (v) direct λ -conversion is not possible: y is not free for x in $\forall y(A(x)(y))$. However, λ -conversion is possible if $\forall y(A(x)(y))$ is first transposed into $\forall z(A(x)(z))$, in which case one obtains $\forall z(A(y)(z))$. (vi) $M(j)$. (vii) Here too, direct λ -conversion is not possible. By transposing $\forall x(Y(x))$ into $\forall z(Y(z))$, one first obtains $\forall z(\lambda y(A(x)(y))(z))$ and then $\forall z(A(x)(z))$.

Chapter 5

Exercise 1

- (a) (i) yes, t (viii) no
 (ii) no (ix) no
 (iii) no (x) no
 (iv) yes, t (xi) yes, $\langle s, t \rangle$
 (v) no (xii) no
 (vi) yes, t (xiii) yes, $\langle s, \langle s, e \rangle \rangle$
 (vii) yes, t (xiv) yes, t
- (b) (i) α is of type $\langle \langle s, t \rangle, t \rangle$, for $\wedge p$ is of type $\langle s, t \rangle$;
 (ii) $\langle \langle s, e \rangle, t \rangle$;
 (iii) $\vee \alpha(j)$ is of type t , so $\vee \alpha$ is of type $\langle e, t \rangle$, and α of type $\langle s, \langle e, t \rangle \rangle$;
 (iv) $\vee \alpha$ is of type $\langle \langle s, t \rangle, e \rangle$, and so α is of type $\langle s, \langle \langle s, t \rangle, e \rangle \rangle$.

Exercise 2

- (a) $I(j)(w_1) = I(j)(w_2) = a$; $I(j)(w_3) = b$;
 $I(m)(w) = c$ for all $w \in W$;
 $I(M)(w_1)(a) = I(M)(w_1)(b) = 1$;
 $I(M)(w_1)(c) = I(M)(w_1)(d) = 0$;
 $I(M)(w_2)(a) = I(M)(w_2)(b) = 0$;
 $I(M)(w_2)(c) = I(M)(w_2)(d) = 1$;
 $I(M)(w_3)(e) = 0$ for all $e \in D$;
 $I(M)(w_1) = I(M)(w_2) = I(M)$;
 $I(M)(w_3)(w)(e) = 1$ for all $w \in W$ and for all $e \in D$.
- (b) (i) $\llbracket j \rrbracket_{M, w_2, g} = I(j)(w_2) = a$;
 (ii) $\llbracket \wedge j \rrbracket_{M, w_1, g}$ = that $h \in D^W$ such that for all $w \in W$: $h(w) = \llbracket j \rrbracket_{M, w, g}$, i.e., $I(j)$;
 (iii) $\llbracket \wedge j \rrbracket_{M, w_3, g} = I(j)$, as in (ii);
 (iv) $\llbracket M(j) \rrbracket_{M, w_2, g} = \llbracket M \rrbracket_{M, w_2, g} (\llbracket j \rrbracket_{M, w_2, g}) = I(M)(w_2)(I(j)(w_2)) = I(M)(w_2)(a) = 0$;
 (v) $\llbracket \vee M \rrbracket_{w_3} = \llbracket M \rrbracket_{w_3} (w_3) = I(M)(w_3)(w_3) =$ the function from $\{0, 1\}^D$ which yields the value 1 for all $e \in D$;

- (vi) $\llbracket \vee M \rrbracket_{w_1} = \llbracket M \rrbracket_{w_1}(w_1) = I(M)(w_1)(w_1) = I(M)(w_1)$ and $\llbracket \vee M(j) \rrbracket_{w_1} = I(M)(w_1)(w_1)(I(j)(w_1)) = I(M)(w_1)(a) = 1$;
- (vii) $\llbracket \vee M(j) \rrbracket_{w_2} = I(M)(w_2)(w_2)(I(j)(w_2)) = I(M)(w_2)(a) = 0$;
- (viii) $\llbracket M = \wedge M \rrbracket_{w_1} = 1$ iff $\llbracket M \rrbracket_{w_1} = \llbracket \wedge M \rrbracket_{w_1}$. Now, $\llbracket M \rrbracket_{w_1}$ is the function $I(M)(w_1) = I(M)$. On the other hand, $\llbracket \wedge M \rrbracket_{w_1}$ is the function h such that for all $w \in W$: $h(w) = \llbracket M \rrbracket_w$ and this is also exactly $I(M)$. From this it follows that $\llbracket M = \wedge M \rrbracket_{w_1} = 1$;
- (ix) $\llbracket \vee M = M \rrbracket_{w_3} = 1$ iff $\llbracket \vee M \rrbracket_{w_3} = \llbracket M \rrbracket_{w_3}$. Now, as in (v), $\llbracket \vee M \rrbracket_{w_3}$ is the function from $\{0, 1\}^D$ which yields the value 1 for all $d \in D$. On the other hand, $\llbracket M \rrbracket_{w_3} = I(M)(w_3)$, the function which assigns the value 0 to all $d \in D$. Thus $\llbracket \vee M = M \rrbracket_{w_3} = 0$.
- (c) (i) Because of (bviii), $\llbracket M = \wedge M \rrbracket_{M, w_1, g} = 1$ holds. So it follows that $\llbracket \Diamond(M = \wedge M) \rrbracket_{M, w, g} = 1$, for all $w \in W$. So $\Diamond(M = \wedge M)$ is valid in M .
- (ii) Given (bix), $\llbracket \vee M = M \rrbracket_{M, w_3, g} = 0$ holds. This means that for all $w \in W$, $\llbracket \Box(\vee M = M) \rrbracket_{M, w, g} = 0$, and so $\Box(\vee M = M)$ is invalid in M .
- (iii) In all $w \in W$, m refers to c . This means that $\llbracket m = x \rrbracket_{M, w, g[x/c]} = 1$, $\llbracket \Box(m = x) \rrbracket_{M, w, g[x/c]} = 1$, and $\llbracket \exists x \Box(m = x) \rrbracket_{M, w, g} = 1$, for all $w \in W$. This means that $\exists x \Box(m = x)$ is valid in M .

Exercise 5

$\llbracket \wedge \vee \alpha \rrbracket_{M, w, g}$ = the function $h \in D_a^W$ such that $h(w') = \llbracket \vee \alpha \rrbracket_{M, w', g}$, for all $w' \in W$ = the function $h \in D_a^W$ such that $h(w') = \llbracket \alpha \rrbracket_{M, w', g}(w')$, for all $w' \in W$ = the function $h \in D_a^W$ such that $h(w') = g(\alpha)(w')$, for all $w' \in W$. This means that $h = g(\alpha)$, so $\llbracket \wedge \vee \alpha \rrbracket_{M, w, g} = h = g(\alpha) = \llbracket \alpha \rrbracket_{M, w, g}$.

Exercise 7

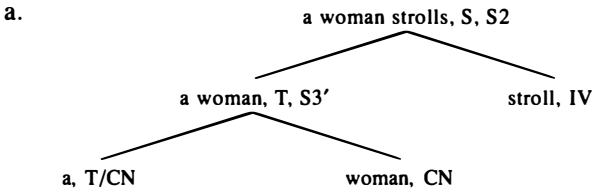
- (i) $\vee M(j)$ (theorem 5);
- (ii) $\lambda x \lambda X (\vee X(x))(j)(\wedge M)$ reduces to $\lambda X (\vee X(j))(\wedge M)$ and then to $\vee \wedge M(j)$ by theorem 5, and this is reduced to $M(j)$ by theorem 2.
- (iii) $\lambda x \lambda X \Box(\vee X(x))(j)(\wedge M)$ does not reduce, because the variable x is in the scope of \Box and $j \notin \text{ICE}$.
- (iv) $\lambda X \lambda x \Box(\vee X(x))(\wedge M)(j)$ reduces to $\lambda x \Box(\vee \wedge M(x))(j)$ by theorem 5, because $\wedge M \in \text{ICE}$. By theorem 2, $\lambda x \Box(\vee \wedge M(x))(j)$ is reduced to $\lambda x \Box M(x)(j)$, which cannot be reduced, since x is in the scope of \Box and $j \notin \text{ICE}$.
- (v) $\lambda x \wedge (M(x) \wedge \vee M(x))(y)$ reduces to $\wedge (M(y) \wedge \vee M(y))$ by theorem 5, since $y \in \text{ICE}$.
- (vi) $\lambda x \lambda y \Box(B(x)(y))(\wedge j)(j)$ reduces to $\lambda y \Box(B(\wedge j)(y))(j)$ by theorem 5, because $\wedge j \in \text{ICE}$. Since y is in the scope of \Box and $j \notin \text{ICE}$, $\lambda y \Box(B(\wedge j)(y))(j)$ cannot be reduced any further.
- (vii) Since $z \in \text{ICE}$, $\lambda y \lambda x \wedge \exists x (B(x)(y) \wedge \wedge x = y)(\wedge \vee z)(\wedge x)$ reduces by theorem 3 to $\lambda y \lambda x \wedge \exists x (B(x)(y) \wedge \wedge x = y)(z)(\wedge x)$, which reduces by

theorem 5 to $\lambda x \wedge \exists x(B(x)(z) \wedge \wedge x = z)(\wedge x)$, again because $z \in \text{ICE}$. The reduction stops here, since x is in the scope of $\exists x$. If we replace this quantifier and its bound variable by $\exists y$ and y , respectively, we can continue the reduction with the help of theorem 5, because $\wedge x \in \text{ICE}$, thus obtaining $\wedge \exists y(B(\wedge x)(z) \wedge \wedge y = z)$.

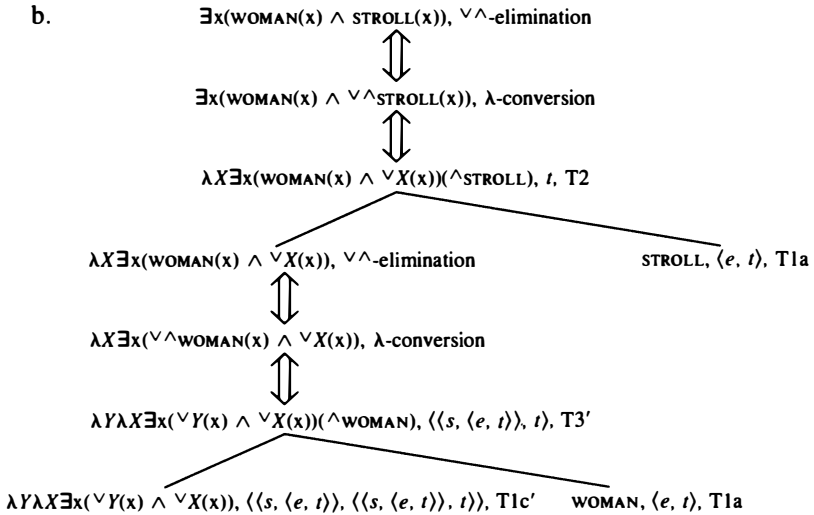
Chapter 6

Exercise 1

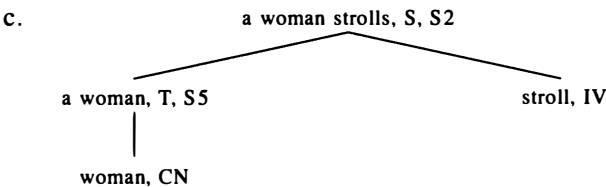
(a) Catecorematically, the analysis tree is given in figure (a):



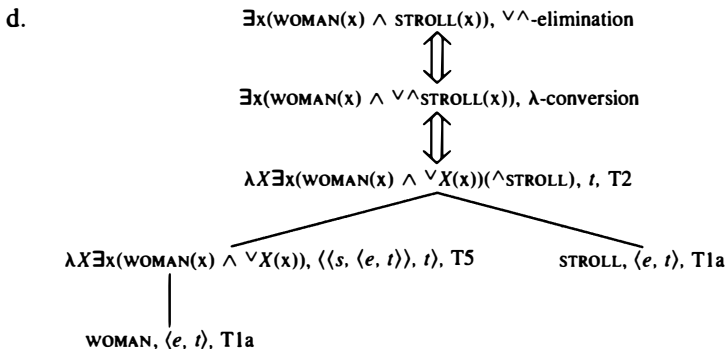
The translation tree is given in figure (b):



Synkategorematically, the analysis tree is given in figure (c):



The translation tree is given in figure (d):

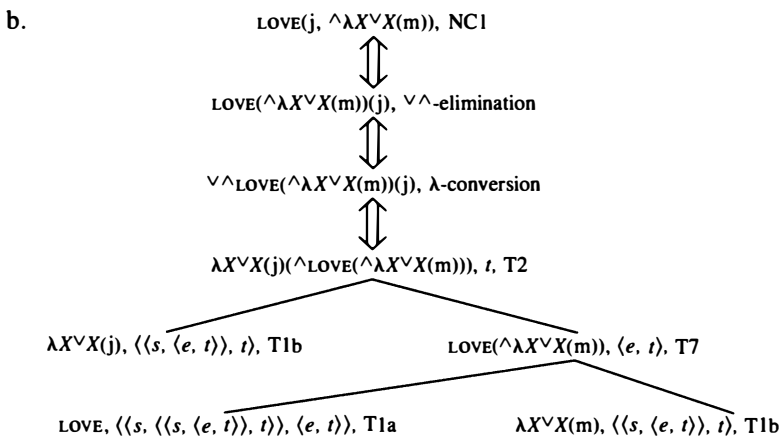
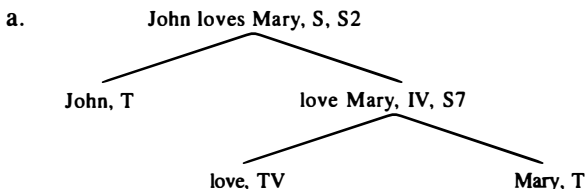


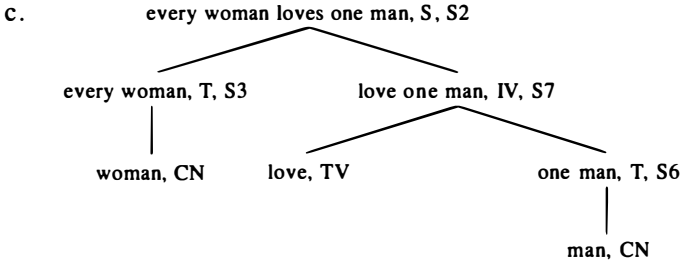
(b) $\lambda Y \lambda X \neg \exists x (\vee^{\wedge} Y(x) \wedge \vee^{\wedge} X(x))$

Exercise 2

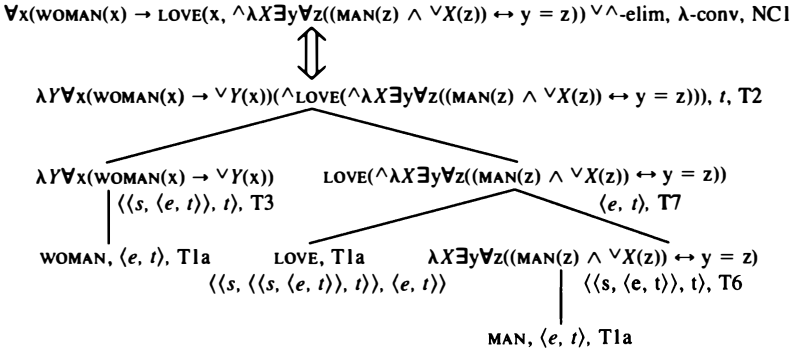
(i) Analysis tree: see figure (a); translation tree: see figure (b).

(ii) Analysis tree: see figure (c); translation tree: see figure (d).





d.



Exercise 3

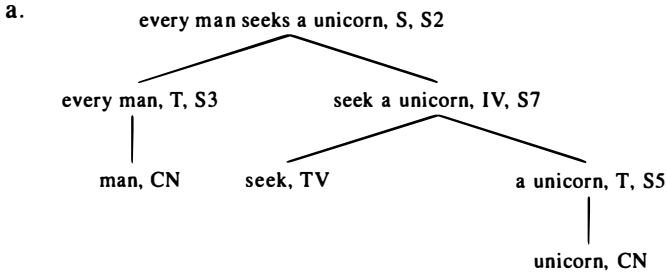
- (i) $\text{LOVE}(j, \wedge \lambda X \forall X(m)) = \text{LOVE}_*(j, m)$ BP1 and NC2
- (ii) $\forall x(\text{WOMAN}(x) \rightarrow \text{LOVE}(x, \wedge \lambda X \exists y \forall z((\text{MAN}(z) \wedge \forall X(z)) \leftrightarrow y = z)))$
- $= \forall x(\text{WOMAN}(x) \rightarrow$
- $\vee \wedge \lambda X \exists y \forall z((\text{MAN}(z) \wedge \forall X(z)) \leftrightarrow y = z)$
- $(\wedge \lambda w \text{LOVE}_*(x, w)))$ Theorem 1
- $= \forall x(\text{WOMAN}(x) \rightarrow \lambda X \exists y \forall z((\text{MAN}(z) \wedge \forall X(z)) \leftrightarrow y = z) (\wedge \lambda w \text{LOVE}_*(x, w)))$ $\vee \wedge\text{-elim.}$
- $= \forall x(\text{WOMAN}(x) \rightarrow \exists y \forall z((\text{MAN}(z) \wedge \vee \wedge \lambda w \text{LOVE}_*(x, w)(z)) \leftrightarrow y = z))$ $\lambda\text{-conv.}$
- $= \forall x(\text{WOMAN}(x) \rightarrow \exists y \forall z((\text{MAN}(z) \wedge \lambda w \text{LOVE}_*(x, w)(z)) \leftrightarrow y = z))$ $\vee \wedge\text{-elim.}$
- $= \forall x(\text{WOMAN}(x) \rightarrow \exists y \forall z((\text{MAN}(z) \wedge \text{LOVE}_*(x, z)) \leftrightarrow y = z))$ $\lambda\text{-conv.}$

Exercise 4

- (a) The sentence *Every man seeks a unicorn* has three different readings. First of all, it has a de dicto reading which does not commit us to the belief that unicorns exist. Second, there is a reading in which every man seeks the

same existing unicorn. And third, the sentence can mean that every man seeks a unicorn which exists; the sentence is indeterminate as to whether or not all men try to find the same unicorn. The last two readings are both *de re*.

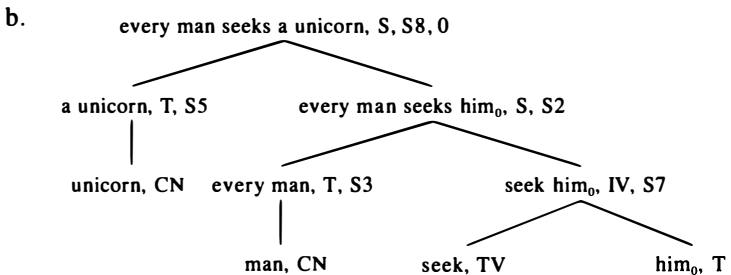
The first reading can be constructed directly, as shown in figure (a).



The translation of (a) is:

- | | |
|---|-------------------------|
| 1. $unicorn \mapsto UNICORN$ | T1a |
| 2. $F_4(unicorn) \mapsto \lambda X \exists x (UNICORN(x) \wedge \forall X(x))$ | T5 |
| 3. $seek \mapsto SEEK$ | T1a |
| 4. $F_6(seek, a\ unicorn) \mapsto SEEK(\wedge \lambda X \exists x (UNICORN(x) \wedge \forall X(x)))$ | T7 |
| 5. $man \mapsto MAN$ | T1a |
| 6. $F_2(man) \mapsto \lambda Y \forall y (MAN(y) \rightarrow \forall Y(y))$ | T3 |
| 7. $F_1(every\ man, seek\ a\ unicorn) \mapsto \lambda Y \forall y (MAN(y) \rightarrow \forall Y(y)) (\wedge SEEK(\wedge \lambda X \exists x (UNICORN(x) \wedge \forall X(x))))$ | T2 |
| 8. $= \forall y (MAN(y) \rightarrow \forall (\wedge SEEK(\wedge \lambda X \exists x (UNICORN(x) \wedge \forall X(x))))(y))$ | λ -conv. |
| 9. $= \forall y (MAN(y) \rightarrow SEEK(\wedge \lambda X \exists x (UNICORN(x) \wedge \forall X(x)))(y))$ | $\forall \wedge$ -elim. |
| 10. $= \forall y (MAN(y) \rightarrow SEEK(y, \wedge \lambda X \exists x (UNICORN(x) \wedge \forall X(x))))$ | NC1 |

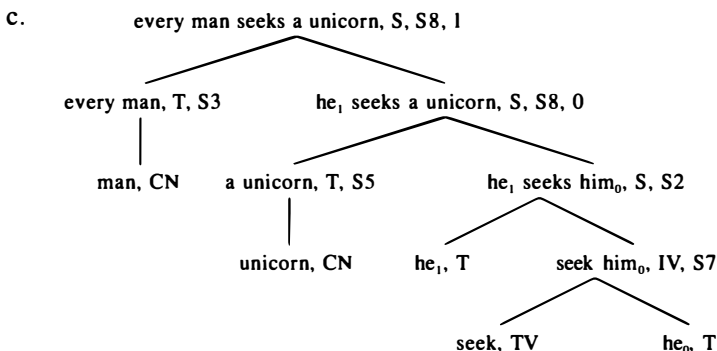
The second reading can be obtained by quantifying *a unicorn* into *every man seeks him₀*, as shown in figure (b):



The translation of the analysis tree (b) is as follows:

1. $he_0 \mapsto \lambda X \vee X(x_0)$ T1b
2. $seek \mapsto SEEK$ T1a
3. $F_6(seek, he_0) \mapsto SEEK(\wedge \lambda X \vee X(x_0))$ T7
4. $man \mapsto MAN$ T1a
5. $F_2(man) \mapsto \lambda Y \forall y (MAN(y) \rightarrow \vee Y(y))$ T3
6. $F_1(every\ man, seek\ him_0) \mapsto \lambda Y \forall y (MAN(y) \rightarrow \vee Y(y)) (\wedge SEEK(\wedge \lambda X \vee X(x_0)))$ T2
7. $= \forall y (MAN(y) \rightarrow \vee \wedge SEEK(\wedge \lambda X \vee X(x_0))(y))$ λ -conv.
8. $= \forall y (MAN(y) \rightarrow SEEK(\wedge \lambda X \vee X(x_0))(y))$ $\vee \wedge$ -elim.
9. $= \forall y (MAN(y) \rightarrow SEEK(y, \wedge \lambda X \vee X(x_0)))$ NC1
10. $= \forall y (MAN(y) \rightarrow SEEK_*(y, x_0))$ NC2
11. $unicorn \mapsto UNICORN$ T1a
12. $F_4(unicorn) \mapsto \lambda X \exists x (UNICORN(x) \wedge \vee X(x))$ T5
13. $F_{7,0}(a\ unicorn, every\ man\ seeks\ him_0) \mapsto \lambda X \exists x (UNICORN(x) \wedge \vee X(x)) (\wedge \lambda x_0 \forall y (MAN(y) \rightarrow SEEK_*(y, x_0)))$ T8,0
14. $\exists x (UNICORN(x) \wedge \vee \wedge \lambda x_0 \forall y (MAN(y) \rightarrow SEEK_*(y, x_0))(x))$ λ -conv.
15. $\exists x (UNICORN(x) \wedge \lambda x_0 \forall y (MAN(y) \rightarrow SEEK_*(y, x_0))(x))$ $\vee \wedge$ -elim.
16. $\exists x (UNICORN(x) \wedge \forall y (MAN(y) \rightarrow SEEK_*(y, x)))$ λ -conv.

In the third reading, *a unicorn* has wider scope than *seek*, but it is in the scope of *every man*. This reading can be obtained by first quantifying *a unicorn* into the sentence he_1 seeks him_0 . This results in he_1 seeks *a unicorn*. Then *every man* can be quantified in, as shown in (c).

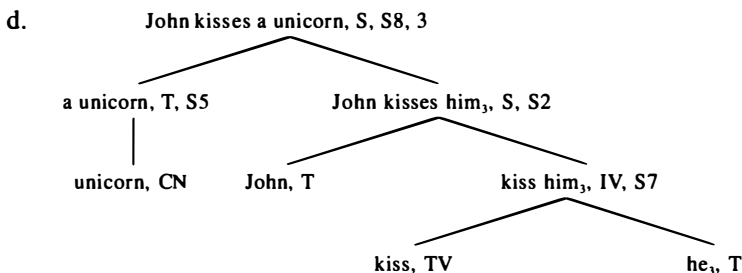


The translation of (c) proceeds as follows:

1. $he_0 \mapsto \lambda X \vee X(x_0)$ T1b
2. $seek \mapsto SEEK$ T1a
3. $F_6(seek, he_0) \mapsto SEEK(\wedge \lambda X \vee X(x_0))$ T7
4. $he_1 \mapsto \lambda Y \vee Y(x_1)$ T1b

5. $F_1(he_1, seek\ him_0) \mapsto \lambda Y \vee Y(x_1)$
 $(\wedge SEEK(\wedge \lambda X \vee X(x_0)))$ T2
6. $= \vee \wedge SEEK(\wedge \lambda X \vee X(x_0))(x_1)$ λ -conv.
7. $= SEEK(\wedge \lambda X \vee X(x_0))(x_1)$ $\vee \wedge$ -elim.
8. $= SEEK(x_1, \wedge \lambda X \vee X(x_0))$ NC1
9. $= SEEK_*(x_1, x_0)$ NC2
10. $unicorn \mapsto UNICORN$ T1a
11. $F_4(unicorn) \mapsto \lambda X \exists x(UNICORN(x) \wedge \vee X(x))$ T5
12. $F_{7,0}(a\ unicorn, he_1\ seeks\ him_0) \mapsto$
 $\lambda X \exists x(UNICORN(x) \wedge \vee X(x)) (\wedge \lambda x_0 SEEK_*(x_1, x_0))$ T8, 0
13. $\exists x(UNICORN(x) \wedge \vee \wedge \lambda x_0 SEEK_*(x_1, x_0)(x))$ λ -conv.
14. $\exists x(UNICORN(x) \wedge \lambda x_0 SEEK_*(x_1, x_0)(x))$ $\vee \wedge$ -elim.
15. $\exists x(UNICORN(x) \wedge SEEK_*(x_1, x))$ λ -conv.
16. $man \mapsto MAN$ T1a
17. $F_2(man) \mapsto \lambda X \forall y(MAN(y) \rightarrow \vee X(y))$ T3
18. $F_{7,1}(every\ man, he_1\ seeks\ a\ unicorn) \mapsto$
 $\lambda X \forall y(MAN(y) \rightarrow \vee X(y))$
 $(\wedge \lambda x_1 \exists x(UNICORN(x) \wedge SEEK_*(x_1, x)))$ T8, 1
19. $= \forall y(MAN(y) \rightarrow \vee \wedge \lambda x_1 \exists x(UNICORN(x) \wedge$
 $SEEK_*(x_1, x))(y))$ λ -conv.
20. $= \forall y(MAN(y) \rightarrow \lambda x_1 \exists x(UNICORN(x) \wedge$
 $SEEK_*(x_1, x))(y))$ $\vee \wedge$ -elim.
21. $= \forall y(MAN(y) \rightarrow \exists x(UNICORN(x) \wedge SEEK_*(y, x)))$ λ -conv.

(b) We begin with the indirect construction of *John kisses a unicorn*. This is represented in the analysis tree (d).

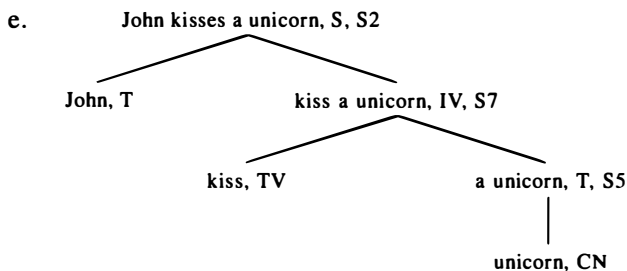


The translation of the tree (d) is:

1. $he_3 \mapsto \lambda X \vee X(x_3)$ T1b
2. $kiss \mapsto KISS$ T1a
3. $F_6(kiss, he_3) \mapsto KISS(\wedge \lambda X \vee X(x_3))$ T7
4. $John \mapsto \lambda Y \vee Y(j)$ T1b
5. $F_1(John, kiss\ him_3) \mapsto \lambda Y \vee Y(j) (\wedge KISS(\wedge \lambda X \vee X(x_3)))$ T2
6. $= \vee \wedge KISS(\wedge \lambda X \vee X(x_3))(j)$ λ -conv.
7. $= KISS(\wedge \lambda X \vee X(x_3))(j)$ $\vee \wedge$ -elim.
8. $= KISS(j, \wedge \lambda X \vee X(x_3))$ NC1

9. $= \text{KISS}_*(j, x_3)$ NC2
10. $\text{unicorn} \mapsto \text{UNICORN}$ T1a
11. $F_4(\text{unicorn}) \mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))$ T5
12. $F_{7,3}(a \text{ unicorn}, \text{John kisses him}_3) \mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)) (\wedge \lambda x_3 \text{KISS}_*(j, x_3))$ T8, 3
13. $\exists x (\text{UNICORN}(x) \wedge \forall \lambda x_3 \text{KISS}_*(j, x_3)(x))$ λ -conv.
14. $\exists x (\text{UNICORN}(x) \wedge \lambda x_3 \text{KISS}_*(j, x_3)(x))$ $\forall \wedge$ -elim.
15. $\exists x (\text{UNICORN}(x) \wedge \text{KISS}_*(j, x))$ λ -conv.

The direct construction is shown in figure (e).



The translation of the analysis tree (e) is:

1. $\text{unicorn} \mapsto \text{UNICORN}$ T1a
2. $F_4(\text{unicorn}) \mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))$ T5
3. $\text{kiss} \mapsto \text{KISS}$ T1a
4. $F_6(\text{kiss}, a \text{ unicorn}) \mapsto \text{KISS}(\wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)))$ T7
5. $\text{John} \mapsto \lambda Y \forall Y(j)$ T1b
6. $F_1(\text{John}, \text{kiss a unicorn}) \mapsto \lambda Y \forall Y(j)$
 $(\wedge \text{KISS}(\wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))))$ T2
7. $= \forall \wedge \text{KISS}(\wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)))(j)$ λ -conv.
8. $= \text{KISS}(\wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)))(j)$ $\forall \wedge$ -elim.
9. $= \text{KISS}(j, \wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)))$ NC1
10. $= \forall \wedge \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x))$
 $(\wedge \lambda y \text{KISS}_*(j, y))$ Theorem 1
11. $= \lambda X \exists x (\text{UNICORN}(x) \wedge \forall X(x)) (\wedge \lambda y \text{KISS}_*(j, y))$ $\forall \wedge$ -elim.
12. $= \exists x (\text{UNICORN}(x) \wedge \forall \wedge \lambda y \text{KISS}_*(j, y)(x))$ λ -conv.
13. $= \exists x (\text{UNICORN}(x) \wedge \lambda y \text{KISS}_*(j, y)(x))$ $\forall \wedge$ -elim.
14. $= \exists x (\text{UNICORN}(x) \wedge \text{KISS}_*(j, x))$ λ -conv.

Exercise 6

The problem is to show that (iii) does not follow from (i) and (ii):

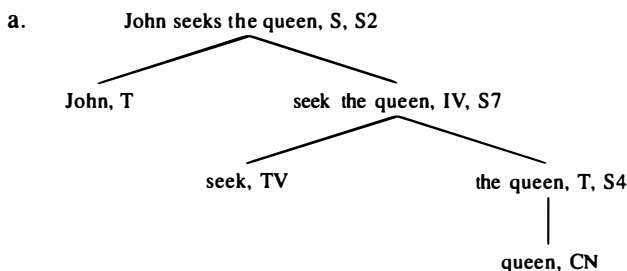
- (i) John seeks the queen.
- (ii) Elsie is the queen.
- (iii) John seeks Elsie.

We do this by analyzing each of these sentences syntactically such that the translation of (iii) does not follow from the translations of (i) and (ii), given these syntactic analyses. Sentences (ii) and (iii) are not ambiguous. They can be constructed in different ways, but they result ultimately in the same translations, (iv) and (v), respectively.

(iv) $\exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge x = e)$

(v) $\text{SEEK}_*(j, e)$

On the other hand, sentence (i) is ambiguous. It has a *de re* reading and a *de dicto* reading. We obtain the latter by constructing (i) directly, as in (a).



This syntactic analysis yields the following translation:

(vi) $\text{SEEK}(j, \wedge \lambda X \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee X(x)))$

Formula (v) does not follow from formulas (vi) and (iv). Assume a model **M** with a world *w* such that (vi) and (iv) are true in *w* given **M**. A counterexample can now easily be produced by showing that it does not follow that (v) is true in **M** in *w* as well.

To see this, observe first that from the definition of SEEK_* it follows that (v) is equivalent to (vii):

(vii) $\text{SEEK}(j, \wedge \lambda X \vee X(e))$

Formula (vii) follows from (vi) only if $\wedge \lambda X \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee X(x))$ refers in **M** to the same function from worlds to sets of first-order properties as $\wedge \lambda X \vee X(e)$ does. (The extension of these two expressions is dependent only on model **M**, not on the world or the assignment. This is because they are intensionally closed expressions containing no free variables. This is the case if in all $w' \in W$ in **M** $\wedge \lambda X \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee X(x))$ refers to the same set of first-order properties as $\wedge \lambda X \vee X(e)$).

This is true just in case (iv) $\exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge x = e)$ is true in all w' . By assumption, (iv) is true in *w* and so both $\wedge \lambda X \exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee X(x))$ and $\wedge \lambda X \vee X(e)$ refer in *w* to the same set of first-order properties.

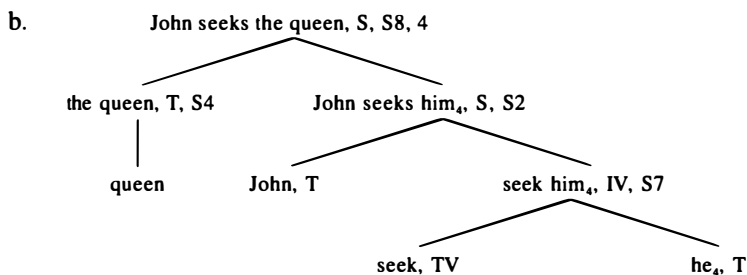
However, this does not say anything about the truth value of $\exists x(\forall y(\text{QUEEN}(y) \leftrightarrow x = y) \wedge x = e)$ in worlds different from *w*. So, without contradicting our original assumptions, we may assume that **M** in addition to

w also contains a world w' in which (iv) is false. In w' $\lambda X \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y \wedge \vee X(x)))$ and $\lambda X \vee X(e)$ refer to different sets of first-order properties.

From this it follows that $\wedge \lambda X \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee X(x))$ and $\wedge \lambda X \vee X(e)$ refer in \mathbf{M} to different functions, and hence also that formula (v) does not follow from (vi), even if (iv) holds. Thus we can construct a counterexample to the statement that (v) follows from (vi) and (iv) by taking a model \mathbf{M} with a world w in which both (vi) and (iv) are true, whereas (v) is not true, and a world w' in which (iv) is not true.

This shows that (iii) does not follow from (i) and (ii). After all, we have given syntactic analyses for (i), (ii), and (iii) such that the translation of (iii) does not follow from the translations of (i) and (ii).

It should be noted, however, that the de re reading makes the argument valid. This reading results from an indirect construction, as given in figure (b):



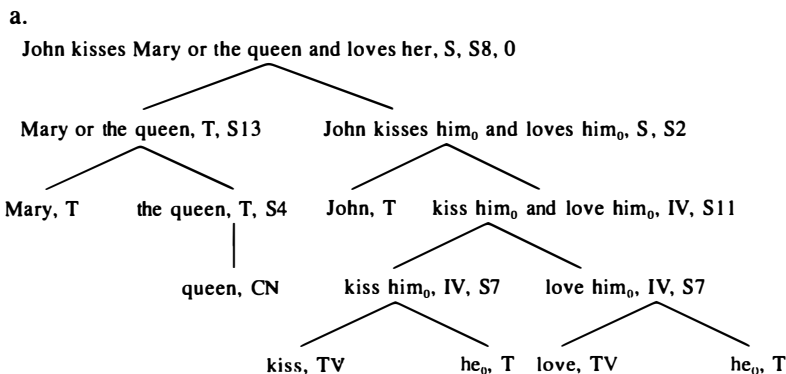
The resulting translation is:

(viii) $\exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge \text{seek}_*(j, x))$

Formula (v) follows directly from (viii) and (iv).

Exercise 7

For the sentence *John kisses Mary or the queen and loves her*, there are two analysis trees, resulting in two nonequivalent translations. The first analysis tree is given in figure (a):



The translation of analysis tree (a) is:

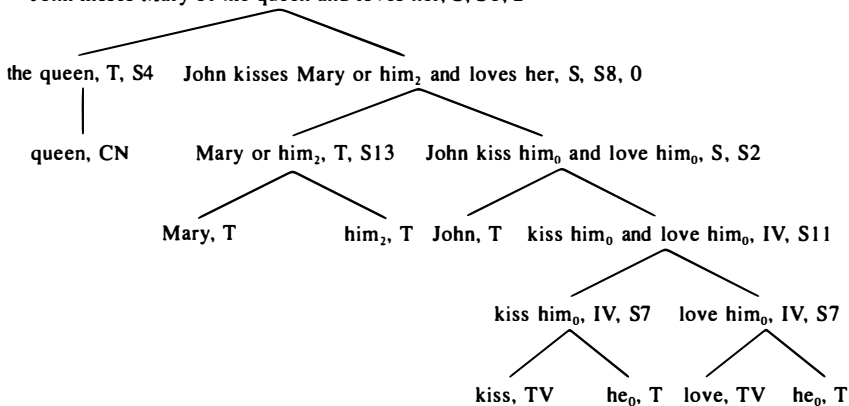
1. $kiss \mapsto KISS$ T1a
2. $he_0 \mapsto \lambda X \vee X(x_0)$ T1b
3. $F_6(kiss, he_0) \mapsto KISS(\wedge \lambda X \vee X(x_0))$ T7
4. $love \mapsto LOVE$ T1a
5. $F_6(love, him_0) \mapsto KISS(\wedge \lambda X \vee X(x_0))$ T7
6. $F_8(kiss\ him_0, love\ him_0) \mapsto$
 $\lambda x(KISS(\wedge \lambda X \vee X(x_0))(x) \wedge LOVE(\wedge \lambda X \vee X(x_0))(x))$ T11
7. $= \lambda x(KISS(x, \wedge \lambda X \vee X(x_0)) \wedge LOVE(x, \wedge \lambda X \vee X(x_0)))$ NC1
8. $= \lambda x(KISS_*(x, x_0) \wedge LOVE_*(x, x_0))$ NC2
9. $John \mapsto \lambda X \vee X(j)$ T1b
10. $F_1(John, kiss\ him_0\ and\ love\ him_0) \mapsto$
 $\lambda X \vee X(j)(\wedge \lambda x(KISS_*(x, x_0) \wedge LOVE_*(x, x_0)))$ T2
11. $= \vee \wedge \lambda x(KISS_*(x, x_0) \wedge LOVE_*(x, x_0))(j)$ λ -conv.
12. $= \lambda x(KISS_*(x, x_0) \wedge LOVE_*(x, x_0))(j)$ $\vee \wedge$ -elim.
13. $= KISS_*(j, x_0) \wedge LOVE_*(j, x_0)$ λ -conv.
14. $Mary \mapsto \lambda X \vee X(m)$ T1b
15. $queen \mapsto QUEEN$ T1a
16. $F_3(queen) \mapsto \lambda Y \exists x(\forall y(QUEEN(y) \leftrightarrow x = y) \wedge \vee Y(x))$ T4
17. $F_6(Mary, the\ queen) \mapsto \lambda X(\lambda X \vee X(m)(X) \vee$
 $\lambda Y \exists x(\forall y(QUEEN(y) \leftrightarrow x = y) \wedge \vee Y(x))(X))$ T13
18. $= \lambda X(\vee X(m) \vee \exists x(\forall y(QUEEN(y) \leftrightarrow x = y) \wedge \vee X(x)))$ λ -conv.
19. $F_{7,0}(Mary\ or\ the\ queen, John\ kisses\ him_0\ and\ loves\ him_0)$
 $\mapsto \lambda X(\vee X(m) \vee \exists x(\forall y(QUEEN(y) \leftrightarrow x = y) \wedge \vee X(x)))$
 $(\wedge \lambda x_0(KISS_*(j, x_0) \wedge LOVE_*(j, x_0)))$ T8, 0
20. $= \vee \wedge \lambda x_0(KISS_*(j, x_0) \wedge LOVE_*(j, x_0))(m) \vee$
 $\exists x(\forall y(QUEEN(y) \leftrightarrow x = y)$
 $\wedge \vee \wedge \lambda x_0(KISS_*(j, x_0) \wedge LOVE_*(j, x_0))(x))$ λ -conv.
21. $\lambda x_0(KISS_*(j, x_0) \wedge LOVE_*(j, x_0))(m) \vee$
 $\exists x(\forall y(QUEEN(y) \leftrightarrow x = y)$
 $\wedge \lambda x_0(KISS_*(j, x_0) \wedge LOVE_*(j, x_0))(x))$ $\vee \wedge$ -elim.
22. $= (KISS_*(j, m) \wedge LOVE_*(j, m)) \vee \exists x(\forall y(QUEEN(y) \leftrightarrow$
 $x = y) \wedge KISS_*(j, x) \wedge LOVE_*(j, x))$ λ -conv.

In this way we obtain a reading in which the sentence can be true in a situation where there is no queen, that is, in that situation in which it is true that John kisses and loves Mary.

The reading which asserts the existence of the queen can be obtained by analyzing the sentence syntactically as shown in figure (b):

b.

John kisses Mary or the queen and loves her, S, S8, 2



The translation of analysis tree (b) is:

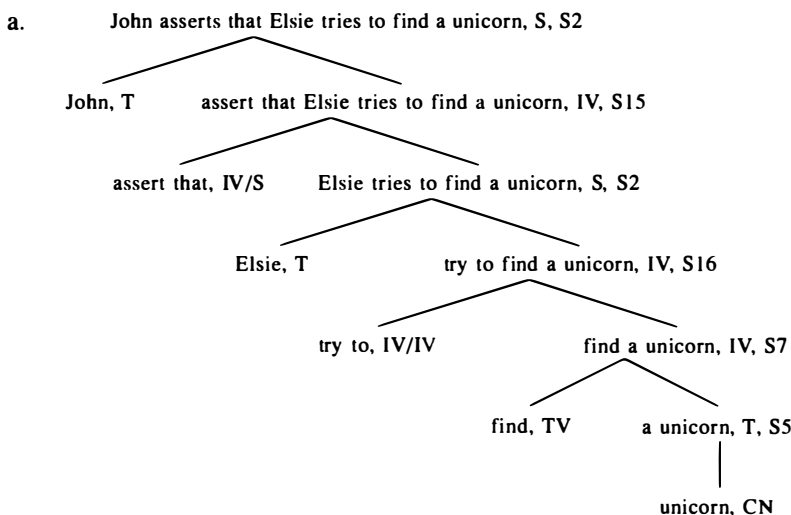
1. $\text{John kisses him}_0 \text{ and loves him}_0 \vdash \dots$ see (1)–(13)
 $= \text{KISS}^*(j, x_0) \wedge \text{LOVE}^*(j, x_0)$ above
2. $\text{Mary} \vdash \lambda X \vee X(m)$ T1b
3. $he_2 \vdash \lambda X \vee X(x_2)$ T1b
4. $F_9(\text{Mary}, he_2) \vdash \lambda X (\lambda X \vee X(m)(X) \vee \lambda X \vee X(x_2))(X)$ T13
5. $= \lambda X (\vee X(m) \vee \vee X(x_2))$ λ -conv.
6. $F_{7,0}(\text{Mary or } he_2, \text{John kisses him}_0 \text{ and loves him}_0) \vdash$
 $\lambda X (\vee X(m) \vee \vee X(x_2)) (\wedge \lambda x_0 (\text{KISS}^*(j, x_0) \wedge$ T8, 0
 $\text{LOVE}^*(j, x_0)))$
7. $= \vee \wedge \lambda x_0 (\text{KISS}^*(j, x_0) \wedge \text{LOVE}^*(j, x_0))(m) \vee$ λ -conv.
 $\vee \wedge \lambda x_0 (\text{KISS}^*(j, x_0) \wedge \text{LOVE}^*(j, x_0))(x_2)$
8. $= \lambda x_0 (\text{KISS}^*(j, x_0) \wedge \text{LOVE}^*(j, x_0))(m) \vee$ $\vee \wedge$ -elim.
 $\lambda x_0 (\text{KISS}^*(j, x_0) \wedge \text{LOVE}^*(j, x_0))(x_2)$
9. $= (\text{KISS}^*(j, m) \wedge \text{LOVE}^*(j, m)) \vee (\text{KISS}^*(j, x_2) \wedge$ λ -conv.
 $\text{LOVE}^*(j, x_2))$
10. $queen \vdash \text{QUEEN}$ T1a
11. $F_3(queen) \vdash \lambda X \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee X(x))$ T4
12. $F_{7,2}(\text{the queen}, \text{John kisses Mary or him}_2$
 $\text{and loves him}) \lambda X \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge$
 $\vee X(x)) (\wedge \lambda x_2 ((\text{KISS}^*(j, m) \wedge \text{LOVE}^*(j, m)) \vee$
 $(\text{KISS}^*(j, x_2) \wedge \text{LOVE}^*(j, x_2))))$ T8, 2
13. $= \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge \vee \wedge \lambda x_2 ((\text{KISS}^*(j, m) \wedge$ λ -conv.
 $\text{LOVE}^*(j, m)) \vee (\text{KISS}^*(j, x_2) \wedge \text{LOVE}^*(j, x_2)))(x))$
14. $= \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge \lambda x_2 ((\text{KISS}^*(j, m) \wedge$ $\vee \wedge$ -elim.
 $\text{LOVE}^*(j, m)) \vee (\text{KISS}^*(j, x_2) \wedge \text{LOVE}^*(j, x_2)))(x))$
15. $= \exists x (\forall y (\text{QUEEN}(y) \leftrightarrow x = y) \wedge ((\text{KISS}^*(j, m) \wedge$ λ -conv.
 $\text{LOVE}^*(j, m)) \vee (\text{KISS}^*(j, x) \wedge \text{LOVE}^*(j, x)))$

Exercise 8

(a) The sentence *John asserts that Elsie tries to find a unicorn* contains two intensional expressions: *assert* and *try to find*. The term *a unicorn* can have narrow or wide scope with respect to both expressions. This results in three readings which can be paraphrased as follows:

- (i) There is a unicorn and John asserts that Elsie tries to find it.
- (ii) John asserts that there is a unicorn and that Elsie tries to find it.
- (iii) John asserts that Elsie tries to find a unicorn.

The third reading, in which *a unicorn* has narrow scope with respect to both *assert* and *try to find*, is the result of constructing the tree directly, as shown in figure (a):

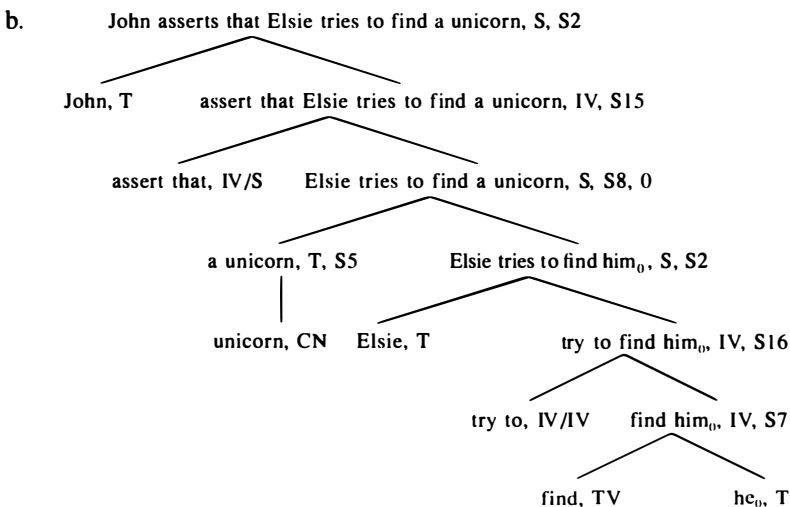


Translation:

- | | |
|---|----------------------|
| 1. <i>unicorn</i> \mapsto UNICORN | T1a |
| 2. $F_4(\textit{unicorn}) \mapsto \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))$ | T5 |
| 3. <i>find</i> \mapsto FIND | T1a |
| 4. $F_6(\textit{find}, a \textit{ unicorn}) \mapsto \text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x)))$ | T7 |
| 5. <i>try to</i> \mapsto TRY | T1a |
| 6. $F_{12}(\textit{try to}, \textit{find a unicorn}) \mapsto \text{TRY}(\wedge \textit{find}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))))$ | T16 |
| 7. <i>Elsie</i> $\mapsto \lambda Y \vee Y(e)$ | T1b |
| 8. $F_1(\textit{Elsie}, \textit{try to find a unicorn}) \mapsto \lambda X \vee X(e) (\wedge \text{TRY}(\wedge \text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))))$ | T2 |
| 9. $= \vee \wedge \text{TRY}(\wedge \text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))))(e)$ | λ -conv. |
| 10. $= \text{TRY}(\wedge \text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))))(e)$ | $\vee \wedge$ -elim. |
| 11. $= \text{TRY}(e, \wedge \text{FIND}(\wedge \lambda X \exists x(\text{UNICORN}(x) \wedge \vee X(x))))$ | NC1 |

12. $= \text{TRY}(e, \wedge \lambda y \text{FIND}(y, \wedge \lambda x \exists x (\text{UNICORN}(x) \wedge \vee X(x))))$ NC1, λ -abstr.
13. $= \text{TRY}(e, \wedge \lambda y (\vee \wedge \lambda x \exists x (\text{UNICORN}(x) \wedge \vee X(x)) (\wedge \lambda z \text{FIND}_*(y, z))))$ Theorem 1
14. $= \text{TRY}(e, \wedge \lambda y \lambda x \exists x (\text{UNICORN}(x) \wedge \vee X(x)) (\wedge \lambda z \text{FIND}_*(y, z)))$ $\vee \wedge$ -elim.
15. $= \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \vee \wedge \lambda z \text{FIND}_*(y, z)(x)))$ λ -conv.
16. $= \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \lambda z \text{FIND}_*(y, z)(x)))$ $\vee \wedge$ -elim.
17. $= \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x)))$ λ -conv.
18. *assert that* $\vdash \text{ASSERT}$ T1a
19. $F_{11}(\text{assert that}, \text{Elsie tries to find a unicorn}) \vdash \text{ASSERT}(\wedge \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x))))$ T15
20. *John* $\vdash \lambda X \vee X(j)$ T1b
21. $F_1(\text{John}, \text{assert that Elsie tries to find a unicorn}) \vdash \lambda X \vee X(j) (\wedge \text{ASSERT}(\wedge \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x))))$ T2
22. $= \vee \wedge \text{ASSERT}(\wedge \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x))))(j)$ λ -conv.
23. $= \text{ASSERT}(\wedge \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x))))(j)$ $\vee \wedge$ -elim.
24. $= \text{ASSERT}(j, \wedge \text{TRY}(e, \wedge \lambda y \exists x (\text{UNICORN}(x) \wedge \text{FIND}_*(y, x))))$ NC1

In the second reading *a unicorn* has narrow scope with respect to *assert*, but wide scope over *try to find*. Of course this reading is obtained by using the quantification rule, as shown in figure (b):

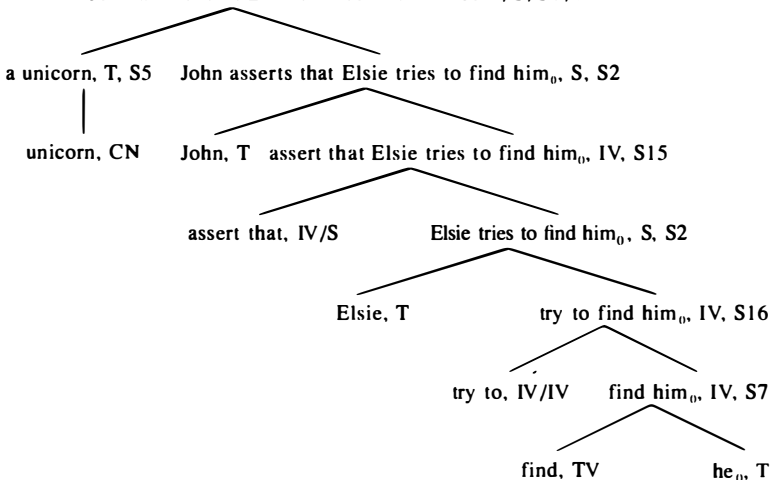


Omitting the steps resulting in the translation of *Elsie tries to find him₀*, which are completely analogous to 1–17 in the translation of (a), we obtain the translation of figure (b) as follows:

1. $F_1(\text{Elsie}, \text{try to find him}_0) \mapsto \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0))$ see above
2. $\text{unicorn} \mapsto \text{UNICORN}$ T1a
3. $F_4(\text{unicorn}) \mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$ T5
4. $F_{7,0}(a \text{ unicorn}, \text{Elsie tries to find him}_0) \mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x)) (\wedge \lambda x_0 \text{ TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0)))$ T8, 0
5. $= \exists x (\text{UNICORN}(x) \wedge \vee \wedge \lambda x_0 \text{ TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0))(x))$ λ -conv.
6. $= \exists x (\text{UNICORN}(x) \wedge \lambda x_0 \text{ TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0))(x))$ $\vee \wedge$ -elim.
7. $= \exists x (\text{UNICORN}(x) \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x)))$ λ -conv.
8. $\text{assert that} \mapsto \text{ASSERT}$ T1a
9. $F_{11}(\text{assert that}, \text{Elsie tries to find a unicorn}) \mapsto \text{ASSERT}(\wedge \exists x (\text{UNICORN}(x) \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x))))$ T15
10. $\text{John} \mapsto \lambda X \vee X(j)$ T1b
11. $F_1(\text{John}, \text{assert that Elsie tries to find a unicorn}) \mapsto \lambda X \vee X(j) (\wedge \text{ASSERT}(\wedge \exists x (\text{UNICORN}(x) \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x))))$ T2
12. $= \vee \wedge \text{ASSERT}(\wedge \exists x (\text{UNICORN}(x) \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x))))(j)$ λ -conv.
13. $= \text{ASSERT}(\wedge \exists x (\text{UNICORN}(x) \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x))))(j)$ $\vee \wedge$ -elim.
14. $= \text{ASSERT}(j, \wedge \exists x (\text{UNICORN}(x) \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x))))$ NC1

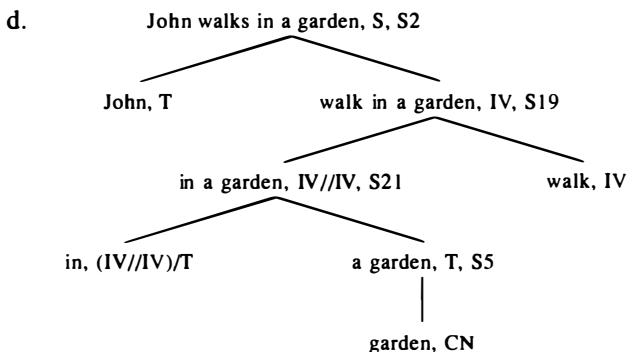
Finally, in the third reading, *a unicorn* has wide scope over both *assert* and *try to find*. This reading can also be obtained with the help of the quantification rule, as shown in figure (c):

c. John asserts that Elsie tries to find a unicorn, S, S8, 0



Again omitting the steps leading to the translation of *John asserts that Elsie tries to find him*₀, because it is completely analogous to the translation of tree (a) with *he*₀ in the place of *a unicorn*, we obtain the translation of tree (c) as follows:

1. *John asserts that Elsie tries to find him*₀ \mapsto
 $\text{ASSERT}(j, \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0)))$ see above
 2. *unicorn* \mapsto UNICORN T1a
 3. $F_4(\text{unicorn}) \mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$ T5
 4. $F_{7,0}(\text{a unicorn, John asserts that Elsie tries to find him}_0)$
 $\mapsto \lambda X \exists x (\text{UNICORN}(x) \wedge \vee X(x))$
 $(\wedge \lambda x_0 \text{ASSERT}(j, \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0))))$ T 8, 0
 5. $= \exists x (\text{UNICORN}(x) \wedge$
 $\vee \wedge \lambda x_0 \text{ASSERT}(j, \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0)))(x))$ λ -conv.
 6. $= \exists x (\text{UNICORN}(x) \wedge$
 $\lambda x_0 \text{ASSERT}(j, \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x_0)))(x))$ $\vee \wedge$ -elim.
 7. $= \exists x (\text{UNICORN}(x) \wedge$
 $\text{ASSERT}(j, \wedge \text{TRY}(e, \wedge \lambda y \text{ FIND}_*(y, x))))$ λ -conv.
- (b) S21: If $\gamma \in P_{(IV//IV)/T}$ and $\alpha \in P_T$, then $F_{16}(\gamma, \alpha) \in P_{IV//IV}$ and $F_{16}(\gamma, \alpha) = \gamma\alpha$.
 T21: If $\gamma \in P_{(IV//IV)/T}$ and $\alpha \in P_T$ and $\gamma \mapsto \gamma'$ and $\alpha \mapsto \alpha'$, then $F_{16}(\gamma, \alpha) \mapsto \gamma'(\wedge \alpha')$.
- (c) The analysis tree in figure (d) represents the direct construction.

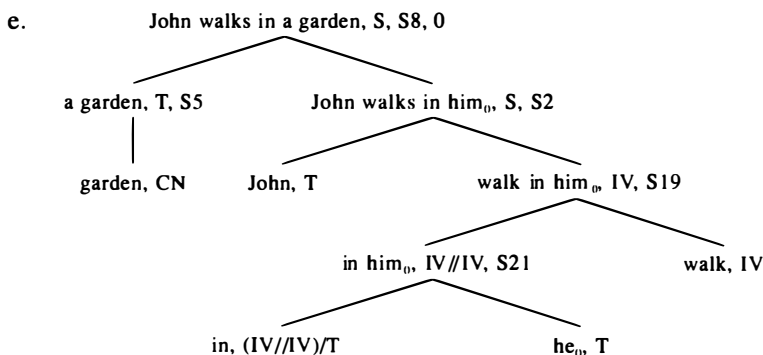


The translation of analysis tree (d) is:

1. *garden* \mapsto GARDEN T1a
2. $F_4(\text{garden}) \mapsto \lambda X \exists x (\text{GARDEN}(x) \wedge \vee X(x))$ T5
3. *in* \mapsto IN T1a
4. $F_{16}(\text{in, a garden}) \mapsto \text{IN}(\wedge \lambda X \exists x (\text{GARDEN}(x) \wedge \vee X(x)))$ T21
5. *walk* \mapsto WALK T1a
6. $F_{15}(\text{in a garden, walk}) \mapsto$
 $\text{IN}(\wedge \lambda X \exists x (\text{GARDEN}(x) \wedge \vee X(x)))(\wedge \text{WALK})$ T19
7. *John* $\mapsto \lambda X \vee X(j)$ T1b
8. $F_1(\text{John, walk in a garden}) \mapsto \lambda X \vee X(j)$
 $(\wedge \text{IN}(\wedge \lambda X \exists x (\text{GARDEN}(x) \wedge \vee X(x)))(\wedge \text{WALK}))$ T2

9. $= \vee \wedge \text{IN}(\wedge \lambda X \exists x(\text{GARDEN}(x) \wedge \vee X(x)))(\wedge \text{WALK})(j)$ $\lambda\text{-conv.}$
 10. $= \text{IN}(\wedge \lambda X \exists x(\text{GARDEN}(x) \wedge \vee X(x)))(\wedge \text{WALK})(j)$ $\vee \wedge\text{-elim.}$

This translation expresses that there is a three-place relationship *IN* between the second-order property expressed by *a garden*, the property of walking, and the individual John. This leaves the existence of gardens indeterminate, which is not in accordance with the meaning of the sentence under analysis. The indirect construction given in figure (e) results in a translation from which the existence of gardens can be inferred.



The translation of (e) is as follows:

1. $\text{John walks in him}_0 \mapsto \text{IN}(\wedge \lambda X \vee X(x_0))(\wedge \text{WALK})(j)$ see above
2. $\text{garden} \mapsto \text{GARDEN}$ T1a
3. $F_4(\text{garden}) \mapsto \lambda X \exists x(\text{GARDEN}(x) \wedge \vee X(x))$ T5
4. $F_{7,0}(a \text{ garden}, \text{John walks in him}_0) \mapsto$
 $\lambda X \exists x(\text{GARDEN}(x) \wedge \vee X(x))$
 $(\wedge \lambda x_0 \text{IN}(\wedge \lambda X \vee X(x_0)))(\wedge \text{WALK})(j))$ T 8, 0
5. $= \exists x(\text{GARDEN}(x) \wedge \vee \wedge \lambda x_0 \text{IN}(\wedge \lambda X \vee X(x_0))$
 $(\wedge \text{WALK})(j)(x))$ $\lambda\text{-conv.}$
6. $= \exists x(\text{GARDEN}(x) \wedge \lambda x_0 \text{IN}(\wedge \lambda X \vee X(x_0))$
 $(\wedge \text{WALK})(j)(x))$ $\vee \wedge\text{-elim.}$
7. $= \exists x(\text{GARDEN}(x) \wedge \text{IN}(\wedge \lambda X \vee X(x)))(\wedge \text{WALK})(j))$ $\lambda\text{-conv.}$

(d) The sentence *John walks in a garden* is not ambiguous: *in* denotes a relation between an entity, a property, and an entity. The following meaning postulate accounts for this:

$$\exists D \forall X \forall Y \forall x \square (\text{IN}(X)(Y)(x) \leftrightarrow \vee X(\wedge \lambda y D(y)(X)(x)))$$

D is a variable of type $\langle e, \langle \langle s, \langle e, t \rangle \rangle, \langle e, t \rangle \rangle \rangle$, the type of three-place relations between entities, properties of entities, and entities. With the help of the following notational convention, we also obtain a notation for the three-place relation whose existence is guaranteed by this meaning postulate:

$$\text{IN}^* =_{\text{def}} \lambda y \lambda X \lambda x \text{IN}(\wedge \lambda Y \vee Y(y))(X)(x)$$

The following theorem is valid because of the meaning postulate and the notational convention:

$$\forall X \forall \lambda \forall x \square (\text{IN}(X)(x) \leftrightarrow \forall y (\lambda y \text{ IN}_*(y)(X)(x)))$$

The result of the direct construction of *John walks in a garden* can now be reduced with the help of this theorem:

10. $\text{IN}(\lambda X \exists x (\text{GARDEN}(x) \wedge \forall X(x)))(\lambda \text{WALK})(j)$ see above
11. $= \forall \lambda X \exists x (\text{GARDEN}(x) \wedge \forall X(x))(\lambda y \text{ IN}_*(y)(\lambda \text{WALK})(j))$ Theorem 1
12. $= \lambda X \exists x (\text{GARDEN}(x) \wedge \forall X(x))(\lambda y \text{ IN}_*(y)(\lambda \text{WALK})(j))$ $\forall \wedge$ -elim.
13. $= \exists x (\text{GARDEN}(x) \wedge \forall \lambda y \text{ IN}_*(y)(\lambda \text{WALK})(j)(x))$ λ -conv.
14. $= \exists x (\text{GARDEN}(x) \wedge \lambda y \text{ IN}_*(y)(\lambda \text{WALK})(j)(x))$ $\forall \wedge$ -elim.
15. $= \exists x (\text{GARDEN}(x) \wedge \text{IN}_*(x)(\lambda \text{WALK})(j))$ λ -conv.

The result of the indirect construction can be reduced to 15 as well. We need only the notational convention to effect this.

7. $= \exists x (\text{GARDEN}(x) \wedge \text{IN}(\lambda X \forall X(x)))(\lambda \text{WALK})(j)(x)$ see above
8. $= \exists x (\text{GARDEN}(x) \wedge \text{IN}_*(x)(\lambda \text{WALK})(j))$ NC

Chapter 7

Exercise 1

- (i) $\{X | j \in X\}$;
- (ii) $\{X | 2 \leq \text{card}(X \cap \llbracket N \rrbracket)\}$ (to which we might add a condition that $\text{card}(X \cap \llbracket N \rrbracket)$ not exceed a contextually specified number ('relatively few N '));
- (iii) $\{X | \emptyset \neq X \cap \llbracket N \rrbracket \neq X\}$;
- (iv) $\{X | X \cap \llbracket N \rrbracket = \emptyset\}$, with the presupposition that $\text{card}(\llbracket N \rrbracket) = 2$;
- (v) $\{X | \text{card}(X \cap \llbracket N \rrbracket) \text{ is finite}\}$.

Exercise 2

- (a) (i), (ii), and (v) are upward monotonic; the others are not.
- (b) Since $\llbracket \text{NP} \rrbracket$ is nonempty and $\llbracket P \rrbracket = E$, it follows that there is an X such that $X \in \llbracket \text{NP} \rrbracket$ and $X \subseteq \llbracket P \rrbracket$. Hence, $\llbracket P \rrbracket \in \llbracket \text{NP} \rrbracket$, if NP is upward monotonic. That this property is not a sufficient condition for upward monotonicity is demonstrated by, for example, *all or no N*, which satisfies this condition yet is not upward monotonic.

Exercise 4

Assume that Q satisfies definition 1, and choose X and Y such that $X \cap Y \in Q$. Since $X \cap Y \subseteq X$ and $X \cap Y \subseteq Y$, it follows by definition 1 that $X \in Q$ and $Y \in Q$, which means that Q satisfies definition 3.

Assume that Q satisfies definition 3 and choose X and Y such that $X \subseteq Y$ and $X \in Q$. If $X \subseteq Y$, then $X \cap Y = X$, and hence $Y \in Q$ by definition 3; so Q satisfies definition 1.

Exercise 6

If $\llbracket \text{boy} \rrbracket_M$ is a singleton, $\llbracket \text{exactly one boy} \rrbracket_M$ is upward monotonic; and if $\llbracket \text{boy} \rrbracket_M = \emptyset$, $\llbracket \text{exactly one boy} \rrbracket_M$ is downward monotonic.

Exercise 7

Consider the following examples:

- (i) John and no woman walked rapidly \neq John and no woman walked.
- (ii) John and no woman walked \neq John and no woman walked rapidly.

The first example shows that *John and no woman* is not upward monotonic; the second one shows that it isn't downward monotonic either.

More formally, choose a model M and sets X , Y , and Z such that $X \in \llbracket \text{John and no woman} \rrbracket_M$, $Y = X - \{j\}$, and $Z = X \cup \{v\}$, with $v \in \llbracket \text{woman} \rrbracket_M$. We know that $\llbracket \text{John and no woman} \rrbracket = \{X \mid j \in X \text{ \& } X \cap \llbracket \text{woman} \rrbracket = \emptyset\}$. Now we have $Y \subseteq X$ and $Y \notin \llbracket \text{John and no woman} \rrbracket$; hence *John and no woman* is not downward monotonic. Also, we have $Z \supseteq X$ and $Z \notin \llbracket \text{John and no woman} \rrbracket$; hence *John and no woman* is not upward monotonic.

Exercise 10

First we consider negation, which we define as follows:

$$\neg D = \{\langle X, Y \rangle \mid \langle X, Y \rangle \notin D\}$$

For arbitrary X and Y we have: $\langle X, Y \rangle \in \neg D$ iff (by definition of $\neg D$) $\langle X, Y \rangle \notin D$ iff (by conservativity of D) $\langle X, X \cap Y \rangle \notin D$ iff (by definition of $\neg D$) $\langle X, X \cap Y \rangle \in \neg D$.

Next, conjunction, which is defined as:

$$D_1 \wedge D_2 = D_1 \cap D_2 = \{\langle X, Y \rangle \mid \langle X, Y \rangle \in D_1 \text{ \& } \langle X, Y \rangle \in D_2\}$$

It follows that: $\langle X, Y \rangle \in D_1 \wedge D_2$ iff (by definition of $D_1 \wedge D_2$) $\langle X, Y \rangle \in D_1$ and $\langle X, Y \rangle \in D_2$ iff (by conservativity of $D_1 \wedge D_2$) $\langle X, X \cap Y \rangle \in D_1$ and $\langle X, X \cap Y \rangle \in D_2$ iff (by definition of $D_1 \wedge D_2$) $\langle X, X \cap Y \rangle \in D_1 \wedge D_2$.

Finally, restriction:

$$D_A(X, Y) = D(X \cap A, Y)$$

So $\langle X, Y \rangle \in D_A$ iff (by definition of D_A) $\langle X \cap A, Y \rangle \in D$ iff (by conservativity of D) $\langle X \cap A, X \cap A \cap Y \rangle \in D$ iff $\langle X \cap A, X \cap A \cap X \cap Y \rangle \in D$ iff by conservativity of D $\langle X \cap A, X \cap Y \rangle \in D$ iff (by definition of D_A) $\langle X, X \cap Y \rangle \in D_A$.

Exercise 11

All but one has the same meaning as *exactly one not*, which in its turn can be viewed as the conjunction *at least one not and at most one not*. Both *at least one* and *at most one* being monotonic determiners, it suffices to show that this type of negation turns a monotonic determiner into a monotonic one. The definition is as follows:

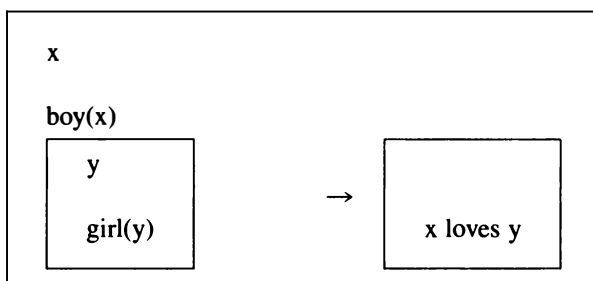
$$D\neg^* = \{\langle X, Y \rangle \mid \langle X, E - Y \rangle \in D\}$$

It can be proven that this negation reverses monotonicity: D is $\text{mon } \uparrow$ ($\text{mon } \downarrow$) iff $D\neg^*$ is $\text{mon } \downarrow$ ($\text{mon } \uparrow$). We show that if D is $\text{mon } \downarrow$, then $D\neg^*$ is $\text{mon } \uparrow$. Suppose D is $\text{mon } \downarrow$. Choose X, Y , and Z such that $\langle X, E - Y \rangle \in D$ and $Y \subseteq Z$. The first guarantees, by the definition of $D\neg^*$, that $\langle X, Y \rangle \in D\neg^*$. The second implies that $E - Z \subseteq E - Y$, whence, by the downward monotonicity of D , $\langle X, E - Z \rangle \in D$. By definition of $D\neg^*$ again, we have $\langle X, Z \rangle \in D\neg^*$, which shows that $D\neg^*$ is $\text{mon } \uparrow$. The other cases are similar.

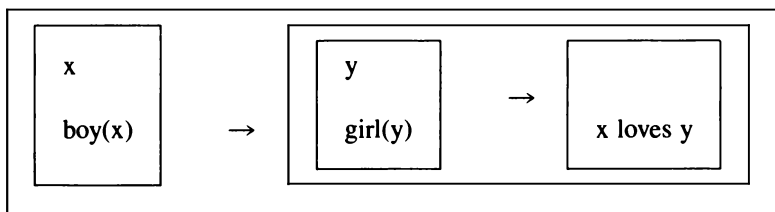
That *an even number of* is not a conjunction of monotonic determiners can be shown as follows. In view of fact 6, it is sufficient to show that this determiner is not continuous. Construct \mathbf{M} as follows: $E_{\mathbf{M}} = \{1, 2, 3, 4\}$, $\llbracket N \rrbracket_{\mathbf{M}} = \{1, 2, 3, 4\}$, $\llbracket V_1 \rrbracket_{\mathbf{M}} = \{1, 2\}$, $\llbracket V_2 \rrbracket_{\mathbf{M}} = \{1, 2, 3\}$ and $\llbracket V_3 \rrbracket_{\mathbf{M}} = \{1, 2, 3, 4\}$. Then we have: $\llbracket \text{an even number of } N V_1 \rrbracket_{\mathbf{M}} = \llbracket \text{an even number of } N V_3 \rrbracket_{\mathbf{M}} = 1$, but $\llbracket \text{an even number of } N V_2 \rrbracket_{\mathbf{M}} = 0$, even though $\llbracket V_1 \rrbracket_{\mathbf{M}} \subseteq \llbracket V_2 \rrbracket_{\mathbf{M}} \subseteq \llbracket V_3 \rrbracket_{\mathbf{M}}$.

Exercise 14

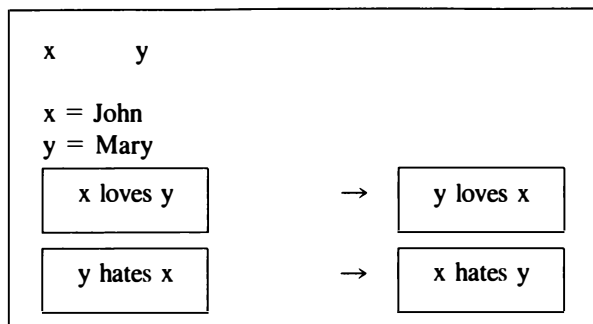
(a)



(b)



(c)

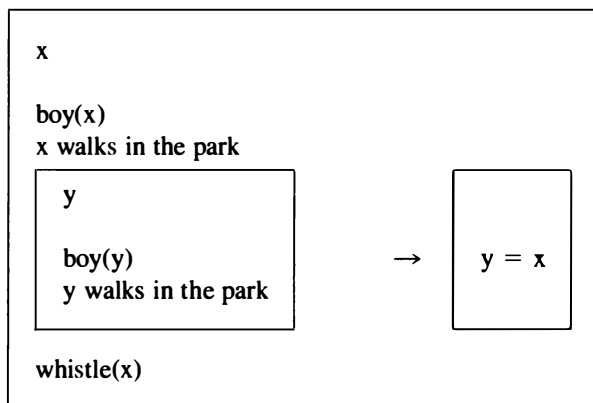


Note that in order to get a correct representation for the meaning of this sequence of sentences, the reference markers x and y and the conditions $x = \text{John}$ and $y = \text{Mary}$ have to be placed in the main DRS and not in the antecedents of the first sentence. Otherwise they could not bind the occurrences of x and y in the second conditional. This is the general way that DRT deals with proper names: they always introduce new markers in the main DRS.

Exercise 15

The DRS construction rule for NPs with the determiner *exactly one* reads as follows: (i) add a new reference marker x to the DRS; (ii) if α is the CN of the NP, then add a condition $\alpha(x)$ to the DRS; (iii) replace the subject NP of the sentence with the reference marker x ; (iv) add a condition consisting of two sub-DRSs connected by \rightarrow to the DRS; (v) repeat steps (i)–(iii) in the box on the left with a new marker y ; (vi) add a condition $y = x$ to the box on the right.

Using this construction rule the sequence of sentences (96) results in the following DRS:



Exercise 16

- (a) $\langle \{x\}, \{ \text{BOY}(x), (\langle \{y\}, \{ \text{girl}(y) \} \rightarrow \langle \emptyset, \{ \text{LOVE}(x, y) \} \rangle) \rangle \rangle$
 (b) $\langle \emptyset, \{ (\langle \{x\}, \{ \text{BOY}(x) \} \rightarrow \langle \emptyset, \{ (\langle \{y\}, \{ \text{GIRL}(y) \} \rightarrow \langle \emptyset, \{ \text{LOVE}(x, y) \} \rangle) \rangle) \} \rangle \rangle$
 (c) $\langle \{x, y\}, \{ x = \text{JOHN}, y = \text{MARY}, (\langle \emptyset, \{ \text{LOVE}(x, y) \} \rightarrow \langle \emptyset, \{ \text{LOVE}(y, x) \} \rangle), (\langle \emptyset, \{ \text{HATE}(y, x) \} \rightarrow \langle \emptyset, \{ \text{HATE}(x, y) \} \rangle) \} \rangle$
 (d) $\langle \{x\}, \{ \text{BOY}(x), \text{WALK IN THE PARK}(x), (\langle \{y\}, \{ \text{BOY}(y), \text{WALK IN THE PARK}(y) \} \rightarrow \langle \emptyset, \{ y = x \} \rangle), \text{WHISTLE}(x) \} \rangle$

If we use our abbreviatory notational conventions, (a)–(d) can be written as:

- (a) $\langle \{x\}, \{ \text{BOY}(x), (\langle \{y\}, \text{GIRL}(y) \rangle \rightarrow \text{LOVE}(x, y)) \} \rangle$
 (b) $\langle \{x\}, \text{BOY}(x) \rangle \rightarrow (\langle \{y\}, \text{GIRL}(y) \rangle \rightarrow \text{LOVE}(x, y))$
 (c) $\langle \{x, y\}, \{ x = \text{JOHN}, y = \text{MARY}, (\text{LOVE}(x, y) \rightarrow \text{LOVE}(y, x)), (\text{HATE}(y, x) \rightarrow \text{HATE}(x, y)) \} \rangle$
 (d) $\langle \{x\}, \{ \text{BOY}(x), \text{WALK IN THE PARK}(x), (\langle \{y\}, \{ \text{BOY}(y), \text{WALK IN THE PARK}(y) \} \rightarrow y = x), \text{WHISTLE}(x) \} \rangle$

Exercise 18

- (143) $\langle \{x\}, \{ \text{MAN}(x) \} \rangle \rightarrow \{ \text{WALK IN THE PARK}(x), \text{WHISTLE}(x) \}$
 (144) $\langle \{x\}, \{ \text{MAN}(x) \} \rangle \rightarrow \{ \neg \text{WALK IN THE PARK}(x), \text{HOME}(x) \}$
 (145) $\langle \{x\}, \{ \text{PLAYER}(x) \} \rangle \rightarrow \langle \{y\}, \{ \text{PAWN}(y), \text{CHOOSE}(x, y), \text{PUT ON SQUARE ONE}(x, y) \} \rangle$
 (146) $\langle \{x\}, \{ \text{CLIENT}(x), \text{ENTER}(x) \} \rangle \rightarrow \{ \text{TREAT POLITELY}(\text{YOU}, x), \text{OFFER COFFEE}(\text{YOU}, x), \text{ASK TO WAIT}(\text{YOU}, x) \}$

Exercise 19

A DRS, a predicate-logical formula, and a DPL formula which correctly represent the meaning of (96) are (a), (b), and (c), respectively, (we avoid the use of material equivalence in (b) to make comparison easier):

- (a) $\langle \{x\}, \{ \text{BOY}(x), \text{WALK IN THE PARK}(x), (\langle \{y\}, \{ \text{BOY}(y), \text{WALK IN THE PARK}(y) \} \rightarrow \langle \emptyset, \{ y = x \} \rangle), \text{WHISTLE}(x) \} \rangle$
 (b) $\exists x(\text{BOY}(x) \wedge \text{WALK IN THE PARK}(x) \wedge \forall y((\text{BOY}(y) \wedge \text{WALK IN THE PARK}(y)) \rightarrow y = x) \wedge \text{WHISTLE}(x))$
 (c) $\exists x(\text{BOY}(x) \wedge \text{WALK IN THE PARK}(x) \wedge \forall y((\text{BOY}(y) \wedge \text{WALK IN THE PARK}(y)) \rightarrow y = x)) \wedge \text{WHISTLE}(x)$

It is only in DPL formula (c) that the translation of the first sentence of (96) appears as a subformula in the translation of the sequence of sentences. In both DRS (a) and predicate-logical formula (b), the translation of the second

sentence has to be brought under the scope of the quantifier $\{x\}$ or $\exists x$, respectively, to get correct semantic results. So the DPL formula is the only compositional translation of the three; the DRS and the translation in predicate logic are equally noncompositional. Note that in the semantics of DPL, (b) and (c) are equivalent.

Exercise 20

Under the interpretation of disjunction given in definition 4, an existential quantifier $\exists x$ in the first disjunct cannot bind free occurrences of x in the second disjunct (nor the other way around). Further, if we continue a disjunction $\phi \vee \psi$ with a new conjunct: $(\phi \vee \psi) \wedge \chi$, an existential quantifier $\exists x$ in ϕ or ψ cannot bind occurrences of x in χ .

Under the proposed alternative definition of disjunction, it is still not possible for a quantifier in the first disjunct to bind variables in the second (nor vice versa). So this alternative interpretation of disjunction can be of no help in accounting for the anaphoric relations in the problematic donkey disjunctions (148) and (149) discussed in §7.4.4.

By the way, these examples are of the form $\neg \exists x \phi(x) \vee \psi(x)$. Since negation blocks binding of variables outside the scope of the negation by quantifiers inside the scope of that negation, no alternative definition of disjunction alone could help to account for the anaphoric relations in these examples. What would also be needed is an alternative definition of negation.

But the proposed alternative interpretation of disjunction differs from the original one in another respect. According to the alternative definition, it is possible for a quantifier $\exists x$ in either of the disjuncts ϕ or ψ to bind free occurrences of x in χ in the conjunction $(\phi \vee \psi) \wedge \chi$. In fact, $(\phi \vee \psi) \wedge \chi$ is strongly equivalent to $(\phi \wedge \chi) \vee (\psi \wedge \chi)$, in the sense that they have the same embedding conditions. So if each of the disjuncts ϕ and ψ contains an occurrence of the same quantifier $\exists x$, both occurrences will bind free occurrences of x in χ simultaneously. $(\exists x \phi \vee \exists x \psi) \wedge \chi$ is equivalent to $(\exists x \phi \wedge \chi) \vee (\exists x \psi \wedge \chi)$, and hence to $\exists x(\phi \wedge \chi) \vee \exists x(\psi \wedge \chi)$.

Therefore we can use the alternative notion of disjunction to account for the anaphoric relations in a sequence of sentences like:

A professor or an assistant professor will attend the meeting. He will report to the faculty.

In the translations of the two indefinite terms *a professor* and *an assistant professor*, we should then use the same quantifier $\exists x$, and the pronoun in the second sentence should be translated by the variable x .

Exercise 21

In DPL there is a *unique* smallest subset of $\{\exists, \forall, \neg, \wedge, \vee, \rightarrow\}$ in terms of which the remaining logical constants can be defined. As is explained in the

text, $\forall x\phi$ can be defined as $\neg \exists x\neg\phi$, but $\exists x\phi$ cannot be defined as $\neg\forall x\neg\phi$. $\exists x\phi$ and $\neg\forall x\neg\phi$ are not strongly equivalent in DPL. They do not have the same embedding conditions, though they do have the same truth conditions.

Similarly, it is possible to define $\phi \vee \psi$ as $\neg(\neg\phi \wedge \neg\psi)$ and to define $\phi \rightarrow \psi$ as $\neg(\phi \wedge \neg\psi)$. But it is not possible to define $\phi \wedge \psi$ as $\neg(\neg\phi \vee \neg\psi)$ or as $\neg(\phi \rightarrow \neg\psi)$. These three formulas have the same truth conditions but not the same embedding conditions. (By the way, the same holds for ϕ and $\neg\neg\phi$.)

So the only possible minimal subset of the total set of logical constants $\{\exists, \forall, \neg, \wedge, \vee, \rightarrow\}$ is the set $\{\exists, \neg, \wedge\}$.

Exercise 22

- (i) $\exists xFx \not\models_a Fx$ $\exists xFx \not\models_a \exists yFy$
 $\exists xFx \not\models_b Fx$ $\exists xFx \models_b \exists yFy$
 $\exists xFx \models_c Fx$ $\exists xFx \models_c \exists yFy$

- (ii) $\phi \models_c \psi$ iff $\models_c \phi \rightarrow \psi$

This does not hold for the notions \models_a and \models_b . Notion \models_c allows for quantifiers in the premise to bind variables in the conclusion, in the same way as a quantifier in the antecedent of an implication can bind variables in the consequent. As a notion of entailment, \models_c makes it possible to account for anaphoric relations in natural language argumentations such as: All human beings are mortal. Socrates is a human being. So *he* is mortal.

- (iii) Unlike \models_a and \models_b , the entailment relation \models_c is neither reflexive nor transitive. A counterexample to reflexivity: $Fx \wedge \exists xGx \not\models_c Fx \wedge \exists xGx$. While the occurrence of x in Fx in the premise is free, its occurrence in the conclusion is bound by the quantifier in the premise. A counterexample to transitivity: although it holds that $\exists xFx \models_c \exists yFy$ and that $\exists yFy \models_c Fy$, it does not hold that $\exists xFx \models_c Fy$.

Bibliographical Notes

These bibliographical notes contain suggestions for further reading, without any pretence at being exhaustive. In general, references to literature in the text are not repeated.

1. The Origins of Intensional Logic

The original papers in which Frege developed his ideas about meaning are Frege 1892a, 1892b, 1918a, 1918b, 1923. They are collected in Frege 1962a, 1962b. English translations of the first two papers can be found in Geach and Black 1960, and of the others in Geach 1975. The reference for a detailed exposition and appraisal of Frege's philosophy of language is Dummett 1973. See also Dummett 1981. A study from a historical perspective is Sluga 1980. A collection of recent papers is Wright 1984.

2. Intensional Propositional Logic

Two excellent textbooks on modal propositional logic are Hughes and Cresswell 1968 and, Chellas 1980. A survey of intensional systems with various applications can be found in van Benthem 1988. For recent computational connections, see also Goldblatt 1987. A useful collection of surveys of various branches of intensional logic is Gabbay and Guenther 1984. A pioneering work in building actual systems of intensional logic is Lewis 1918. For the original work of Carnap, Kanger, Hintikka, and Kripke, see Carnap 1947; Kanger 1957; Hintikka 1961; Kripke 1963.

No introductory works of similar status are available for tense logic (but cf. Rescher and Urquhart 1971). The reader may consult the original work of Prior (1967). An advanced technical study is van Benthem 1983a. Reichenbach's analysis of tense can be found in Reichenbach 1947. An interesting polemical study of the relation between standard logic and tense logic can be found in Needham 1975. For Kamp's analysis of the N-operator, see Kamp 1971.

Lewis's analysis of counterfactuals can be found in Lewis 1973. Also see Veltman 1981, 1985; Kratzer 1981.

3. Intensional Predicate Logic

Most textbooks on intensional logic concentrate on propositional logic. But Hughes and Cresswell 1968 contains an interesting section on modal predicate logic. Philosophically interesting are Hintikka 1969 and Plantinga 1974. A number of classical

papers in this field by Quine, Kaplan, Kripke, and Hintikka, among others, are collected in Linsky 1971.

For the theory of rigid designation the reader may consult, besides Kripke's paper referred to in the text, the work by Kripke, Putnam, Donellan, and Kaplan that can be found in Schwartz 1977 and in French, Uehling, and Wettstein 1979.

Quine expressed his ideas about intensional logic in a number of papers; see Quine 1961. For Lewis's counterpart theory, see Lewis 1968, 1973.

4. The Theory of Types and Categorical Grammar

Accessible textbooks on the theory of types are not available. An early formulation can be found in Church 1940. For Russell's theory of finite types, see Russell 1908. The reader may also consult Hindley and Seldin 1986.

The original formulation of categorical syntax was given by Lesniewski (1929). See also Adjuciewicz 1935. For the proposals made by Bar-Hillel, see Bar-Hillel 1953; for those of Lyons, see Lyons 1968. Other relevant literature is Lewis 1972; Montague 1970a, 1973; Cresswell 1973; Geach 1972; Bartsch and Vennemann 1972; and Bartsch 1976b. See also §7.3 for references to more recent literature.

5. The Intensional Theory of Types

An extensive logical study of the intensional theory of types and of two-sorted type theory is Gallin 1975. See also Janssen 1986.

For the theory of questions used in §5.5 see Groenendijk and Stokhof 1982, 1984, 1988b.

6. Montague Grammar

The last section of this chapter contains references to introductory texts and to other literature. See also the references in chapter 7.

The ideas of Tarski referred to in §6.1.2 can be found in Tarski 1935, 1944. For Kripke's theory, see Kripke 1975; for Gupta's alternative, see Gupta 1982. Both are reprinted, along with other relevant papers, in Martin 1984. See also Barwise and Etchemendy 1987.

7. Recent Developments

See the references given in the text. One important and influential development not treated in chapter 7 is that of 'situation semantics'. See Barwise and Perry 1983 and the special issue of *Linguistics and Philosophy*, 8 (1985).

References

- Adjuciewicz, K. 1935. Die syntaktische Konnexität. *Studia Philosophica* 1.
- Asher, N. 1986. Belief in discourse representation theory. *Journal of Philosophical Logic* 15.
- . 1987. A typology for attitude verbs and their anaphoric properties. *Linguistics and Philosophy* 10.
- Bach, E. 1979a. Control in Montague grammar. *Linguistic Inquiry* 10.
- . 1979b. Montague grammar and classical transformational grammar. In M. Mithun and S. Davis, eds., *Linguistics, Philosophy, and Montague Grammar*. Austin: University of Texas Press.
- Bar-Hillel, Y. 1953. A quasi-arithmetical notation for syntactic description. *Language* 29.
- Bar Hillel, Y., C. Gaifman, and E. Shamir. 1960. On categorial and phrase structure grammars. *Bulletin of the Research Council of Israel* 9F.
- Bartsch, R. 1976a. *The Grammar of Adverbials*. Amsterdam: North-Holland.
- . 1976b. The role of categorial syntax in grammatical theory. In A. Kasher, ed., *Language in Focus*. Dordrecht: Reidel.
- . 1979. The syntax and semantics of subordinate clause constructions and pronominal coreference. In F. Heny and H. Schnelle, eds., *Selections from the Third Groningen Round Table*. New York: Academic Press.
- Bartsch, R., and Th. Vennemann. 1972 *Semantic Structures*. Frankfurt: Athenäum.
- Barwise, J., and R. Cooper. 1981. Generalized quantifiers and natural language. *Linguistics and Philosophy* 4.
- Barwise, J., and J. Etchemendy. 1987. *The Liar: An Essay in Truth and Circularity*. Oxford: Oxford University Press.
- Barwise, J., and J. Perry. 1983. *Situations and Attitudes*. Cambridge, Mass.: Bradford Books.
- Bäuerle, R., U. Egli, and A. von Stechow, eds. 1979. *Semantics from Different Points of View*. Berlin: Springer.
- Bäuerle, R., Ch. Schwarze, and A. von Stechow, eds. 1983. *Meaning, Use, and the Interpretation of Language*. Berlin: de Gruyter.
- Belnap, N. 1982. Questions and answers in Montague grammar. In Peters, S. and E. Saarinen, eds., *Processes, Beliefs, and Questions*. Dordrecht: Reidel.
- Bennett, M. 1977. A guide to the logic of tense and aspect in English. *Logique et Analyse* 80.
- Benthem, J. van. 1983a. Determiners and logic. *Linguistics and Philosophy* 6.
- . 1983b. *The Logic of Time*. Dordrecht: Reidel.
- . 1984a. Questions about quantifiers. *Journal of Symbolic Logic* 49.

- . 1984b. The logic of semantics. In F. Landman and F. Veltman, eds., *Varieties of Formal Semantics*. Dordrecht: Foris.
- . 1986. *Essays in Logical Semantics*. Dordrecht: Reidel.
- . 1987. Semantic automata. In J. Groenendijk, D. de Jongh, and M. Stokhof, eds. *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Dordrecht: Foris.
- . 1988. *A Manual of Intensional Logic*. Chicago: University of Chicago Press.
- Benthem, J. van, and A. ter Meulen, eds. 1984. *Generalized Quantifiers in Natural Language*. Dordrecht: Foris.
- Buszkowski, W., W. Marciszewski, and J. van Benthem, eds. 1988. *Categorical Grammar*. Amsterdam: Benjamins.
- Carlson, G., 1977. Reference to kinds in English, Ph.D. diss., University of Amherst.
- . 1982. Generic terms and generic sentences. *Journal of Philosophical Logic* 11.
- Carnap, R. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.
- Chellas, B. 1980. *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.
- Chierchia, G. 1982. Nominalization and Montague grammar: A semantics without types for natural languages. *Linguistics and Philosophy* 5.
- . 1984. Topics in the syntax and semantics of infinitives and gerunds. Ph.D. diss., University of Amherst.
- Chierchia, G., and R. Turner. 1988. Semantics and property theory. *Linguistics and Philosophy* 11.
- Church, A. 1940. A formulation of a simple theory of types. *Journal of Symbolic Logic* 5.
- Cooper, R. 1979. The interpretation of pronouns, In F. Heny, and H. Schnelle, eds., *Selections from the Third Groningen Round Table*. New York: Academic Press.
- . 1984. *Quantification and Syntactic Theory*. Dordrecht: Reidel.
- Cooper, R., and T. Parsons. 1976. Montague grammar, generative semantics, and interpretive semantics. In B. Partee, ed., *Montague Grammar*. New York: Academic Press.
- Cresswell, M. 1973. *Logics and Languages*. London: Methuen.
- . 1976. The semantics of degree. In B. Partee, ed., *Montague Grammar*. New York: Academic Press.
- . 1979. Adverbs of space and time. In F. Guenther and S. Schmidt, eds., *Formal Semantics and Pragmatics for Natural Languages*. Dordrecht: Reidel.
- . 1985. *Structured Meanings*. Cambridge, Mass.: Bradford Books.
- Davidson, D., and G. Harman, eds. 1972. *Semantics of Natural Language*. Dordrecht: Reidel.
- Davis, S., and M. Mithun, eds. 1979. *Linguistics, Philosophy, and Montague Grammar*. Austin: University of Texas Press.
- Dowty, D. 1979. *Word Meaning and Montague Grammar*. Dordrecht: Reidel.
- Dowty, D., R. Wall, and S. Peters. 1981. *Introduction to Montague Semantics*. Dordrecht: Reidel.
- Dummett, M. 1973. *Frege: Philosophy of Language*. London: Duckworth.
- . 1981. *The Interpretation of Frege's Philosophy*. London: Duckworth.
- Engdahl, E. 1985. *Constituent Questions: The Syntax and Semantics of Questions with Special Reference to Swedish*. Dordrecht: Reidel.
- Eyck, J. van. 1985. Aspects of quantification in natural language. Ph.D. diss., Department of Philosophy, University of Groningen.

- Frege, G. 1879. *Begriffsschrift*. Halle: Verlag Louis Nebert. English translation in J. van Heyenoort, ed., *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879–1931*. Cambridge, Mass.: Harvard University Press, 1967.
- . 1892a. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100.
- . 1892b. Über Begriff und Gegenstand. *Vierteljahrsschrift für wissenschaftliche Philosophie* 16.
- . 1918a. Der Gedanke: Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus* 1.
- . 1918b. Die Verneinung: Eine logische Untersuchung. *Beiträge zur Philosophie des deutschen Idealismus* 1.
- . 1923. Logische Untersuchungen. Dritter Teil: Gedankengefüge. *Beiträge zur Philosophie des deutschen Idealismus* 3.
- . 1962a. *Funktion, Begriff, Bedeutung: Fünf Logische Studien*. ed. G. Patzig. Göttingen: Vandenhoeck.
- . 1962b. *Logische Untersuchungen*. ed. G. Patzig. Göttingen: Vandenhoeck.
- French, P., Th. Uehling, and H. Wettstein, eds., 1979. *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesota Press.
- Gabbay, D., and F. Guenther, eds. 1984. *Handbook of Philosophical Logic*. Vol. 2: *Extensions of Classical Logic*. Dordrecht: Reidel.
- Gallin, D. 1975. *Intensional and Higher-Order Modal Logic*. Amsterdam: North-Holland.
- Gazdar, G. 1982. Phrase structure grammar. In P. Jacobson and G. Pullum, eds., *The Nature of Syntactic Representation*. Dordrecht: Reidel.
- Gazdar, G., E. Klein, and I. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.
- Gärdenfors, P., ed. 1988. *Generalized Quantifiers: Linguistic and Logical Approaches*. Dordrecht: Reidel.
- Geach, P. 1972. A program for syntax. In D. Davidson and G. Harman, eds., *Semantics of Natural Language*. Dordrecht: Reidel.
- Geach, P., ed. 1975. *Logical Investigations*. Oxford: Blackwell.
- Geach, P., and M. Black, eds. 1960. *Translations from the Philosophical Writings of Gottlob Frege*. Oxford: Blackwell.
- Goldblatt, R. 1987. *Logics of Time and Computation*. Chicago: University of Chicago Press.
- Groenendijk, J., T. Janssen, and M. Stokhof, eds. 1981. *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre.
- . 1984. *Truth, Interpretation, and Information*. Dordrecht: Foris.
- Groenendijk, J., D. de Jongh, and M. Stokhof. 1987a. *Foundations of Pragmatics and Lexical Semantics*. Dordrecht: Foris.
- . 1987b. *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Dordrecht: Foris.
- Groenendijk, J. and M. Stokhof. 1982. Semantic analysis of *wh*- complements. *Linguistics and Philosophy* 5.
- . 1984. Studies on the semantics of questions and the pragmatics of answers. Ph.D. diss., Department of Philosophy, University of Amsterdam.
- . 1988a. Context and information in dynamic semantics. In B. Elsendoorn and H. Bouma, eds., *Working Models of Human Perception*. New York: Academic Press.

- . 1988b. Type-shifting rules and the semantics of interrogatives. In G. Chierchia, B. Partee, and R. Turner, eds., *Properties, Types, and Meaning*. Dordrecht: Reidel.
- . 1990. Dynamic predicate logic: Towards a compositional, non-representational semantics of discourse. *Linguistics and Philosophy* 13.
- Groenendijk, J., M. Stokhof, and F. Veltman, eds. 1988. *Proceedings of the Sixth Amsterdam Colloquium*. Amsterdam: ITLI.
- Guenther, F., and C. Rohrer, eds. 1978. *Studies in Formal Semantics, Intensionality, Temporality, Negation*. Amsterdam: North-Holland.
- Guenther, F., and S. Schmidt, eds. 1979. *Formal Semantics and Pragmatics for Natural Languages*. Dordrecht: Reidel.
- Gupta, A. 1982. Truth and paradox. *Journal of Philosophical Logic* 11.
- Halvorsen, P., and W. Ladusaw. 1979. Montague's 'Universal grammar': An introduction for the linguist. *Linguistics and Philosophy* 3.
- Hamblin, C. 1973. Question in Montague English. *Foundations of Language* 10.
- Hausser, R. 1979. How do pronouns denote? In F. Heny and H. Schnelle, eds., *Selections from the Third Groningen Round Table*. New York: Academic Press.
- . 1983. The syntax and semantics of English mood. In F. Kiefer, ed., *Questions and Answers*. Dordrecht: Reidel.
- . 1984. *Surface Compositional Grammar*. Munich: Fink.
- Hausser, R., and D. Zaffner. 1978. Questions and answers in a context-dependent Montague grammar. In F. Guenther and S. Schmidt, eds., *Formal Semantics and Pragmatics for Natural Languages*. Dordrecht: Reidel.
- Heim, I. 1982. The semantics of definite and indefinite noun phrases. Ph.D. diss., University of Massachusetts, Amherst.
- . 1983. File change semantics and the familiarity theory of definiteness. In R. Bäuerle, Ch. Schwarze, and A. von Stechow, eds., *Meaning, Use, and Interpretation of Language*. Berlin: De Gruyter.
- Hendriks, H. 1988. Type change in semantics: The scope of quantification and coordination. In E. Klein and J. van Benthem, eds., *Categories, Polymorphism, and Unification*. Edinburgh/Amsterdam: CCS/ITLI.
- Hindley, J., and J. Seldin. 1986. *Introduction to Combinators and λ -Calculus*. Cambridge: Cambridge University Press.
- Hinrichs, E. 1986. Temporal anaphora in discourses of English. *Linguistics and Philosophy* 9.
- Hintikka, J. 1961. Modality and quantification. *Theoria* 27.
- . 1969. *Models for Modalities*. Dordrecht: Reidel.
- Hoeksema, J. 1984. Categorical morphology. Ph.D. diss., Department of Linguistics, University of Groningen.
- Hughes, G., and M. Cresswell. 1968. *An Introduction to Modal Logic*. London: Methuen.
- Jacobs, J. 1982. *Syntax und Semantik der Negation im Deutschen*. Munich: Fink.
- Janssen, T. M. V. 1986. *Foundations and Applications of Montague Grammar*. Amsterdam: Centre for Mathematics and Computer Science.
- Jongh, F. de, and H. Verkuyl. 1984. Generalized quantifiers: the properness of their strength. In J. van Benthem and A. ter Meulen, eds., *Generalized Quantifiers in Natural Language*. Dordrecht: Foris.
- Kadmon, N. 1987. On unique and non-unique reference and asymmetric quantification. Diss., Department of Linguistics, University of Amherst.

- Kamp, H. 1971. Formal properties of 'Now'. *Theoria* 31.
- . 1975. Two theories about adjectives. In E. Keenan, ed., *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press.
- . 1980. Some remarks on the logic of change: Part 1. In C. Rohrer, ed., *Time, Tense, and Quantifiers*. Tübingen: Niemeyer.
- . 1981a. A theory of truth and semantic representation. In J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre. Reprinted in J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Truth, Interpretation, and Information*. Dordrecht: Foris, 1984.
- . 1981b. Événements, représentations discursives et référence temporelle. *Langues* 64.
- Kamp, H., and C. Rohrer. 1983. Tense in texts. In R. Bäuerle, Ch. Schwarze, and A. von Stechow, eds., *Meaning, Use, and Interpretation of Language*. Berlin: De Gruyter.
- Kanger, S. 1957. The morning star paradox. *Theoria* 23.
- Kaplan, D. 1978. Dthat. In P. Cole, ed., *Syntax and Semantics 9: Pragmatics*. New York: Academic Press.
- . 1979. On the logic of demonstratives. In P. French, Th. Uehling, and H. Wettstein, eds., *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesota Press.
- Kaplan, R., and J. Bresnan. 1982. Lexical-functional grammar: a formal system for grammatical representation. In J. Bresnan, ed., *The Mental Representation of Grammatical Relations*. Cambridge, Mass.: MIT Press.
- Karttunen, L. 1976. Discourse referents. In J. McCawley, ed., *Syntax and Semantics 7*. New York: Academic Press.
- . 1977. Syntax and semantics of questions. *Linguistics and Philosophy* 1.
- Katz, J. 1966. *The Philosophy of Language*. New York: Harper & Row.
- . 1981. *Language and Other Abstract Objects*. New York: Harper & Row.
- Keenan, E. 1987. Lexical freedom and large categories. In J. Groenendijk, D. de Jongh, and M. Stokhof, eds., *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Dordrecht: Foris.
- Keenan, E., ed. 1975. *Formal Semantics of Natural Language*. Cambridge: Cambridge University Press.
- Keenan, E., and L. Faltz. 1985. *Boolean Semantics for Natural Language*. Dordrecht: Reidel.
- Keenan, E. and L. Moss. 1984. Generalized quantifiers and expressive power of natural language. In J. van Benthem and A. ter Meulen, eds., *Generalized Quantifiers in Natural Language*. Dordrecht: Foris.
- Keenan, E. and J. Stavi. 1986. A semantic characterization of natural language determiners. *Linguistics and Philosophy* 9.
- Klein, E. 1980. A semantics for positive and comparative adjectives. *Linguistics and Philosophy* 4.
- Klein, E. and J. van Benthem, eds. 1988. *Categories, Polymorphism, and Unification*. Edinburgh/Amsterdam: CCS/ITLI.
- Klein, E., and I. Sag. 1984. Type-driven translation. *Linguistics and Philosophy* 8.
- Kratzer, A. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic* 10.
- Kripke, S. 1963. Semantical considerations on modal logic. *Acta Philosophica Fennica* 16.

- . 1972. Naming and necessity. In D. Davidson and G. Harman, eds., *Semantics of Natural Language*. Dordrecht: Reidel.
- . 1975. Outline of a theory of truth. *Journal of Philosophy* 72.
- Ladusaw, W. 1980. *Polarity Sensitivity as Inherent Scope Relations*. Indiana University Linguistics Club.
- Lambek, J. 1958. The mathematics of sentence structure. *American Mathematical Monthly* 65.
- Landman, F. 1988. Groups, plural individuals, and intentionality. In J. Groenendijk, M. Stokhof, and F. Veltman, eds., *Proceedings of the Sixth Amsterdam Colloquium*. Amsterdam: ITLI.
- Landman, F., and F. Veltman, eds. 1984. *Varieties of Formal Semantics*. Dordrecht: Foris.
- Lesniewski, S. 1929. Grundzüge eines neues Systems der Grundlagen der Mathematik. *Fundamenta Mathematicae* 14.
- Lewis, C. 1918. *A survey of Symbolic Logic*. Berkeley: University of California Press.
- Lewis, D. 1968. Counterpart theory and quantified modal logic. *Journal of Philosophy* 65.
- . 1972. General semantics. In D. Davidson and G. Harman, eds., *Semantics of Natural Language*. Dordrecht: Reidel.
- . 1973. *Counterfactuals*. Oxford: Blackwell.
- Link, G. 1979. *Montague-Grammatik*. Munich: Fink.
- . 1983. The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, Ch. Schwarze, and A. von Stechow, eds., *Meaning, Use, and Interpretation of Language*. Berlin: De Gruyter.
- Linsky, L., ed. 1971. *Reference and Modality*. Oxford: Oxford University Press.
- Löbner, S. 1976. *Einführung in die Montague-Grammatik*. Kronberg: Scriptor.
- . 1987. Quantification as a major module of natural language semantics. In J. Groenendijk, D. de Jongh, and M. Stokhof, eds., *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Dordrecht: Foris.
- Locke, J. 1690. *An Essay Concerning Human Understanding*. Abridged and ed. by A. Woozley. London: Collins, 1964.
- Lyons, J. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- McCloskey, J. 1979. *Transformational Syntax and Model-Theoretic Semantics*. Dordrecht: Reidel.
- Martin, R., ed. 1984. *Recent Essays on Truth and the Liar Paradox*. Oxford: Oxford University Press.
- Meulen, A. ter, ed. 1983. *Studies in Model-Theoretic Semantics*. Dordrecht: Foris.
- Montague, R. 1970a. English as a formal language. In B. Visentini, et al., eds., *Linguaggi nella società e nella tecnica*. Milan: Edizioni di Comunità. Reprinted in Montague 1974.
- . 1970b. Universal grammar. *Theoria* 36. Reprinted in Montague 1974.
- . 1973. The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravcsik, and P. Suppes, eds., *Approaches to Natural Language*. Dordrecht: Reidel. Reprinted in Montague 1974.
- . 1974. *Formal Philosophy: Selected Papers of Richard Montague, Edited and with an Introduction by Richmond H. Thomason*. New Haven: Yale University Press.

- Moortgat, M. 1988. *Categorial Investigations: Logical and Linguistic Aspects of the Lambek Calculus*. Dordrecht: Foris.
- Needham, P. 1975. *Temporal Perspective*. Uppsala: Philosophical Studies.
- Oehrle, R., E. Bach, and D. Wheeler, eds. 1988. *Categorial Grammars and Natural Language Structures*. Dordrecht: Reidel.
- Partee, B. 1973. Some transformational extensions of Montague grammar. *Journal of Philosophical Logic* 2.
- . 1975. Montague grammar and transformational grammar. *Linguistic Inquiry* 6.
- . 1979a. Constraining transformational Montague grammar: A framework and a fragment. In S. Davis, and M. Mithun, eds., *Linguistics, Philosophy and Montague Grammar*. Austin: University of Texas Press.
- . 1979b. Montague grammar and the well-formedness constraint. In F. Heny and H. Schnelle, eds., *Selections from the Third Groningen Round Table*. New York: Academic Press.
- . 1984a. Compositionality. In F. Landman and F. Veltman, eds., *Varieties of Formal Semantics*. Dordrecht: Foris.
- . 1984b. Nominal and temporal anaphora. *Linguistics and Philosophy* 7.
- Partee, B., ed. 1976. *Montague Grammar*. New York: Academic Press.
- Partee, B., and E. Bach. 1981. Quantification, pronouns, and VP-anaphora. in J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre. Reprinted in J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Truth, interpretation and Information*. Dordrecht: Foris. 1984.
- Partee, B., and M. Rooth, 1983. Generalized conjunction and type ambiguity. In R. Bäuerle, Ch. Schwarze, and A. von Stechow, eds., *Meaning, Use, and Interpretation of Language*. Berlin: De Gruyter.
- Plantinga, A. 1974. *The Nature of Necessity*. Oxford: Oxford University Press.
- Prior, A. 1967. *Past, Present, and Future*. Oxford: Oxford University Press.
- Pullum, G., and G. Gazdar. 1982. Natural languages and context-free languages. *Linguistics and Philosophy* 4.
- Quine, W. 1961. *From a Logical Point of View*. Cambridge, Mass.: MIT Press.
- Reichenbach, H. 1947. *Elements of Symbolic Logic*. New York: Macmillan.
- Rescher, N., and A. Urquhart. 1971. *Temporal Logic*. Berlin: Springer.
- Roberts, C. 1987. Modal subordination, anaphora, and distributivity. Ph.D. diss., Department of Linguistics, University of Amherst.
- . 1989. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy* 12.
- Rohrer, C., ed. 1980. *Time, Tense, and Quantifiers*. Tübingen: Niemeyer.
- Russell, B. 1908. Mathematical logic as based on the theory of types. *American Journal of Mathematics* 30.
- Savitch, W., E. Bach, W. Marsh, and G. Safran-Navek, eds. 1987. *The Formal Complexity of Natural Language*. Dordrecht: Reidel.
- Scha, R. 1981. Distributive, collective, and cumulative quantification. In J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre. Reprinted in J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Truth, Interpretation, and Information*. Dordrecht: Foris. 1984.
- Schwartz, S., ed. 1977. *Naming, Necessity, and Natural Kinds*. Ithaca: Cornell University Press.

- Seuren, P. 1985. *Discourse Semantics*. Oxford: Blackwell.
- Sluga, H. 1980. *Gottlob Frege*. London: Routledge & Kegan Paul.
- Stalnaker, R., and R. Thomason. 1973. A semantic theory of adverbs. *Linguistic Inquiry* 4.
- Stevenson, R. 1886. *Dr. Jekyll and Mr. Hyde*. London.
- Tarski, A. 1935. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica* 1. English translation in A. Tarski, *Logic, Semantics, Metamathematics*. Oxford: Oxford University Press. 1956.
- . 1944. The semantic conception of truth. *Philosophy and Phenomenological Research* 4.
- Thijssen, E. 1983. On some proposed universals of natural language. In A. ter Meulen, ed., *Studies in Model-Theoretic Semantics*. Dordrecht: Foris.
- . 1984. Counting quantifiers. In J. van Benthem and A. ter Meulen, eds., *Generalized Quantifiers in Natural Language*. Dordrecht: Foris.
- Turner, R. 1983. Montague semantics, nominalization, and Scott's domains. *Linguistics and Philosophy* 6.
- Veltman, F. 1981. Data semantics. In J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre. Reprinted in J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Truth, Interpretation, and Information*. Dordrecht: Foris.. 1984.
- . 1985. Logics for conditionals. Ph.D. diss., Department of Philosophy, University of Amsterdam.
- Verkuyl, H. 1981. Numerals and quantifiers in X-bar syntax and their semantic interpretation. In J. Groenendijk, T. Janssen, and M. Stokhof, eds., *Formal Methods in the Study of Language*. Amsterdam: Mathematical Centre.
- . 1987. Aspect, quantification and negation. In J. Groenendijk, M. Stokhof, and F. Veltman, eds., *Proceedings of the Sixth Amsterdam Colloquium*. Amsterdam: ITLI.
- . 1989. Aspectual classes and aspectual composition. *Linguistics and Philosophy* 12.
- Wittgenstein, L. 1921. *Tractatus Logico-Philosophicus*. In *Annalen der Naturphilosophie*. English translation by D. Pears and B. McGuinness. London: Routledge & Kegan Paul, 1961.
- Wright, C. ed. 1984. *Frege: Tradition and Influence*. Oxford: Blackwell.
- Zeevat, H. 1987. A treatment of belief sentences in discourse representation theory. In J. Groenendijk, D. de Jongh, and M. Stokhof, eds., *Studies in Discourse Representation Theory and the Theory of Generalized Quantifiers*. Dordrecht: Foris.
- Zwarts, F. 1981. Negatief polaire uitdrukkingen I. *GLOT* 4.
- . 1983. 'Determiners: A relational perspective. In A. ter Meulen, ed., *Studies in Model-Theoretic Semantics*. Dordrecht: Foris.
- . 1986. Categoriale grammatica en algebraïsche semantiek. Ph.D. diss., Department of Linguistics, University of Groningen.

Index

- a(n)*, 154, 164
- abstraction over a free variable, 104
- accessibility relation, 18
- accidental property, 47
- adjective, 77, 195f.; intersective, 199; relative, 199
- adverb, 171f., 195f.; predicate, 77, 200; sentence-modifying, 200
- alethic modality, 30
- alleged*, 46, 197
- always*, 40
- ambiguity: de dicto/de re, 147, 170; derivational, 147f.; scope, 165, 170; structural, 147
- analogs, 3
- analysis: structural, 148; syntactic, 148; tree, 155
- anaphora, 221
- anaphoric relation, 266ff.
- anomalists, 3
- anti-persistent, 243
- any*, 224
- argument lowering, 260
- argument schema, xv
- Aristotle's Sea Battle argument, 32
- article. *See* determiner
- assignment, 65
- at least*, 243, 249
- at most*, 242
- autonomy of syntax, 141, 215
- axiom, 28
- axiom scheme, 29
- Barcan formula, 61, 63, 66
- basic: expression, 151; rule, 153; type, 79
- be*, 187ff.
- Bedeutung* (Frege), 7; *ungerade*, 12
- behavioristic theories of meaning, 2
- believe*, 16, 27, 45, 73
- bidirectional categorial grammar, 93ff.
- bound occurrence of a variable, xiv
- box, 272
- bracket, 104
- branching time, 25, 41f.
- can*, 19
- cap, 120–28
- Cartesian product, xv
- CAT, 151, 205
- categorematic introduction, 114, 155, 163, 195
- categorial syntax, 150
- category: basic, 92, 150, 205; derived, 92, 150
- category change rule, 258
- certainly*, 19
- characteristic function, 83f.
- characterization of frames, 25
- CN, 151, 205, 226
- composition of predicates, 105
- compositionality, 12, 214, 265; of meaning, 5, 140; of translation, 112
- CON, 80, 119
- concatenation, 92f., 150f.
- conceptualism, 3
- condition in a DRS, 274, 278
- conjunction, xiv, 101, 106
- conjunction reduction, 224
- conjunction rule, 191f.
- connected relation, 35
- connectives, 'strong' and 'weak' interpretation of, 56
- conservativity, 245, 263
- constant, individual, xiv
- constituent: discontinuous, 98; intransitive verbal, 151; transitive verbal, 152
- constraint, 226, 240, 245
- construction: direct way of, 181, 202; indirect way of, 181
- context, 14, 16f.; conceived as moments of time, 16; hyperintensional, 73; intensional, 157; of use, 67; opaque, 45; transparent, 45

- context dependence, 13, 16
 context-free grammar, 95, 99, 150
 contingent, 19
 continuous determiner, 250
 contradiction, xv, 109
 conventionalism, 2
 conversion. *See* λ -conversion
 coordination, 259
 copula, 187
 coreference, 193, 202
 correspondence theories of meaning, 1ff.
 counterexample, 25
 counterfactual, 10, 52
 counterpart relation, 72
 cup, 120–28

 DRS, 272; condition, 274, 278; construction rule, 272; subordination between, 276; truth of, 275; verifying embedding, 275
 DRT, 264–97; and compositionality, 285–96
 decreasing domains, 63, 65
 de dicto/de re, 170, 196, 202
 definite description, xv, 49f., 158, 173, 230
 definition, recursive, xiv
 deitic pronoun, 67
 density, 36
 deontic logic, 16
 derivability: axiomatic, xv; syntactic, xv
 derivational ambiguity, 147f.
 derivational history, 184
 derived category, 92, 150
 derived type, 79
 DET, 226
 determiner, 226ff.; antipersistent, 243; conservative, 245, 263; continuous, 250; having extension, 250; having variety, 250; homogeneity of, 255; intensional, 247; logical, 229, 251ff.; persistent, 242; possessive, 252; quantitative, 251; simple, 246; uniformity of, 255
 direct interpretation, 230
 discontinuous constituent, 98
 discourse, 264ff.
 discourse, referent, 269
 discourse representation structure, 272
 discrete time, 36
 disjunction, xiv
 disjunction rule, 191f.
 domain, 4; of expressions of type *a*, 84
 domain function, 53f.
 domains of NP's: increasing, 57, 62; decreasing, 63, 65
 donkey-sentence, 270ff.
 double indexing, 70
 downward monotonic, 234
 dual of a quantifier, 238
 dynamic predicate logic, 292

earlier than relation, 33f.
 elimination: $\wedge \vee$, 128
 empty set, xv
 epistemic: alternative, 16; modality, 30
 equivalence, xiv; logical, xv; material, xiv; strong, 95; weak, 95
 equivalence relation, 29
 essentialism, 47
every, 113f., 164, 179, 227f.
exactly, 115, 239, 243, 250, 268
 existence predicate, 63f.
 existential: quantifier, xiv, 89; reading, 185
 expression, intensionally closed, 130, 137
 extension, 14, 123, 132f.
 extensional verb, 189
 extensional logical systems, 5, 145
 extensionality, principle of, 5, 133

Färbung (Frege), 7
 first order predicate logic, 76
 flexible categorial grammar, 257–64
 formula: of predicate logic, xiv; propositional, xiv
 frame, 22; irreflexive, 25; reflexive, 25; transitive, 26
free for x, 110
 free occurrence of a variable, xiv, 109
 Frege's principle, 140. *See also* compositionality
 function, characteristic, 83f.
 functional abstraction, 105
 functional application, 79, 104, 153, 263
 functor category, 151
 future tense, 32, 39f.

 Geach's rule, 258f.
Gedanke, 10
 generalized quantifiers, 223–57
 global constraint, 245
 grammar: context-free, 95, 99; generalized phrase structure, 100, 220; generative, 92, 99, 141, 144, 214; universal, 225
 grammatical form, 214–20

he_n, 151, 180
 hearer, 67

higher order quantification, 89
hyperintensional context, 73

I, 67ff.

idea (Vorstellung), 8

identity, 57, 187, 263

identity relation, 101

implication, xiv; material, xiv, 20; strict, 20

implicational logic, 260

increasing domains requirement, 57, 62

index, 67

indirect speech, 45

indiscernibility of identicals, 5

individual, 49

individual concept, 14, 49, 59, 68, 203–14

individual constant, xiv

infinitival complement, 195

initial baptism, 53

integer, 36

intensional, 14; context, 157; determiner,
247; operator, 131; semantics, 14; types,
119, 126, 263; verb, 165

intensional construction, 7

intensionally closed expression, 130, 137

intensional verb, 45, 186

interpretation function, xv

interpretation of natural language: direct, 148,
230; indirect, 148

intersection, xv

intransitive verb phrase, 151, 205, 226

introduction of an individual, 269

irreflexive relation, 25, 34

IV, 151, 205, 226

K (minimal modal propositional logic), 29
know, 45

Kraft (Frege), 7

Kripke model, 18

λ -abstraction, 104, 111, 263

λ -conversion, 109–12, 131ff.

λ -operator, 102–8

Lambek calculus, 260

Leibniz's law, 5

lexical element, 151

lexical meaning, 173

lexicon, 150

linear order, 36

live on, 246

Logic: deontic, 16; epistemic, 16; many-
valued, 56; minimal, 29; modal proposi-
tional, 16–32; propositional tense, 32–44;

second order, 76; second order predicate,
103, 159

logical equivalence, xv

logical form, 214–20

logically necessary, 79

look for, 45

many, 115, 239, 243, 246

many-valued logic, 56

material implication, 20

meaning: and reference, 7; and sense, 7f.;
compositionality of, 5, 140

'meaning is use' theory, 2

meaning postulate(s), 171ff.; MP1, 173; MP2,
175, 213; MP3, 190; MP4, 190; MP5, 197;
MP6, 199; MP7, 201; MP8, 210; MP9,
212; MPI0, 212

meaning theories: behavioristic, 2; correspon-
dence, 1ff.; meaning is use, 2; referential,
4ff.

mental representation, 145

mentalism, 141, 145

metalanguage, 142

minimal modal logic (K), 29

Misleading Form Thesis, 214

model: for intensional propositional logic, 18;
for intensional type theory, 122; for modal
predicate logic, 53; for modal propositional
logic, 22; for theory of types, 85f.

modal operators, stacked, 21

modal propositional logic, 16–32

modality: athletic, 30; de dicto/de re, 45;
epistemic, 30

moments of time, 16, 67

monotonical decrease. *See* downward
monotonic

monotonical increase. *See* upward monotonic
monotonicity, 232ff., 262; constraint, 240;
downward, 234; upward, 232

Montague rule, 259f.

morning star/evening star paradox, 6, 9

most, 115

N, 226

NP, 226

narrow scope, 179

natural deduction, 28

naturalism, 2

naturalism-conventionalism, 2

necessary, 19, 200, 208

necessary: logical, 17, 29; physical, 17

negation, xiv, 257

- negation of quantifiers: external, 236; internal, 236
- negation rule, 191f.
- negative polarity, 224
- nominalization, 259
- notational conventions: NC1, 167; NC2, 176, 213; NC3, 211
- now*, 37f, 69f.
- object language, 142
- obligatory*, 16, 20
- one*, 115, 179; *at least*, 243; *exactly*, 115, 239
- only*, 239, 246
- opaque context, 45ff.
- operator, 16; modal, 21ff.; tense, 32ff.; unary, 16
- order, linear, 36
- ordered *n*-tuple, xv
- paradox(es), 6, 9, 20, 78f., 142
- partiality of interpretation, 56
- past tense, 32, 39f.
- persistence, 242
- personal pronoun, 67
- physical necessity, 17
- picture theory of meaning, 4
- Polish notation, 101
- Plato's beard*, 51
- Platonism, 3
- point: of event, 39; of reference, 39; of speech, 39
- possible, 18
- possible individuals, 185
- possible world, 22, 134
- power set, xv, 84
- pragmatics, 73
- predicate adverbials, 77, 200
- predicate: existence, 63f.; reflexive, 103; second order, 103
- preposition, 78
- present perfect tense, 40
- principle of semantic compositionality, 5, 12
- problem of transworld identity, 52, 71
- pronoun: anaphoric, 184, 221; deictic, 67; personal, 67; reflexive, 180
- proper name, 8, 49f., 158, 173, 206, 238, 263, 273
- property, 47, 121
- proposition, 14, 121
- propositional attitude, 45
- propositional formula, xiv
- quantification: interactive, 292; vacuous, 181
- quantified term, 154, 159
- quantifier: existential, xiv; universal, xiv
- quantifying into, 180, 193, 201f., 267, 271
- quantitative determiner, 251
- rational number, 36
- reading: specific, 185; non-specific, 185
- realism, 3
- recursive, xiv; definition, xiv
- reduction, λ . *See* λ -conversion
- reduction step, 110
- reference, 7, 144; indirect, 12; marker, 273; multiple, 14, 144; self-, 143
- reflexive predicate, 103, 107
- relation: irreflexive, 25, 34; reflexive, 24; connected, 35; symmetric, 26; transitive, 26, 35
- relative adjective, 77, 199
- relative clause, 195
- rigid designator, 10, 49f., 73, 173, 184
- rule of quantification, 178–87
- Russell paradox, 78f.
- S, 151
- S5, 29f.
- scope, xiv
- scope ambiguity, 165, 170
- Sea Battle argument, 32
- second order: logic, 76; predicate, 103, 159
- seek*, 157, 165f, 197
- self-reference, 143
- semantic automaton, 255
- semantic paradoxes, 142
- semantic universal, 226, 239f.
- semantic closure, 142
- semantics: extensionalist, 145; intensional, 14, 145; lexical, 172
- sense, 7f.
- sentence: conjunction, 191; semantics, 172
- sentential complement, 195
- set abstractor, 105
- similarity relation, 43
- simple past tense, 32, 39f.
- Sinn* (Frege), 7, 118
- situation semantics, 144
- some*, 115, 229, 243, 251
- sometimes*, 40
- speaker, 67
- specific reading, 185
- square of opposition, 238
- stacked modal operators, 21

- strict implication, 20
 'strong' interpretation of the connectives, 56
 structural ambiguity, 147
 structure, 22
 surface structure, 214
 symmetry, 26
 syncategorematic introduction, 101, 155
 synonymy, 145
 syntactic operation: F_1 , 153; F_2 – F_5 , 154; F_6 , 166; $F_{7,n}$, 180, 202; F_8 , 191; F_9 , 192; F_{10} , 194; F_{11} , 195; $F_{12,n}$, 199; F_{13} , 200
 syntactic rules, 92; S_1 , 153; S_2 , 153; S_3 – S_6 , 154; S_7 , 166; S_8 , n, 180; S_9 , 191; S_{10} – S_{13} , 192; S_{14} , 194; S_{15} , 195; S_{16} , 196; S_{17} , 198; S_{18} , n, 199; S_{19} , 200; S_{20} , 200; S_{20} , n, 202
 syntactic variable, 151
 syntax: autonomy of, 141, 215; bidirectional, 93ff.; categorical, 150; context-free, 95, 99; unidirectional, 93ff., 151

 T2, 133f.
 tautology, xv
 tense (of a verb), 33, 39f.
 tense and modality, 40f.
 tense logic, 32–44
 tense operator, 32
 term, 152f.
 text, 264
the, 164
there insertion, 224f.
 time axis, 33ff.; branched, 35
 transformational-generative grammar, 92, 99
 transitive verb phrase, 152
 transitive verb, 165
 transitivity, 26, 35
 translation: compositionality of, 112; function f , 156, 206; function g , 158, 207
 translation rules: $T_1(a)$, 158, 207; $T_1(a')$, 190; $T_1(b)$, 162, 207; $T_1(c)$, 188, 208; $T_1(c')$, 165; T_2 , 163; T_3 – T_6 , 160, 207; T_3' , 165; T_7 , 167; T_8 , n, 181, 208; T_9 , 191; T_{10} – T_{13} , 192; T_{14} , 194; T_{15} , 195; T_{16} , 197; T_{17} , 198; T_{18} , n, 199; T_{19} , 200; T_{20} , 200; T_{20} , n, 202
 translation tree, 163

 transparent context, 45
 transworld identity problem, 52, 71
 tree of numbers, 251f.
 true, in a model, xv
 truth, 144
 truth definition, 4; for intensional type theory, 123; for modal predicate logic, 54, 60, 64; for modal propositional logic, 23; for theory of types, 86, 105
 truth of a discourse, 265
 truth of a DRS in a model, 275
 truth predicate, 142
try, 197
 TV, 152, 226
 two-sorted type theory, 133–38
 type, 79f., 119, 126; basic, 79, 100; derived, 79; intensional, 119, 126; two-sorted, 132–38

 unidirectional categorical grammar, 93ff., 151
 union of sets, xv
 universal, 225, 246; grammar, 225; quantifier, xiv, 89; validity, xv
 upward monotonic, 232

 validity, xv; in a model, 24; on a frame, 25; syntactic, 28; universal, xv
 valuation, xv
 VAR, 80, 119
 variable, xiv
 variable: free, 110; syntactic, 151
 variety, 249
 verb: extensional, 189; intransitive, 151, 205, 226; transitive, 152, 226
 verb phrase: intransitive, 151, 205, 226; transitive, 152, 226
 verifying embedding of a DRS, 275
 vocabulary, 80
Vorstellung (Frege), 8
 VP, 226

 WE, 80, 119
 weak equivalence, 95
 wide scope, 179
 word order, 98
 word semantics, 172

Although the two volumes of *Logic, Language, and Meaning* can be used independently of one another, together they provide a comprehensive overview of modern logic as it is used as a tool in the analysis of natural language. Both volumes provide exercises and their solutions.

Volume 2, *Intensional Logic and Logical Grammar*, begins with an introduction to the various principles of intensional semantics, and then provides a treatment of various intensional logics, such as modal propositional logic and modal predicate logic, and tense logic. It also introduces type theory, lambda-abstraction, and categorial syntax. Type theory and intensional logic are combined in the well-known intensional type logic employed by Montague. After the necessary technical logical machinery is thus presented, a detailed introduction to Montague Grammar as a specimen of logical grammar follows. The final chapter is dedicated to an introduction of three new developments in logical grammar: the theory of generalized quantifiers, flexible categorial syntax, and discourse representational theory.

L. T. F. Gamut is a collective pseudonym for **J. F. A. K. van Benthem**, professor of mathematical logic; **J. A. G. Groenendijk**, senior lecturer in the department of philosophy; **D. H. J. de Jongh**, associate professor in the department of philosophy and mathematics; **M. J. B. Stokhof**, lecturer in the department of computational linguistics, all at the University of Amsterdam, and **H. J. Verkuyl**, professor of linguistics at the University of Utrecht.



The University of Chicago Press

ISBN 0-226-28088-8



9 780226 280882

90000>

