The Language Of Space: The Acquisition And Interpretation of Spatial Adpositions In English

Francesco-Alessio Ursini
M.Phil. Linguistics

Macquarie Centre for Cognitive Science
Faculty of Human Sciences, Macquarie University
Sydney, Australia

This thesis is presented for the degree of Ph.D. in Cognitive Science

June, 2011
Contents

Summary iv
Acknowledgements vi

1 General Introduction: Scope and Goal(s) of This Thesis 1

I Adpositions, Space and The Language-Vision Interface 12

2 Adpositions, Space and the Language-Vision Interface: a Model-Theoretic Approach 13

2.1 Introduction: what We talk about, when We talk about Space . . . . . . . . . 13

2.2 The Relation between Spatial Vision and Language . . . . . . . . . . . . . . 16

2.2.1 Basic Notions of Space . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

2.2.2 Previous Literature . . . . . . . . . . . . . . . . . . . . . . . . . . . . 19

2.3 The Nature of Spatial Vision and Language, and a Formal Analysis . . . . . 23

2.3.1 Classical and Modern Varieties of Object Recognition . . . . . . . . . . 23

2.3.2 A Logic of Vision, Part I: Static Vision . . . . . . . . . . . . . . . . . . 29

2.3.3 Theories of Dynamic Vision . . . . . . . . . . . . . . . . . . . . . . . . 36

2.3.4 A Logic of Vision, Part II: a Model of Visual Logical Space . . . . . 41

2.3.5 A Logic of Language: Discourse Representation Theory and Space . . 46
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4</td>
<td>A Theory of the Vision-Language Interface, and beyond</td>
</tr>
<tr>
<td>2.4.1</td>
<td>The Vision-Language Interface: a Formal Approach</td>
</tr>
<tr>
<td>2.4.2</td>
<td>Testing the Theory against the Data</td>
</tr>
<tr>
<td>2.4.3</td>
<td>What is unique to Language, and why</td>
</tr>
<tr>
<td>2.5</td>
<td>Conclusions</td>
</tr>
<tr>
<td>II</td>
<td>The Grammar Of Adpositions And Spatial Sentences</td>
</tr>
<tr>
<td>3</td>
<td>The Grammar of Adpositions and Spatial Sentences, Part I: Syntax</td>
</tr>
<tr>
<td>3.1</td>
<td>Introduction: the Problem of Ps and their Structure</td>
</tr>
<tr>
<td>3.2</td>
<td>Previous Literature, and Basic Facts about Ps</td>
</tr>
<tr>
<td>3.3</td>
<td>The Structure of Ps, I: a Proposal, and some Empirical Coverage</td>
</tr>
<tr>
<td>3.4</td>
<td>The Structure of Ps: a Derivational Account, and more Empirical Coverage</td>
</tr>
<tr>
<td>3.5</td>
<td>Conclusions</td>
</tr>
<tr>
<td>4</td>
<td>The Grammar of Adpositions and Spatial Sentences, Part II: Semantics</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>4.2</td>
<td>What Ps denote, I: Previous Proposals</td>
</tr>
<tr>
<td>4.3</td>
<td>What Ps denote, II: a Proposal for Ontological Simplicity</td>
</tr>
<tr>
<td>4.4</td>
<td>What Ps denote, II: a Novel Proposal</td>
</tr>
<tr>
<td>4.5</td>
<td>The Semantics of Ps: Analysis of the Data</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Locative Ps</td>
</tr>
<tr>
<td>4.5.2</td>
<td>Directional Ps: the Facts</td>
</tr>
<tr>
<td>4.5.3</td>
<td>“Intermediate” Ps, and other Phenomena</td>
</tr>
<tr>
<td>4.5.4</td>
<td>Basic Facts of Clause Architecture and Sentence Interpretation</td>
</tr>
<tr>
<td>4.6</td>
<td>Conclusions</td>
</tr>
</tbody>
</table>
SUMMARY

This thesis by publication presents a study on English adpositions (e.g. to, in, at, from, in front of, through). It attempts to offer a solution to the following three outstanding problems, which are presented in each of the three parts making up the thesis, preceded by a general introduction (chapter 1) and followed by the general conclusions (chapter 7). The first part includes chapter 2, and discusses the problem of What is the relation between adpositions and the non-linguistic, visual content they represent. The second part includes chapters 3 and 4, and discusses the problem of what is a proper compositional theory of the Syntax and Semantics of adpositions. The third part includes chapters 5 and 6, and discusses the problem of what is the psychological reality of this theory, regarding adults and children’s data.

The following three solutions are suggested. First, the relation between adpositions and their corresponding visual information is an isomorphism: adpositions capture how we “see” possible spatio-temporal relations between objects, at a flexible level of fine-grainedness. Second, a proper compositional treatment of adpositions treats each syntactic unit (in front, of) as offering a distinct semantic contribution, hence spelling out a restricted instance of a spatio-temporal part-of relation. Third, this compositional treatment of adpositions can also stand as a theory of on-line interpretation in adults and a theory of their acquisition in children.

These three answers are couched within a single theoretical approach, that of Discourse Representation Theory, and offer a unified solution to three apparently distinct problems regarding spatial adpositions and their linguistic properties.
DECLARATION

I, Francesco-Alessio Ursini, declare that this thesis titled, “The Language Of Space: The Acquisition And Interpretation of Spatial Adpositions In English” and the work presented in it are my own. I confirm that this work was done wholly or mainly while in candidature for a research degree at this University. Where I have consulted or quoted from the published work of others, this is always clearly attributed and the source always given. With the exception of such quotations, this thesis is entirely my own work. I have acknowledged all main sources of help. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself. I have sought and obtained Ethics Committee approval, protocol number HE25JUL2008-D05968L&P.

Some of the material appearing in this thesis has already been accepted for publication. It is adapted and presented as an integral part of the present thesis. Chapter 2 will appear in the journal Biolinguistics 5(3), pp. 500-553, with minor formatting revisions. Chapter 3 will appear in a different format in the journal Lingue & Linguaggio X(I), pp. 57-87, with various formatting and prose revisions. Chapter 4 and 6 are currently in preparation for submission. All references are collated in the “Bibliography” section.

The full references for these publications are:


Signed:

Francesco-Alessio Ursini (Student Number: 41145755)
Acknowledgements

Ph.D. theses are written for one purpose: to thank all the people that have supported the author in the long road to the thesis’ completion. I will be no exception, although I will be extremely concise in acknowledging everyone. I shall start by writing a few words in Italian, since part of my relevant audience may not be able to read the acknowledgements, if written in a Language they don’t speak that well.

Un grazie veramente sentito alla mia famiglia (genitori, nonni, zii, cugini) per il supporto finanziario, nel momento del bisogno. Senza di voi non ce l’avrei fatta, quindi: Grazie!

Thanks to my Supervisors: Stephen Crain, Drew Khlentzos and Rosalind Thornton (“Rozz”). Their supervision has always been enlightening, especially when I had no clue on what I had to do, in order to meet my goals. I would also like to thank them for their long-term research. I think that, without the pioneering work of Stephen and Rozz, I would not have had the honor of testing whether the rigorous tenets of Logic indeed represent “real” linguistic processes. Thanks for making this science so fun to do, guys. I also think that, without Drew’s insights and suggestions about how to properly present logic-bound arguments, I would have not been able to present my arguments in a clear and concise way. Thanks for making the right suggestion at the right time, Drew.

Thanks for the participants of my experiments, adults and children alike. The data presented here existed because you enjoyed my crazy stories about Tank Engines, the amnesiac Mr. Little Bears and Godzilla as a good guy. Thanks for your answers, which made me discover the joys of checking out whether my predictions were borne out or not. Thanks, of course, to Godzilla,
Mr. Little Bears and Thomas and the other tank engines, who worked hard as my experimental
props.
Thanks to Gene Roddenberry for “Star Trek”, to Leiji Matsumoto for “Captain Harlock”, to
Eiichiro Oda for “One Piece”, to Hideaki Anno for “the wings of Honnemaise”, and to Hayao
Miyazaki for “Laputa, castle in the sky”. Without the burning desire for discovering bold new
worlds, it is hard to do proper science. These works constantly acted as a memento of this fact,
via some kind of peculiar process of abduction. Thanks to Hideaki Sorachi for “Gintama”: a
heart-felt laugh is the best cordial for the soul.
Thanks to Shinobu Yagawa for programming “Battle Garegga”, and to Tsuneki Ikeda for pro-
gramming “Batsugun”. I played little during my Ph.D., but I always managed to find time for
these two shmups, every once in a while. Both games constantly acted as a memento that, with-
out method and practice, no result can be obtained, scientific or ludic alike; and that Rank must
be manipulated wisely, for great success!
Thanks to Zuntata for composing so many soundtracks, and to Philip Glass for his early works.
Every human is, at some point in life, an Albert Einstein wondering about the clouds on his
tomorrow, in the darkness of one’s bed. Every arcade gamer is, at some point in life, a fan of
Taito games and the Darius series OSTs, because these soundtracks are worth the price of one
play.
Finally, thanks to my Princess, whose constant support and endearing love has been the key
factor to overcome all the obstacles I have faced. This thesis is for you.
Chapter 1

General Introduction: Scope and Goal(s) of This Thesis

This thesis presents a novel study on English spatial adpositions, such as \textit{in}, \textit{at}, \textit{to}, \textit{from}, \textit{in front of}, \textit{ahead of}, and similar others. The goal of this study is to offer an account on the interpretation of adpositions by English native speakers which meets three empirical goals still in need of a thorough solution.

The first goal is to offer a thorough account of the relation between adpositions and the non-linguistic information they are matched with. The aim is to analyze and account \textit{what} kind of visual spatial information we try to capture when we use an adposition such as \textit{in front of}, \textit{what} is the relation between the visual, spatial non-linguistic information, “what we see”, and spatial linguistic information, “what we say”.

The second goal is to offer a thorough account of the structure and interpretation of adpositions which is inherently compositional in nature, and which explains why certain semantic distinctions among adpositions appear to be systematic. The aim is to analyze and account \textit{what} are the differences in interpretation between \textit{in front of} and \textit{ahead of}, since we can say that one constituent remains constant (i.e. \textit{of}) while others vary from adposition to adposition (i.e. \textit{in front vs. ahead}); and \textit{why} adpositions appear to be distinguishable among adpositions express-
ing motion and change (e.g. to, from), and adpositions expressing stasis and location (e.g. at, in front of); and what is the logical relation between these adpositions, e.g. why certain adpositions appear to be connected in certain entailment patterns (e.g. to and at).

The third goal is to offer empirical evidence regarding the psychological reality of this interpretive process. The aim is to analyze and account how English speakers interpret on-line (i.e. in real time) adpositions such as in, at, to and from and the logical relations holding among these adpositions, as well as offering an account of how and why these adpositions emerge in children’s Language in a certain order (e.g. in before to).

In reaching these three goals, this thesis attempts to solve three outstanding problems regarding adpositions that are still in need of a solution, and on which there is little consensus in the relevant literature. The following examples will help us in illustrating the problems at stake. Adpositions are in cursive:

(1) The boy is sitting in front of the desk
(2) The boy is sitting on top of the desk
(3) The boy is sleeping in the bed
(4) The boy went to the desk
(5) The boy was sitting at the desk
(6) The boy has arrived from the room

Each sentence in (1)-(6) captures the position of one entity, a certain boy, with respect to a given desk (or bed, room), acting as a “reference point”, and thus depicts a slightly different scenario than the ones captured by the other sentences. Although each adposition appears to convey at a minimum a “core” spatial relation holding between boy and desk, each adposition also conveys some more specific information, which makes more precise the position of the boy with respect to the desk. For instance, in front of in (1) conveys the information that the boy is sitting in a position which is aligned with the front section of the desk, whereas on top of in (2) conveys the information that the boy is using the main surface of the desk as a sitting ground. In (3),
in conveys the information that boy and bed virtually occupy the same position, so that the boy is in a sense “contained” by the bed, while the boy is taking some rest. In (4), to conveys the information that the boy was somewhere close to the desk after moving in its direction; at in (5) conveys the information that the boy was sitting somewhere close to the desk, possibly as a result of moving to the desk. In (6), from conveys the information that the boy was somewhere around (or in) the room before moving.

Sentences (1)-(6) are also connected via certain logical relations that hold among them, because of the information that adpositions convey. For instance, the sentence in (5) can be understood as expressing the “logical” consequence of (4): if the boy went to the desk, he was (sitting) at the desk, i.e. somewhere in the proximity of this object, as a consequence of this “event” of motion. If one wants to be more accurate about the boy’s location, then the sentences in (1)-(2) may be thought as conveying a more specific “part” of the information expressed by (5): if Mario is sitting on top of the desk, he is certainly somewhere in proximity of (or at) the desk. More in general, sentences including adpositions can be related one another via the entailment and the sub-set relation. Respectively, the truthful interpretation of one sentence can be accessed once the truthful interpretation of another sentence is accessed (entailment, e.g. the relation between (4) and (5)); the truthful interpretation of a sentence can be accessed as a more specific part of the truthful interpretation of another sentence (sub-set, e.g. the relation between (1) and (5)). In both cases, adpositions play an active part in establishing this relation, because they may capture relations between different positions that an entity may occupy with respect to another reference entity, as in the described cases involving boy and desk.

The intuition that spatial adpositions capture how we conceive spatial relations between objects is to an extent uncontroversial: although adpositions such as to suggest that these parts of speech may capture a “richer” notion of spatial relations than a purely geometrical one, their role as the chief part of speech expressing “where” things are, is intuitively correct and more or less uncontroversial. The analysis of these facts, however, is far from uncontroversial, with the following three problems being particularly pressing.
Our first outstanding problem is that we still do not have a clear picture of the relation between Spatial Vision and Spatial Language, between our ability to see and find things in the world, and our ability to exchange information about these things and their positions. We still do not have a clear picture of what type of visual information we represent and process when we see boy and desk in each of the visual scenarios corresponding to the sentence (1)-(4), and thus of what kind of spatial information adpositions such as in front of, on top of, to and at convey. Also, we still do not have a clear picture of what kind of relation exists between this visual information and linguistic information: whether an adposition such as in front of expresses all or just a part of the non-linguistic spatial information we can access from the corresponding scenario, and what is the precise relation between these two levels of information-processing.

The lack of a precise picture, and its corresponding problem, can be illustrated by the main controversy found in the literature regarding this topic. Several proposals attempt to solve this problem, offering for the most part partial solutions. Some proposals can give an accurate analysis of how visual information is processed, but not of how this information is matched with adpositions (e.g. Coventry & Garrod 2004). Other proposals can give an accurate analysis of how adpositions express spatial information, but not of the visual information that these adpositions capture (e.g. Landau & Jackendoff 1993). No proposals deal in detail with “dynamic” scenarios and related sentences: sentences such as (3) are beyond the empirical coverage of both types of proposals. More in general, no proposal can successfully gives a unified account of both sets of data, nor establish a precise relation between these sets: what kind of visual information an adposition such as in front of corresponds to.

Our second outstanding problem is that we still do not have a clear picture of subtle syntactic and semantic differences between adpositions, or why both in front of and on top of appear to express the same type of “static” relation expressing the boy’s sitting position, in (1) and (2). Intuitively, the subtle difference in meaning between these two adpositions stems from the difference in interpretation between in front and on top: however, we still do not have a clear picture of how this difference can be captured in a principled way. We still do not have a clear
picture of *how* to give a *compositional* treatment of the Syntax and Semantics of adpositions: *what* are the underlying syntactic and semantic structures that these adpositions express in Language, *how* they are combined together to form complex sentences, and what is the exact nature of the logical relations holding among these sentences. The lack of such a precise picture has also one important consequence: we do not know if, by proposing an *off-line* linguistic theory of adpositions, we are in position to test the *on-line* status of this theory, thus offering experimental evidence that support this approach.

The precise shape of this problem is the following. From a syntactic perspective, we still do not have a clear picture of *what* kind of syntactic structure adpositions such as *in front of* or *on top of* correspond to, and whether this structure has a fixed position in clausal structure. We also don’t know *how* this structure emerges as the result of a syntactic process, or how *of* can combine with either *in front* or *on top*, intuitively yielding the same syntactic structure. From a semantic perspective, we still do not have a clear picture of *what* kind of semantic content these elements express, as the result of this compositional process; *what* is the semantic contribution of *in front* and *on top*, once they combine with *of* and with parts of speech in a sentence; and *why* adpositions such as *to* and *in front of* appear to express different types of spatial relations (respectively, “stasis” vs. “change”), or adpositions such as *to* and *at* appear to be logically connected, as sentences (4) and (5) suggest.

The lack of a precise picture, and its corresponding problem, can be illustrated by analyzing the lack of a unified theory of the syntax and semantics of adpositions. Some syntactic proposals suggest that Ps consist of several syntactic positions, but do not offer an account of what semantic contribution each position offers to the interpretation of a sentence, what is the contribution of *in front* or *on top* as opposed to *of* (e.g. Hale & Keyser 2002; den Dikken 2010). Other proposals give an accurate treatment of the semantic contribution of adpositions, but do not spell out in detail which aspect of meaning corresponds to each distinguishable syntactic unit, as well as presenting these treatments via rather different sets of assumptions. While in some proposals *to* and *in front of* denote relations between eventualities (e.g. Parsons 1990), in other proposals
these adpositions denote model-theoretic objects such as “vectors” or “regions” (e.g. Zwarts & Winter 2000; Kracht 2002; respectively). No proposal attempts to explain why there is an apparently systematic distinction in meaning between adpositions such as to and in front of, or at, and thus to explain in a more systematic way why adpositions have distinct interpretations and semantic properties, such as the entailment patterns holding between adpositions and the sentences they are part of, or their distribution. Given this systematic lack of a compositional treatment of adpositions, no proposal appears to offer an account that can be considered as psychologically plausible, i.e. that it explains the syntax and semantics of Ps as the result of mental, linguistic processes.

Our third outstanding problem is that we still do not have a clear picture of the on-line process of interpretation of adpositions in native speakers of English (children and adult alike), and thus whether our theoretical proposals can actually account empirical data with a psychological, experimental nature. In particular, we still do not have a clear picture regarding the interpretation of certain “core” adpositions that express general spatial relations (i.e. in, to, at and from), and the logical relations holding between these adpositions: whether speakers can accept that a sentence such as (5) (i.e. the boy is sitting at the desk) is true in discourse as a consequence of (4) (i.e. the boy has gone to the desk) being true as well. We still do not have a clear picture of what is the interpretation of these core adpositions, and the sentences they occur in, by adult speakers. We also do not have a clear picture of what is the interpretation of these adpositions by English-speaking children, whether they significantly differ from adults, and why they acquire them in certain way.

The lack of this precise picture, and its corresponding problem, can be illustrated by the experimental evidence found in the literature, and the understudied topics regarding the interpretation of Ps. Several experimental studies have been carried out, their findings suggesting that adults interpret adpositions such as in front of or above as expressing restricted relations between locations. These studies do not offer any empirical evidence regarding certain “general” adpositions, such as in, at, to or from, and the logical relations holding among them, e.g.
entailment patterns (e.g. Richards & Garrod 2005). Other proposals suggest that children access adpositions in increasing order of complexity, e.g. from in to across (e.g. Slobin 2004); and that children interpret adpositions in an adult-like way, when they can access their interpretation (e.g. Stringer 2005). However, little is known about this group of adpositions which are poorly studied in adults’ experiments, as well as the logical relations holding between these adpositions, and their emergence in children’s language. Furthermore, little is known regarding their relevance with respect to theories of grammar and adpositions: although these experiments offer evidence regarding the interpretation of adpositions, the relation between this on-line evidence and off-line analysis of adpositions is far from clear.

Each part of the thesis aims to offer a solution to each of these outstanding problems, in the proposed order. Each solution forms a distinct part of the thesis, for a total of three parts. Each problem includes a very basic question, what is known about the problem at hand and what solutions have been offered so far. Consequently, each part contains a detailed review of the relevant literature, and suggests not only which solutions are at our disposal, but also why these solutions require further extensions in order to meet the empirical desiderata. Each part focuses on its respective problem, and aims to offer the following solutions to each outstanding problem, which I shall present part by part.

Part I includes chapter 2, and offers a solution to the first problem, the relation between adpositions and the non-linguistic content they are matched with. I first offer a solution to the what-problem regarding previous solutions on the first problem, and their pros and cons. I propose a solution to the how-problem and the what-problem that make up the first problem, and which consists in offering a formal treatment of visual and linguistic processes based on a minimal fragment of Discourse Representation Theory (Kamp, van Genabith & Reyle 2005). I will propose that visual processes can be represented via a “visual” fragment of Discourse Representation Theory, which I shall call Visual Representation Theory; and that the representations computed by this fragment can stand in a one-to-one relation with the representations computed by the “linguistic” fragment. Via these fragments, I offer a way to represent the visual and lin-
guistic processes that occur when we observe objects and their position in the world, how we see things in the world and say something about their position. I offer a solution to the first what-problem by suggesting that adpositions express how we see things in the world, by instantiating in Language how we relate the position of one object with respect to another, as it changes over time; and a solution to the second what-problem by suggesting that we can “say” exactly what we can “see”, via the use of adpositions of increasing level of precision, possibly with a one-to-one matching between what spatial relations we can see and what adpositions we can use in a sentence.

Part II includes two chapters, chapter 3 and chapter 4, and offers a solution to the second problem, how we can offer a compositional approach to the syntax and semantics of adpositions. This solution can be thought as an off-line model of the linguistic processes underlying adpositions. I start chapter 3 by proposing a solution to the what-problem regarding previous solutions on the syntactic component of the second problem, and their pros and cons. I then propose a solution to the what-problems, the how-problem and the why-problem that make up the second outstanding problem. I offer a solution to the first what-problem by suggesting that adpositions correspond to syntactic heads that can combine with other adpositions and adposition phrases, in a restricted but recursive manner. I offer a solution to the how-problem by suggesting that adpositions are combined together into larger units, and thus that in front of is the result of combining in, front and of together. These two solutions are offered in chapter 3, and form the syntactic part of the solution of the global how-problem.

In Chapter 4 I offer the semantic part of the solution to this global question. In this chapter, I first offer a solution to the what-problem regarding previous solutions on the semantic component of the second problem, and their pros and cons. I then offer a solution to the second what-problem by suggesting that the different elements constituting adpositions express different parts of their underlying interpretation, e.g. that in and front contribute the specific positions involved in the spatial relation, which is captured by of. I then offer an answer to the why-problem. I will suggest that there is a transparent mapping between syntactic and semantic representations,
an instance of the *Curry-Howard isomorphism*; and that, as a consequence, adpositions always denote *relations* between *situations*, intended as spatio-temporal particulars of various “form” and “nature”. The difference between *in front of* and *to* is that the second adposition expresses a relation involving an *event of change*, which is not present in the relation denoted by *in front of*; and that this difference is a reflection about how the domain of spatial relations is organized in “smaller”, distinct sub-domains, which are expressed by semantically distinct adpositions. Both chapters are based on an extension of our *Discourse Representation Theory* fragment that can correctly represent the fine-grained details of syntactic structure and semantic interpretation in a fully compositional way (informally, constituent by constituent), and which can be thought as offering a psychological (but off-line) model of adpositions.

Part III includes two chapters, chapter 5 and 6, and offers a solution to the third problem: *what* is the on-line interpretation of the understudied adpositions *in*, *to*, *at* and *from* by adults and children. In Chapter 5 I propose a preliminary solution to the *what*-problem regarding our knowledge of Language Processing and Language Acquisition phenomena: *what* we know about Language Processing, and *what* we know about Language Acquisition. I focus on importing this knowledge within the theoretical framework used in part I and II, before moving to chapter 6. In Chapter 6 I offer a solution to the three problems making up the global *what*-problem, by suggesting that adults and children interpret the core adpositions *in*, *at*, *to* and *from*, as well as the logical relations holding across these adpositions, according to the proposal laid out in part II. I also offer a solution to the *why*-problem by suggesting that children acquire adpositions in increasing order of complexity, being in a sense guided by the *logical relations* holding between these adpositions (i.e. from *in* to *at* passing through *to* and *from*). Both solutions are couched in a further extension of *Discourse Representation Theory* that can capture not only how speakers interpret sentences, but also how this process may successfully allow the acquisition and retention of these interpretations over time. Consequently, both solutions offer experimental evidence in support of the model of adpositions outlined in part II.

After these three parts, I will offer my conclusions. I shall do so by offering an overview of
the results obtained in the thesis, hence evaluating the overall empirical import of the proposed solutions. I shall suggest that the solutions adopted in this thesis can cover a broader set of data than the one covered by the various theories reviewed in the thesis, and it offers this coverage under a unified perspective of our extended Discourse Representation Theory fragment, when compared to the other reviewed theories. I shall suggest three key arguments in favor of the solutions proposed in this thesis.

First, I shall suggest that our treatment of visual processes and their relation with adpositions, and the sentences they are embedded in, allows a straightforward analysis of e.g. the relation between *in front of* and a visual scenario this adposition describes (e.g. a boy sitting in front of the desk) which is not within the reach of previous proposals. I shall suggest that via our proposal, we can give a better account of the relation between “what we see” and “what we say”, which can also cover previously untreated data (e.g. the analysis of visual scenarios matching with *to* or *from*, *inter alia*).

Second, I shall suggest that our treatment of the syntax and semantics of adpositions can capture the fine-grained details and “reason” of their linguistic properties in a straightforward manner, which is not within the reach of previous proposals. I shall suggest that via our proposal, we can give a better account of the various types of (logical) relations holding between e.g. *ahead of* and *in front of*, or *to* and *from*, or *to* and *at*, and explain why *to* denotes a change of position, while *at* denotes a lack of this change.

Third, I shall suggest that our treatment of adpositions can be seen as psychologically plausible proposal, since it is supported by experimental data involving both adults and children, which is not within the reach of previous proposals. I shall suggest that via our proposal, we can verify whether adults and children interpret *to*, *in*, *at* and *from* and their logical relations according to our theory, and how children acquire these adpositions during their acquisition of English.

I will suggest that our unified theory of adpositions will successfully account how English speakers can understand sentences (1)-(6), match them against the scenarios they are possibly
observing, and how they can acquire this understanding as they learn English adpositions such as *in front of*, *to*, *at* and several others over developmental time. I will consequently suggest that, in doing so, our unified theory of adpositions offers an account of this category which has a much broader empirical coverage, since it can cover data across at least two modules of cognition (Vision and Language); it can give a fully compositional analysis of this category (Syntax and Semantics of adpositions); and can give straightforward account of experimental data (processing and acquisition). Hence, it will offer a theory of adpositions which will consistently extend and improve what is already known about this still poorly understood category.
Part I

Adpositions, Space and The Language-Vision Interface
Chapter 2

Adpositions, Space and the Language-Vision Interface: a Model-Theoretic Approach

2.1 Introduction: what We talk about, when We talk about Space

In this chapter, which coincides with the first part of the thesis, I shall address the first outstanding problem regarding adpositions, the problem of the Vision-Language interface: informally, what is the exact relation between “what we see” and “what we say”, (or: “how much space gets into Language?”: Bierwisch 1996:7). This problem can be formulated via the following (and slightly different) global research question:

Q-A: What is the relation between Vision and Language?

I shall suggest that the problem of the Vision-Language interface and its nature is not much a problem of “quantity” but of “quality”: in order to solve this problem, we need to address not

---

“how much” information belonging to spatial representations (“what we see”) find its way in Language (and vice-versa), but how this process comes by and how it is possible that visual information can be realized in Language in a rather flexible way. I shall argue that in order to understand how sentences such as:

(7) Mario sits in front of the chimney

(8) Mario has gone to the rugby match

Can convey non-linguistic spatial information, we need to first solve the problem of how the relation between “what we see” and “what we say” comes about, in the first place, and then apply the solution of this problem to the specific problem of spatial information, visual and linguistic alike.

This problem can be solved by a divide et impera research strategy. I shall first split the problem in three smaller problems (the divide part), and solve each of them, integrating these solutions in a “global” solution (the impera part). The three problems that constitute our central problem are the following.

First, we have a foundational problem, since previous proposals in the literature make different assumptions on the nature of “what we see” and “what we say”. Some assume that Language expresses only shapes of objects (as nouns) and geometrical configurations (as adpositions) (e.g. Landau & Jackendoff 1993); others that we directly express perceptual information “as we see it”, without an intermediate level of processing (i.e. Language, e.g. Coventry & Garrod 2004). Hence, we don’t have a clear (theoretical) picture regarding spatial Vision and spatial Language, and to what extent they are distinct modules of Cognition, let alone a strong, clear theory of their interface.

Second, we have a descriptive and logical problem, since previous proposals only cover inherently “static” aspects of Space, but not “dynamic” aspects. Informally, these theories can account where things are, but not where things are going. Hence, we do not know what visual information adpositions such as to and from stand for, nor whether this information should be
considered as “spatial” or not.

Third, we have a theoretical and a philosophical problem, since we must define a novel theory that is built upon the solutions to the first and second problem and can explain all the data. Then, we must assess the consequences of this theory with respect to a broader theory of Vision and Language as part of Cognition, and their unique aspects, or: what is information (and properties thereof) is found in Vision but not in Language, and vice-versa.

These three “smaller” problems can be reformulated as the following research questions, which have been already foreshadowed in the main Introduction, although in a slightly different form. The questions are the following:

**Q-1:** What do we know so far regarding spatial Vision, Language and their interface;

**Q-2:** What further bits of spatial knowledge we must include in our models of (spatial) Vision and Language, and which formal tools we must use to properly treat these bits;

**Q-3:** What is the nature of the Vision-Language interface, and which aspects are unique to Language;

Anticipating matters a bit, I shall propose the following answers. First, we know that previous literature tells us that (spatial) Vision and Language express internal models of objects and their possible spatial relations, and that nouns and adpositions respectively represent objects and possible relations in Language. Second, we must include any type of relations in our models of Vision and Language, insofar as they allow to establish a relation between entities, since the emergent notion of “Space” we will obtain from our discussion is quite an abstract one. Hence, we can use a model-theoretic approach, such as Discourse Representation Theory (henceforth: DRT, Kamp, van Genabith & Reyle 2005), to aptly represent these models. Third, the Vision-Language interface consists of the conscious processes by which we may match visual representations with linguistic ones and vice-versa, though some linguistic representations do not represent visual objects, rather “processes” by which we may reason about these visual
objects. Consequently, Vision and Language can be represented as distinct models sharing the same “logical structure”, which may be connected or “interfaced” via an opportune set of functions, representing top-down processes by which we may (consciously) evaluate whether what we see accurately describes what we say (or hear), but need not to do so.

This chapter is organized as follows. In section 2.2, I introduce some basic notions and review previous proposals, offering an answer to the first research question. In section 2.3, I review theories of “static” and “dynamic” object recognition (section 2.3.1, 2.3.3), and propose a model-theoretic approach to Vision (section 2.3.2, 2.3.4); I then focus on Language and offer a DRT treatment of spatial Language (section 2.3.5). In section 2.4, I integrate the two proposals in a novel theory of the Vision-Language interface (section 2.4.1) and offer empirical evidence in support of this theory (section 2.4.2). I then focus on some of the broader consequences of the theory, by sketching an analysis of what properties emerge as unique to language from my theory, thus suggesting a somewhat novel perspective to the nature of the narrow faculty of Language (FLN: Hauser, Chomsky & Fitch 2002; Fitch, Hauser & Chomsky 2005; section 2.4.3). I then offer my conclusions.

2.2 The Relation between Spatial Vision and Language

In this section I shall outline notions of spatial Vision and Language (section 2.2.1); and review previous approaches to their interface, consequently offering the first research answer (section 2.2.2).

2.2.1 Basic Notions of Space

Our daily life experiences occur in space and time, as we navigate our environment by analyzing spatial relations between objects. A basic assumption, in cognitive science, is that we do so by

---

2Here and throughout the thesis, I shall focus my attention (and use of labels) to “Space”, although it would be more accurate to think of our topic as being about spatio-temporal Vision and Language, i.e. how we process location and change of location of objects. I hope that the lack of precision will not confuse the reader.
processing (mostly) visual information about such objects and their relations as they may evolve over time, e.g. a toy which is on top of a table, and that we internally represent this information via a corresponding “mental model” (e.g. Kraik 1943; Johnson-Laird 1983, 1992; O’Keefe & Nadel 1978).

Another basic assumption is that, when we share this information with other fellow human beings (i.e. when we speak), we do so by defining a sub-model of space in which one object acts as the “centre” of the system, as in (9):

(9) The toy is on top of the table

With a sentence such as (9), we convey a state of affairs in which, informally, we take the table as the origin of the reference system, take one portion of the table (its top) and assert for the toy to be more or less located in this “area” (Talmy 1978, 2000). Our Cognition of Space is thus (mostly) based on the information processed and exchanged between our Vision\(^3\) module (“what we see”) and our Language module (“what we say”). It is also based on an emerging type of information, the structural relations that may be defined between these two modules, our ability to integrate together visual and linguistic units (“what we see and what we say”) into coherent representations, over time.

The exact nature of these types of information, however, is a matter of controversy. Some say that spatial Vision amounts to information about objects, their parts and shape, and the geometrical relations between these objects as when an object is on top of another (e.g. Landau & Jackendoff 1993; O’Keefe 2003). Another series of proposals offers evidence that other aspects, such as mechanical interactions (a table supporting a toy) and more abstract properties play a crucial role in how we mentally represent space (Coventry & Garrod 2004 and references therein).

We can thus observe that there is a certain tension between “narrower”, or purely geometrical, approaches and “broader” approaches to both Vision and Language; as a consequence,\(^3\)The notion of spatial Vision and Cognition are somewhat interchangeable for most authors. In this chapter I shall use the term “spatial Vision” and “spatial Language” to avoid this confusion.

17
there is also a certain tension between theories that consider spatial Vision “richer” than spatial Language (e.g. Landau & Jackendoff 1993), and theories that do not assume such difference, often by simply collapsing these two modules into “Cognition” (e.g. Coventry & Garrod 2004). We thus do not have a clear picture of what information is spatial Language, and what is spatial Vision.

The problem of the exact type of spatial information, however, takes an even more complex nature when we look at another way in which we process spatial information, which can be loosely labeled as “change”. Take a sentence such as (10):

(10) Mario is going to the rugby stadium

Intuitively, this sentence describes a state of affairs in which the locatum(s) changes position over a certain amount of time of which we are aware. Mario can start at some unspecified starting point, move for a while, and then stop once he’s at his planned destination (the rugby stadium). While there are theories of “dynamic” Vision, or how we keep track of objects changing position, as well as theories of “dynamic” Language and more specifically adpositions like to, no one has attempted to integrate these theories into a broader theory of spatial Vision and Language, let alone in a theory of the Vision-Language interface.

Another challenge comes from purely linguistic facts, and what kind of information is in a sense “unique” to a linguistic level of representation. Take a sentence such as (11):

(11) Every boy is going to a rugby field

In this case, we can have a certain number of boys involved in the corresponding state of affairs, and each of them is described as moving in direction of a rugby field. Yet, if there are several fields at which the children can arrive (Paul goes to Manly’s oval, Joe to Randwick field, etc.), the sentence may describe slightly different states of affairs, since they informally describe a “collection” of more specific relations, and what they have in common. As these facts show, we need to take a broader and more flexible perspective in order to address the issue of the Vision-Language interface than the one usually assumed in the literature, as well as assessing
in detail what elements of previous proposals we can maintain in our novel approach. Hence, I am also suggesting that the solution to this problem will offer us a quite different, but hopefully correct, answer to the “problem of Space”. Before offering this answer, however, I shall review the previous literature.

2.2.2 Previous Literature

Previous proposals on the Vision-Language interface can be divided into a “narrower”, geometric approach (or: “spatial Language expresses geometric relations”) and “broader”, “functional” approach (or: “spatial Language also expresses extra-geometrical relations”). One well-known and influential example of the geometric approach is Landau & Jackendoff 1993, while a well-known and influential functional approach is the Functional Geometric Framework (FGF, Coventry & Garrod 2004). I will offer a review of both, highlighting their features and shortcomings, with respect to the topic of this chapter, starting from Landau & Jackendoff’s proposal.

Landau & Jackendoff offer evidence that, at a visual level, objects and their relations are captured using “spatial representations”, chiefly expressed by adpositions. Size, orientation, curvature and other physical properties all conspire for an object to be recognized as more than a sum of its parts: a “whole” entity, or what the object is. Whole objects or “whats” can also be related one to another: if we have two objects, one will be conceived as a landmark object (or ground), while the other will be the “located” entity (or figure, Talmy 1978, 2000).

They also argue that the rich and variegated layers of visual-cognitive information are processed and then clustered together and associated with “conceptual labels” (or just “concepts”) and hierarchically organized within the Conceptual System (CS, Jackendoff 1983, 1990, 1991, 2002), the interface between non-linguistic modules and (linguistic) domain of semantics. This proposal and further extensions assumes that nouns are the main category representing objects in Language, whereas adpositions represent spatial representations/relations (e.g. van der Zee 2000). In line with other literature, Landau & Jackendoff propose that spatial expressions mostly involve “count” nouns, which can be seen as labels for objects with a given “shape” (e.g. “cylinder...
der” or the fictional “dax”; Carey 1992, 1994; Soja et al. 1992; Bloom 2000; Carey & Xu 2001; Carey 2001). Adpositions, on the other hand, are argued to express core geometrical properties such as overlap, distance and orientation (e.g. *in, in front of*: Landau et al. 1992).

Recent inter-disciplinary research has shown that the picture is somewhat more complex. A rich body of evidence has been accumulated suggesting that adpositions can also convey information which is not necessarily geometric in nature. Look at the examples:

(12) The book is on the table

(13) Mario is beside the table

(14) #The table is beside Mario

(15) Mario is taking the moka machine to the kitchen

If a book is “on” the table (as conveyed by (12)), the table will also act as a mechanical support to the book, i.e. it will prevent the book from falling. We can say that Mario is “beside” the table (as in (13)), but saying that the table is beside Mario will be pragmatically odd (as in (14)): figures tend to be animate entities (or at least conceived as such), whereas grounds tend to be inanimate entities.

These mechanical properties can also be seen as extra-linguistic or “spatial” properties associated to nouns. Informally, if a count noun as “book” is associated to an object with definite shape, and which can (and should) be involved in causal physic relations (e.g. support, or containment: Kim & Spelke 1992, 1999; Smith et al. 2003; Spelke & Hespos 2002; Spelke & van der Walle 1993; Spelke et al. 1994; Shutts & Spelke 2004; van der Walle & Spelke 1996).

Dynamic contexts offer similar evidence for the relevance of extra-geometric information to be relevant. For instance, in a scenario corresponding to (15), we will understand that the Moka machine\(^4\) brought to the kitchen by Mario will reach the kitchen because of Mario’s action (Ullman 1979, 1996; von Hofsten et al. 1998, 2000; Scholl 2001, 2007). We will also take for granted that the machine’s and handle beak will reach the kitchen as well, as parts of the

\(^4\)The traditional Italian machine for espresso coffee.
machine, unless some problem arises in the meanwhile. If Mario trips and the Moka machine falls mid-way to the kitchen, breaking in many pieces, we may not be able to recognize the Moka machine as such (Keil 1989; Smith et al. 1996; Landau et al. 1998). Spatial relations, and thus adpositions that express these relations, can implicitly capture the (potential) causal relations or affordances between different objects (e.g. Landau 1994, 2002; Munnich & Landau 2003).

For these reasons, Coventry & Garrod (2004) propose their FGF framework, according to which mechanical, geometrical and affordance-oriented properties form the mental model or schema (in the sense of Johnson-Laird 1983) of adpositions that we store in long-term memory. This model can be seen as the “complete” representation of an adposition’s meaning, which can then only partially correspond to its actual instantiation in an extra-linguistic context (see also Herskovits 1986).

According to this theory, speakers can then judge a sentence including a spatial adposition more or less appropriate or felicitous, depending on whether the adposition’s content is fully or partially instantiated in an extra-linguistic scenario (e.g. van der Zee & Slack 2003; Coventry & Garrod 2004, 2005; Carlson & van der Zee 2005; Coventry et al. 2009; Mix et al. 2010). Two examples are the following:

(16) The painting is on the wall

(17) The painting is in the wall

A sentence such as (16) can be considered more appropriate than (17) when used in an extra-linguistic context in which a certain painting is just hanging on the wall, but less appropriate when the painting is literally encased in the wall’s structure.

Other theories take a perspective which is either close to Landau & Jackendoff’s theory or to FGF. The Vector Grammar theory (O’Keefe 1996, 2003) treats English adpositions as conveying information about vector fields, the graded sequence of vectors representing the minimal “path” from ground to figure, and thus conveying purely geometric information. Another theory which is based on similar assumptions is the Attentional Vector Sum model (AVS, Regier &
Carlson 2001; Regier & Zheng 2003; Carlson et al. 2003, 2006; Regier et al. 2005). In this theory, “vectors” represent features of objects that can attract the speaker’s attention once he interprets a spatial sentence, and can thus include mechanical and functional aspects, as well as environmental (“reference frames”) information.

These theories thus predict that a sentence such as (18):

(18) The lamp is above the chair

is interpreted as a “set of instructions” that informs us about where to look at, in a visual scenario, but they differ with respect to these instructions being purely geometrical or not. Furthermore, AVS predicts that “above” will be considered more appropriate if used in an extra-linguistic context in which the lamp is above the chair also with respect to three possible systems of orientation or reference frames, e.g. if the lamp is above the chair with respect to some environmental landmark (e.g. the floor: absolute reference frame); with respect to the chair’s top side (intrinsic reference frame); and with respect to the speaker’s orientation (relative reference frame: e.g. Carlson-Radvansky & Irwin 1994; Carlson 1999).

Although the insights from these theories are quite enlightening and consistent with various approaches to Vision, their approach to Language is inherently a “blurry” one, as each of these theories says virtually nothing about the specific contribution of nouns and adpositions. Since these theories tend to reduce Language to general Cognition, this is not surprising. Aside from this problem, no theory really attempts to analyze “dynamic” spatial expressions, such as (15). The same holds for Landau & Jackendoff (1993) and FGF: examples such as (10) and adpositions such as to are still a mystery, with respect to the Vision-Language interface. Nevertheless, both sides of the debate offer at least two important points regarding the nature of spatial Vision and spatial Language.

These aspects form the answer I shall propose to the first research question:

A-1: Previous literature offers a clear mapping between Vision and Language (Landau & Jackendoff 1993) and evidence that spatial Vision and Language express...
Because of these previous proposals I shall assume, based on the literature on the topic, that spatial Vision and spatial Language are not just about geometrical relations, and thus suggest that both modules can express the same “amount” of spatial information, although in (quite) different formats. I shall also assume that there is one precise, although flexible, correspondence between units of Vision and units of Language. Visual objects find their way in Language as nouns, and spatial relations as adpositions, at least for English cases I shall discuss here. In the next section, I shall offer a justification to these assumptions and propose a richer theory of spatial Vision and Language.

2.3 The Nature of Spatial Vision and Language, and a Formal Analysis

In this section I shall offer an analysis of “static” and “dynamic” vision (2.3.1 and 2.3.3); and a Logic of Vision of these theories (section 2.3.2 and 2.3.4); I shall then analyze (specific aspects of) spatial Language via DRT (section 2.3.5).

2.3.1 Classical and Modern Varieties of Object Recognition

In highly schematic terms, we can say that spatial information is processed via visual perception, for most human beings. Light “bounces” off an object and the surviving wave-length is processed by the eyes. This information is then transmitted to the optic nerve, to be further processed in various parts of the brain, like the primary and secondary visual cortex. Once the perceptual inputs are processed, their corresponding (internal) representations become the basic chunks or atoms of information processed by higher cognitive functions, such as vision and

---

5A specific Language may lack a term for a certain visual object, so the correspondence between visual objects and nouns on the one hand, and spatial relations and adpositions on the other hand, may be subject to subtle cross-linguistic variation. Informally, if a Language has a term for a certain visual object, this term will be a noun, syntax-wise: the same holds for spatial relations. I thank an anonymous reviewer for bringing my attention to this point.
One of the earliest schools of research that attempted to investigate the nature and properties of these units of information was the Gestalt school of psychology. This school assumed that our unconscious processes of visual recognition allow us to individuate objects from the background via the following four principles: Invariance (“sameness” of an object); Emergence (parts making up a whole); Reification (interpolation of extra information); Multi-stability (multiple “good” images of an object).

These principles converge into underlying principle of Pragnanz or conciseness, our ability to form discrete visual units from different and perhaps contradictory “streams” of perceptual information. This process may not necessarily be “veridical” in nature: if we look at a car in motion and we do not notice its radio antenna, we may consider the two objects as one, as long as there is no visual cue that they are indeed distinct objects (e.g. the antenna breaks and flies away).

The Gestalt school’s thrust in the study of invariant properties lost momentum after the end of World War II, until Gibson (1966) re-introduced the study of Vision as a process of “information-processing” (and integration), which sparked the interest of various researchers including David Marr and his model of Vision, and which had an ever-lasting influence in Vision sciences and in some linguistic literature (e.g. van der Does & Lambalgen 2000).

Marr’s initial research started from the physiological bases of Vision (collected in Vaina 1990). His interest slowly shifted from the neurological and perceptual facts to cognitive aspects of visual processes, which culminated in Marr (1982). The core assumption in Marr’s theory is that Vision can be best understood and represented as a computational, algebraic model of information processing. It is a bottom-up and cognitively impenetrable process, since it is mostly realized without the intervention of conscious effort.

Marr proposed that any model, and thus any mental process or structure it represents, should

---

6J.J. Gibson would come to reject his stance in favor of an “ecological” or “externalist” approach, in Gibson (1979). More information about perceptual and historical aspects can be found in Farah (2004); Bruce et al. (1996); Scholl (2001, 2007); inter alia.
be defined at three levels of understanding: *computational* ("why" of a model), *algorithmic* (the "how" of a model) and *implementational* (the "what" of a model). Marr proposed that our Vision developed with a perhaps very abstract computational nature, that of "grouping" any type of visual information (geometric and not) into implementable units, which can be retrieved and stored in memory. Regardless of its purposes, Marr proposed that the computational system of human vision is assumed to have three intermediate levels of representation, or "sketches".

At the *Primal Sketch* level, boundaries ("zero crossings") and edges are computed, so that the continuous stream of perception is partitioned into discrete units of attention, or "receptive fields". Photo-receptive cells detect the change of light in the receptive fields, and split it in two parts: an "on-centre" and an "off-centre". In "on-centre" cells, the cell will fire when the centre is exposed to light, and will not fire when the surround is so exposed. In "off-centre" cells, the opposite happens. When both types of cells fire at the same time, they are able to represent an entity like an edge, its adjacent "empty" space and the boundary between the two partitions. The change of polarity between these two partitions is defined as a *zero-crossing*. A zero-crossing represents change in terms of opposed polarities: if an edge is marked as +1 in value, then the adjacent "empty" part will have value −1, and a border will be represented as a 0, or as a "boundary".

At the 2 1/2-*D sketch* level, these elements are integrated in the computation of surfaces and their distance from the observer. For instance, a triangle represents three lines whose edges coincide in a certain order, forming a connected contour, the triangle itself. Other information, such as depth or orientation, is computed via the integration of information about, respectively, the distance of the single surfaces from the observer (hence, an *egocentric* perspective), and integrated in a mean value, the normal "vector" from those surfaces. Missing information can here be interpolated: if part of the triangle’s side is occluded, we may just “infer” it from the orientation of the visible sides.

At the 3-*D model* level, the recognized parts and portions are integrated into one coherent whole. At this level, vision becomes an object-centred (or *allocentric*) process, which allows for
shape recognition to be viewpoint-invariant. The computation of a full 3-D model (object recognition) is crucially based on how the computation evolves from the 2 1/2-D sketch to its final level. If the various 2 1/2-D sketches can be integrated into a coherent unit, and this computed unit matches with a corresponding unit in memory, then the process of “object” recognition is successful (see also Marr & Nishihara 1978).

Marr’s model, given its algebraic nature, can be informally stated as a model in which basic information units or indexes can represent single parts of an object: $a$ and $b$ can stand for head and torso of a human figure, represented as the index $c$. If the unification or merging$^7$ of the two more “basic” information units $a$ and $b$ into a single unit is identified with a whole, then object recognition occurs. Simply put, from head and torso (and other parts) we obtain a human figure, a process that can be represented as $(a + b) = c$, $c$ standing for the human figure index.

This quite informal exposition should already made clear that two basic principles can be identified as being part of spatial Vision. One is the need to “chunk” the perceptual stream into discrete, computational units; and the other possibility to “merge” and identify these units in a rather abstract way, which allows us to establish part-of relations, according to Marr, among different information units.

After Marr’s seminal work, theories of object recognition roughly distributed between a more representational and a more derivational stance. While representational theories stress relations between different objects and parts (or, rather, representations thereof), derivational theories stress the processes by which these representations come into being. I will start from the representational stance, introducing Recognition By Components theory (henceforth: RBC, Biederman 1987; Hummel & Biederman 1992), probably the most influential theory for the representational stance.

RBC offers an approach which is substantially similar to Marr’s original proposal, although it is postulated that object recognition occurs via seven sketches of representation, rather than three. One important difference is that, after the first two sketches are computed, each (part

$^7$Here I use the term “merge” in a pre-theoretic way, but I will offer a more precise definition in section 2.3.3.
of an) object is conceptualized as a geon \textit{(generalized ion, Biederman 1987)}, a primitive shape or visual “ur-element”\textsuperscript{8}. The combination of various geons allows to define complex forms: for instance, an ice-cream can be idealized as a semi-sphere connected to a cone, consequently capturing complex relations between the parts they represent. Whenever an object is successfully recognized, it can be and stored in memory as a distinct entity (Hummel & Stankiewicz 1996, 1998; Stankiewicz & Hummel 1996).

An important aspect of RBC is that it addresses how different information units are combined together over the time of a computation, a phenomenon defined as \textit{dynamic binding}. Informally, if we recognize a sphere shape \( a \) and a cone shape \( b \) at a(n interval) time \( t \) in the computation, their integration as integrated units \( a + b \) will occur at a time \( t + 1 \). In this perspective, object recognition can be seen as a dynamic process of binding different units of information together, so that “new” objects emerge from this process: by dynamically binding edges and lines together in a coherent representation we have surfaces, and by dynamically binding surfaces together have three-dimensional objects, at an interval \( t + n \).

An alternative view to this representational approach may be exemplified by the derivational model \textit{H-MAX} (short for “Hierarchical MAXimization” of input) of Poggio and associates (Edelman & Poggio 1990; Riesenhuber & Poggio 1999a, 1999b, 2000, 2002; Serre, Wolf & Poggio 2005). In this model, objects can be any parts of which we receive visual input, via their luminosity, and of which we compute possible visual candidates (e.g. different possible representations of the same dog). No intermediate levels of representation are however assumed to exist, since the flow of information is constrained via a pair of simple principles, \textit{SUM} and \textit{MAX}, which are in turn defined over vectors as sequences of minimal parts and boundaries of an object.

An example is the following. Suppose that we look at our pet Fido, starting from his tail. At this initial step, our visual system first computes parts and boundaries, such as the tail’s

\textsuperscript{8}Geons are not exactly primitives \textit{per se}, but represent the (finite) set of combinations (36 in total) of five binary or multi-valued properties that combine together to define a shape. These five properties are: \textit{Curvedness} (if a component is curved or not); \textit{Symmetry}; \textit{Axis} (specifically, the number of axes); \textit{Size}; \textit{Edge type} (if the edges define an abrupt or smooth “change of direction”).
tip, which can be badly lighted or “stilted”, if we are observing it by an odd angle. From this “vector”, we access other possible memorized images of Fido’s tail and combine them with other visual features (vectors) we recognize about Fido. In case the image is somehow poor, we may compare it as a “noisier” version of Fido’s tail.

All these vectors are then summed together in the sum vector, the averaged sum of the vectors corresponding to the various visual inputs. If this sum exists, then a “standard” (or allocentric) view will be defined, which corresponds to the final step of the process of object recognition. In keeping track of these different views, “feature clusters”, edges of a surface or other easily observable points play a vital role.

In more formal terms, the $SUM$ takes two visual objects and unites them together into a new visual object: if $a$ and $b$ are Fido’s head and torso, then $a + b = c$ is Fido’s body. The $MAX$ operation minimally differs from the $SUM$ operation in two subtle ways. First, it may sum together two visual objects and obtain one of the two objects as the result, i.e. $a + b = b$. This is possible when one object “includes” the other, i.e. when one visual object contains all the features of another object. Hence, their union will be the “strongest” object. Second, it may average visual objects representing the same entity, i.e. it may sum objects which have common features. In formal terms, this can be consequently represented as $(a + b) + (b + c) = a + b + c$, a novel visual object (the “average” image) obtained out of previous objects. These processes are dynamic, so if two visual objects are $SUMmed$ ($MAXed$) at a time $t$, the result will hold at a time $t + 1$.

While these two theories show a substantial convergence in their treatment of object recognition, their assumptions about the nature of “objects” is quite different. Representational theories consider an “object” as the end result of a visual computation, while derivational theories consider an “object” as any unit that is manipulated by a computation. This difference may appear purely theoretic, but it has its own relevance once we take in consideration how this information is mapped onto linguistic units.

Consider, for instance, the following examples:
In (19), the spatial relation is defined over a book and a rather specific part of a blue table, the tip of its left edge, whereas such level of detail is left implicit in (20). Note that this relation also informs us that the book is supported by one part of the table (the tip of the left edge), which in turn may be seen as not so ideal for supporting books (tips are intuitively worse “supports” than centers).

For the time being, though, I shall leave aside adpositions and spatial relations, and concentrate on objects and nouns. In both sentences, any object or part thereof (“edge”, “tip”) finds its linguistic realization as a noun: if there is a difference between different layers of visual representation, this difference disappears at a linguistic level, since both visual objects are represented in Language as nouns. Consequently, a theory of object recognition that makes no difference between parts and whole objects, such as H-MAX, offers an easy counterpart to these simple linguistic facts, while other theories are less suitable for my goal of offering a theory of the Vision-Language interface. I shall base my formal proposal on Vision by offering a logical treatment of H-MAX, in the next section.

2.3.2 A Logic of Vision, Part I: Static Vision

The core aspects shared by the models of Static Vision (object recognition) we have seen in the previous section are the following. First, Vision involves the explicit, internal representation of perceptual stimuli in terms of discrete information units, or visual objects (of any size and shape, so to speak). Second, these units are combined together via one underlying principle, which we can temporarily label as “sum”. Third, the result of this process defines more complex objects, but also relations between these objects, which can be seen as instances of the part-of relation. These three aspects can be easily represented in one (preliminary) unified Logic of Vision, which I shall define as follows, and which I shall expand in more detail in section 2.3.4.
First, I shall assume that Vision includes a set of visual objects, the (countably infinite) set $V = \{a, b, c, ..., z\}$. Each of these objects represents a minimal information unit, an output which is activated (instantiated) when some perceptual input exceeds a threshold level. Hence, each information unit in a computation represents an instance of transduction, since it represents the (automatic) conversion from one type of (input) information to another type of (output) information (Pylyshyn 1984; Reiss 2007). I shall assume that each object can be represented as a singleton set, via “Quine’s innovation”: hence, $a$ is shorthand for $\{a\}$. Consequently, our operations will be defined over sets (cf. Schwarzschild 1996: appendix). I shall use the label “object” when it will make the presentation of the arguments more immediate.

Second, I shall assume that one syntactic operation can be defined over these units, the sum operation “$+$”, an operation that I will call merge. An example of merge is $a + b = c$, which reads: “$c$ is the merge of $a$ and $b$”. It is a binary operation, which is also associative, commutative, and idempotent. Associativity means that the following holds: $a + (b + c) = (a + b) + c$. In words, and using again Fido’s example, Fido’s head with Fido’s body (torso and legs) corresponds to the same object as Fido’s upper body and legs: Fido. Commutativity means that the following holds: $a + b = b + a$. In words, Fido’s head and body form Fido, much like Fido’s body and head. Idempotence means that the following holds: $b + b = b$. Fido’s head and Fido’s head give us Fido’s head, i.e. we can repeat information. Since our objects are singleton sets, this operation is basically equivalent to Set Union. The intuition behind the merge operation is that it takes two “old” distinct objects and creates a “new” object as a result, in a sense distinct from the basic sum of original parts. For instance, our Fido can be conceived as the new visual object that is obtained when the visual objects corresponding to Fido’s body and Fido’s head are merged together into an integrated representation, Fido as a “whole” entity.

Third, I shall assume that one semantic relation can be defined between objects, the part-of relation, represented as “$\leq$”. An example of the part-of relation is $a \leq b$, which reads: “$a$ is part of $b$”. Since I am using Quine’s innovation, the part-of relation is roughly equivalent to set
membership\(^9\). This relation is also binary, and it is reflexive, transitive and antisymmetric. It is reflexive, since the following holds: \(a \leq a\). It is transitive, because the following holds: if \(a \leq b\) and \(b \leq c\), then \(a \leq c\). It is antisymmetric, because the following holds: if \(a \leq b\) and \(b \leq a\), then \(a = b\). In words, each Fido’s part is part of itself (reflexivity); if Fido’s leg is part of Fido’s body and Fido’s body is part of Fido, then Fido’s leg part of Fido (transitivity); If Fido’s body parts are part of Fido, and Fido consists of Fido’s body parts, then these body parts are recognized as the same entity (antisymmetry). The intuition behind the part-of relation is that it establishes a relation between “old” objects and a “new” object, as a result of the merge operation. For instance, if Fido is the result of merging Fido’s legs and Fido’s body into a “new” visual object, then Fido’s legs will be part of Fido. If we recognize Fido, then we will also recognize Fido’s legs as well as other parts that make up Fido, as a consequence of the relation between parts and whole.

The resulting model of (object) Vision emerging from these basic definitions is the triple \(S = \langle V, +, \leq \rangle\), a simplified variant of a structure known as join Lattice, a type of full Boolean Algebra (e.g. Keenan & Faltz 1985:ch.1; Landman 1991:ch. 2; Grätzer 2008:ch.1-2). A join lattice can be seen as a set with at least one binary operation of “composition” and one relation defined over its elements, which also has the following property: if \(a \leq b\), then \(a \cap b = a\) and \(a \cup b = b\). In words, if \(a\) is part of \(b\), then the intersection of \(a\) and \(b\) is \(a\), while the union of \(a\) and \(b\) is \(b\). Informally, if the merge of two objects creates a novel object, the part-of relation establishes that this novel object includes the old objects as its (proper) parts. Because of these properties, this type of Boolean algebra is a complete structure, i.e. it will have one maximal object including every other object (i.e. \(V\)) and one minimal object which is included in every other object, which we will call “0”, and which represents any instance in which we “fail” to recognize objects\(^{10}\).

---

\(^9\)The subtle but very important differences between the notion of “set membership” and the part-of relation are not important for our current discussion. Chapter 4, section 4.2 will also briefly touch this topic in some more detail. The interested reader is deferred to e.g. Link (1983); Landman (1991:ch.1); Schwarzschild (1996:ch.1); for a more thorough discussion.

\(^{10}\)Note that the operation MAX can be now reconstructed as a special instance of SUM (i.e. our merge). I shall
Since we mostly operate on individuals-singleton sets via \textit{merge} and the \textit{part-of} relation, the Logic of Vision I define here is substantially a \textit{first order logic}. Since this logic allows us to define an algebraic model of objects and their interpretation and relations, it is a \textit{model-theoretic approach} to Vision. Anticipating matters a bit, the discussion of the Vision-Language interface will coincide with the discussion on how this model and the model defined by Language are related, via the more basic discussion on how visual information bits are related to linguistic bits, and \textit{vice-versa}.

These logical/algebraic properties represent the following facts: the visual “integration” of Fido’s leg and Fido gives us Fido, i.e. Fido’s leg is “recognized” as part of Fido’s whole image (union). If from Fido’s whole image we focus on Fido’s leg, then the other parts will be ignored (intersection). This latter interpretation of “attention as intersection” can be found in RBC and Ullman (1996), and is based on one simple intuition: if \textit{merge} represents object recognition, as the process that brings together different visual inputs into more complex objects, then its complementary operation represents the process by which we focus on a single visual object out of an array, i.e. attention. Furthermore, the sum of objects forms the full “library” of our model of Vision (the maximal object \(V\)), and there can be cases in which we cannot recognize any object whatsoever, for instance when we fail to focus our attention on something (the empty object).

This brief and semi-formal excursus suffices for our discussion of object recognition. The important aspect is that we can now define a tight relation between the syntax and semantics of our Logic of Vision: for each instance of the \textit{merge} operation, the result will define another visual object and a \textit{part-of} relation between this object and its constituent parts. Informally, we are able to recognize the legs of Fido as Fido’s, because we first integrate Fido’s legs with other Fido’s body parts into Fido’s whole image, and then retrieve this relation between legs and Fido.

The merging of visual objects does not occur in a temporal void, as we have seen, but is

---

leave to the reader the simple proof of this fact. Also, note that given our definition of the sum operation, visual objects can be either \textit{atomic}, i.e. they only include themselves as proper parts (e.g. \(\{a\}\)), or \textit{non-atomic}, i.e. they may have other objects as their proper parts (\textit{plural/sum objects}): e.g. \(\{a,b\}\), including \(\{a\}\) as its part. See e.g. Schwarzschild (1996); Link 1983, 1998). The import of this subtle distinction will be explored in chapter 4.
dynamically realized over discrete intervals of time. In RBC, this is represented via *dynamic binding*, i.e. the explicit representation of derivations as they occur over time. Before defining dynamic binding, I shall define the structure of the Index Set that represents intervals of time. This structure is the duple \( I = \langle t, + \rangle \), a set of intervals of time with an operation of *addition*. Although I represent this operation via “+”, it is a slightly different operation than *merge*, since it is only associative but not *commutative* nor *idempotent*. Intuitively, from a starting interval \( t \) we can “move forward” to other intervals, e.g. \( t + 1, t + 2 \) and so on, via the simple iteration of this “asymmetric” *merge*.

The corresponding type of structure is a simpler algebra, a *strict order*, i.e. a structure in which each element is a distinct object. Intuitively, this structure represents the directed flow of the logical processes underpinning visual computations, the “arrow of time” that tells us how visual objects are integrated together during distinct intervals of time, but which cannot “remember” any relations between the objects manipulated in these operations.

The explicit integration of this structure with Vision is the duple \( S_d = \langle I, S \rangle \), the “dynamic” Logic of Vision and object recognition. Its dynamic nature stems from the ability to represent visual computations as they occur over derivational times, in a simple format similar to standard proof-theoretic (i.e. syntactic) component of various Logical systems (e.g. Landman 1991 for discussion). One example is the following:

\[
\begin{align*}
(21) \quad t. \quad a & \quad \text{(visual object instantiation, e.g. Fido’s head)} \\
& t + 1. \quad b & \quad \text{(visual object instantiation, e.g. Fido’s body)} \\
& t + 2. \quad a + b & \quad \text{(merge introduction)} \\
& t + 3. \quad (a + b) = c & \quad \text{(Fido as “sum” of Fido’s parts)} \\
& t + 4. \quad a \leq c & \quad \text{(part-of introduction, Fido’s head as part of Fido)}
\end{align*}
\]

This derivation roughly captures how the process of recognizing Fido may occur a dynamic (and bottom-up) way, modeling the actual processes described in the reviewed theories. The various objects are first recognized (“instantiated” in the derivational space) one by one and then merged.
via the introduction of this operation. Once this process is over, we can also access the relation between Fido’s head and Fido’s whole image, since we can establish that one is part of another.

This simple example of a derivation in our Logic of Vision may not capture all the aspects involved in visual computations and, to an extent, it is quite idealized: for instance, an individual may consciously assume, and thus exert a top-down choice, that he is seeing Fido’s body, since he can partially recognize it as a visual entity connected to Fido’s head. In this and other examples, I shall leave these matters aside, as they are not crucial, for our discussion. This example, however, introduces one important advantage of my theory over the theories I reviewed so far: it makes fully explicit the structural relations between the various components of the object recognition process, including its unfolding over time.

This Logic of Vision is still a preliminary proposal, since for one thing, it does not allow us to make a distinction between objects (individual constants such as e.g. a) and the properties they instantiate (e.g. constant functions such as “dog”). It also cannot represent spatial representations, and thus the visual content of adpositions, but this is a void that will be filled in section 2.3.4, along with a theory of visual properties. However, it already allows us to give a compact definition on how we see things in the world, at least with respect to static objects.

Now we can explicitly represent (visual) objects in a very preliminary logical Space, and we can also define how these objects are mapped onto their corresponding linguistic labels, nouns. I shall assume, differently from previous proposals such as Landau & Jackendoff (1993), that this mapping is an isomorphism, a one-to-one correspondence between objects of different types (i.e. visual objects to noun labels). The reasons for this assumption are the following. The discussion of examples (19) and (20), and the intuition that each visual object may (potentially) have a corresponding “noun” label, has one important theoretical consequence. If we define a function mapping visual objects to nouns, then this function will be injective, it will find at least a label n’ for each visual object v: a noun such as table, for instance, stands for the corresponding visual object, a table. Furthermore, it is possible that several visual objects can correspond to one linguistic label: a noun such as table also stands for the sum of legs, surface, edges and other
visual objects making up a table. Hence, this mapping function will be surjective as well.

A function which is injective and surjective is a bijective function, hence a function that defines an isomorphism. More formally, for each visual object \( v \), for each noun label \( n' \), there will be a function \( f \) such that : \( f(v) = n' \). Since this function is surjective, the following holds: given \( a + b + c = v \) and \( f(v) = n' \), then \( f(a + b + c) = n' \). In words, we have the “lexical” identity \( \text{edge} + \text{legs} = \text{table} \), which can be also indirectly represented as \( f(a + b) = f(a) + f(b) \), with \( f(a) = \text{edge} \), \( f(b) = \text{legs} \) and \( f(a + b) = \text{table} \). Furthermore, this isomorphism preserves relations, so if one object is part of another, one corresponding noun will be lexically related to another. We have \( f(a) \leq f(b) \), which in words says that \( \text{edge} \) is (lexically) related to \( \text{table} \).

This isomorphism can be interpreted as follows. Our Logic of Vision is a partial, yet very fine-grained model of object recognition, with a simple and yet rich hierarchical structure, defined by the part-of relations that can be established between the objects in this domain. The function \( f \) tells us that such structure can also be connected with other structures, provided that they are governed by the same (logical) principles. Informally, it allows us to potentially define a correspondence between nouns in Language and visual objects in Vision, on a one-to-one basis. Although a language may lack a specific lexical item for each visual object, it is at least possible to define such a tight correspondence between nouns on the one hand, and visual objects on the other hand.

This function can be thus thought as a representing a top-down, conscious (and optional) process, which occurs when we consciously match visual information against linguistic information. It allows defining a correspondence between simple and complex visual objects and the nouns that represent these objects at a linguistic level, e.g. to establish that a noun such as \( \text{table} \) can indeed refer to a visual object we may observe, and which is made of four legs, a surface and other relevant parts. With this notion in place, then, we have introduced enough “machinery” to handle the static side of Vision and its Logic; we need to focus on the neglected

\[\text{Reference} \text{ is not equivalent to the one commonly employed in the literature. A standard assumption is that reference is the relation between a term and the “real world” object that corresponds to a given term. Here and for the rest of the thesis, I shall assume that linguistic terms can refer to extra-linguistic but internal information, such as visual objects.}\]
dynamic side, and propose a full Logic of Vision, by which we can also analyze spatial rep-resentations/relations. I shall do so in the next two sections.

2.3.3 Theories of Dynamic Vision

In the discussion in the two previous sections I have introduced a view of spatial Vision in which the ability to explicitly represent objects and their relations plays a crucial part in “static” scenarios, i.e. cases in which we “find” objects which are not changing position over time. One aspect missing from this discussion is how we establish relations between objects, especially when they change their position over time: how dynamic spatial Vision comes about.

A preliminary step to answer these questions is to define how we can keep track of objects over time. For this purpose, I shall review a theory about dynamic object tracking: Multiple Object Tracking (MOT), introduced in Pylyshyn (1989) and developed in a number of successive works (e.g. Pylyshyn 1994, 2001, 2003, 2004, 2006; Pylyshyn & Annan 2006; see Kanheman et al. 1992 for the roughly equivalent Object File Theory).

MOT offers a theory about object recognition in dynamic scenarios by analyzing how we are able to individuate and form mental representations of objects in the way they instantiate some properties (e.g. being yellow in color), and by how we maintain or change these representations over time and the unfolding of events. MOT is probably best presented via a preliminary example. Imagine that we look at the panorama: we detect trees, clouds, buildings, and so on. If we focus our attention on a flying black swan, we can do so because we are first able to detect a mysterious object (call it “x”), which instantiates the properties “swan”, “black” and “flying”, among others.

With some imagination, we can assume that “swan” is the primitive and most basic property which allows us to recognize the mysterious entity as such, the equivalent of an imaginary finger stretching from our eyes to the object itself. Such a finger allows us to define the mysterious object in terms of what property it instantiates, and it is thus defined as Finger of Instantiation, or FINST. The very act of this process is usually defined as FINSTing in the literature and, since
it can be defined for any entity that can be so individuated, it makes no distinction between types of objects: everything which can be FINSTed is an object, simply enough.

It is useful to illustrate MOT’s notation for the basic process of FINSTing, as well as the addition of further features. I will follow Pylyshyn’s (1989) notation, for ease of exposition. Aside from the basic process of FINSTing, we can imagine a situation in which the black swan is flying above a cloud. The process of FINSTization is illustrated in (22), while (23) illustrates the more complex “above” case:

\[
\begin{align*}
\text{(22)} & \quad a.\text{FINST}[x], [\text{swan}] = (x : \text{swan}) \\
& \quad b.\text{FINST}[x : \text{swan}, [x : \text{black}] = (x : \text{swan}, x : \text{black}) \\
\text{(23)} & \quad \text{ABOVE}(x : \text{swan}, x : \text{black}, x : \text{flying}, y : \text{cloud})
\end{align*}
\]

In (22-a), a basic property like “swan” is mapped onto a visual object, acting as the FINST that tracks the visual object. In (22-b), the combination of two properties acting as FINSTs creates a new, more complex FINST, which identifies the visual object \(x\) as a black swan. In the case of (23), we can observe that such “internal fingers” can also define relations between simpler “fingers”, hence expressing a relation between different instances of the same underlying process.

This relation is, in turn, a description or property of an event of motion, in which the swan is the moving figure, while the cloud is the contingent ground. Further information can be stacked up via dynamic binding: informally, each individuating property for \(x\) can be in a temporally incremental fashion (e.g. “black” at time \(t\), “flying” at time \(t + 1\)), which in turn is realized via the iterated application of the FINST operation.

One problem emerging from the presentation of MOT is that this theory cannot easily be used to analyze how the temporal relations between properties can be defined and represented in their own right. While “black” may be instantiated after “swan”, we cannot explicitly represent that the corresponding “fingers” can be taken as entities in their own right, the events during which these properties are instantiated and combined together, or defined in terms of their order.
of occurrence.

One theory that aims to fill this conceptual void is Event Segmentation Theory (henceforth: EST), a theory of events and psychological events first outlined in Zacks & Tversky (2001) and Zacks *et al.* (2001). In this theory, an original philosophical intuition by Quine (1960) and further developed in Davidson (1967) acts as the basic insight and ontological assumption: that our understanding of the world includes not only objects, but also the events in which these objects are involved.

At one level of comprehension, our mind represents objects as “things that are in the world”, such as birds and apples and cups. Once we add a temporal level of comprehension, and thus we observe how things change or preserve their own identity through time, we also keep track of what causes may change the properties of an object. The focus of EST is on events, which are treated as “pegs”, basic computational units or “slots” on which we stack up information, and which stand for relations and order among relations in which objects are involved, as they unfold over time (Speer 2003; Zacks 2004; Zacks *et al.* 2007; Zacks & Swallow 2007; Reynolds *et al.* 2007; Tversky *et al.* 2008).

EST assumes that, at a minimum, we can observe objects at two levels. One basic level is that of their structure and how it is realized in space (*partonomy*) and one of an object and its relation to an abstract class (e.g. a chair as part of furniture: *taxonomy*). Once we take in consideration a temporal dimension, in which objects can have different properties in different intervals of time, we will have “dynamic” objects or events. Events are conceived as discrete information units derived (i.e. *transduced*) from perceptual information, i.e. the “indexes” attributed to the combination of a (rather abstract) visual property and the object that instantiates it.

For instance, if someone throws a cup on the floor, then the cup will likely be shattered into pieces because of this action. The temporary relation between an individual and the cup will bring about a new state of affairs in which the cup will be a new property of some sort, that of being shattered. At the same time, we represent this change via the temporal and causal relation between the two state of affairs, one involving an event of me shattering the cup, and another
in which the cup will be shattered, which is separated by a boundary event, an interval of time
in which neither the cup is shattered nor it is still intact, and in which we will need to “update”
our model of events. Events can also be combined together: if someone is stacking pillows,
each single pillow-stacking event can be combined into a “bigger” pillow-stacking event, and
possibly “merged” with other events, forming a more complex event e.g. “pillow-ordering”.

Such complex sequences of events can be seen as event models or schemata, structures
of events and their causal/temporal connections, as they are represented in short-term memory
(models) or stored in long-term memory (schemata in “semantic memory”: see e.g. Tulving
1972, 2000a, 2000b, 2002 and references therein for an introduction). Events can be dynamically
bound together: the “throwing” event occurs at a time $t + 1$, a boundary event is formed at a time
$t + 2$ and the “shattering” event occurs at a time $t + 3$, then there will be a causal, as well as
temporal relation between these events.

Both MOT and EST are theories that offer a detailed picture of how dynamic Vision can
occur, defining in detail the mechanisms by which we track objects in motion, and the com-
plex spatial representations that arise from this process, or events. One alternative view to these
approaches that offers some further important insights on spatial representations is the Hipp-
started as a study of rats’ navigational system, the way they represent objects and their places
in the environment, and how this information is memorized and accessed or updated at later
stages. According to this theory, humans (and rats) build up a complex spatial representations
of the environment via two parallel systems: the place and the misplace system. The place sys-
tem records information about objects’ position in the environment, and “checks” whether this
information is correct when visual information is processed. If an object has changed position,
then the misplace system records the change of position and updates the object’s new position,
accordingly.

This model has been further extended over the years. O’Keefe (1983, 1990, 1991; Burgess
& O’Keefe 1996, 2003) show that information about objects and their relations is processed, in
“real time”, by the navigational system. This system computes the location of a figure in terms of polar angle $\phi$ and distance $d$ from the ground as the relation $\theta(\phi, d)$, computed via $\theta$—rythem signals, which mostly originate in the Hippocampus.

The result of these computations can be modeled as a vector, a sequence of cells (Boundary Vector Cells) that fire when an observer visually tracks relevant entities in an environment, and can also allow to compute geometrical properties of objects. Hence, the place and misplace systems build up complex spatial representations over time, or Cognitive Maps (O’Keefe & Burgess 1999; Burgess et al. 2002; Burgess 2006a, 2006b; see Arsenijević 2008 for a linguistic proposal).

These theories give us a basic insight on the nature of dynamic spatial Vision. When we keep track of objects in motion, we do via the properties that objects may have over time, whether they are geometrical, functional or “functional”, insofar as they allow us to track objects in Space. At the same time, we also keep track of the relations between these properties and their order of causal/temporal occurrence: spatial representations have an inherent temporal dimension, which represents the structural relations between the events making up these representations.

Adpositions, as the chief part of speech expressing these relations, must also have such an abstract nature. Look at the examples:

(24) Mario has fallen onto the floor
(25) Mario has gone into the room
(26) Mario is sitting near the patio

A scenario which is more ore less depicted by (24) and (25) is one in which Mario is respectively on the floor and in the room as a consequence of a particular event, one of falling and one of going. A scenario depicted by (26) is one in which Mario is involved in an event of sitting, which occurs at some distance from the patio. He may be involved in other events, although these events are in a sense “backgrounded”, in the sentence.

In all three cases, the spatial relation holding between Mario and different grounds holds at
some moment of time because of some previous event, and involves more than just geometrical 
information. If we conceive Mario and the floor as inherently visual objects, then the adposition 
onto will capture not only that these two objects are currently related one another via a “support” 
relation, but also that such relation has come into being because of a previous falling event. Since 
adpositions seem to express the “logical” structure behind the events described by a sentence, 
the kind of spatial representations they capture are representations in logical Space, and define 
possible relations between objects and how they are represented in this logical Space. I shall 
offer the precise logical details of this enriched logical Space in the next section, in the proposal 
I shall call Visual Representation Theory.

2.3.4 A Logic of Vision, Part II: a Model of Visual Logical Space

In the previous section, we have been able to define a richer notion of visual object (i.e. things 
and their spatio-temporal properties), as well as sketching the nature of the relations holding 
between these objects. I shall integrate these results in our Logic of Vision as follows.

I shall assume that the set $V$ of visual objects is now made of “structured” entities, the 
combination of events $e$, objects $o$, and properties $pr$. The complex visual object that is made of 
these elements is the triple $v = \langle e, o, pr \rangle$, a basic entity which I shall call Visual Representation 
Structure (henceforth: VRS). Importantly, the set of events $E$ (with $e \leq E$) is disjointed from 
that of objects $O$ (with $o \leq O$) and the union of the two sets forms the whole set of (basic) 
visual objects, i.e. $E \cap O = \emptyset$ and $E \cup O = V$. Properties $pr$ form a set of properties by which 
these visual objects can be individuated, i.e. we have $pr \leq PR$. In words, Visual Representation 
Structures (VRSs) are made of basic objects (e.g. “x”), the properties by which we individuate 
them (e.g. “swan”), and the events in which these properties are instantiated, i.e. their position 
in logical space with respect to other events. The following will hold: $v \leq V$, i.e. each VRS is 
part of the set of VRSs. I shall represent a VRS as $v : pr(o)$, which reads: an event $e$ instantiates 
a property $pr$ of an object $o$. This format follows the format of DRT, which I shall introduce in 
full in section 2.3.5.
As we discussed, VRSs can be combined together via *merge*. The sum of two VRSs can be seen as a complex, novel event in which different properties of the same object can be combined together into a more complex, novel property. If \( e : \text{grab}(x) \) and \( h : \text{order}(x) \) are respectively an event of (pillow) grabbing and (pillow) ordering, then the complex event of (pillow) clean up can be formally defined as: \( (e : \text{grab}(x) + h : \text{order}(x)) = i : \text{clean-up}(x)). \)

The structural properties of *merge* (associativity, commutativity, idempotence), are defined over VRSs as well, although the apparent “temporal” nature of VRSs as representing “objects in motion” requires some discussion. I shall focus on events, to make the discussion simple. An event of (pillow) clean up can be organised in different ways (associativity); while we usually first grab pillows and then order them, when we clean up, an event of (pillow) clean up consists of both events, regardless of their linear order (commutativity); several events of pillow-grabbing are still a (complex) event of pillow-grabbing (idempotence). Although VRSs are more complex objects, their combinatorics can be nevertheless defined via one basic operation, that of *merge*, which represents how complex VRSs are created from the union of basic VRSs.

The *part-of* relation is also defined over VRSs and events, and allows defining how events are structured. Reflexivity and transitivity allow to establish order/overlap among complex sequences of VRSs, straightforwardly enough. Antisymmetry allows establishing whether two VRSs (or parts thereof) are really the same, and thus to establish the identity between a complex VRS and the sum of its constituting VRSs. It also allows us to reconstruct their consequential/temporal relation as well: if \( e : \text{grab}(x) \leq i : \text{clean-up}(x)), \) then \( e : \text{grab}(x) \cup i : \text{clean-up}(x) = i : \text{clean-up}(x) \) and \( e : \text{grab}(x) \cap i : \text{clean-up}(x) = e : \text{grab}(x). \) Since an event of pillow-grabbing is a proper part of (pillow) cleaning up, then it must precede the realization of a cleaning up event. The structural relations between events thus represent their causal/temporal relations: “new” events come into being as the result of “old” events being combined together, in an incremental fashion. If we don’t grab and order pillows, we won’t have an event of pillow-cleaning up: the existence of this event is a consequence of the combination of previous events. The tight relation between the syntax and semantics of our Logic of Vision thus allows us to
capture one aspect of “dynamic” space by simply looking at how events are computed, without introducing further principles of analysis.

Our updated logic of vision can be thus represented as $S = \langle V, +, \leq \rangle$, with $V$ being a shorthand for $V = \langle E, O, PR \rangle$. This new Logic of Vision is a fully dynamic Logic of Vision when combined with an index set $I$, i.e. when we have $S_d = \langle I, V \rangle$, with $I$ being the index structure $I = \langle t, + \rangle$, since it allows us to explicitly represent how we integrate VRSs together. One example is the following:

\begin{align*}
(27) & \quad t. & e: \text{grab}(x) & \quad \text{(VRS instantiation)} \\
& \quad t + 1. & h: \text{order}(x) & \quad \text{(VRS instantiation)} \\
& \quad t + 2. & e: \text{grab}(x) + h: \text{order}(x) & \quad \text{(Merge introduction)} \\
& \quad t + 3. & e: \text{grab}(x) + h: \text{order}(x) = i: \text{clean} - up(x) & \quad \text{(Sum of events)} \\
& \quad t + 4. & e: \text{grab}(x) \leq i: \text{clean} - up(x) & \quad \text{(part-of relation introduction)}
\end{align*}

In words, the merging of two VRSs yields a more complex VRS as the result, and allows establishing structural relations between VRSs. As we can see, the use of dynamic binding also allows us to bring out one aspect of the temporal nature of events: if we grab a pillow at a time $t$ and then put it in order at a time $t + 1$, then the resulting pillow-cleaning up event will be realized as a later time $t + 3$, in a progressive way.

At this point, we have a quite rich and thorough Logic of Vision which allows us to model spatial representations/relations in a rather elegant and simple way, and which turns out to be somewhat similar to other logical calculi of events/situations proposed in e.g. AI literature (e.g. Event Calculus, see Hamm & van Lambalgen 2005 and references therein; see also van der Does & van Lambalgen 2000), and non-linguistic applications of situation semantics (e.g. Barwise & Seligman 1997). One example of the elegance behind our logic is the notion of “location”. VRSs explicitly represent the spatio-temporal “location” of some event and its participants by representing the properties that individuate these entities. Geometric or mechanical properties are not any different from “grabbing” properties, with respect to how this process occurs over
time: we can thus represent e.g. the notion of inclusion as the VRS $e : in(x)$, that of support as $e : on(x)$ and so on.

We can then represent the notion of “motion”, or more appropriately the notion of change, as an ordering (part-of) relation between VRSs and the events they represent. So, if Mario goes in direction of the room and then stops once he’s in the room, he will be in the room, as a consequence of this event of motion. This can be represented as $e : go(r) < i : in(r)$, i.e. an event of going into the room as expressing the relation holding between one event and its consequence. In a scenario in which Mario is sitting near the patio, instead, other events may be going on at the same time, but at least these properties allow us to individuate Mario. We can represent this as $n : near(p) \leq z : gen(p)$, an event of (sitting) near the patio as part of a more generic event.

These relations between VRSs and the events they represent may find their way into language chiefly as adpositions via the function $f$, the isomorphism between Vision and Language. I shall re-define this function as follows. If $f$ takes a pair of visual object and property as an input, it will return a noun as an output: we have $f((o, pr)) = n'$. If $f$ takes a pair of event and property as an input, it will return a verb as an output: we have $f((e, pr)) = v'$. If it takes a full VRS as an input, it will return an adposition as a result: we have $f((e, o, pr)) = p'$.

The intuition is that “partial” VRSs find their way in Language as (common) nouns, labels for objects; as verbs, labels for “actions”; both individuate some entities, but do not express relations between these entities. Adpositions, instead, express the structural relations between VRSs, ultimately complex VRSs. Intuitively, nouns (and verbs) “find” objects in logical Space; adpositions “find” the relations between these objects, which in turn represent a very abstract notion of space. Landau & Jackendoff’s syntactic proposals are still maintained, to an extent.\(^{12}\)

The function $f$, as an isomorphism, preserves structure on VRSs as well: an adposition such as *between*, for instance, is usually analyzed as the “sum” of two simpler adpositions, such as “to the left or to the right of” some ground (e.g. Zwarts & Winter 2000). This can be represented as $f(\text{r–of} + \text{l–of}) = f(\text{r–of}) + f(\text{l–of})$, i.e. the adposition representing the “between”

\(^{12}\)In this chapter (and the rest of the thesis), I shall not propose an explanation on why the function $f$ seems to operate such distinctions in the labeling process, and leave such a complex topic for future research.
relation is lexically equivalent with the adpositions representing the relations “to the left of” and “to the right of”. Generalizing a bit, from basic spatial representations we can build more complex spatial relations; the complex structure defined by this process, the model of logical Space defined by our Logic of Vision, may be represented in Language up to isomorphism, via process of progressive refinement and specificity of relations (cf. also Levinson & Meira 2003). Hence, the mapping function $f$ may assign different labels to its outputs, depending on the level of fine-grainedness and with some consistent cross-linguistic variation, but in a quite fine-grained and structurally regular way (cf. again Talmy 2000; see also section 2.4.2). Again, via this function we represent the possibility that we can match for each VRS a corresponding linguistic unit, and that the structural or “spatial” relations between VRSs can find their way into language, chiefly as adpositions, at least in a language such as English.

Before moving to Language, however, I shall make one observation regarding the nature of this process. According to the HCM proposal, when we mentally represent visual objects, these objects can be seen as output to some previous visual, perceptual input, which is then transduced as a visual object. This process occurs over discrete intervals of time, which in turn may be seen as minimal cycles of the θ-rythm, and which may actually occur independently of the presence of external stimuli. In the absence of external stimuli, our brain still partitions the perceptual stream into minimal, discrete units. Very informally, our Vision faculty will organize the perceptual stream into minimal units even if we are not observing finite objects such as tables, or if we look at the same portion of sky for quite a long interval of time. When external stimuli are tracked, then it is possible to check whether they stand for some “new” or “old” information, i.e. whether their internal representation matches previous visual computations.

Hence, the underlying properties of these computations do not crucially hinge on external stimuli, but on the possibility (perhaps, necessity) to integrate these different forms of information together in an effortless way, and in a coherent, “synchronized” model (e.g. O’Keefe 2003; Buzsáki 2006). Our Logic of Vision thus represents an internal model of logical Space, and represents the properties and relations defined over this model. By this point, our discussion of
the Logic of Vision should be thorough enough: I shall concentrate on spatial Language and its Logic.

2.3.5 A Logic of Language: Discourse Representation Theory and Space

The study of meaning in natural Language as a psychological phenomenon has long been ad-versed in model-theoretic approaches, traditionally rooted in an “anti-psychologist” philosophy (e.g. Davidson 1967; Montague 1973; Cresswell 1985). Some modern research, however, broke with this tradition and attempted to study whether the models defined in this approach can be seen as mental structures and processes of some sort, represented via varieties of dynamic logic (e.g. Kamp 1981; Heim 1982; Chierchia & Turner 1988; Kamp & Reyle 1993; Chierchia 1995).

Among these different approaches, Discourse Representation Theory (DRT) represents the most important theory with such a “cognitive” stance, and offers a set of tools which will allow us to (easily) treat all the linguistic phenomena we shall address via a single set of formal tools. For instance, it includes detailed treatments of the semantics of nouns and temporal expressions, which can be extended to treat our adpositional data (e.g. theories of noun reference and plurality: Link 1983,1998; treatments of events, Parsons 1990; Landman 2000, 2004). It also allows us to take a perspective to sentence interpretation as a dynamic process, since it aims to model how sentences are interpreted and “used” to form models in a compositional and incremental and on-line fashion, as in models of parsing such as e.g. Crain & Steedman (1985).

For this reason, it is also compatible with syntactic theories that take a derivational, psychological stance to syntactic derivations, such as the generative framework offered in Phillips (1996). Consequently, it offers an ideal theoretical framework not only to analyze the data discussed in this part, but also to address the problems that will be addressed in part II and part III, as we are going to see. The version I shall use here is also fully compositional and thus allows analyzing the contribution of each word to a sentence (i.e. Kamp et al. 2005, based on Muskens 1996; van Eijk & Kamp 1997). However, I shall focus on the contribution of nouns and adpositions for the most part, being somewhat sloppy on other parts of speech (e.g. verbs), and
leaving a more precise analysis of adpositions for part II. Although the structural equivalences with my Logic of Vision should be immediately obvious, I will defer a thorough discussion to section 2.4.1, and focus here on the linguistic bits.

The most basic bits of information in DRT are Discourse Representation Structures (henceforth: DRSs). A DRS can be thought, at a minimum, as a linguistic information state containing a set of discourse referents (or $U$, for “Universe”), an “object” in discourse, and the conditions (or $CON$) which allow us to individuate such objects in discourse. While basic (“extensional”) DRSs are at a minimum a duple of discourse referents (or individuals, for the sake of clarity) and their associated conditions, they “become” information states when a third set of objects is taken in consideration, possible worlds (the set $W$). Hence, a DRS or information state is the triple $\langle W, U, CON \rangle$ or $\langle \{w\}, \{x\}, \{\text{con'}(x)\} \rangle$, in which a discourse referent is paired with a “world” referent and a condition, and which can be seen as a mental representation or (mini-)model that a speaker entertains, when he parses chunks of sentences, incrementally.

The nature of this world “coordinate” deserves a few words of discussion. In classical logic, possible worlds are seen as quite real Leibnizian entities, such as the world we live in (e.g. Lewis 1986). Many versions of DRT, however, propose a different approach, partially based on Stalnaker’s work (i.e. Stalnaker 1973, 1999\textsuperscript{13}), in which possible worlds are mental objects, and represent nothing else than possible scenarios in which referents are involved, those for instance expressed by a sentence or a more complex text. Consequently, possible worlds can vary in “size” and structure, and may be intuitively related one another according to the same principles definable over individuals, DRSs or other model-theoretical objects, as also assumed in Situations Semantics (e.g. Barwise & Etchemendy 1990; Kratzer 1989, 2007) or modern Modal Logic (Hughes & Cresswell 1996; Blackburn et al. 2006).

Let us now turn to formal matters. As a standard convention, I write conditions in boldfaced characters and by adding a prime, e.g. “$\text{con'}$”. Hence, conditions in DRT are roughly equivalent

---

\textsuperscript{13}This is true insofar as we look at the “raw mechanics” of the underlying logic. Stalnaker’s position is not a mentalist/internalist one: for him, “possible worlds” are those of classical logic. DRT offers a much stronger mentalist perspective: very informally, “worlds” in DRT are roughly equivalent to possible thoughts or beliefs, information states ascribed to (thinking) agents. See Maier (2006:ch.1) for discussion.
to non-logical constants of first-order logic, and thus they represent “concepts” or “thoughts” as they are expressed in natural language, together with the distinction between intension and extension (cf. Margolis & Laurence 1999; Gärdenfors 2000; Winter 2008). The obvious consequence of this assumption is that our concepts/conditions will thus be invariably complex and definable in terms of their internal structure, unlike assumed in atomistic theories of concepts such as Fodor (1998, 2003). While an interesting topic per se, its discussion would lead us too far afield from our main topic of discussion, so I shall leave it aside for the time being.

For our purposes, worlds and eventualities (i.e. events, properties changing over time; and states, properties holding over time) are basically the same (model-theoretic) objects, as in some variants of Situation Semantics (e.g. Barwise & Etchemendy 1990; Kratzer 2007). Very informally, if individuals represent objects, then eventualities represent the relations in which individuals are involved\(^{14}\). I shall use the term “events” and avoid making any distinction between events and states, procrastinating a more accurate distinction to chapter 4.

Once I have defined the basic structures of DRSs, I shall focus on the combinatorial and interpretative apparatus, i.e. how DRSs can be used to represent linguistic expressions. Here I shall use a variant of the “linear” notation, rather than the more popular “box” format, to enhance readability (as in e.g. Geurts 1999). I shall roughly match one syntactic phrase with one DRS, although more precise accounts are possible (see Kamp et al. 2005:ch.0 for discussion). Look at the example:

(28) A man walks quickly. He whistles.

When a sentence like (28) is parsed, the parser builds up a bottom-up, left-to-right syntactic representation and, for each constituent and phrase, it builds up the corresponding DRS. For instance, a man is parsed as noun phrase/determiner phrase, and interpreted as the DRS \[\{x\} : \text{man}'(x)\], a DRS representing a referent \(x\) and a condition individuating him.

\(^{14}\)Note that, informally speaking, events and states are included in intervals of time, within the DRT architecture, with intervals of time forming up the main “temporal structure” of a discourse. I shall diverge from DRT and use intervals of time in a different way, as I shall show in the remainder of the section, and return to this topic in chapter 4, section 4.2.
The next step consists in combining the predicate walks with the noun phrase a man. This is obtained via the syntactic operation merge, which shall represent as “+”\textsuperscript{15}. Merge in DRT is a binary (associative, commutative, idempotent) operation that takes two DRSs and gives a “bigger” (or new) DRS as the output, by unifying the universes and conditions of each DRS. In more formal terms, we have:

\[
(29) \quad [\{x\} : \text{con'}(x)] + [\{y\} : \text{con'}(y)] = [\{x,y\} : \text{con'}(x), \text{con'}(y)] \quad \text{(merge introduction)}
\]

In words, the merging of two DRSs forms a bigger DRS in which the universes and the conditions are merged pair-wise. Merged conditions are interpreted as being conjoined. If we were to translate conditions from our DRT language to first order logic, merged conditions would be interpreted as being conjoined, whereas each referent in the universe of discourse can be translated as an existentially quantified variable. We would have “∃x∃y[\text{con'}(x)&\text{con'}(y)]” for the two conditions in (29) (Kamp et al. 2005:768-770). I shall use brackets to mark the universe, and thus enhance readability (e.g. \{x,y\}), as in van Eijk & Kamp (1997); Kamp et al. (2005).

The verb walks can now be simply represented as \([\{e\} : e : \text{walk'}(x)]\), i.e. a DRS which introduces no new (object) referents but a novel spatio-temporal referent, the event of walking. The merging of the two resulting DRSs can be represented, in a piece-meal fashion, as:

\[
(30) \quad t. \quad [\{x\} : \text{man'}(x)] + [\{e\} : e : \text{walk'}(x)] = [\{e,x\} : \text{man'}(x), e : \text{walk'}(x)] \quad \text{(m.intr.)}
\]

\[
\begin{align*}
t + 1. & \quad [\{e,x\} : \text{man'}(x), e : \text{walk'}(x)] + [\{e\} : \text{quickly'}(e)] = [\{e,x\} : \text{man'}(x), e : \text{walk'}(x), e : \text{quickly'}(e)] \\
\end{align*}
\]

In words, we obtain the DRS representing the first sentence in (28) (A man walks quickly), by merging the DRSs representing its main constituting phrases. The DRS for a man acts as the context DRS, which is then updated via merge by the DRS for walks, acting as the context change potential DRS. The dynamic aspect of meaning is thus represented by the ability for

\textsuperscript{15}Kamp et al. (2005) use a different symbol (i.e. “♀”), but this difference is immaterial, for our purposes. Note also that the properties of merge (associativity, commutativity, idempotence) stem from its definition as a (complex) form of set union, with idempotence allowing to “reduce” universes whenever they are identical (see e.g. (28), i.e. \{e,x\} + \{e,x\} = \{e,x\}).
new phrases/words to add more information regarding referents and events represented by each sentence, and thus define a “broader” model representing facts. This is also represented via the explicit use of an index set in the derivations, which allows us to explicitly represent how DRSs are combined together (as in e.g. Muskens 1996; van Eijk & Kamp 1997).

The merging of DRSs has also one important consequence: it defines a semantic level of relations between DRSs and their universes/conditions, the accessibility/part-of relation. The accessibility/part-of relation is a transitive, antisymmetric, reflective relation which allows us to define one DRS d as part of another DRS d’, i.e. $d \leq d'$. While transitivity and reflexivity intuitively define how DRSs are connected over the flow of discourse, antisymmetry allows establishing what relation holds between two referents/events/DRSs. One example is pronoun resolution: intuitively, a pronoun such as he in (28) denotes one whistling individual as being a specific individual out of those who are walking quickly in the park. If at least part of the content expressed by two DRSs can be the same, then the two DRSs individuate the same object, a condition which expresses an anaphoric relation and is usually represented as $x = y$. When the accessibility relation is restrained to discourse referents or events, it is usually called part-of relation (e.g. Kamp et al. 2005:760). Consequently, I shall just use the part-of label for a semantic relation holding between DRSs, in order to make the exposition of the arguments clearer.

This is shown in the remainder of the derivation for (26):

(31) \[ t + 3, \{e, x\} : \text{man}'(x), e : \text{walk}'(x), e : \text{quickly}'(x) \] + \[ s, y = ?. e : \text{whistle}'(y) \] = \[ \text{(merge introduction)} \]

\[ t + 4, \{e, x, y\} : \text{man}'(x), e : \text{walk}'(x), e : \text{quickly}'(e), y = x, e : \text{whistle}'(y) \]

---

16 In the “Dynamic Semantics” literature, the notion of “dynamic binding” has a more restricted (semantic) application, and it is restricted to inter-sentential merge, i.e. the binding of information units over the sentence boundary (e.g. Chierchia 1995; Stockhof et al. 1997).

17 Pronoun resolution is sensi to features, like gender and number or temporal/aspectual values. I just ignore these aspects here, for the sake of clarity. In DRT, pronoun resolution also involves presupposition resolution, what could be (very) informally defined as the integration of implicit information in a DRS, together with the resolution of the anaphoric relations associated with this implicit information. See e.g. van der Sandt (1988,1992); Geurts (1999); Kamp et al. (2005:ch.1-2) for discussion and references on this very complex and rich topic.
In words, the merging of the first and second sentence will also establish an identity relation between first walking man and second whistling man: there is really one man we are talking about, in (26). The resolution of the open anaphoric relation (i.e. \( x = ? \)) amounts to identifying two referents by stating that the properties by which these referents are individuated converge to the same result.

After this brief introduction to the relevant aspect of DRT, I shall focus on a compact treatment of adpositions, which diverges from the standard DRT treatment of this category (cf. Kamp et al. 2005: ch.2-3) and introduce a more thorough analysis of these terms, based on the vast literature on the topic. My basic assumption will match the non-linguistic considerations I offered in the previous section: adpositions denote relations between DRSs, by expressing how the events denoted by these relations are ordered (e.g. Kamp 1979a,b; Jackendoff 1983; Parsons 1990; Bierwisch 1991; Nam 1995; Fong 1997; Kracht 2002; Kratzer 2003; Landman 2004; Zwarts 2005; Svenonius 2006; Ramchand 2008).

I shall thus assume that adpositions denote anaphoric relations between events/DRSs. Differently from pronominal anaphora, though, they may express “asymmetric” relations, i.e. relations in which events are not necessarily identical. In this perspective, adpositions are akin to the “duplex conditions” of DRT, which are used to represent quantifiers such as every, but also conditionals (e.g. donkey sentences), temporal adverbs and other temporal/logical relations. The main reason for this assumption can be motivated by the following entailment patterns in the examples (adapted from Parsons 1990):

(32) A delegate walked into the park \( \models \) A delegate was in the park

(33) A delegate is near the park \( \models \) A delegate is near the park

In (32), the sentence “a delegate…” entails that the relevant delegate was in the park as a consequence of this event of motion. In (33), the sentence “a delegate…” entails itself, in the sense that it the delegate’s position is not an explicit consequence of some previous event of motion, but also holds for possibly more specific states (e.g. the delegate being currently near
the park). The symbol “\(\models\)” represents the entailment relation between the two pairs of sentences.

The intuition behind these patterns is simple: adpositions, as they mirror relations between VRSs in language, also denote equivalent relations between DRSs and the events included in these DRSs. They do so by explicitly stating how events are ordered one another, thus explicitly representing the causal/temporal structure of (parts of) a sentence, possibly restricting this relation to certain events (e.g. those being “in” the park). I shall thus translate into as the complex DRS \([\{e,s,x,y\} : e < s, s : \text{in}'(x,y)]\) and near as the complex DRS \([\{s,s',x,y\} : s \leq s', e : \text{near}'(x,y)]\). The DRSs represent in a compact manner the Parsonian entailments, as part-of relations between the events denoted by the merged sentences. Informally, if a delegate walked into the park, then he was in the park as a consequence. If a delegate is near the park, he may have arrived there because of some other events, or may stay there for some unspecified interval of time.

The interpretation of (32), at the relevant step and abstracting away from tense, is the following:

\[
(34) \quad t. \quad [\{e\} : \text{delegate}'(x), e : \text{walk}'(x)] + [\{e,s,y\} : e < s, s : \text{in}'(x,y), \text{park}'(y) ] = \\
\text{(merge introduction)} \\
t + 1. [\{e,s,x,y\} : \text{delegate}'(x), e : \text{walk}'(x), e < s, s : \text{in}'(x,y), \text{park}'(y) ]
\]

In words, (34) says that a delegate walked and, as a consequence of this event of walking, he ended up in the park. The interpretation of (29) would be similar, except that the contribution of near would yield the following (slightly informal) DRS: \([\{e,s,x,y\} : \text{delegate}'(x), s : \text{be}'(x), s \leq s', s : \text{near}'(x,y), \text{park}'(y) ]\).

This treatment of English adpositions is by no means exhaustive. Previous literature suggests us that, from a cross-linguistic perspective, we would probably need a more accurate analysis of data (e.g. Talmy 1978, 2000; Svenonius 2006; Higginbotham 2009; Zwarts 2010 for discussion). Part II will focus on solving part of this problem, by offering a more detailed analysis of the English data at hand. For the moment, however, this treatment allows us to represent in a rather simple what kind of contribution adpositions (and nouns) offer to a sentence, as well as
introducing a rather compact theory of linguistic representation, in the guise of DRT. I shall thus collect all the crucial aspects of DRT and present them as parts of DRT’s underlying logic.

DRT can be treated as a *Logic of Language*, which can be represented as the model $L = \langle \mathbf{D}, +, \leq \rangle$. The set $\mathbf{D}$ of *DRSs* is in turn a set of triples, defined as $d = \langle w, u, \text{CON} \rangle$, and with $d \leq \mathbf{D}$ holding for each $d$. The model (of Discourse) defined by DRT is a Lattice, which has a structure entirely equivalent to that defined for Vision\(^{18}\). The “dynamic” incarnation of this model is $L_d = \langle I, \mathbf{D} \rangle$, the duple formed by DRSs and intervals of time at which they are combined together, with $I$ again being defined as $I = \langle t, + \rangle$.

The mapping from this model of Language to other models, most specifically our Logic of Vision, can be easily defined via the function $g$, which is usually known as the *anchor function* in DRT (Kamp et al. 2005:ch.4; Maier 2006:ch.3 for discussion). This function is defined as an isomorphism mapping each linguistic information unit onto a non-linguistic unit, in this case a visual unit, i.e. $g(d') = v$: In our case, it matches DRSs (linguistic information units) with VRSs (non-linguistic, visual information units).

Since it is an isomorphism, it maps at least one DRS onto one VRS, and at most one DRS onto one VRS. It preserves structure, so a mini-discourse like (26) can be seen as the description of a complex scenario, made of two connected, simpler scenarios. Formally, we have $g(d' + k') = g(d') + g(k')$, which in words says: the scenario corresponding to the mini-discourse in (26) corresponds to the scenario matched by the first sentence (a man is walking in the park) followed by the scenario matched by the second sentence (this man is whistling). Much like the function $f$ the function $g$ can, but needs not to, find a VRS for each mapped term. In this regard, the function $g$ can also be thought as representing a top-down process, since it represents how we can consciously match a sentence (and its content) with an extra-linguistic scenario it refers to.

Now that both sides of the isomorphism are defined, we have a good understanding of how information flows from Vision to Language and from Language to Vision, and thus we are ready

\(^{18}\)In DRT or similar approaches (e.g. Krifka 1998), events and referents are part of (structurally) different structures; here I follow Link (1983,1998) and assume one common type of structure for all types of object. I shall expand this topic in chapter 4.
to tackle the problem of the Vision-Language interface in an explicit way. However, before doing so, I shall offer an answer to the second research question, which is now within our reach. The answer is the following:

A-2: Our models of spatial Vision and Language must include any possible property and relation that can “connect” two entities; these models can (must) be treated via a model-theoretic approach;

Spatial Language and Vision, then, can be seen as two systems representing different aspects of the same underlying phenomena: how we build up and maintain complex “maps” of the objects we keep track of, over discourse. At this point, we can explore the common Space generated by these two structures, and thus focus on the Vision-Language interface.

2.4 A Theory of the Vision-Language Interface, and beyond

In this section I shall offer a logical theory of the Vision-Language interface based on the results of the previous section (section 2.4.1); I shall offer empirical evidence in support of this approach (section 2.4.2); and sketch some broad consequences of my approach with respect to theories of the Language Faculty (section 2.4.3).

2.4.1 The Vision-Language Interface: a Formal Approach

A theory of the Vision-Language interface, given the discussion so far, must be a theory about the two-way information flow between two structures which represent (external) spatial information in a principled and highly organized way, the logical Space defined by the logics of Vision and Language. As section 2.3 has been a relatively long analysis of how these notions emerge from the basic bits of Vision and Language, I shall re-state my basic assumptions, and then focus on the Vision-Language interface problem.

I have assumed that both Vision and Language can be represented via a precise Logic, which I respectively called the Logic of Vision and the Logic of Language (or, equivalently, VRT
and DRT: *Visual Representation Theory* and *Discourse Representation Theory*). These logical calculi share the same underlying structure: VRT is defined as triple \( S = \langle V, +, \leq \rangle \) and DRT as the triple \( L = \langle D, +, \leq \rangle \). These models are lattices, partially ordered sets, which minimally differ in having different types of elements, rather than in their structure.

The basic elements in these domains are respectively *VRSs* and *DRSs*: for each *VRS* \( v \), the relation \( v \leq V \) holds; for each *DRS* \( d \), the relation \( d \leq D \) holds. For each *VRS* \( v \), the following identity holds: \( v = \langle e, o, pr \rangle \), i.e. each *VRS* is a triple of an event, an object and a property that identifies an object in an event. For each *DRS* \( d \), the following identity holds: \( d = \langle w, u, \text{CON} \rangle \), i.e. each *DRS* is a triple of a world/event, a referent and a condition that identifies a referent in a world/event. While *VRSs* are discrete units (possibly) representing perceptual stimuli from the visual apparatus, via transduction, *DRSs* may be seen as discrete units representing other types of information units (e.g. “concepts” or “thoughts”). They may be connected to *VRSs* via a slightly different type of transduction, but do not have a direct “external” grounding: they represent purely “internal” information.

While the two structures have different elements, their operations and relations are basically the same. A syntactic operation, *merge*” (ultimately, set union), allows us to define each element as the sum of other elements, possibly only itself. We represent it as “+”. Its definition is simple: it is a binary operation taking two inputs of the same type (e.g. *DRSs*: \( a + b \)), yielding an output of the same type as the inputs (a *DRS*: \( a + b = c \)). It is associative, commutative and idempotent: it allows combining the same elements in different ways (associativity: \( (a + b) + c = a + (b + c) \)), regardless of their order of occurrence (commutativity: \( a + b = b + a \)), and can be “repeated” on the same input (idempotence: \( a + a = a \)).

A semantic relation, the *accessibility/part-of* relation (represented as “\( \leq \)”), integrates this syntactic operation and establishes how the results of the *merge* operation are “connected”. It is binary, as it establishes a relation between two objects of the same type (e.g. *VRSs*: \( a \leq b \)), and it is reflexive, asymmetric and transitive: it allows us to establish that objects are part of themselves (i.e. \( a \leq a \)); that objects can be identified (i.e. if \( a \leq b \) and \( b \leq a \), then \( a = b \)); and
that multiple relations can be compressed as a single relation (i.e. if $a \leq b$ and $b \leq c$, then $a \leq c$).

The *merge* operation and the *part-of* relation are connected via the following properties, which I shall again represent via set-theoretic notation. If $a \leq b$, then $a \cup b = b$ and $a \cap b = a$. In words, if one object is part of another, then their merging will correspond with the “bigger” object (union), and their product will correspond with the “smaller” object (intersection). Semantic relations can be seen as the result of previous instances of syntactic operations, in a sense recording the successful *merge* of two objects into a more complex, novel object. The structures defined by these operations are complex Lattices, i.e. partially ordered sets with a syntax and a corresponding semantics, and thus models of the phenomena they represent.

Although other operations can be defined (e.g. set intersection standing for attention), these “minimal” logical systems allow us to aptly model how information units are processed and integrated together into more complex units, in a bottom-up way. They also allow us to define how one logic can be tightly connected to another via two functions, $f$ and $g$, which respectively define an isomorphic mapping from VRSs to DRSs and from DRSs to VRSs. These functions are isomorphic because they map at least one input and at most one input to the same output, i.e. they are respectively injective and surjective, thus they are bijective.

The function $f$ is defined as: $f : v \mapsto d$, i.e. a function that maps each visual structure $v \leq V$ onto a discourse structure $d \leq D$, whereas the function $g$ is defined as: $g : d \mapsto v$, i.e. a function that maps each discourse structure $d \leq D$ onto a visual structure $v \leq V$. Via these functions, we have the identities $f(v) = d'$, and $g(d') = v$. Note, now, that these two functions are one the inverse of the other: their composition (represented via the symbol “$\circ$”) will yield the identity function, e.g. we have $f \circ g = i$, with “$i$” being the identity function. This latter property tells us that e.g. each noun may act as the linguistic label for a visual object, and thus that each visual object may a have noun as a linguistic label.

These isomorphisms allow us to explicitly represent how we “translate” one type of objects into another, while for logical operators (i.e. *merge* and the *part-of* relation), they offer evidence that these operations are the same across models/logics. The reason is simple: while objects
define non-logical constants, merge and the part-of relation define logical constants, elements of a logic that receive the same interpretation on any model, whether it represents Vision or Language. In words, merge is interpreted as the union of two objects, whether these sets stand for visual structures or discourse structures, and so is the part-of relation interpreted as a relation between objects.

This is explicitly represented via a structure-preserving condition on our isomorphisms: 
\[ f(a + b) = a' + b', \] given that \[ f(a) = a' \] and \[ f(b) = b'. \] In words, the noun for the object corresponding to the merge of the objects “legs” and “surface” (table) corresponds to the superordinate noun that stands for the objects “legs” and “surface”, the noun for the object “table”. The merge symbol is the same on both sides of the identity, while the merged objects are different. The same holds for the part-of relation: since if we have \[ a \leq b, \] then we have \[ f(a) \leq f(b). \] In words, if a leg is part of a table, then the noun/concept “leg” is part of the noun/concept “table”. The same considerations hold, mutatis mutandis, for the function g. In words, Vision and Language may differ as models representing different “things”, but they are equivalent as models sharing the same structure.

The definition of these two isomorphisms has one important consequence: it allows us to outline a simple and yet very precise theory of the Vision-Language interface. The main assumption I shall make is that the Vision-Language interface is defined as a Galois connection between these two structures. A Galois connection is defined as follows: given two lattices \( \langle A, \leq \rangle \) and \( \langle B, \leq \rangle \), \( f(a) \leq b \) if (and only if) \( a \leq g(b) \). In our case and with some notational fantasy, given the Lattices \( \langle D, \leq \rangle \) and \( \langle V, \leq \rangle \), we have \( g(d') \leq v \) if and only if \( d' \leq f(v) \). In words, if Vision and Language are connected via a Galois connection, then the VRS corresponding to a DRS is part of a larger VRS, and a DRS corresponding to a VRS is part of a larger DRS. In words, Vision and Language representations are connected if each linguistic term is matched by a visual entity, which is part of “larger” scenario; and if each linguistic term expressing a visual object is part of a sentence. Informally, a Galois connection is a method of defining an isomorphism between structures in which weaker relations can also be defined: it allows us to express not
only that structures “look the same”, but also to compare the relation between many elements of one structure to an element of the other structure (e.g. Ganter & Wille 1998:ch.1).

The strength of this proposal is that it allows us to define a degree of accuracy by which a certain sentence describes a state of affairs and vice-versa. For instance, an adposition matches a spatial representation when the two following conditions hold: \( f(v) = d' \) and \( g(d') = v \). In words, if we take in consideration a scenario in which a book is supported by the top vertical surface of a computer, then the adposition on top of is quite ideal match for this scenario, since we intuitively have \( f(on − top) = on-top' \), but also \( g(on-top') = on − top \).

While identity cases are in a sense trivial, cases of partial matches allow us to grasp the crucial strength of the proposal. For instance, an adposition such as on expresses only support of the book by the computer, and thus is intuitively less accurate than on top of, which expresses the specific surface of the computer offering this support. This because it will represent only a part of the spatial representation in which book and computer are involved: if \( g(on') = on \) and \( on \leq on − top \), then we will have \( g(on') \leq on − top \) to hold. In words, on represents only a part of a certain extra-linguistic scenario, and thus will be less accurate than on top of. Conversely, the relation \( on' \leq f(on − top) \) also holds, i.e. on is less accurate than the adposition which would perfectly match the said scenario. Hence, the part-of relation, when it is defined on “mixed” objects by means of a Galois connection, can be interpreted as relation expressing a degree of accuracy of a sentence, an adposition or any part of speech, with respect to the extra-linguistic context.

This proposal on the Vision-Language interface makes two main predictions. First, it predicts that the “amount” of spatial (visual) information expressed by a sentence is flexible, and may be as accurate as the corresponding visual scenario, but also that the same scenario can be described by adpositions of different “accuracy”. Second, it predicts that, since the “binding” between the two layers of information may go in both directions, there is no “causal” relation between these different computations, so one type of information is processed independently of the other. We are quite able to evaluate whether what we see refers to (or matches with) what we say.
and vice-versa, but both mental processes need not a constant, unconscious feedback between the two levels of comprehension to occur. In words, we can produce sentences about “where” things are (including, but not limited to geometric relations), but need not to limit ourselves to what we see.

I will analyze how these predictions are borne out in the next section, after giving a formal treatment of this “parallel” processing:

\[
\begin{align*}
I & \quad V & \quad V \Leftrightarrow D & \quad D \\
\text{t.} & \quad (a + b) & \quad (a' + b') \\
\text{t + 1.} & \quad (a + b) = g(a' + b') & \quad f(a + b) = (a' + b') \\
\text{t + 2.} & \quad (a + b) = g(a' + b') \Leftrightarrow f(a + b) = (a' + b')
\end{align*}
\]

In words, at some interval in a computation, the two types of information are first mapped onto the other domain, and then (dynamically) bound together if the two “flows” of the process yield the same result, possibly compared in terms of accuracy in a common logical space, which is represented as “\( V \Leftrightarrow D \)”. Informally, we check if what we see matches with what we say and vice-versa, hence obtaining a “broader” picture of facts. Since what we see needs not to match with what we say, the binding relation between these two types of information is entirely optional and, as we have discussed so far, it ultimately represents a top-down translation process, which can be more or less accurate.

One important thing to note that this formal treatment is modular also because the binding of two types of information is explicitly represented as a distinct result of a matching operation. If we would have assumed that the binding occurs by the simple co-synchronous occurrence of these operations, our architecture would actually have been connectionist, in nature. While the two processes are isomorphic and can be tightly connected, they are nevertheless two distinct processes, and a third process is their matching relation (i.e. binding). See Marcus (2001) for discussion. Now that we have gone through the formal details and their predictions, we can focus on their empirical support, which I shall analyze in the next section.
2.4.2 Testing the Theory against the Data

The theory I have proposed in the previous section is consistent with general assumptions about Vision and Language as parts of a cognitive and modular architecture (cf. e.g. Jackendoff 1997, 2002), and possibly offers a more fine-grained and formally precision analysis and representation of these modules and their processes. In this section I shall explain more in detail why this theory is consistent with previous proposals and offer an “improved” model of their insights, and why it is consistent with general assumptions about cognitive architecture, i.e. why the two main predictions I offered in the previous section hold. I shall focus on four topics, offering evidence that confirms these predictions.

The first topic pertains the “amount” of Space found in Language. Let me repeat (19) and (20) as (36) and (37) below, to illustrate the point:

(36) The book is on the tip of the left edge of the blue table
(37) The book is on the table

The crucial difference between (36) and (37) is that both sentences may be used to convey information about the same extra-linguistic scenario, but (37) is definitely more accurate than (36). Vision-wise, a scenario in which the book is supported by the tip (of the edge) of the table is also a scenario in which a book is supported by the table: hence, the relation on ≤ on-top holds. Language-wise, the DRS representing (37) is part of the DRS representing (34), so the relation on’ ≤ on-tip’ holds. Hence, the following identities g(on-tip’) = on − top and g(on’) = f(on) hold, as well as on-tip’ = f(on − top) and on’ = (on). We can then observe that the relation g(on’) = on − top holds, i.e. that (37) is a partial representation of the same scenario that (36) is a total representation of, and thus a less accurate description of facts. Conversely, the relation on’ ≤ f(on − t) holds, i.e. (37) expresses part of the information expressed by (36), and thus of the scenario that (36) represents.

A second topic pertains the different degree of accuracy that two sentences can have in describing a certain scenario, when involving different adpositions. If the meaning of two adpo-
sitions overlaps or stands in an entailment relation, then speakers may favor one over another, when they need to associate it to visual information. The entailment cases are quite intuitive, and can be seen as a general case of the relation between (36) and (37). In a situation in which a book is supported by the upper part of a drawer, on top of may be judged as a “perfect” adposition to describe this situation while on, that is entailed by on top of, may be considered as less appropriate, with respect to the scenario it purports to match with.

The cases in which adpositions overlap in meaning require some more discussion. Let me repeat (16) and (17) as (38) and (39) to illustrate the point:

(38) The painting is on the wall
(39) The painting is in the wall

In a scenario in which a panting is literally encased in the wall, (39) may be a more accurate sentence to describe this scenario than (38), because it may express in a more precise way the matching extra-linguistic scenario. Intuitively, if a painting is in the wall, it is certainly supported by it, and actually part of the wall’s surface, rather than just adjacent to it (as for on). Formally, we can say that in is more accurate than on with respect to the aforementioned scenario if the following holds: if in ≤ on, then in ∩ on = in, i.e. in is a part of on and its meaning; hence, if g(on) = g(in), then g(on) ∩ g(in) = g(in), i.e. in describes a more specific scenario than on, and hence is considered more accurate.

The treatments I discussed in the first and second topic are consistent with results like those of Coventry & Garrod (2004), Regier et al. (2006) and much aforementioned literature on (spatial) sentence processing, which also cover the relations between e.g. above and on, in and under, and so on. It is also consistent with research on the relation between visual stimuli and the interpretation of sentences. For instance, it captures the fact that participants may focus their attention on visual stimuli representing food, when parsing sentences involving verbs such as eat, since this verb will act as a “visual” cue to look for edible objects (e.g. Altmann 2001, 2004). It is also consistent with Levinson & Meira (2003) cross-linguistic results, which are
indeed based on how adpositions can be conceptually organized in terms of increasing accuracy and specificity of their use in (implicit) context\textsuperscript{19}. This literature also offers indirect evidence of the validity of my proposal: most experiments aim to test how participants consciously match visual stimuli with linguistic stimuli, evaluating how accurate sentences can be in describing a scenario. Hence, it indirectly supports the view that the functions $f$ and $g$ represent conscious processes.

A third topic regards a complex case, that of the relation between Vision and Language with respect to reference systems and their computation. As we have seen in section (2.2), works like Carlson-Radvansky (1994), Carlson (1999) or Regier \textit{et al.} (2006) show that, when speakers interpret axial terms such as \textit{to the left of}, their accuracy can be measured with respect to different reference frames, e.g. whether a chair is to the left of a table with respect to the observer (relative frame), the chair itself (intrinsic frame) or an environmental cue like the floor (absolute frame). What I have suggested for “standard” adpositions can be extended to these “axial” adpositions as well, with no need to make any further assumptions. Furthermore, although some proposals conjecture that the “cognitive” procedures by which “absolute” spatial relations are computed dramatically differ from other visual procedures (e.g. Levinson 2004), their mapping onto linguistic unit seems to be rather “ordinary”. Whether we may compute a polar direction such as the one corresponding to \textit{North} via an entirely different set of cognitive resources than the ones involved in e.g. computing the support relation corresponding to \textit{on}, the two adpositions share the same underlying grammar, and seem not to reflect this “cognitive difference”, if it exists.

From these three topics we can observe that the first prediction of my novel interface approach, the flexibility of this interface, is substantially borne out. This allows making a further general comment regarding the “how much Space” problem, and how we may choose the degree of accuracy we want to express. The literature gives us the relevant answer regarding how this

\textsuperscript{19}A conjecture is that classical results of prototype theory (e.g. Rosch 1975) may actually find a formally precise account, if we e.g. pursue the intuition that a noun such as \textit{robin} may be seen as the perfect linguistic label for the sum of all visual/cognitive information we ascribe to birds, rather than \textit{penguin}. This intuition is actually pursued in Ganter & Wille (1998) and especially in van Eijk & Zwarts (2004) in thorough detail.
process comes about, in the guise of theories of sentence planning and production. For instance, in a theory of sentence-planning (Speaking) like Levelt (1993), speakers are assumed to decide, at a pre-linguistic level, both which basic “facts” and the relations between these facts they wish to convey (Levelt’s level of macro-planning), and consequently which language-specific rules (syntactic and semantic alike) to use in order to convey these facts (Levelt’s level of micro-planning).

For our discussion macro-planning represents the relevant aspect of production, since it indirectly defines “how much” we may express about extra-linguistic information. In slightly more formal terms, macro-planning may be treated in the following way. A speaker may look at a certain general visual context $V$ and may decide to express part of this scenario via the selection of a certain VRS $v$. Given a selection function $s$, this process can be represented as e.g. $s(V) = v$. For instance, a speaker may look around a room and may decide to say that a certain specific book is on the tip of the left edge of the blue table. The selected VRS $v$ would actually stand for the complex VRSs representing book, blue table, edges and tips, and the relations holding between these VRSs.

The sentence corresponding to this VRS, which we can represent as $f(v) = \text{S}'$ and thus as $f(s(V)) = \text{S}'$, indirectly represents which pre-linguistic facts are chosen by the speaker as finding their way into language. The amount of Space finding its way into Language roughly corresponds to the speaker’s intentions to be more or less accurate in describing a scenario and his eventual desire to express one outstanding aspect over another. Although he may do so via different micro-plans, i.e. via the choice of different words and sentences formed out of the merging of these words, this choice is inherently flexible, rather than dictated by constraints on what type of spatial information finds its way in Language. This is captured by the function $f$ taking the function $s$ as its input. Informally, we may decide to say something about the scene we are paying attention to and, in doing so, we selectively (and consciously) pick out visual information about this scene, then “convert” it into the corresponding sentence, thus effectively deciding how much “Space” gets into Language.
A fourth topic is the relation between Vision and Language, in case of cognitive impairment in one of the two modules. The intuition is the following: if my theory can predict how the Vision-Language interface works, it should also make predictions about the problems that could arise when these modules are not properly interfaced: it should be *breakdown-compatible* (e.g. Grodzinsky 1990). The following three examples suggest that this is indeed the case.

A well-known fact is that people affected by *Williams Syndrome* may have relatively good language skills, including a good understanding of spatial language, but are usually unable to assess even basic spatial relations, from a visual perspective. They may be able to understand an adposition such as *in front of*, but may not be able to evaluate what is the front of an object, such that another object can be aligned with this portion (e.g. Landau & Hoffmann 2005 and references therein). An obvious account, in the current proposal, is that since spatial vision is quite impaired, it will not be possible to have a visual input that will correspond to a linguistic output, i.e. the function \( f(v) \) will be undefined since it will have no input, and so the \( g(d') \) will be undefined, too. As a consequence, it may not be possible for patients with Williams syndrome (to make one example), to relate what they see with what they say. As it stands, our proposal seems to be consistent not only with a general modular approach to Cognition, but also with a general approach to Cognition and its disorders.

Another well-known case of a cognitive disorder affecting one side of the “Space” interface is *Aphasia*. In *Broca’s aphasia*, prepositions’ omission (among other functional words), while spatial Vision is usually (completely) spared. Adposition omission in Aphasia may be gradual, and patients with this cognitive disorder tend to omit more general adpositions (e.g. *at*) rather than less general adpositions (e.g. *in front of*: see e.g. Trofimova 2009 for a recent review). Regardless of their degree of Language Impairment, aphasics usually lose their ability to *produce* adpositions, but not their ability to *comprehend* adpositions and, more in general, Language; hence, they are able to understand whether adpositions correctly describe a scenario or not. While one aspect of spatial Language can be dramatically impaired (e.g. production), all other aspects of both spatial Language and Vision, including their interface, are substantially spared.
in line with the assumptions outlined so far.

A similar account may be extended to another cognitive disorder, that of Dyslexia\textsuperscript{20}. Models like the Dual Route Cascaded Model of reading aloud (DRC, e.g. Coltheart \textit{et al.} 2001; see also Beaton 2004), the processing of (“reading”) a single word is assumed to occur via three parallel processes, one in which we visually recognize a written word (non-lexical route), and one in which we (may) retrieve its lexical entry, its phonology and syntax-semantic properties (lexical/sub-lexical route). Although one process can be faster than the other, the full recognition of a word occurs when both processes converge to the same output, but fails if the “visual” process is damaged (failure to read graphemes and words, or shallow dyslexia) or the “linguistic” process is damaged (failure to understand words’ meaning or deep dyslexia).

As per the other cognitive disorders, our theory of the Vision-Language interface is consistent with this analysis of dyslexia, without the need of adding any further assumptions. Although for dyslexia we would certainly need a more accurate and specific analysis of both sides of the problem, the intuition seems to be correct: we may not be able to see certain visual objects correctly, but we may still retrieve their corresponding linguistic labels, and \textit{vice-versa}. We can also observe that the second prediction is borne out, since these cognitive disorders show that spatial computations can occur both at the visual and linguistic level and can be bound together, but also that this binding process is not necessary. In fact, even if one side of this process may be completely impaired, the other side will be still able to work independently.

Summing up, the discussion of these four topics suggests that our Vision-Language interface theory can have theoretical value and can withstand empirical scrutiny, even once we look beyond the topic of Space. As we have seen, visual and linguistic represented can be matched in a quite precise way, but the processes regulating this matching of information is inherently conscious, i.e. based on a speaker’s top-down thought processes. Speakers may wish to be more or less accurate in describing a scenario and may evaluate sentences with respect to their de-

\textsuperscript{20}Dyslexia can be informally defined as a cognitive disorder which influences our ability to successfully read, i.e. to either successfully decode the sequence of graphemes (“letters”) making up a written word, or to properly interpret a word, and access syntactic information about it. See Beaton (2004) for a thorough introduction.
scriptive accuracy. They may be able to understand spatial Language even if they can’t navigate the environment and, for complex tasks such as reading (i.e. the codified matching of visual and linguistic stimuli), they require conscious and protracted effort to establish the proper mappings, provided that this mapping is not impaired by cognitive deficits.

These facts are somehow hard to explain in previous accounts of the Vision-Language interface, but fall out as predictions of the theory I have sketched so far, thanks to its flexibility. This theory also presents in detail the convergences between space Vision and Language, offering a view in which these two modules are remarkably similar; as such, it may appear that there is little or no difference between the two modules, both from a structural and content-bound point of view. I shall focus on these differences in the next section.

2.4.3 What is unique to Language, and why

The discussion I have offered so far has sketched the strong similarities between Vision and Language as modules of Cognition, and has offered an attempt to explain how these two modules exchange information, for instance via the synchronization of their processes. (Spatial) Vision and Language seem to be remarkably similar modules, and it is not surprising that in some quarters they are considered as contiguous modules, if not the same module, in some respect (e.g. Talmy 2000; Coventry & Garrod 2004).

There are, however, a number of properties of Language which seem rather hard to reduce to general, non-linguistic features, and which inherently involve the possibility in Language to convey information about “unbounded” quantities. Much of our discussion up until this point has focused on defining the properties that can be ascribed to the Broad Faculty of Language (FLB), since I have mostly been concerned with the relation between Language and Vision, and with those properties that are shared by both computational system (Hauser, Chomsky & Fitch 2002; Fitch, Hauser & Chomsky 2005). In this section, I shall sketch a very preliminary proposal, stemming from the discussion offered so far, on what properties are unique to Language and thus may be possible candidates to form the Narrow Faculty Of Language (FLN) kernel. I
shall do so by focusing, for the most part, on spatial Language. I shall discuss these properties in a less formally rigorous way, focusing on speculative aspects of the discussion.

Look at the examples:

(40) Mario has gone to the store *three times*

(41) Mario *may* go to the store

(42) *All* the boys have gone towards the store

(43) *Every* boy will go toward the fence

(44) A boy may come to the party

(45) *Some* boy may come to the party

(46) Mario *always* goes to the store

(47) Mario *seldom* goes to the store

(48) *Where* are you going?

(49) I am going *there*, too

(50) Mario *lends* a book to Luigi

(51) Luigi *borrows* a book from Mario

In (40), Mario’s going to the store is described as occurring *three times* or instances, but little is said about when this happens: it may occur one time right now, one time yesterday, and one time when he was a young lad. Two of the events that the adverb denotes cannot be mapped onto visual inputs, because two of them cannot correspond to current facts, but rather to “memory traces” we have recorded of them. Language allows us to merge together pieces of information which do not necessarily correspond to one modality, into a unified type of information.

In (41), Mario’s *possible* event of going to the store is something that we conceive as occurring in, say, a few more minutes, or if he feels like it, or perhaps tomorrow. In the case of the non-current events of (40), the modal auxiliary may simply denotes a linguistic unit which
hardly can find a visual unit as its counterpart. In (42) and (43), the amount of boys that have
gone to the store may vary, and may involve pairs or triples (or bigger quantities), but each of
these possible combinations of boys will go to the store, without any exceptions.

In (44) and (45), instead, we may not know the identity of who is going to come to the
party, except that it is likely a single boy, someone who we may have not mentioned so far and
we may never come to know, let alone see. These cases may already show that the mapping
from Vision to Language can be quite partial (i.e. not always defined), but the following cases
should give even stronger evidence. Adverbs such as always and seldom, as in (46) and (47),
suggest that we may even convey linguistic information about several (infinite) situations (sets
of events) in which Mario goes to the store, or say that such situations are rare, but they do occur
(i.e. seldom).

Examples such as (48) and (49) show that we may actually rely on someone else’s ability to
access information, in order to retrieve information of Mario whereabouts: if someone answers
our question, we will be able to know Mario’s location without actually see this location, and if
someone has already told us where Mario is going, we may say that we are going there, although
we may not be able to see “where” is “there”. In (48) and (49), the same set of events is presented
under two different, and in a sense complementary, perspectives: while the visual scenario is in a
sense the same (a book is temporarily exchanged, between Mario and Luigi), the two sentences
express these facts from Mario or Luigi’s perspective, respectively.

There are two generalizations that we can make, from these examples. One generalization
is that Language may convey information which can be multi-modal, in the sense that linguisti-
c units may bring and represent together information which comes from different cognitive
sources, and may have no extra-linguistic instantiation whatsoever. This is not surprising if we
look at Language at a module that only processes internal information, stripped of any percep-
tual or modal-specific aspects (unlike Vision), but it is also consistent with various theories of
memory as a “mental” model in which we record and organize memory.

One way to look at this aspect is the following, and it is based on theories of memory like
Cowan (1988, 1995, 2005). In this theory, Long-term memory is seen as the model representing all the information we may have stored about the world, whether it is veridical or not (i.e. whether it is represented in episodic memory or not\(^{21}\)). Short-term memory, on the other hand, can be seen as the current part of long-term memory which is accessed and evaluated at a given time. In our Logic, Long-term memory can be seen as a static model \(\langle D \rangle\) or \(\langle V \rangle\), while short-term memory can be seen as the dynamic counterparts of these models, \(\langle I, D \rangle\) or \(\langle I, V \rangle\).

For instance, we may have observed Mario going to the store in three very different moments of our life, but if we use a sentence like (40), we represent these otherwise separate events of time in the same representation (ultimately, a DRS) in our short-term memory. Language allows us to define a “common space” in which “displaced” events form a may form a consistent representation insofar as they share the same formal properties (e.g. being three instances of a “walking” event), and thus are stripped of any constraints on perceptual information, but may also be bound with other “portions” of short-term memory (e.g. Visual computations, cf. previous section). Informally, an adverb such as three times says that there are three contiguous intervals in a derivation in which three events of going to the station become logically contiguous, i.e. we have \(a + b + c\) at an interval \(t + n\).

Another generalization is that Language can express relations and quantities which are not necessarily finite (or bounded), and is not limited to offering one perspective. This latter, (quite) rough intuition is based on our last pair of examples, but several other similar examples could be made: think of any active sentence and its passive counterpart, for instance. If we think in slightly more formal terms, we may think of (44) as representing a scenario in which Mario’s actions as an “agent” operates onto Luigi as a “patient”, and can be very schematically represented as \(a \rightarrow p\). We can then assume that (51) can be represented as the inverse type of relation, which can be represented as also \(\neg(a \rightarrow p)\). In very informal words, we can represent that the sequence of events expressed by (51) flows in the opposite direction of (50), as the informal use of negation aims to represent, although we express the order of relevant entities in the same way.

\(^{21}\)Episodic memory is a component of memory which “records” perceptual information regarding the first time we observe a given event.
as in (50).

This is possible because in Language we can express the same underlying conceptual structures under different “perspectives”, but via virtually the same logical apparatus (cf. also Landman 1991: ch.3; Landman 2004: ch. 7-8). Again, if we think of Language as defining a conceptual “Space” not constrained by perceptual limits, then the same underlying information can be expressed in two apparently opposite ways, which however underlie the same logical principles and processes (e.g. merge, the part-of relation). Although (50) and (51) describe the same extra-linguistic event, their interpretations represent two possible ways by which Language can structure this information.

Another form of “unboundedness”, its linguistic realization as well, is ultimately represented by the interpretation of quantifiers and other “expressions of quantity”, as the examples show. Informally, in Language we can express information about numbers of individuals which are far greater than the amount of individuals we can “see”, and can be structured in rather complex and fine-grained ways, as adverbs like seldom and always suggest.

This can be illustrated via a detailed analysis of (42) and (43). Note here that I shall depart quite dramatically from DRT and treat a quantifier like every as represented by a logical operator, rather than a duplex condition. In both sentences, it is possible to represent the contribution of all and every to the sentence in terms of the universal quantifier, which I shall here represent in its Boolean incarnation, “$\forall$” (e.g. Montague 1973; Keenan & Faltz 1985). This symbol can be informally interpreted as a form of unbounded coordination: informally, the sentence every boy has gone to the store can be interpreted as the equivalent “Mario has gone towards the store and Luigi has gone to the store and...”, i.e. as if we were to state each possible boy in a large, perhaps infinite domain of discourse, one by one.

Suppose then that we take the set of boys as a list (sequence) of boys in discourse. The DRS representing all the boys is equivalent to the merging of the DRS representing the sum of the last boy with the sequence of boys occurring before him, in this infinite list. We define the interpretation of a universally quantified noun phrase (its DRS) via the sum of the interpretation
of its parts, via induction (its constituting \textit{DRS}s).

This can be represented as:

\begin{equation}
(52) \quad t. \quad [(\{x - 2\} : \text{boy'}(x - 2)] + [(\{x - 1\} : \text{boy'}(x - 1)] = \text{(merge introduction)}
\end{equation}

\begin{equation}
(53) \quad t + n. \quad [(\{x - 2, x - 1\} : \text{boy'}(x - 2), \text{boy'}(x - 1)] =
\end{equation}

\begin{equation}
(54) \quad t + n + 1.\lceil \bigwedge x : \text{boy'}(x) \rceil
\end{equation}

With the referent/individual “(x - 2)” representing the list of boys preceding the last boy (i.e. the second-to last complex referent), “(x - 1)” representing the last boy, and “\bigwedge x” representing the “new” referent obtained from the merging of the two “old” referents. This is a \textit{recursive, inductive} definition of the universal quantifier, in terms of an unbounded form of \textit{merge}, and its interpretation. In words, we interpret the universal quantifier as the result of taking each referent in discourse via one common condition. This result is another \textit{DRS}, the \textit{DRS} representing the result of taking each referent which can be identified as a “boy” one by one, i.e. via the product of each condition merged in a \textit{DRS}, here represented as “\bigwedge”.

These considerations can be also extended to other quantifiers with the proper \textit{provisos}, but also to adpositions, and suggest that spatial Language is also “unbounded” in its interpretive range. For instance, the relational component of any adposition (e.g. \textit{near}) can be recursively defined as the merging of two opportune relations. Abstracting away from the specific condition on proximity (i.e. \textit{near'}) and with some notational fantasy, \textit{near} can be represented as:

\begin{equation}
(55) \quad [(s, s') : s \leq s'] = [(s, (s' - 2)) : s \leq (s' - 2)] + [(s, (s - 1)) : s \leq (s' - 1)]
\end{equation}

Here the “geometry” approach to adpositions is quite useful to illustrate the intuitive meaning of (53). If a figure is the ground when it occupies a certain region, then it will be near the ground if it occupies any region which is included in the bigger region. Conversely, once we sum all the (sub)-regions in which a figure is near a ground, then we will obtain the “general” region which can be labeled as \textit{near}.

This way of representing the universal quantifier and in general of representing quantified noun phrases, as well as the interpretation of \textit{near} and other adpositions, is informally based on
one recursive function, the *Fibonacci series*, which allows defining one object (e.g. a natural number) as the sum of its direct predecessors. Intuitively, it may be extended to all of the other functional words I have discussed in examples (38)-(49), and to any expression that captures a form of quantification.

Several authors have argued that the Fibonacci series can represent how recursion is expressed in Language (e.g. Soschen 2008 and references therein), but one may also assume that the *successor function* may be a recursive function that can also be used to represent the recursive nature of syntactic processes (Landman 1991:ch.1 for discussion). The crucial aspect is that since Language is different from Vision by being fully recursive at a syntactic level, it will also be different in having terms which directly express the result (interpretation) of this unboundedness, and thus will be fully recursive at the semantic level.

An indirect way of capturing this difference is by enriching our Logic representing Language, so that we have the tuple $L = \langle \mathcal{D}, +, \leq, \cap \rangle$. This tuple represents the “structure” of language as including not only a minimal syntax (*merge, “+”*) and semantics (the *part-of* relation, “$\leq$”), but also a set of operators, here represented by the universal quantifier, that denote *the result* of linguistic processes. We have an indirect reconstruction of the distinction between FLN and FLB. This reconstruction is indirect only because recursion is a resulting property of the “logic of Language”, but it nevertheless represents one element of distinction (maybe *the* element of distinction) between Language and Vision.

Informally, it tells us that Language has certain recursive *closure principles* which allow to label not only objects from other models (e.g. nouns for objects), but also to express the processes by which we collect together these objects into abstract structures. Adpositions represent one case, and quantifiers represent a more language-specific case, but the same reasoning could be applied to any functional word in Language. We are able to talk about e.g. *all the past boys and apples* because we are able to compute a referent that stands for the combination of two different, and possibly representing entities “displaced” in time and space, sets of entities (i.e. boys and apples), even if these sets may include an infinite amount of “smaller” referents (i.e.
each single boy and apple).

It is also indirect because, as we have seen, visual and linguistic information are processed as distinct, although potentially connected up to isomorphism, types of information. While there can be an intimate relation between what we see and what we say, Language is not bound by other modules of cognition, in its expressive power, although the entities that make up the Universe of Discourse denoted by Language must be the result of previous processes of interpretation, as the closure principle entails.

One important aspect, however, is again that Vision can be represented via a similar, although less “powerful”, logical structure: as observed in Pinker & Jackendoff (2005); Jackendoff & Pinker (2005), a number of “structural” or hierarchical properties are domain-general, and thus not unique to Language, because they represent domain-general logical principles by which we process, retain and organize different types of information. Vision represents here one important case, but the phonological component of Language also offers a similar case, and other examples also abound (e.g. the “grammar of action” analyzed by Fujita 2009, the “grammar of music” of Jackendoff & Lerdahl 2006; “the grammar of phonetics” of Reiss 2007).

The intuition behind these considerations is the following. Each module of Cognition that is properly definable can be represented via the same underlying Logic, which I have presented here in two slightly different “incarnations”. The structures defined by this logic are models of the “things” they represent, for instance visual objects. These models can be infinite, since they can potentially represent e.g. the infinity of objects we can recognize, or events we can witness, and so on. The models defined by each module can be mapped onto a “common” logical space, that of Language: we can talk about what we see, smell, think, believe, etc.

This very informal discussion can be made more precise via the discussion of a well-known theorem of model-theoretic semantics, the Löwenheim-Skolem theorem. This theorem can be very roughly paraphrased in the following way: if a first order logic has infinite models, then it is a corresponding countable infinite model. In our case, its import can be see as follows. We may define several logical systems, each of them representing a single module of Cognition.
Each logic has the same underlying (and thus domain-general) syntactic and semantic principles. Each logic can define an infinite model: we may be able to recognize an infinity of (moving) objects, an infinity of sounds, realize an infinite of possible actions, and so on. Defined in this way, each logic/module appears to be an independent system, an internal model that potentially allows us to represent how we can interact with the external world, but needs not to rely on “external” inputs for these computations.

The Löwenheim-Skolem theorem tells us that, even if we have an infinity of such logics, it is possible to define a more general logic which includes all of these modules in a “common” logical space. More precisely, the downward part of the theorem tells us that, if a model is (countably) infinite, then this model may include an infinity of possible sub-models, themselves infinite. The upward part of the theorem tells us that for each (infinite) sub-model, we can find an extension of this model that includes the sub-model and some other elementary statements. So, if our “main” model represents Language, then it will include models of other modules as proper sub-models (downward part); if a module such as Vision can be represented via a model, then this model can be integrated inside the main model of Language (upward part).

The conceptual import of this theorem can be dynamically interpreted as follows. We can assume that, for each (well-formed) visual computation, we can have a matching VRS in our model of Vision. Each visual information unit can then be mapped onto the language model, and thus can be part of a general model that includes other types of information (upward part). Conversely, for each linguistic unit so defined, a corresponding non-linguistic unit can be found, so that from the general model, we can move to the more specific model (downward part). This process can unravel over time: for each thing we see, we may have a corresponding noun, which we then associate to any object that has that shape, to put it in a very informal way. The same principle of closure can be defined for adpositions (and verbs): for each type of spatio-temporal relation between objects we “see”, we may have a corresponding adposition, which we associate to any relation that has that spatio-temporal structure, or “shape”.

Both model (Language) and sub-model (Vision) will thus be expanded or updated over time,
but the underlying (Boolean) structure representing these processes and their results will retain the same basic structure, as this update process will be guided by the same basic principles. Informally, these models can become quite “rich” over time, but the basic structural principles by which their growth occurs remain the same, as a consequence of their recursive definition. In this regard, recursion represents the possibility for Language to apparently expand ad infinitum, representing any type of information in a common space. Similarly, the relation between the Language model and its sub-models, which takes the shape of interface relations/conditions, represents the possibility that language (recursively) emerges as a “general” model, generated by the projection of all models of Cognition into a “neutral” logical space.22

I shall thus propose the following answer to the third research question:

A-3: The nature of the Vision-Language interface is that of a bijection; recursive closure principles and interface conditions define what is unique to Language;

What distinguishes Language from other modules of Cognition is not the type of underlying structure, but two properties emerging from this structure and its ability to represent other structures in common space, recursive closure principles and interface conditions, i.e. the “mapping” relations between linguistic and non-linguistic inputs (in this case, visual ones). The answer I offered so far is virtually the same offered in Hauser, Chomsky & Fitch (2002), although the argument on which I have based my answer is relatively different, and perhaps places a greater emphasis on the interaction between recursion and interface conditions and their inherent “logicality”, as the kernel properties of FLN. This answer is also consistent with the considerations made by Pinker & Jackendoff (2005) and similarly-minded contributions to the FLN/FLB debate, since it suggests that Language and other modules of Cognition are quite more similar than it may appear at first glance.

The answer I offered so far might also offer an insight with respect to one important Linguistic problem, the emergence of Language from an evolutionary perspective. I tentatively suggest

---

22 This assumption leaves open the problem of “how many” models make up our Cognitive architecture, that are integrated in this model. I leave open this question, but I assume that we can leave out the “massive modularity” hypothesis typical of some evolutionary psychology literature. See Fodor (1998, 2000) for further discussion, however.
the following way to look at this problem, assuming in advance that what I shall say in this paragraph is nothing more than a wild (and perhaps wrong) conjecture. If we take a logical perspective and compress the evolutionary millennia into a conceptual space, then the emergence of a FLN kernel, from an evolutionary perspective, occurs when the integration of different types of information into a common format emerges.

Pushing this speculative line to its logical limit, we might assume that at some point, roughly 200,000 years ago, our ancestors were (suddenly?) able to compare what they saw with what they said, and vice-versa. From this initial step, which could be called the first step of induction, we might as well as assume that the \((n + 1)^{th}\) subsequent steps followed suit over the next few hundred of thousand years, taking shape as the unraveling of the Gamut of Languages we can currently attest in the world (e.g. Piattelli-Palmarini & Uriagereka 2005; for discussion and some references).

This single and yet very powerful emergent property could have arisen as the possibility (perhaps, necessity) to integrate different bits of information into an internally coherent (and perhaps optimal) representational/computational system. It might have been the case that Language arose as the “proof” that it is possible for different cognitive processes/modules to combine together into a unified, coherent cognitive architecture; thus it emerged entirely because of internal, structural pressures (again, Buzsáki 2006), although it become one tool (out of many) for humans to grasp and represent facts about the world, including the position of the things we see around us.

I shall leave these complex topics aside, and focus my attention back to our much more modest topic of discussion. Given the discussion I offered so far, I shall propose the following answer to the global research question:

\[ A-A: \text{The relation between spatial Vision and spatial Language is an isomorphism, as both models represent the same “amount” of information via different types of information;} \]

This answer sums up the results of this section. Note that, while in this section I have suggested
that Language, broadly defined, describes a model which includes a model of Vision as one of its proper parts, if we focus on spatial Language, then this portion of Language has the same structure and properties of Vision; consequently, it (correctly) appears that Vision and Language are more similar than it seems, as observed in much literature. Much more could be said about this topic, as the discussion I have offered in this section can only be thought as a very preliminary attempt at refining a notion of FLN (and FLB, for that matter) and its emergence, from the point of view of “Space”. Such discussion will be left for future research, as I shall move to the conclusions to this part, and move to less speculative topics in the remainder of the thesis.

2.5 Conclusions

In this chapter I have offered a novel proposal regarding the relation between Vision and Language with respect to “Space”, our understanding of things and their place in the world. I have argued that our spatial Vision and Language are quite abstract in nature, as they involve the processing of various types of information and their ability to individuate objects and the events they are involved in, as well as the “structural” relations that emerge from this process of individuation.

In doing so, I have offered a number of innovations on a number of closely related topics, including an updated review of the debate, a model-theoretic approach to Vision which covers data usually ignored in the debate on “Space” (via the VRT proposal), a novel DRT treatment of adpositions, as well as novel analysis of the Vision-Language interface, and what consequences this analysis has, for a general theory of the Language faculty.

The general picture I offered is one in which different models of cognitive processes can be formally defined in detail, and then embedded into a more general model of “knowledge”, modeled via a particular approach to Fodor’s (1975) notion of “Language of Thought”, DRT, and the “modularity of mind” hypothesis (Fodor 1983), although taken from a definitely more logical
stance (as in e.g. Crain & Khentzos 2008, 2009). Informally, Language represents a “neutral” logical space, a model of knowledge representation in which different concepts can be freely combined together, since they are already stripped of their “external” constraints when they are represented in the corresponding models (e.g. Asher & Pustejovsky 2004; Asher 2011 and references therein). A similar reasoning holds for the articulatory-perceptual side of Language. While we need to organize e.g. speech streams into coherent units, the result of this process must then be organized into a coherent structure of syllables, words and utterances, which may be organized according to processes and relations not unlike those of other modules, and which are then mapped onto concepts, and thus lose their “external” part. See Reiss (2007); Hale & Reiss (2008); Samuels (2009) for discussion.

In this regard, Language is the model that comes into being when all other “sub-models” expressed by other modules of Cognition are joined in a common logical space, and which might have emerged as the “projection” of different cognitive modules into this common logical space. With respect to this neutral logical space, then, spatial Language represents that fragment of space which represents spatial Vision, i.e. our abstract representation of things in the world, whether this representation is veridical or not. As a consequence, the proposals I have made here, although still very preliminary in their nature, can be seen as offering a better picture not only on what is the nature of spatial representations in Vision and Language, but also on the logic behind the processes by which we combine together these representations, and what this tells us about the general architecture of Mind and Language.

Once we have a solution to our first outstanding problem, we can concentrate on addressing the second and third problem: how to obtain a compositional approach to the syntax and semantics of Ps rather than the basically non-compositional one offered in this chapter (the second problem), and what is the psychological reality of such a fully compositional approach to Ps (the third problem). For this purpose, we will leave aside the more speculative topics of this part such as the nature of the Language faculty and its emergence, and focus in detail on the fine-grained analysis of English adpositions, thus moving our conceptual focus fully on Language, in partic-
ular on the (compositional) Syntax and Semantics of adpositions (part II), and their processing and acquisition in native speakers of English (part III).
Part II

The Grammar Of Adpositions And Spatial Sentences
Chapter 3

The Grammar of Adpositions and Spatial Sentences, Part I: Syntax

3.1 Introduction: the Problem of Ps and their Structure

In this chapter and the next I shall address the topic of the Syntax and the Semantics of spatial adpositions in more thorough detail, as the main topic of part II. The goal of the second part of the thesis is to offer a solution to the second outstanding problem, which can be captured via the following global research question:

Q-A: How adpositions can combine together and with other parts of Speech, and express distinct but related types of spatio-temporal relations;

As for the case of the first global research question, this question can be solved via a divide et impera strategy. In this case, I shall adopt a strategy which consists in first addressing the syntactic part of the problem in this chapter, and then address the semantic part in the next...

---

1 Portions of this chapter (i.e. part of section 3.2, section 3.3 and part of section 3.4) appear in a different format in Lingue & Linguaggio X(1), 64-87 (i.e. sections 2-4). The full reference for the published version is “Ursini, Francesco-Alessio. (2011a). On the Syntax and Semantics of Spanish Ser and Estar. Lingue & Linguaggio X(1), 57-87”. References to the paper are collected in the “Bibliography” section.

2 I shall use the less precise notion of “spatial adpositions” here onwards, when this use should not lead to confusion, following the intuition that our enriched notion of “Space” includes causal/temporal aspects as well.

81
The syntactic problem I shall address in this chapter consists of three sub-problems. The first problem consists in outlining what is known in the current literature about the syntactic properties of adpositions. The second, purely empirical, problem is that the treatment outlined in chapter 2 is in part non-compositional, and thus fails to capture important distinctions in the interpretation of adpositions: as it stands, it does not offer an improvement with respect to previous proposals on adpositions, just a (more precise) DRT-bound formalization. The third problem stems from the second problem: without a precise analysis of the internal syntactic structure of adpositions, we can only “approximate” how adpositions are merged with the rest of a clause.

These problems can be illustrated as follows. Take the following examples which present the syntactic phenomenon of argument demotion (e.g. den Dikken 2006; Svenonius 2010; and references therein):

(54) The boy sits behind the girl
(55) The girl is sitting behind the tree. The boy sits behind (the girl)
(56) The boy goes into the room
(57) The girl is waiting for the boy in the room. The boy goes in(to the room)

Under our current approach, the adposition behind in (54) is merged in a sentence as a single syntactic unit, which is then merged with the DP the girl. A similar consideration can be extended to (55). According to this approach, the example in (55) cannot be accounted for, since in this case only a “part” of this adposition is merged with the combination of subject DP and verb, the boy sits. The common phenomenon involving the omission of an adposition’s argument (here, the girl), argument demotion, is outside the reach of the theory. A similar consideration can be offered for the adpositions into and in in examples (56) and (57), with the further proviso that, in (57), part of the adposition is demoted too (i.e. the demoted “part” is -to the room). More in general, argument demotion can easily occur if the DP denoting the reference location
has been already introduced in discourse, as it can be retrieved from the explicit context (i.e. the first sentence in (55) and (57)).

The specific goal of this chapter is thus to solve three outstanding problems regarding the syntactic properties of adpositions, and thus offer an answer to each of the following questions:

Q-1: What do we know so far regarding the syntactic properties of adpositions;
Q-2: What is the syntactic structure of English adpositions, and the sentences they occur in;
Q-3: How this structure is realized as the result of syntactic, derivational and psychologically plausible processes;

In order to answer this questions, I will review the literature on adpositions couched within the syntactic framework of Generative Syntax, in its current incarnation known as the Minimalist Programme (e.g. Chomsky 1995, 2004), and incorporate certain proposals within this programme in our fragment of DRT. There are at least two reasons for choosing this theory over other syntactic frameworks (e.g. constructionist grammars such as Tomasello 2003; Goldberg 2006; Culicover & Jackendoff 2005; HSPG, Sag, Wasow & Bender 2003; inter alia).

The first reason is that these theories, behind inherently non-compositional, face the same empirical problems as our current approach: the examples in (54)-(57) thus represent an obstacle for these theories as well. The second reason is that these theories cannot correctly account the following phenomena involving adpositions, among others: locative inversion (e.g. den Dikken 2006:ch.2, 2008) and preposition stranding (e.g. Koopman 2000). Look at the examples:

(58) In front of the car, the men are having a beer
(59) The city we flew over is Laputa

In (58), the Prepositional Phrase (PP) in front of the car occurs “to the left” the rest of the sentence it combines with, unlike in standard sentences. In (59), the DP the city occurs “to the left” the rest of the sentence, and not as a syntactic complement of the adposition over. Since these frameworks assume a non-compositional approach to the syntax of adpositions and the
sentences they are part of, they cannot account how PPs, or the acting as complement of DPs, can be combined with other parts of speech in such a flexible way. Anticipating matters a bit, I shall offer a solution to these problems by incorporating the analysis of adpositions structure offered in Hale & Keyser (2002) (i.e. the “P-within-P” hypothesis), and the Parser Is Grammar (henceforth: PIG) theory of syntactic derivations offered in Phillips (1996) in my DRT fragment. In this way I will be able to spell out which minimal syntactic units adpositions are made of, and how they are merged into sentences in a piece-meal fashion, thus offering a psychologically plausible model of the syntactic process.

Before focusing on the analysis, I shall briefly introduce some basic (generative) syntactic notions that I shall use for the reminder of the chapter and the thesis. First, I shall assume that Minimalism’s \textit{merge} is the same operation defined in our DRT calculus, and thus that both operations can be seen as an instance of set union, represented via the symbol “+”. I will thus assume that \textit{merge} is basically an instance of set union (cf. Chomsky 1999:2 definition, roughly representable as: \textit{merge}(X, Y) = \{X, Y\}, \textit{X}, \textit{Y} being sets). Second, I will follow cartographic approaches (e.g. Shlonsky 2010), rather than minimalist assumptions (e.g. Kayne 1994; Chomsky 1995), and use the X-bar format to represent standard phrase structure. Consequently, I will represent the merge of a \textit{head} and a \textit{complement} as a \textit{bar-constituent} (e.g. we have \textit{X} + \textit{YP} = \textit{X’}), and the merge of a \textit{bar-constituent} and a specifier as a phrase (i.e. we have \textit{XP} + \textit{YP’} = \textit{YP}). Third, I shall use the standard linear format of representations (i.e. “square brackets”) rather than trees, given that this way of representing syntactic structures is virtually identical with the syntactic approach used so far in our DRT fragment.

This chapter is organized as follows. In section 3.2, I shall introduce basic facts about adpositions and their structure, thus offering an answer to the first research question. In section 3.3, I shall present current proposals regarding the structure of adpositions, and offer a reason on why Hale & Keyser’s approach is more adequate, thus offering an answer to the second research question. In section 3.4, I shall offer an integration of the PIG proposal in our DRT fragment, and thus offer a fully compositional, derivational approach to the syntactic structure of Ps.
3.2 Previous Literature, and Basic Facts about Ps

Spatial relations are usually realized by different parts of speech. In English, they are realized for the most part by *Prepositions* (e.g. *to, from, at, in front of, etc.*) and *Particles* (e.g. *up as in go up*), but also by *verbal Prefixes* (e.g. *in- as in in-sert*) (e.g. Svenonius 2004, 2008, 2010). A fourth possibility is that they are expressed via *case morphology*, as Finnish *talo-ssa* (i.e. lit. “sea-at”, e.g. Caha 20093), and a fifth possibility is that they are expressed by *Postpositions*, or adpositions following their complement (e.g. Japanese, Asbury 2006; *inter alia*). Prepositions and postpositions are usually known as *adpositions*, and are often considered to be the more representative member of this category (e.g. Svenonius 2004). These parts of speech have been suggested to form a uniform category, usually labeled as “P”, given their ability to express spatial information (e.g. Asbury 2006). Since I will maintain my focus on English, I shall mostly focus on prepositions, since they are the most common and perhaps the most prototypical instance of “spatial P”; hence, I shall analyze in detail their syntactic properties in the remainder of the chapter. I shall sloppily use the label “P” to indicate only adpositions, when this should not generate confusion.

I shall start by discussing some basic cross-theoretical concepts, which were introduced in Talmy’s “Cognitive Linguistics” framework, before focusing on generative-based proposals. Talmy (1978, 2000) introduced a elements of Gestalt theory to the semantics of Natural Language. For instance, the two notions of *figure* and *ground*, as we have seen in 2, section 2.2.2, respectively express the object which is the focus of attention in a visual scenario and the background in which this object is situated. In Talmy’s approach, they are respectively used to mark the noun phrase that refers to the entity made salient by a sentence, and the noun phrase that refers to some other “backgrounded” entity.

An example is the following:

---

3In this specific case, the root *talo-* is combined with the allative case marker *-ssa*, roughly translatable as “at”. It is safe to say that, for the rest of this thesis, no further considerations on case will be necessary. See e.g. Asbury *et al.* 2006.
a. Mario is under the bed

b. [[figure] Verb Satellite [ground]]

Assume that the interpretation of the sentence in (60) gives us a set of coordinates by which we can find Mario, if we start our “query” from the bed. In this case, the bed acts as the centre of the reference system (or “ground”), while Mario is the object which we find at the end of this “visual” query (or “figure”). The adposition under (labeled as “Satellite” in his framework), together with the copula is, this expresses the type or (spatial) relation which holds between the two entities. Specifically, the figure is in the region which is defined via the negative vertical axis of the ground (under), and is located in this region at the moment of the utterance (the present form is). Talmy also introduced the notion of satellite- vs. verb-framed languages. The former type of language includes English, in which most of the spatial information is expressed by satellites (e.g. a P such as across), whereas verb-framed languages convey this information by combining Satellites with verbal constituents (e.g. a verb such as to cross). He assumed that “spatial” verbs in verb-framed languages represent a typical instance of the phenomenon of conflation, defined as the possibility that a verb forms a single unit with the satellite (and in general, of any two syntactic units), hence conflating different levels of information in a single syntactic unit. For the remainder of the thesis, I shall adopt these two labels to refer to the entities involved in spatial sentences, for the sake of clarity.

A rather similar analysis can be found in other syntactic and typological theories (e.g. Croft 1991; Stassen 1997; Lieber 2004; inter alia). Within the generative framework, these notions have been borrowed in various measure. One influential proposal is Jackendoff (1983, 1990) and his “Conceptual Semantics” (CS), which I have briefly mentioned in chapter 1. Jackendoff argues that the principal part of speech expressing spatial information is the (syntactic) head “P”, with “P” standing for “Preposition”. Jackendoff further argues that Ps can be decomposed in two distinct syntactic heads. One expresses a PLACE concept, a 1-argument function that expresses a spatial region with respect to a ground, while the PATH function is a relation (2-argument
function) between the theme and the PLACE output, i.e. a PLACE-function combined with the ground.

A variant of Jackendoff’s classical example and its structure is in (61):

(61)  

a. The mouse appears from under the table  

b. \([\text{event} \text{Appear}[[\text{Theme} \text{mouse}] \text{Path} \text{from} \text{place} \text{under} [\text{ground} \text{the table}]]]]\)

In words, the path-function *from* takes the “theme” (Talmy’s “figure”) and relates it to the place-function *under*, assessing that it is the initial point or region in which the theme is located, during an event of appearing. The place-function, in turn, takes a noun phrase such as *the table* and returns the region (of the table) that acts as a reference point in the event of appearing. Jackendoff then proposes that these two concepts are usually mapped onto one syntactic constituent, that of Ps, which thus acts as the main predicate of a sentence, since it expresses a relation between figure/theme and ground (or, more appropriately, the relevant “region” of a ground). Examples of this conflation phenomenon include *into*, *onto*, and similar others. Verbs, on the other hand, act as modifiers to the PPs headed by a P, as the example in (55) shows.

Works by Henk Riemsdijk introduced these ideas in generative syntax in a somewhat different manner. For instance, van Riemsdijk (1978, 1990, 1998) proposed that the concept PATH and PLACE are mapped onto the heads “Pdir” and “Ploc” (“directional” and “locative” component, respectively), with Pdir acting as the complement of a verb, and taking Ploc as its complement. Further works on prepositions, such as e.g. Wunderlich (1991, 1993); Nam (1995); Kracht (2002, 2004); Riemsdijk & Huysbregts (2007) mainly follow this tack, and thus contend that the general syntactic structure of Ps (in particular, adpositions) is approximately the following 4:

(62) \([\text{VP}[\text{DP The mouse}][\text{VP appears[PdirP from[PlocP under[DP the table]]]]]}\)

The main innovation in (62) is that the two spatial heads are considered as possibly distinct syntactic positions, since they can be realized by two different adpositions (as *from* and *under* in the

4Here and in the rest of the chapter, I shall gloss over the fine-grained structure of verbal projections, i.e. whether a VP combines with other functional heads expressing temporal or aspectual information, and located higher in the clausal structure (e.g. the \(v\) (“small \(v\)”) head). This simplification is not crucial to the discussion of our data.
example); Ps such as *into*, instead, may consist in the conflation of the “PDir” and “Ploc” heads, as also assumed in Jackendoff’s theory. Furthermore, the verb is considered as the main predicate in the sentence, whereas Ps when they are combined in one position, act as the complement of the verb.

An alternative formulation of the structure of Ps, within generative syntax, can be found in the work of Emonds (1976, 1985). Emonds suggests that Ps pattern with other parts of speech, usually known as *complementizers* in the literature (i.e. the category “C”), in introducing a relation between (parts of) sentences, whether it can be treated as a form of subordination or coordination. Spatial Ps such as *in, to* and several others thus form a wider category, including modal “relators” such as *that, or relative pronouns* such as *who, where* (e.g. Caponigro 2003 and references therein), temporal Ps (e.g. *before, after* and so on: Kamp *et al.* 2005:ch.3), and the “Boolean” Ps *and, or*, as well as of (Emonds 1985:ch.10). Emonds thus suggests that Ps/Cs are a general syntactic category, which at its core has a purely predicative function, and thus denotes an abstract relation between entities denoted by the specifier and complement it combines with.

Emonds also proposes that morphological features combine with this abstract syntactic head to spell out a more restricted “sort” of relation. For instance, a spatial P such as *in front of* may have been specified with features such as [±locative], [±directional]. He also suggests that morphological features are “invisible” to syntactic derivations, and thus that they do not project their own position in the clausal structure. The syntactic structure proposed by Emonds (1985) for e.g. “spatial” sentences is the following:

(63) The ball is in front of the table
(64)  \([DP]V[P_{[+\text{horizontal, locative}]}[DP]]\)

The structure in (64) shows that in front of can be represented as an abstract P, which includes features such as “+horizontal”; “+locative” represented as subscripts to the P head, which is then represented as a single syntactic element. Such a P head acts as a complement of the copula is, represented here as the category V. The main difference between this approach and the
“standard” Jackendovian approach is thus that Ps are not necessarily divided in two heads, often conflated together; hence, the two approaches offer a convergent picture regarding the structure of Ps.

At this point, we are able to offer an answer to the first research question of this chapter, which is as follows:

A-1: Adpositions, as the chief members of the category “P”, are constituted of two syntactic heads often conflated together, usually labeled as “PathP” (“DirP”) and “PlaceP” (“LocP”).

This answer offers a relatively straightforward picture regarding Ps and their position within clausal architecture, which also suggests a transparent relation between syntactic categories and semantic content. The “Path” head denotes a certain notion of “path”, a stretch of space that connects figure and ground. The “Place” head denotes a certain notion of position, a location occupied by a figure with respect to a ground. It apparently offers an answer to the second research question as well, since it spells out what are the “basic” constituents that make up a P, and how they are combined with other parts of speech to form “spatial” sentences. However, recent research on the syntax of Ps has suggested that this picture is perhaps not appropriate, with distinct proposals suggesting that it is either too coarse-grained, or to an extent inappropriate. The next section will discuss these proposals.

3.3 The Structure of Ps, I: a Proposal, and some Empirical Coverage

There are at least two main proposals that offer an alternative view of Ps than the one offered in the previous section. A first is the Cartographic approach, which proposes a more complex structure for Ps by introducing several other positions labeled as “Ps”. A second is the “theory of argument structure” introduced in Hale & Keyser (2002), which instead offers a simpler structure
for Ps, known as the *P-within-P* hypothesis. I shall clarify their assumptions in the remainder of the section.

The Cartographic approach is an approach within the minimalist programme that seeks to develop a thorough “map” of functional categories\(^5\) and their position in clausal structure, such as Ps, adverbs, question and discourse particles, and so on. In a manner similar to classical structuralist accounts, it seeks to study which are the possible positions for these different parts of speech across languages, as well as identifying an ideal list of all parts of speech that can find a position in clausal structure. Consequently, it studies in detail the properties of *functional elements* of clausal structure and how these elements express grammatical properties (“features”) ascribed to the basic content expressed by *lexical categories* (e.g. verbs, nouns, adjectives and prepositions). It proposes that such functional elements are arranged in a fine-grained *Functional SEQuence* of syntactic heads, which are stacked on a core lexical component (*FSEQ*, e.g. Cinque 1999).

One further assumption in this framework is that morphological features may appear as syntactic units and project their own syntactic structure, so they may combine together as a result of syntactic processes. Hence, for potentially each morphological feature, one should postulate a corresponding syntactic head (and position) (Shlonsky 2010:422). This entails that for Ps such as *in front of*, each morpheme (i.e. *in*, *front* and *of*) should project its own syntactic head, whereas under and similar other Ps correspond to only one position. The exact status and “amount” of elements constituting the fine-grained structure of Ps is subject of considerable (and ongoing) debate.

In the specific case of Ps, it has been proposed that heads such as “Pdir” and “Ploc” represent functional features that are stacked on a basic lexical element expressing a general notion of location (e.g. English word “place”), which is usually labeled as p, and appears to be overtly realized in the P structure of languages such as Persian or Japanese (e.g. Pantcheva 2008; Cinque

\(^5\)Functional categories are those parts of speech which are defined as forming a “closed class”, as opposed to “open class”, lexical categories (e.g. Nouns). For instance, Ps across languages tend to be few in number; new Ps cannot be added in a principled way to a Language’s lexicon (e.g. via compounding, borrowing, etc.). Although there is a long and interesting debate on whether Ps are functional or lexical, this debate is not crucial to our analysis of Ps.
2010). It has also been proposed that several other distinct P positions can be identified, within the “P field”, the hypothesized sequence of heads falling under the P macro-category (see e.g. Asbury et al. 2008; Cinque & Rizzi 2010). Influential examples include Koopman (2000); den Dikken (2010); Svenonius (2010). Here I shall present the basic intuition behind this syntactic theory via a relatively simple example, based on Svenonius (2006):

(65) to in front of the car

(66) [\text{to}[\text{in}[\text{front}[\text{of}[\text{the car}]]]]

(67) [\text{Path}[\text{Place}[\text{Axpart}[\text{Kase}[\text{DP}]]]]]

The two heads “Path” and “Place” express, as the names suggest, the syntactic realization of the PATH and PLACE functions. The heads “Axpart” and “Kase” are two “new” heads that are suggested to project from \text{front} and \text{of}, respectively. Intuitively, these constituents respectively express the axial orientation of the figure with respect to the ground, and the “type” of basic relation holding these two entities, on which other constituents are stacked, contributing a form of semantic “restriction” on this more abstract relation.

While morphologically complex Ps such as \text{to in front of} seem to offer evidence in support of such a cartographic approach, less complex Ps seem not to be easily analyzable in this theory. For instance, a P such as \text{under} is assumed to correspond only to one position, “Axpart”, whereas a P such as in corresponds to a “Ploc” position. Consequently, these Ps realize only part of the rich structure assumed to exist for the “P field”. However, a common assumption is also that single morphemes in highly inflecting languages may be represented as projecting onto syntactic structure: one such case is the treatment of spatial cases morphology in Finnish and similar other Languages (see e.g. Caha 2009; Kracht 2008; Pantcheva 2006, 2008, 2010 for discussion). Consequently, case morphemes in case-marked nouns such as e.g. \text{talo-ssa} have been analyzed as projecting at least two syntactic positions (e.g. “Ploc” and “Pdir”), conflated in one case marker: we have \text{talo-s-sa}, with -s and -sa as distinct morphemes).\footnote{A brief discussion in Svenonius (2006:91-94) suggests that, to a limited extent, “axial” Ps such as beside appear
Although cartographic works offer a wealth of data and a fine-grained analysis of Ps and their functional structure, the assumption of a FSEQ structure for functional elements is far from uncontroversial. If one assumes the existence of a fixed sequence of functional heads, then the existence of mono-morphemic Ps and Ps having an internal order that is not in line with this FSEQ (e.g. up to or from under, or the data from case-based languages) appears to be mysterious. In general, the FSEQ appears to have little predictive power, and tends to run into “empirical trouble” whenever data suggest that a Language lacks a rigid ordering of heads/syntactic positions and respective labels, such as the English examples just discussed (see e.g. Collins 2002; Nilsen 2003; Boeckx 2010 for discussion). I shall thus leave aside this approach and discuss a more flexible approach to syntactic structure, Hale & Keyser (2002) and their theory of argument structure.

The basic proposal in Hale & Keyser (2002) is the that the four lexical categories (i.e. nouns, adjectives, prepositions, verbs) represent four possible configurations by which a head can project its syntactic properties, chiefly its ability to combine with other constituents (i.e. other phrases acting as specifiers and complements). These four configurations are abstract in nature, so different languages may have different superficial categories to realize them. For instance, a certain configuration may be realized by Verbs in English, and by Ps in Italian. Hale & Keyser (2002) call these configurations “(a)-type”, “(b)-type”, “(c)-type” and “(d)-type”.

Two important types, for our discussion, are the (b)-type and (d)-type, which are presented in (68) and (69) (from Hale & Keyser 2002:13):

\[
(68) \quad [\text{Head}][\text{Specifier}][\text{HeadHead}][\text{Complement}] \quad (\text{(b)-type, from Hale & Keyser 2002 ex. (24)})
\]

\[
(69) \quad [\text{HeadHead}] \quad (\text{(d)-type, from Hale & Keyser 2002 ex. (24)})
\]

to be the result of a past process of conflation (i.e. the conflation of Old English be- and -sidan, respectively “by” and “side”). Whether this allows the actual decomposition of beside (and other similar Ps) into distinct morphemes is a complex matter that bears no crucial relevance to our discussion, so I shall leave the topic aside (but see e.g. Svenonius 2006, 2010 for a thorough discussion).

Subsequent proposals on argument structure include, among others, den Dikken (2006, 2008); Ramchand (2008). The proposal in den Dikken’s works is to an extent similar to the one advocated here, so the differences are immaterial for the discussion. The one in Ramchand (2008) proposes a radical form of the Cartographic approach (“first phase syntax”), which can be rejected on the same grounds of the Cartographic approach.
(70) Mario is blonde

(71) \[VP[DP\text{Mario}]is[AP\text{blonde}]\]

(72) Mario calls (The boy in the garden)

(73) \[PP[DP\text{the boy}]in[DP\text{the garden}]\]

(74) The boy

(75) \[[DP\text{The boy}]\]

(76) Mario goes into the tavern

(77) The ball appears from under the table

(78) Mario waits at the end of the tunnel

These configurations should be read as follows. A (b)-type configuration involves a head that combines with a complement and with a specifier, for instance two DPs, and determines the type of resulting phrase. For instance, the sentence in (70) involves a copula construction in which the copula combines with a DP (i.e. \textit{Mario}) and an A(djective)P (i.e. \textit{blonde}) to form a full sentence (here, a VP). The adjective \textit{blonde}, in this case, instantiates a (d)-type configuration, since it appears as a phrase without its own specifier or complement.

Ds and DPs are also indirectly treated as realizing one of these four possible configurations in the nominal domain, to an extent being considered as the “quantified” Noun Phrases (NPs) of traditional syntactic approaches (e.g. Emonds 1985; Bruening 2009). Definite description DPs, such as \textit{the boy} in (74), usually receive a “referential” interpretation and denote unique referents in discourse, like proper nouns (e.g. \textit{Mario}). Proper nouns, in turn, can be seen as instantiations of DPs, as they do not combine with a distinct specifier and a distinct complement. For this reason, definite DPs can be treated as instantiating the (d)-type. Although quite coarse-grained, this syntactic treatment of DPs is nevertheless appropriate for our purposes, since we shall mostly be concerned with Ps (but see Hale & Keyser 2002:ch.1 and ch.4 for discussion). I shall discuss D cases within a few paragraphs.
The analysis of these sentences allows us to define the basic components we need to account for the structure of spatial sentences, such as the sentences in (76), (77) and (78). In Hale & Keyser (2002) it is assumed that spatial sentences involve the combination of two (b)-type constituents, instantiated by Vs and Ps. Whenever a verb takes a PP complement, the PP may in turn combine with two DPs (as in (72) and its corresponding structure in (73)), or with another PP and a DP. When a PP acts as the specifier of complement of another P, it will instantiate the (d)-type, since it will correspond to a phrase without its own specifier and complement (e.g. from in from under the table, example (77)). An example of this second structure is the PP in (77), which receives the following structure (adapted from Hale & Kayser 2002:ch.9):

(79) \[\text{[from[under[the bed]]]}\]

(80) \[\text{[PP[PP[P[DP]]]]}\]

The resulting PP thus becomes the complement of the verb appears, and denotes the possible “termination” or “location” of the event structure denoted by a sentence, in combination with the verb. Importantly, Hale & Keyser (2002) suggest that Ps in structures such as (79) can conflate, as it appears to be the case for Ps such as into, onto and similar others. One consequence of this assumption is that, for Ps expressing simple location, the head capturing the underlying relation may apparently be omitted. For instance, Hale & Keyser (2002:ch.8) suggest that Ps such as in or under may be understood as Ps appearing in the specifier position of a possibly phonologically null head expressing a “location” relation, represented as (at).

They also suggest that Ps such as in front of or ahead of may be understood as being the conflation of these two heads into a single constituent, with of being an instantiation of an abstract P (cf. also the “Relator” head of den Dikken 2006). Consequently, in this theory all the lexical items belonging to the category of Ps share the same underlying structure, although the phonological realization of this structure may vary to some extent, depending on the lexical item at hand. While in front of and behind may appear to be structurally different, their underlying syntactic structure is assumed to be the same. I propose to account for the internal structure of
Ps in more precise detail, via two rather simple assumptions.

First, I assume that adpositions occurring in specifier position are realizations of a (d)-type, much like DPs (e.g. *the boy* in (73)). Consequently, I shall represent such elements as Phrases, whether they are composed of one or more words. Examples include PPs such as *in* in (76) and *from* in (77). In the case of more complex combinations of Ps, such as *in front* in *(in front of)*, that a phonologically null head will combine with the two PPs *in* and *front*, to form a PP that ends up in the specifier position of the head *of*. For Ps such as *at the end of*, found in examples such as (78) and similar others (e.g. *in the front of*, cf. Svenonius 2006) the definite article *the* appears to overtly realize the P head in the “P-within-P” structure. Since I have rejected the FSEQ assumption and its assumptions regarding rigid positions in clausal structure, there are no specific reasons to assume that this definite article is in the “wrong” syntactic context. On the contrary, the definite article offers indirect evidence that the “P-within-P” hypothesis is sound, since it predicts the possible realization of a head in this “position”, although it superficially takes the form of a “nominal” element, in this case.

This assumption is simply a natural extension of the “P-within-P” hypothesis, which allows us to capture the internal structure of more complex Ps in a rather principled way. Other P structures can be represented as well. The structure on the left side, in (82), represents the structure associated to a more complex P such as *at the end of*, but it can also be instantiated by Ps such as *in front of*, modulo a phonologically null head between *in* and *front* (i.e. we have *in (P) front*). “Simpler” Ps such as *ahead of* can be represented via the “simpler” central structure, while “basic” Ps such as *behind* can be represented via the basic structure on the right side of (82). As

---

8Cross-linguistically, there are several languages in which Ps and Ds intermingle in rather interesting ways. The distribution of other quantifier-like elements in this position appears to be limited, however (e.g. we can have *in every front of the car*, but apparently not *in most front of the car*). This aspect is immaterial, for our discussion. See Svenonius (2006); Asbury (2008:ch.4) for further discussion.
we can see, the possibility that some Ps may have “more” structure than other Ps corresponds to the instantiation of recursion within this structure, as per assumptions; and the possibility that some Ps appear to only involve one position corresponds to the absence of some phonologically realized material. With this approach, we can easily account the apparent structural differences among Ps as combinatoric possibilities stemming from one syntactic structure, a welcome result.

Second, since I dropped the FSEQ assumption, I assume that the exact “label” of these constituents is not crucial: if we consider in- as an instance of a Ploc category, and from as one of a Pdir category, then we would have a problem of linearization, i.e. to establish a single linear order of heads, given the examples in (76) and (77). I will thus assume that not only different types may be realized by different morpho-syntactic categories, but also that single lexical items may display this pattern (cf. Hale & Keyser 2002:ch.2 on de-verbal nouns, inter alia). The intuition behind these two assumptions is quite simple: complex adpositions may include simple adpositions as part of their structure, so the syntactic role of these simple Ps is not “fixed”, but determined by the syntactic environment they occur in. Consequently, the use of rigid labels to designate fixed position is unnecessary. Whether in is a head or a phrase, as a basic lexical item, depends from the elements it combines with in a sentence, whether they are phrases (e.g. in the garden) or heads (e.g. in front of). Similarly, whether the instantiates a P-like or a D-like head depends on the syntactic environment it occurs in, rather than its “position”.

I shall thus assume that a standard spatial sentence such as (76) and (77) has the following structure:

\[ (83) \quad [\text{VP}[\text{DP}[\text{V}[\text{PP}[\text{PP}[\text{P}[\text{DP}]]]]]] \]

In words, (83) says that a P expresses a relation between a DP (the “ground” entity), and a certain spatial configuration, expressed by the PP in specifier position, which may be endowed with its own complex structure. The PP resulting from the merge of this P with its argument phrases will act as the complement of a V, a head which also combines with the “figure” DP. The combination of V with its argument phrases will thus form a VP. The PP in specifier position may have its own internal structure, which is here omitted for the sake of clarity.
Part of this structure needs not phonologically realized. Recall that a P can be omitted, as we have seen. For instance, the P to in (76) captures a relation between a certain location, denoted by the DP the tavern, and the inner region of this entity, denoted by in- (as part of the P into). A similar reasoning can be extended to the complex P from under, which is assigned the structure in (80). So, we may safely assume that the following sentences all share the same underlying syntactic representation:

(84) Mario is in (P) the bedroom  
(85) Mario sleeps under (P) the bed  
(86) Mario jumps into the lake  
(87) Mario arrives from the lake

The sentences in (84)-(87) thus differ not in how the single constituents combine to form a sentence, but rather in the “sort” of semantic relations between events that they introduce in a sentence, e.g. whether these relations convey an implicit relation between locations (e.g. in and under, plus a (P) head); two complementary orders of events (e.g. into vs. from), or they are based on the lexical content associated to these relations (e.g. the difference between runs and jumps).

By this point, we may offer an answer to the second research question:

Q-2: Ps either instantiate syntactic heads or phrases, and thus include other constituents in their structure, or be part of other constituents;

This answer captures the intuition that Ps do not represent a fixed position or category in clausal structure, and can thus occur as either predicative elements (that is, heads) or as argument-like heads (that is, phrases). Consequently, Ps can appear to occur at any point in a sentence, or be “split” into several distinct units. By adopting Hale & Keyser’s proposal, we can capture the structure of Ps such as into, from under, in front of and underneath in a relatively straightforward way. If we were to adopt the Cartographic approach, these data would be outside our reach, since
we would need to merge Ps (and PPs) in a fixed order, contrary to evidence. The next section will present how these insights can be integrated in our DRT fragment.

3.4 The Structure of Ps: a Derivational Account, and more Empirical Coverage

In chapter 2, I followed the standard DRT approach to syntactic derivations and semantic interpretation, and assumed that syntactic derivations proceed “left-to-right”\(^9\). Informally, I assumed that the grammar generates full sentences by first combining subject and verb and thus combining the interpretation (in the form of DRSs) into a single constituent, itself a DRS; and that the same procedure can be applied to combine the result of this first step with the direct object. Informally, in DRT we first merge *John* and *saw* to form the temporary constituent *John saw*, then we merge this constituent with *Mario*, to form the sentence *John saw Mario*.

There is a rich literature offering ample evidence that sentence production and sentence processing, as inherently psychological processes, proceed in such a “left-to-right” fashion (e.g. Crain & Steedman 1985; Altmann & Steedman 1988; Frazier 1987; Frazier & Reyner 1990; Levelt 1993; Grosz *et al.* 1995; Tanenhaus *et al.* 1995; Gibson *et al.* 1996; Phillips 1996; Phillips & Gibson 1997; Braine & O’Brien 1998; Poesio *et al.* 2000, 2004; Altmann 2001; Mulders 2002; Reinhart 2006; Sadeh-Lecht 2007; Stenning & van Lambalgen 2008). However, the generative literature I reviewed so far assumes that syntactic derivations proceed in the opposite direction, e.g. that a verb V combines with its direct object (e.g. a DP) to form a verb phrase (a VP), and then this verb phrase is combined with a subject, to form a full sentence. Informally, we first merge *Mario* and *saw* to form *saw Mario*, then we merge *John* to form *John saw Mario*.

There is an obvious tension between the approach to syntactic derivations offered in these two (families of) frameworks. This tension disappears once we compare DRT to one frame-\(^9\)This informal label refers to the fact that, if we write down sentences in their step-by-step realization, it will appear that we are writing them from left to right. Since this characterization does not take in consideration the hierarchical structure of sentences, I shall just use it as an intuitive approximation of the syntactic processes it refers to.
work within the minimalist programme, the PIG proposal of Phillips (1996, 2003, 2006), which precisely offers a “left-to-right” approach. Before introducing this framework, let me note that a proposal with a somewhat similar philosophy can be found in Pesetsky’s *Zero Syntax* (e.g. Pesetsky 1995). Although the two proposals differ in a quite substantial manner in their assumptions regarding various syntactic phenomena, their treatment of the topic(s) discussed here is substantially the same, so I shall focus my expository efforts on Phillips’ theory.\(^{10}\)

Phillips assumes that the parsing and production of sentences can (and should) be thought as the same underlying process, since they involve the merging of different constituents into a unified syntactic structure, in an apparently sequential manner. He also assumes that a driving principle of syntactic derivations is “merge right”: constituents are (asymmetrically) merged on the right as soon as they are introduced in a derivation (or parsed). So, the DP *John* and the V *saw* form the V’ *John saw*; then, the merging of the DP *Luigi* allows forming the VP *John saw Luigi*. Given this principle, at some point in the derivation a constituent like *John saw* can combine with another constituent of the same type via coordination, yielding a structures such as the following:

(88) (John saw and Mary greeted) Luigi in the garden

In (88), the bracketed part of the sentence captures a single syntactic unit, since the coordination and takes two verb phrases (*John saw, Mary greeted*) and merges them into one constituent, *John saw and Mary greeted*. This complex constituent is then merged another DP, in this case *Luigi*. Other constituents may be similarly merged: for instance, the PP in the garden can be merged to the right of the DP *Luigi*, thus forming the sentence in (88).

The derivation resulting from this process thus builds a tree in a left-to-right, top-down and overall sequential fashion, in a manner substantially similar to the derivations proposed in chapter 2. For instance, Phillips (2006) proposes the following cyclic derivation for the sentence

\(^{10}\)Syntactic derivations in “Zero Syntax” are distinguished between *Layered syntax* and *Cascade syntax* derivations, occurring at the same derivational time. Cascade syntax derivations are left-branching, allow to analyze the “standard” structure and properties of syntactic derivations (e.g. binding), and are virtually equivalent to Phillips’ model of derivations. Layered syntax derivations are right-branching, and allow to analyze the patterns of a limited range of phenomena (e.g. VP-fronting) which are not important, for our discussion.
in (89):

(89) Wallace saw Gromit in the kitchen  (=\(12\), Phillips 2006)

(90) \[\begin{array}{ll}
   a. & [VP[DPWallace][Vsaw]] \\
   b. & [VP[DPWallace][Vsaw[DPGromit]]] \\
   c. & [VP[DPWallace][Vsaw[PP[DPGromit][in]]]] \\
   d. & [VP[DPWallace][Vsaw[PP[DPGromit][in][DPthe kitchen]]]]
\end{array}\]

The derivation I offer in (90) minimally differs from the one in Phillips’ work, since I do not introduce some aspects (i.e. the presence of inflectional heads) in the derivation which are not crucial to our discussion. In order to make the presentation compact, I have also written the type of constituent as a subscript of the corresponding brackets. The “merge right” principle can be informally described as “expanding” a syntactic derivation on its right, in our linear representational format. Note, in step \(b\), that the DP \( \text{Gromit} \) is temporarily attached to the verb \( \text{saw} \), so the syntactic structure formed at this point of the derivation corresponds to a V’ constituent.

With the merging of the P in, in \(c\), this structure is in a sense “destroyed”, and a new one is formed, a VP (i.e. \( \text{Wallace saw Gromit in the kitchen} \)). As Phillips argues, intermediate steps of a syntactic derivation may involve structures that are not part of the final syntactic structure and which are thus changed over the derivational time (e.g. step \(c\), \( \text{Wallace saw Gromit in} \)), but the result of this derivational process must offer, as a result the standard syntactic structure assigned to a well-formed sentence.

This theory can accurately describe many complex syntactic data, and give a principled account of phenomena such as verb ellipsis, pronoun binding, and many other similar data: however, its chief importance for our topic of discussion lies in its elegant way of accounting for syntactic derivations as a dynamic process, and thus in allowing us to bridge our initial DRT approach with the analysis of Ps offered in the previous section. This bridging process is defined as follows.
I shall rehearse a number of notions from chapter 2, section 2.3.5, to make the presentation of the argument clear. A DRT calculus corresponds to a model of language $L = (D, +, \leq)$, i.e. a model which includes a set of expressions (DRSs), a syntactic operation (merge, “+”) and a semantic relation (the part-of relation “$\leq$”). Furthermore, the dynamic counterpart of this formal language, by which we represent how we combine DRSs together, is the duple $L_d = (I, D)$, a pairing of derivational intervals at which syntactic processes occur and the objects they combine. The minimal assumptions I shall make, in order to properly “import” our syntactic data within DRT, are the following.

I shall first introduce some assumptions regarding the categorial status of syntactic constituents, by using some (very) basic notions of Combinatorial Categorial Grammars, since they will easily allow us to represent the syntactic processes at hand, without the need to make any supplementary assumptions with respect to our DRT approach. My assumptions are not particularly theory-specific, so they are compatible with several “dialects”, such as e.g. Jacobson (1999, 2004); Steedman (2000); Jäger (2005); Veermat (2005); Moortgat (2010); and DRT-based approaches such as e.g. Zeevat (1989); Muskens (1994, 1996); van Eijk & Kamp (1997); Bos (2010).

I shall assume that the set $D$ of DRSs consists of a set of atomic syntactic types, defined as $D = \{n, v, p\}$. I shall assume that the type $p$ corresponds to Hale & Keyser’s (b)-type (i.e. the type of heads), and thus that V(erb)s and Ps receive this type accordingly. DPs correspond to the type $n$, the type of DPs, since for the most part I shall treat these constituents as argument-like elements. I shall not use the type $v$ in my derivations, which is informally definable as the type of one-place predicates. I shall leave open the option that this type can correspond to one of the other configurations found in Hale & Keyser’s proposal. These types may be reduced to a more primitive, universal type, that of (basic) DRSs, out of which these three “basic” types are constructed in a recursive fashion (call it “x”; e.g. Partee 2009 and references therein). However, the use of these three primitive types allows us to capture the intuition that, in general, lexical items may represent the three atomic constituents in sentences: heads, modifiers, arguments.
I shall then assume that the duple $LEX = \langle D, + \rangle$ recursively defines all (and only) the possible combinations of syntactic constituents into more complex constituents, the basic Lexicon of our DRT. For instance, the type $pp$ of prepositional phrases can be defined as: $pp = ((n) + p + (n))$, which reads: a PP is the merge of a preposition and two DPs, in a sequential manner. Since the principle “merge right” forms constituents on the right, the “final” two constituents in a derivation always form an X-bar constituent via associativity, i.e. we have $(n+p)+n=n+(p+n)$, with $(p+n)=p'$. The “destruction” and “creation” of syntactic structures is a reflection of one well-known property of merge, associativity.

The general rule for complex types is defined as: if $a$ is a type and $b$ is a type, then $(a+b)$ is a type. Under this definition, both bar- and phrase-elements are shortcuts to represent syntactic structures as the result of syntactic processes, i.e. we have $(a+b)=b'$ and $(b'+a)=bb$. The intuition is the following. The merge of $in$ and $the$ $kitchen$, the P’ in the kitchen, to which we respectively assign type $p$ and $n$, will have the properties of a $p$ element, since the P will be the head constituent. This can be represented as $p+n=p'$, i.e. the merge of a P and a noun gives a syntactic constituents which has the same properties of a P. The same principle holds if we merge this constituent with another noun (e.g. the boy), so that we have the PP Gromit in the kitchen. This PP corresponds to the structure $((n)+p+(n))$, so the identity $((n)+p+(n))=pp$ holds.

In this way, we can reconstruct the X-bar schema in a rather effortless way. We can then have full recursion over our structures by assuming that syntactic constituents can be polymorphic, i.e. they can be the realization of different types. This is a standard assumption in categorial grammars, but also in Hale & Keyser (2002), given that they allow for a category to instantiate different types across languages (e.g. (b)-type or (d)-type for Ps). The minimal form of polymorphism we will need is that phrases may be re-interpreted as arguments: for instance, that PP Gromit in the kitchen may be treated as the complement of the verb saw, so that we may obtain the structure at the end of the derivation in (90), or equivalently that PPs in specifier position may be treated as syntactically “simple” constituents, as entailed by the “P-within-P” treatment of Hale & Keyser (2002) (i.e. we have $pp=(((pp)+p+(pp))$ as a possible syntactic rule).
In order to make derivations more intuitive, I shall add a subscript to represent verbs via their abstract type. An English verb will thus be typed as $p_v$, hence the combination of a verb and a PP will be typed as $p'_v = p_v + \text{pp}$, which reads: a verb bar-constituent is the merge of a verb and a PP (e.g. the merge of *goes* and *into the room*). Consequently, a verb phrase (VP) will be typed as $\text{pp}_v = (n + p'_v)$, corresponding to the merge of subject and V’ constituent (e.g. *John* and *goes into the room*). Note that, if DRT can be seen as an incarnation of first order logic, then our lexicon represents a quite simplified incarnation of a standard lexicon for first order logic (cf. Landman 1991:ch.1.).

These basic considerations certainly do not exhaust the topic of discussion on how syntactic structures can be generated from these basic principles, but the basic idea should be clear. We can represent how syntactic derivations come about in our DRT calculus by simply making more explicit assumptions regarding the nature of our DRSs, and the type of information they represent. Intuitively, we can now incorporate the basic syntactic rules of generative grammar, e.g. $X + Y P = X'$, as categorial rules in our enriched DRT calculus, which we represent in the slightly different format illustrated so far (e.g. $p + \text{pp} = p'$). Consequently, we can now seamlessly incorporate the insights of Hale & Keyser’s theory and PIG into our theory, thus preparing our DRT calculus to become fully compositional, as we are going to see in the remainder of part II.

I shall thus assume that DRSs can be labeled according to the following definition. A syntactically labeled DRS is the pairing of a syntactic object (a “lexical item”) and the interpretation associated to this label, such that for each $d \leq \text{LEX}$ and $d \leq D$, $\langle d, d \rangle$ is a labeled DRS. In words, a labeled DRS is the pairing of a syntactic category and a semantic representation: for instance, a word such as “cat” represents a noun (type $n$), and will be interpreted as the $\text{DRS} [\{x\} : \text{cat}^d(x)]$ (i.e. $\langle n, [\{x\} : \text{cat}^d(x)] \rangle$). In representing the merging of two constituents, I shall explicitly represent the two merged constituents (and DRSs) on the right hand of the derivations, and use a subscript on the brackets to mark the type of formed constituent\(^\text{11}\). I shall use round

\(^\text{11}\)Let me note that, in certain variants of DRT, similar approaches to syntactic labeling are discussed, although they do not generally employ the same set of syntactic categories (e.g. Asher & Pustejovsky 2004; Bos 2010; Asher 2011; but also Segmented DRT, Asher 1993; Asher & Lascarides 2003).
brackets for syntactic constituents, and squared ones for DRSs. A sample derivation is the following:

\[(91) \quad \langle n + p = p'(n), p), [d] + [d'] = [d, d'] \rangle \quad (merge \ introduction)\]

\[t + 1. \langle p'(n), p) + n = pp((n), p, (n)), [d, d'] + [d''] = [d, d', d''] \rangle \quad (merge \ introduction)\]

The sample derivation in (91) can be read as follows. The merging of a noun and a predicate yields a predicate phrase, with its corresponding interpretation. The successive merging of this predicate phrase with another noun yields a full predicate phrase, and its corresponding interpretation. The minimal cosmetic difference between this notation and standard generative notation is the use of commas which aims to distinguish the merged constituents. Again, the reinterpretation of constituent structure is a reflex of the associativity property of \textit{merge}. Hence, the logical connection between Phillips’ theory and our DRT calculus can be defined without any supplementary assumptions, which is a welcome result. The important consequence of this labeling process is that now we can explicitly represent syntactic derivations in such a way that the important empirical considerations presented in the previous section can be accurately represented, without any supplementary assumptions.

I shall now offer some sample derivations of a sentence, in order to present how this integration can be realized. I will offer two sample derivations based on sentences (92) and (93), in order to represent the subtle syntactic differences between the insertion of an implicit vs. an explicit P head. I shall leave aside the semantic component and only use “words” as representations of the corresponding DRSs, as I shall address in more detail the exact interpretation of our terms in the next chapter. The derivation is the following:

\[(92) \quad \text{The ball is underneath the table}\]

\[(93) \quad \text{Mario goes into the room}\]

\[(94) \quad \langle n, [\text{The ball}] \rangle \quad (assumption)\]

\[t + 1. \langle p_v, [is] \rangle \quad (assumption)\]

\[t + 2. \langle (n) + (p_v) = p_v((n), p_v), [\text{the ball}] + [is] = [\text{the ball is}] \quad (merge \ introduction)\]

104
The two derivations in (94) and (95) differ between one another (and with respect to (90)) in a number of cosmetic differences. While the ones between this treatment and Phillips’ should
be obvious (e.g. use of an indexed set), the differences between (94) and (95) require some minimal discussion. As the steps from $t + 4$ to $t + 6$ display, if we follow Hale & Keyser’s assumption, then a standard syntactic derivation involving a P may either involve the merge of a phonologically null element (e.g. “P”, in (88)), or the merge of a P head such as to, which may undergo conflation in cases such as into, but not in cases such as up to. Importantly, this treatment needs not to be “the” approach to Ps and their syntactic processes. It is equally possible that Ps can be treated as “pure” relational heads, so that underneath is the main head in (92) (e.g. Svenonius 2006): in such a case, we can postulate a phonologically null specifier, which may be overtly realized in cases such as from under and up to.

The approach I have outlined so far allows us to offer a compositional treatment of adpositions, since now we can easily account for the examples that led us to expand our DRT fragment in the first place, by being able to represent “P-within-P” structures, and more in general sentence structures and their corresponding derivations in a simple way. Examples (54)-(57), repeated here as (96)-(99), can now be easily accounted for:

(96) The boy sits behind the girl

(97) The girl is sitting behind the tree. The boy sits behind (the girl)

(98) The boy goes into the room

(99) The girl is waiting for the boy in the room. The boy goes in(to the room)

The syntactic structure of the PP structure associated to behind the girl in (96) is, in our formalism, $((p)(p)++(n))$. The P head is in this case phonologically null, as per assumptions. The $n$ represents the DP the girl. The example in (97) involves argument demotion of the ground in the second sentence. In this case, this “final” $n$ is instead a phonological null phrase, an element that undergoes a form of ellipsis, as it is possible to recover its interpretation from the previous explicit context. This is in line with standard assumptions about argument demotion and ellipsis (e.g. den Dikken 2006; Phillips 2003, 2006). A similar reasoning can be applied to (98) and (99). Since now we can treat these structures, and the processes by which they come about, in
a uniform and compositional way and without any relevant supplementary assumptions, these
examples do not represent a problem for our theory any more. On the contrary, they offer evi-
dence suggesting that our (new) DRT fragment has a wider empirical import, which is not found
in other syntactic theories.

Our approach can also easily account the examples that motivated the discarding of other
non-compositional theories of syntax. I shall repeat (57) and (58) as (100) and (101), to illustrate
this point:

(100)  a. In front of the car, the men are having a beer
        b. The men are having a beer in front of the car

(101)   The city we flew over is Laputa

In (100-a), the PP in front of the car appears “to the left” of the sentence the men are having a beer, displaying a different order of constituents than the one found in standard sentences (cf. 100-b). I will idealize things a bit and assume that the structure of (100-b) is $\text{pp}_v = (p, +(\text{pp}))$. I assume then that the structure for (100-a) is $((\text{pp})+p)=\text{pp}_v$, a structure in which the underlying order of its constituents is inverted. In our approach, this is just an instance of commutativity: the merging of two constituents can occur in either order, although one order may appear less canonic, or correspond to a slightly different interpretation, because of this “non-canonic” order. For instance, “locative inversion” is suggested to be a phenomenon by which the displaced PP is marked as being in “topicalized” position (e.g. Moro 2000; den Dikken 2006). The preposition stranding phenomenon in (101) can receive an equivalent treatment, with the difference that the stranded or “fronted” element, the city, is a DP (here, an n) rather than a PP. In this case, too, the interplay between syntactic and information structure seems to play a part in which surface structure is expressed, when two constituents may be merged in either order.

The exact details are not important, since we are not concerned with topics of information structure. The important aspect is that these assumptions allow us to reconstruct the treatment of these and other related phenomena offered in PIG (Phillips 1996, 2003, 2006), as well as
other proposals dealing with argument demotion, locative inversion and preposition stranding, couched within the minimalist programme (e.g. den Dikken 2006). They also offer further evidence that the empirical adequacy of the new DRT fragment I have outlined in this chapter is superior not only to the fragment offered in chapter 2, but also to non-minimalist syntactic proposals, since it can capture “displacement” phenomena (e.g. locative inversion, preposition stranding) and and ellipsis-like phenomena (e.g. argument demotion) regarding Ps via part of the basic assumptions of the theory: the commutativity of \textit{merge}. This empirical coverage must be combined with the empirical coverage of P’s internal structure offered in this section and the previous one, which allows us to treat Ps of indefinite complexity without resorting to the proliferation of positions and labels based on the category P. Our new compositional fragment of DRT can cover two sets of empirical phenomena regarding Ps that most other theories cannot easily account, since it includes the “P-within-P” theory of Hale & Keyser (2002) in a relatively straightforward way.

It is possible now to offer an answer to the third research question:

Q-3: \textit{The structure of Ps is realized via the cyclic merge of these elements over derivational time, which represents how the syntactic process dynamically occurs in the mind;}

As in the case of the second research answer, this apparently simple answer captures the intuition that Ps are not combined in a sentence in “one fell swoop”, but in a compositional way, according to the derivational processes that are represented via our DRT fragment. This answer captures the basic intuition that, since speakers will produce a sentence such as (96) by first producing \textit{the boy}, then \textit{sits}, \textit{behind}, and then \textit{the girl}, the underlying mental (linguistic) processes will also follow this order of production, and will produce a syntactic structure which corresponds to the sentence \textit{the boy sits behind the girl}. Our DRT fragment can now give an accurate rendition of the syntactic structure of Ps and the sentences they occur in, as a dynamic syntactic process that unfolds over time in the minds of speakers, since can capture the standard order of production of words in sentences (informally, first subject, then verb). At this point, the answer to the global
research question is still outside our reach, since we need to investigate in detail the semantics of Ps before answering this question. However, the answers offered to the three syntactic research questions have laid out the syntactic background that will allow us to answer this question, in the next chapter.

3.5 Conclusions

Let us briefly summarize the results of this chapter. I have suggested that adpositions/Ps are the main part of speech expressing the abstract type of spatial relations discussed so far, in turn part of the broader category of Cs. I have offered evidence that they constitute a single syntactic unit, combining with Vs (verbs) and DPs to form spatial clauses. I have then analyzed how Ps are merged in syntactic computations via Phillips’ PIG model, which is quite compatible with our DRT calculus, via a modicum of categorial/type-logical assumptions\textsuperscript{12}. The result of this theoretical synthesis is that we can now offer a more fine-grained and compositional analysis of the syntactic properties of Ps, which does not include any extra theoretical apparatus than the one introduced in chapter 2.

This analysis is to an good extent novel, since it brings together syntactic theories which are usually not directly compared (i.e. DRT, Hale & Keyser’s theory, \textit{PIG}). The resulting synthesis consistently extends the syntactic power of DRT, as well as presenting an extension of Hale & Keyser’s theory of Ps which appear to be empirically superior, and theoretically simpler, to other proposals such as the Cartographic approach. It also offers a dynamic perspective on syntactic derivations, thus allowing Phillips’ PIG theory to be extended to various P phenomena still in need of a precise explanation (e.g. argument demotion, locative inversion, the emerging of Ps’ structure). At this point, we can easily the apparent syntactic differences between \textit{in front of} and \textit{ahead of}, or \textit{ahead of} and \textit{into}, under one underlying syntactic theory. Hence, it represents a

\textsuperscript{12}The main difference is that I do not use a perhaps more well known type-forming operator, the “slash” operator (represented as “/” or “\”), and that my set of types is rather different from the one(s) used in this family of frameworks. It would ideally be possible to bring these approaches closer, but this is something I shall leave for future research.
solid step ahead with respect to previous approaches to this category, as well as a powerful tool
to analyze the semantic nature of Ps in compositional detail. Once we have a clear picture of
syntactic derivations, we can proceed to analyze semantic matters, the topic of the next chapter,
and thus be able to finally offer an answer to the second research question.
Chapter 4

The Grammar of Adpositions and Spatial Sentences, Part II: Semantics

4.1 Introduction

In this chapter I shall address the topic regarding the interpretation of Ps that stems from the syntactic proposal offered in chapter 3. In chapter 2, I have assumed that Ps denote part-of relations among situations/events, and thus express the abstract structure among the different scenarios involving referents and their (changing) positions over time. In order to illustrate the problems that arise from this lack of compositionality, look at the following examples:

(102) Mario sits in front of the blonde girl

(103) Mario sits behind the blonde girl

(104) Mario went to the shop

(105) Mario has arrived from the shop

(106) Mario was at the shop

(107) Mario is in the shop

(108) Mario has moved in front of the blonde girl
(109) Mario has gone to the pub and is sitting at the counter
(110) Mario sits ten meters in front of the desk
(111) *Mario sits ten meters at the desk
(112) Mario has gone ten meters towards the desk
(113) *Mario has gone ten meters to the desk
(114) Mario is sitting here

The sentences in (102) and (103) denote that Mario’s position with respect to the girl is on opposite “verses” of the same axis, since (102) includes in front as a P, while (103) includes behind. The sentences in (104) and (105) share a similar relation: while to denotes that Mario’s final position after moving is the shop, from denotes that the shop is Mario’s initial position. The sentence in (106) can be accepted as a consequence of the sentence in (104), as discussed by Parsons (1990:78-84). If we do so, then we know that Mario has gone to the shop at some earlier moment in the past, so we also know that Mario was at the shop as a consequence of this event of motion. These sentences have been amply treated in the literature, in their more specific details, although often via treatments that are for to an extent non-compositional, much like our treatment in 2 (e.g. Fong 1997; Zwarts 2005). Consequently, they represent an empirical challenge for our treatment and previous works.

The next set of sentences present a further set of challenges not only for our theory, but for other semantic theories of Ps as well. The sentences in (106) and (107) appear to be in a kind of logical relation: if we know that Mario is in the shop, we also know that Mario is some “place” related to the shop, or “at” the shop. The sentence in (108) includes a P, in front of, which usually expresses the location of a given figure, as in (102): in this case, instead, it expresses the final position of Mario’s event of motion. The sentence in (109) denotes that Mario reaches the pub and, once there, he sits at the counter and perhaps grabs a beer. The sentences in (110)-(113) display a rather subtle distributional pattern: although all four sentences denote that Mario’s distance (i.e. (110) and (111)) or “journey” in direction of the desk (i.e. (112) and (113)) is ten
meters long, (111) and (113) are ungrammatical: a measure phrase such as ten meters cannot combine with a P such as to. The sentence in (114), in particular the indexical here, denotes that Mario’s position may either have been mentioned earlier in discourse or can be inferred from the implicit context, hence acting as a special type of “P anaphora”.

In our current proposal on the semantics of Ps, virtually all of these phenomena are outside our reach. The goal of this chapter is to present a compositional semantics that will be able to cover these examples in a principled way. Consequently, the resulting theory will not only offer an empirical coverage which is equal or even broader than current theories on the semantics Ps, since it will cover under a unified theory not only relatively well-known but still problematic data (i.e. sentences in (102)-(106)), but also more recalcitrant data, which have received little or no attention so far in the literature (i.e. sentences in (106)-(114)). Since this semantic theory will be couched in the more dynamic, processing-oriented fragment of DRT we have presented so far, it will offer an account of the interpretation of Ps which can have a psychologically plausible base, and thus form the platform for the studies offered in the third part.

The specific goal of this chapter is thus to offer a solution to three outstanding problems regarding the semantics of Ps, which can be formulated as the following research questions:

Q-1: What do we know so far about the interpretation of Ps;

Q-2: What is the specific semantic contribution of Ps and how it is related to their syntactic status;

Q-3: Why Ps can be distinguished in different types;

In order to solve these problems, I will first review the relevant literature regarding the interpretation of Ps, and outline which basic facts will form the platform for offering a solution to the second and the third problem. Anticipating matters a bit, I will propose that there is a transparent mapping between the syntax and semantics of Ps (an instance of the Curry-Howard isomorphism). Hence, the semantic contribution of Ps and PPs can be directly “read off” from the syntax, as I will propose that Ps (as heads) will always denote relations between situations, while PPs (as phrases) will always denote the situations that are involved in these relations. I
shall propose that e.g. *of, to or from*, whether they are phonologically realized or not, denote a basic relation between situations, and thus that *in front or ahead* specify which exact situations are involved in this relation. I will then propose that the distinction between “static” and “dynamic” Ps is a reflection of how the domain of situations is organized. If we assume that situations are distinguished between “static” and “dynamic” ones, then we will predict that Ps will mirror this distinction. We will also predict that the distinctions between these different types of relations will turn out to be predictions about the properties that these different types of Ps have. For instance, we will be able to explain *why* some Ps can combine with measure phrases such as ten meters (e.g. *ten meters in front of the car*) but others don’t (e.g. *ten meters to the car*); what is the exact interpretation of coordinated Ps (e.g. *above and below*); and *why* most Ps can be ambiguous in interpretation, being able to combine with either “dynamic” or “static” verbs, as the examples (102) and (108) respectively show.

Once we will answer these questions, we will have a clear picture of the semantics of Ps, how this semantics is related to their syntax, and how Ps contribute to the interpretation of sentences in a fully compositional way. We will also be able to answer the global research question for this chapter, which I shall repeat below:

Q-A: How Ps can combine together and with other parts of Speech, and express distinct but related types of spatio-temporal relations;

The answer to this global research question will thus offer a solution to our second outstanding problem, how to offer a thoroughly compositional treatment of Ps that can offer superior empirical coverage than previous treatments on this category.

The chapter is organized as follows. In section 4.2, I shall review previous proposals on the semantic of Ps, and suggest a simplified ontology motivated on these data. In sections 4.3 and 4.4, I shall define the formal details of the proposal, also spelling out the details of the syntax-semantics interface. In section 4.5, I shall test the theory against the data, and offer some conclusions in section 4.6.
4.2 What Ps denote, I: Previous Proposals

In this section I shall review previous proposals regarding the semantics of Ps, and offer a thorough explanation on why we must assume that we only have one type of referents in our universe of discourse (situations/possible worlds), which can be distinguished into different sorts, subsets of referents. I will explain why vectors and other “geometrical” referents which have been invoked to be denoted by Ps, in the literature, can be reduced to a sub-set of situations rather than members of a distinct domain, and why this treatment is more empirically adequate than previous proposals. Let us focus on the literature, then.

The semantics of Ps has been studied in several works, as briefly mentioned in chapter 2. Although there is a general consensus that Ps denote relations/properties between implicit referents, distinct from the individuals denoted by e.g. DPs, there is a little consensus on the exact status of these referents within a broader universe of discourse. Certain proposals consider Ps to denote “geometric” types of referents such as vectors (e.g. Zwarts & Winter 2000), while other theories consider Ps to denote eventualities or situations, much like Vs (e.g. Parsons 1990).

A very preliminary and indirect treatment of Ps can be found in Davidson (1967), the foundational paper in the rich event semantics research programme. Davidson proposes, for a sentence such as (115), the logical form in (116):

(115) John buttered the toast slowly in the bathroom with a knife

(116) \( \exists e [\text{butter}'(e, j, t) \land \text{slowly}'(e) \land \text{in}'(e, b) \land \text{with}'(e, k)] \)

In words, (116) says that there is an event of buttering a toast, occurring in the bathroom, and realized via a knife\(^1\), by John. Furthermore, this complex event is also defined as occurring at slow pace. Here the event’s implicit argument allows defining the entity which slowly and other adverbs denote properties of, and on which subsequent properties can be superimposed. The contribution of the P in is not defined in any specific way, since this P denotes a relation between

---

\(^1\)I will not say anything about with and other Ps, which can be indirectly traced to a spatial interpretation (see e.g. Jackendoff 1983, 1990), although the proposal I will present could be extended to cover these cases as well, at least in theory.
the event argument and the ground denoted by *the bathroom*, much like the “non-spatial” P with
denotes a relation between event and (a) knife.

A refinement of this proposal can be found in Parsons (1990), partially presented in chapter
2, section 2.3.5, and which I shall present in some more precise detail here. Parsons assumes
that there are three sub-types of eventualities: events, processes and states (cf. also Bach 1986).
Events and processes are denoted by “dynamic” verbs (e.g. *go, push*, etc.), states by “static”
verbs (e.g. *sit, know*, etc.). States hold at a time \( t \), processes are events that hold at a time \( t \)
and can temporally overlap with states, events (proper) hold at a time \( t \) and *culminate* at a time
\( t + 1 \), at which a resulting state comes into being. Eventualities can be related to individuals
via *thematic roles*, such as “Agent”, “Patient”, “Theme”, “Location”, etc. Parsons also assumes
that Ps denote both temporal-causal relations between eventualities, and one or more location
involved in this relation. A partial translation of the sentence in (115) is:

(117) \( \exists s \exists e \exists t [theme(s, j) \land butter^*(e, t) \land Hol(e, t) \land Loc(s, b) \land in'(s) \land Hol(s, t)] \rightarrow \exists s' \exists e' \exists t [theme(s, j) \land butter^*(e, t) \land Hol(e, t) \land Loc(s', b) \land in'(s) \land Hol(s', t + 1)] \)

In words, the partial logical form in (117) says that the theme “Mario” will occupy (be in) the
bathroom location while buttering (the toast), during a state s holding at a time \( t \), and he will do
so at distinct states (i.e. \( s' \)) holding at later times (i.e. \( t + 1 \)). Further event semantics approaches
offer a rather fine-grained picture of the neo-Davidsonian theory of events couched in a lattice-
theoretic approach, but do not treat Ps in a more fine-grained way than Parsons (1990) (e.g.
2008).

*Situation semantics* approaches have offered treatments of Ps as well. Two main perspec-
tives can be distinguished. The situation semantics approaches rooted in philosophical logic
and cognitive sciences assume a richer ontology, which includes locations as referents. They
suggest that spatial Ps denote a relation between an individual referent and an implicit location
referent: in our case, the denotation of *in* would be \( \text{in}'(l, x) \), and the implicit referent would be
\( l \), an implicit referent denoting a spatial location (e.g. Barwise 1981; Barwise & Perry 1983,
The accounts rooted in Linguistics, such as Kratzer (1989), suggest instead that situations can be understood as spatio-temporal referents (much like events), which however are equipped with a part-of relation. The logical form for (115), according to this proposal, is the following:

\[\lambda s'. \exists s [\text{butter}'(s, j, t) \land \text{slowly}'(s) \land \text{in}'(s, b) \land \text{with}'(s, k) \land s \leq_p s']\]

In words, (118) says that there is a situation of John buttering a toast in the bath, which is one of the possible instances of this type of situation. The exact semantic contribution of Ps, as this example shows, is still rather coarse-grained, as the introduction of the part-of relation says little about the precise meaning of in and its contribution to the interpretation of sentences. I shall defer discussion of the “new” operator used in (118), the \(\lambda\)-operator, until the next section, and focus on the geometric family of approaches.

There are two different variants of the geometric approach: one originates entirely within Montagovian (and thus intensional) works as Bennett (1975) and Cresswell (1978), the other in a formalization of Jackendoff’s Conceptual semantics. Cresswell’s assumptions can be best presented via an example:

\[(119) \text{The band is playing excerpts from HMS Pinafore across the meadow from here}\]

In (119), the band playing excerpts is understood to be located at the end-point of a “virtual” path that begins from the speaker’s location, covers a (given) meadow and reaches the band’s location and referred to by the indexical here. A path is a three-place relation \(p(a, t, w)\), including a region of space \(a\) occupied by the ground, an interval of time \(t\) and a possible world \(w\). A P such as \textit{across} denotes a relation between a path and a moving referent, and thus will be translated as \textit{across}'\((p(a, t, w), f)\), an “across” relation holding between a (moving) figure \(f\) and a ground.

Jackendoff’s Conceptual semantics, as we have seen in example (61) (chapter 3), also introduces the notion of PLACE. Jackendoff observes that the distinct ontological status of these concepts is supported by sortal restrictions on questions: where questions can only be answered via a (spatial) PP, such as in the garden (e.g. Jackendoff 1983, 1990, 1991, 1997). Several
subsequent model-theoretic works make these notions (formally) precise, giving a formal treatment of PLACE as “spatial” predicates (e.g. \( \text{in}'(y) \) for in) and of PATH as “spatial” relations (e.g. \( \text{Goal}'(x, \text{in}'(y))(i) \) for into, \( \text{Goal}' \) being a sub-type of path) (e.g. Wunderlich 1991, 1993; Verkuyl & Zwarts 1992; Zwarts & Verkuyl 1994). Refinements of this approach include Nam (1995), who proposes that English Ps denote a part-of relation defined over the domain of regions \( R \): for instance, in is translated as \( r \leq r' \), with \( r, r' \leq R \). Another is Kracht (2002, 2004, 2008), which spells out a detailed model-theoretic semantics for Ps as denoting regions in which events occur during intervals of time. For instance, in denotes the complex relation \( e : \forall t[e(t)(\text{loc}(e))] \leq r : t(r, \text{loc}(x), t)] \), an event holding for all intervals of time \( t \), and located in an internal \( t \) region \( r \) of a ground\(^2\).

A different proposal, based on the notion of “vector”, is found in Zwarts (1997); Zwarts & Winter (2000). Their argument for vectors is based on the fact that adverbs of direction and “magnitude” (distance) can occur with (some) spatial Ps, as in the example:

\[(120) \quad \text{The lamp is diagonally one meter above the table}\]

In (120), the adverbs diagonally and one meter respectively denote the direction/orientation and distance of the lamp with respect to the table, further restricting the “vertical” relation between the two entities, denoted by above. These two adverbs, the authors suggest, define the properties of a vector, an ordered pair of points within a Euclidean (i.e. three-dimensional) model of Space, representable as \( v = (p, p') \), and being the type of referent denoted by Ps, which can also be indexed with respect to a set of (temporal) indexes, to denote a path (e.g. \( p(i), p(j) \), etc.).

Summing up, previous proposals on the semantics of Ps tend to converge on two assumptions: that Ps have an inherently relational semantics, and that some implicit referent is involved in this relation semantics. The divergences among these theories pertain the exact nature of this referent, since different authors assume that different implicit referents are involved in these spatial relations, often mentioning the same underlying type of argument, the argument from

---

\(^2\)I partially misrepresent both theories, since both frameworks use the subset relation symbol, “\( \subseteq \)”, to represent the part-of relation. This misrepresentation has one advantage, however, since it avoids the use of different symbols for the same relation, hence avoiding sources of confusion.
modification. At this point, we are able to offer an answer to the first research question, which is as follows:

Q-1: \textit{Ps denote relations over implicit “spatio-temporal” arguments, whose nature is still subject to controversy;}

This answer is justified by the heterogeneous set of proposals that I have reviewed in this section, and highlights that, although most theories treat Ps as being inherently relational, in nature, there is little consensus on what are the model-theoretic entities that they bring together, and whether this lack of consensus is empirically justified or not. The answer to the second question will also provide an answer to this emerging question.

4.3 What Ps denote, II: a Proposal for Ontological Simplicity

Several of the works reviewed so far implicitly suggest that, although distinct, the domain of eventualities and that of “geometric” entities are connected. The explicit proposals on this connection can be divided into two broad groups. Works such as Link (1998) suggest that regions and eventualities are distinct domains, sharing the same lattice structure, and forming the “AETHER” model\footnote{In which Agents, Events, THEMatic roles, Event types and Regions are defined as the basic semantic domains.}. Other proposals follow Jackendoff’s work, and suggest that different domains (e.g. events and paths) have different structures (e.g. partial orders for events, total orders for paths), a reflection of the non-linguistic information they represent in Language (e.g. Zwarts & Verkuyl 1994; Krifka 1998; Zwarts 2005, 2006). Informally, for most authors, events are grouped together in a different way than paths or regions, although it is possible to define how events are related to the paths they take shape on.

Among this second group, we can find DRT proposals (e.g. Kamp & Reyle 1993:ch.5; Kamp et al. 2005:ch.3; Asher 1993:ch.2; Asher & Lascarides 2003:ch.2). The standard treatment of ontology in DRT, which is different from my proposal in chapter 2, stems from the definition of the sorts, sub-types of referents in $U$, the Universe (of Discourse). Aside possible worlds
and discourse referents, $U$ includes Eventualities $EV$, split between mutually exclusive events and states, i.e. $EV = E \cup S$ (with $E \cap S = \emptyset$), and temporal intervals $T$. This is represented as $U = W \cup D \cup EV \cup T$. While referents and eventualities are partially ordered sorts, time intervals are strictly ordered (i.e. we have $(T, <)^4$). Proposals such as Asher & Sablayrolles (1995); Maillat (2001) include spatial intervals (i.e. regions/locations) as sorts within the general domain of (space and time) intervals, based on a euclidean model of space, the one found in Zwarts & Winter (2000). DRT and other “interface” theories assume that the different domains are connected via location functions $Loc$ (trace functions $\tau(e)$ in Link/Krifka’s theories), which map eventualities onto their corresponding time intervals/spatial regions (i.e. we have $Loc(ev) = t$, or $\tau(e) = r$).

A common thread in proposals arguing for structural differences between domains (DRT, Parsons’, Krifka’s) is the appeal to extra-linguistic information to motivate this choice. For instance, in DRT intervals of time and space represent the way humans are assumed to internally represent the flow of time in the external world (see e.g. Kamp et al. 2005:108-109). Given our discussion in chapter 2, this appeal appears to be remarkably weak, if not unmotivated. Intuitively, while all mental processes occur over internal intervals of time (and probably space, i.e. brain regions), these intervals do not represent a type of information which is unique to language, or that determines the structure of linguistic computations, especially semantic ones. Consequently, there is no need to include them in linguistic derivations, if not indirectly.

Aside the extra-linguistic argument against time and space intervals, there is a linguistic argument, which is based on the modifier argument. Intuitively, if measure phrases, direction adverbs and indexicals offer an argument for distinct spatial referents, then this argument is not valid in case these modifiers can occur cross-categorically. If these modifiers combine with e.g. event-denoting verbs, then they must define properties of the events denoted by verbs, rather than purely geometric entities of some sort. Consequently, if these modifiers combine with verbs, then they do not denote properties of e.g. vectors as entities distinct from events, but they

---

4Note that in DRT, pre-orders are usually represented via the symbol “$<$”. This difference is immaterial, here.
denote certain specific (“spatial”) properties of events themselves. The same reasoning against a domain of spatial referents can be applied to nouns and adjectives, also assumed to denote eventualities (e.g. Landman 2000 for verbs; Chierchia 1995a, 1998 for nouns; Kennedy 1999, 2007; Winter 2005; Piñon 2008 for adjectives; *inter alia*).

The relevant examples are the following:

(121) The road is *ten meters* long (absolute adjective)

(122) The main road is *ten meters* longer than the secondary road (comp. adjective)

(123) Mario sits *straight* (verb)

(124) Mario runs *one kilometer* (verb)

(125) A road *one kilometer* towards the woods (noun)

(126) A road *deep* through the woods (noun)

(127) The boy *here* is blonde (noun)

(128) The boy sits/runs *there* (verb)

(129) The road is narrow *here* (adjective)

The examples in (121)-(126) show that both measure and direction/orientation adverbs, e.g. *one kilometer* or *deep* or *straight*, can also combine with verbs, adjectives and nouns (or corresponding phrases), without particular restrictions, and possibly changing their precise syntactic status, when occurring as “bare” complements. The examples in (127)-(129) suggest that indexicals such as *here* and *there* may also combine with the three main lexical categories, although with a rather more restricted distribution, possibly influenced by the implicit context (e.g. Kayne 2004 for discussion). We cannot distinguish the modifier argument for vectors from the modifier argument for eventualities: this because Vs, Ps and other parts of speech denote only one type of implicit “spatial” referent, eventualities.

For this reason, I shall assume that our universe of discourse can be reduced to $U = W \cup EV \cup D$. Informally, this universe of discourse lacks a distinct sort for time/space intervals, since
it turns out not to be empirically motivated. This Universe of Discourse, however, does not come with a clear ontological status to situations. However, a standard assumption in Situations Semantics is that situations do not represent a distinct set of referents in discourse, but rather an “alternative” label for possible worlds (e.g. Zucchi 1993; Landman 2000; Steedman 2004), or possible worlds (e.g. Barwise & Etchemendy 1990; von Fintel 1994; Kratzer 2007). I will thus assume that situations and possible worlds are interchangeable labels for the same domain of referents, something that I shall represent as the identity $W = SIT$, $SIT$ being the set of situations.

With this notational identity in mind, we apparently have a relatively simpler ontology for our universe of discourse, since it only involves possible worlds, referents and eventualities. Once we look at more data regarding $Ps$, however, we can push our ontological simplification further, defining a simpler and empirically motivated Universe of Discourse. The intuition is the following: if $Ps$ denote eventualities/situations and can freely combine with other parts of speech denoting other types of referents, for instance parts of speech denoting possible worlds (or relations thereof), then there is no need to maintain a distinction between these sets of referents within our Universe of Discourse, since it would create an unmotivated distinction. One syntactic argument for this reduction is the following: if we can consider situations/eventualities as equivalent with possible worlds, then we can treat $Ps$ as members of the broader categories of Cs, as suggested in chapter 3, section 3.3. Four semantic arguments in favor of this reduction, seldom mentioned in the literature, are the following.

A first argument is based on the interaction between $Ps$ and modal expressions. Look at the examples:

(130) Mario is possibly in front of the cinema

(131) Mario has necessarily arrived straight from the station

(132) Mario is possibly sitting ten meters right behind the last chair, on the floor

(133) I order you to sleep
I prohibit you from sleeping

At least since Kratzer (1977), it is generally assumed that modal adverbs and auxiliaries, *inter alia*, denote forms of quantification over possible worlds, ordered by an *ordering source* relation, ultimately a *part-of* relation over sets of possible worlds (e.g. Kratzer 1977, 1981, 1991). Simplifying matters somewhat, the modal adverb *necessarily* can be roughly translated as $\forall w P(w)$, whereas *possibly* can be translated as $\exists w P(w)$: respectively, a property $P$ holds in every or at least in one world. The examples in (130)-(134) offer evidence that Ps can directly combine with such adverbs and propositional attitude verbs to express e.g. commands/prohibitions5, as well as “spatio-temporal modifiers” (e.g. *ten meters, right*), while they may also combine with more “spatio-temporal” modifiers, such as *ten meters* and *right*; or they may actually directly denote the (type of) ordering source associated to an attitude verb, since they denote “ordering” relations (Fong 1997; Kratzer 2006 on *that* and other complementizers).

A second argument pertains the semantics of *indirect questions* and *relative clauses* (e.g. Kartunnen 1977; Groenendijk & Stockhof 1984; Aloni 2001; Caponigro 2003). For instance, Groenendijk & Stockhof (1984) discuss the following entailment patterns:

(135) I know who killed N.6 (premiss)
Mario killed N.6 (premiss)
I know that Mario killed N.6 (conclusion)

(136) I know where Mario is (premiss)
Mario is in the kitchen (premiss)
I know that Mario is in the kitchen (conclusion)

The common entailment pattern to (135) and (136) is roughly the following: if I know *who* killed N.6 or *where* Mario is, I know the set of propositions that include a specific proposition (respec-

5Here I shall make matters simple and assume that modal adverbs always take a *de dicto* reading, hence describing a possible, not “inherent” property of some referent, as in *de re* readings.
tively, that Mario killed N.6, and is in the kitchen), and thus conclude that Mario’s position is in the kitchen, or N.6’s killer is Mario. Hence, “wh-” words, including who and where, denote relations between propositions, in this case treated as sets of possible worlds.

A third argument pertains the closely discourse-related phenomena of fragments and ellipsis, of which (incomplete) answers and telegraphic speech are one incarnation:

(137) Q: Where is Mario?
A: In his room

(138) (Extra-linguistic context: two people in the room, one is clearly looking for something.
The other says:)
Near the chair, I think

In (137), a possible answer to the “where” question is the PP in this room. In (138), the PP near the chair can be offered as a “partial” sentence by one agent, which may help the other agent to find the object he’s looking for. One proposal is the “ellipsis” approach to questions, offered in Merchant (2001, 2004, 2008). This proposal assumes that “incomplete” sentences are obtained by first generating a full syntactic structure, and then “canceling” the redundant material, and that the “surviving” material is a proposition entailed by the full sentence’s proposition; thus it must maintain the structural relations introduced by a sentence, e.g. his referring to Mario. Hence, the two PPs in this room and near the chair denote propositions, sets of worlds. In these cases, these propositions are sortally restricted, since they express a specific location in which a certain figure is located.

A fourth argument pertains another discourse-oriented phenomenon, that of stress and focus:

(139) Mario is BEHIND the car

The sentence in (139) is understood as expressing the location of Mario as being mutually exclusive with other possible locations: Mario is behind the car, and not somewhere else. A

---

6One syntactic detail is that the remaining material must “move” in a special syntactic position, for this purpose. For our concerns, this aspect is irrelevant.

124
general consensus about the effect of stress on interpretation, here marked via capital letters, is that it introduces an implicit set of *alternatives*, modeled as mutually exclusive possible worlds, i.e. worlds that represent distinct states of affairs (e.g. Rooth 1985, 1992; Krifka 1999; Schwarzschild 1999; Beck 2006). In our more restricted case, the stressing of the focused P *behind* corresponds to the implicit reference of other possible locations/alternatives that could have been occupied by Mario, e.g. the left or right side of the car, its front; but which are not occupied by him, since he is behind the car.

These four arguments, although not exhaustive, offer solid evidence that Ps can combine with possible worlds/situations-denoting parts of speech (e.g. such as modal auxiliary verbs), or participate in syntactic/semantic phenomena that involve possible worlds, or relations over these worlds. Consequently, as in the case of the argument from modification, Ps and these other parts of speech denote the same general type of referents, possible worlds/situations, although in subtly different shades. I shall thus assume that our universe of discourse consists of the union of discourse referents and possible worlds/situations, ordered by the *part-of* relation, represented by the identity \( U = D \cup W \): since possible worlds/situations and eventualities are the same type of entities, we can just maintain one general sort for this category, i.e. we have \( W = EV = SIT \), and thus we have the identities \( U = (D \cup EV) = (D \cup SIT) \).

Since the *who*-types of questions also seem to suggest that discourse referents can be reduced to situations sorts (see e.g. von Fintel 1994; Kratzer 2007), I shall push this reduction further. I shall assume that the universe of discourse actually coincides with the set of *situations/possible worlds*, i.e. that we have the identity \( U = W \). Our universe of discourse thus contains only the (semantic) type of possible worlds/situations, which however may have several sub-types or sorts. One distinct example we have discussed so far includes a sort of *spatio-temporal situations*. The intuition is that Ps denote a sub-set of the Universe of Discourse, including the properties and relations that can be defined over this sub-set, as we have seen in chapter 2 via the analysis of *into* and *near*. Since the Universe of Discourse is made of possible worlds/situations, then a proper sub-set of the Universe of Discourse includes possible worlds/situations, which at
the same time can also be “spatio-temporal” situations, as also advocated in Cresswell’s classical work (Cresswell 1978). This can be formally represented as $R \leq W$, or informally: the lattice of “spatio-temporal” situations denoted by Ps, $R$, is part of the lattice of situations denoted by the parts of Speech included in a Language. Several other sorts could be proposed and be empirically motivated (e.g. the sort of discourse referents $D$, with $D \leq W$, which would possibly form a lattice of sorts making up the domain of possible worlds (e.g. Asher & Pustejovsky 2004; Asher 2011).

The finer details of this analysis are not our concern, here. What matters is that the use of a universal type of situations captures the very strong intuition that syntactic flexibility is mirrored into the semantic domain, as recently observed in the literature (e.g. Partee 2009). If a DP such as the boy or a PP such in front can combine with a P in their specifier position, then they will likely denote the same basic type of referent. In this case, the referent denoted by the boy is a special type of discourse referent, the “figure” of a sentence: informally, it is an individual identified with its possible spatio-temporal location, which is then related to another possible spatio-temporal location, the external, frontal position denoted by in front. We can now capture in the semantics both the syntactic similarities and the syntactic differences between parts of speech, by simplifying our set of original assumptions, motivated by the data discussed so far regarding Ps and other parts of speech.

We are now able to offer an answer to the second research question, which is as follows:

**Q-2:** Ps denote relations over situations, while PPs denote (complex) situations which are part of these relations;

This answer tells us that Ps such as in, at, to and from invariably denote some instance of the part-of relation, while PPs such as in front, ahead, behind, across invariably denote situations that are included in this part of relation, as we will motivate in the remainder of the chapter. Consequently, PPs such as in front of the car or to the beach invariably denote a part-of relation between situations, one representing the ground, and the other representing the specific situation
coinciding with the figure, as related to the ground. The next section will spell out the formal
details of this idea.

4.4 What Ps denote, II: a Novel Proposal

In this section I shall spell out in detail the semantic properties of Ps, by analyzing the structure
of the “new” domain of situations/possible worlds, and the syntax-semantic interface, i.e. how
these semantic properties are mapped onto syntactic structure. Since I shall only focus on spatial
situations, I shall use the label “situations”, when this use does not create confusion. Furthermore, I will identify the universe of discourse $W$ with the domain of spatio-temporal situations $R$, to make the presentation simpler. When useful in the presentation of the arguments, I shall use the label “region” to refer to this type of spatio-temporal situations. The goal of this section is to explicitly define the semantic properties of (spatial) referents the universe of discourse, and thus make explicit predictions regarding the interpretation of Ps, as denoting relations over these referents.

Let us start from basic definitions. I shall follow the DRT approach and assume that the set
of situations $R$ can be split in two mutually exclusive domains, states $S$ and events $E$: again, we
have $R = S \cup E$, $S \cap E = \emptyset$. I assume that $S$ and $E$ are both lattices, which however differ with
respect to the set of their free generators, the sets of basic (atomic) elements making up a lattice.
The generator of the set $S$ contains a (denumerably) infinite set of states, which can be “split” in several sorts, whereas the generator of the set $E$ contains only a singleton set, the event of change $e$. The reason behind this assumption is the following.

In DRT, temporal/spatial expressions respectively denote ordering relations over intervals
of time and points in space. For instance, the temporal PP after the exam denotes a relation
between a certain eventuality and the interval of time corresponding to the exam, in the latter
precedes the former (i.e. the conditions $[t_{ex} < t_{ev} \leq s]$), and a similar treatment is used for spatial

\footnote{I misrepresent DRT’s formalism by using the part-of relation to represent inclusion between time intervals, i.e. a sub-set relation. This difference is immaterial here.}
Temporal morphology and temporal expressions introduce special conditions, anaphoric relations between an utterance time \( n \) (for “now”) and some interval of time: the present tense is interpreted as the condition \( t_p = n \), the indexical here as the condition \( t_s = p \), \( p \) being the “location” of utterance (e.g. Zeevat 1999:3; Maier 2006:ch.1). Without time/space intervals, we need to reconstruct these notions via situations, and thus to assign to one designated element this special “referential” status. Two specific linguistic reasons for this assumption are the following.

First, as suggested by Parsons (1990); Fong (1997, 2001), dynamic/directional Ps such as to and from denote a relation between eventualities, in which at some point the figure is neither at his starting point nor at his destination, but in some temporary and changing position. Fong (1997) suggests that Ps such as to denotes a sequence (of events) \( \neg e < e_i < e \): in words, a moving figure changes his location from an event of “not being” at one destination to one of being there (i.e. from \( \neg e \) to \( e \)), once a “transition” event occurs (i.e. \( e_i \)); from denotes the opposite order of events, from \( e \) to \( \neg e \). Once we move to the many other dynamic/directional Ps, the need for a distinct “event of change” acting as a reference event appears to be quite evident.

Second, virtually all temporal/aspectual phenomena, such as the interpretation of Vs and the interaction of Vs and Ps, suggest that a distinct element acting as the “reference event” is necessary, as also proposed in several other works in the neo-reichenbachian mould, including DRT (e.g. Reichenbach 1947; e.g. Kamp 1979a, 1979b; Giorgi & Pianesi 1997; Borik 2008). While we may leave aside the notion of “time interval” and thus of a “now” interval, we still need a spatio-temporal referent that represents the moment of change, and thus allows the interpretation of past, present and future morphology, among other things.

Let us focus on the precise formal details. The two sorts forming the domain of situations, \( S \) and \( E \), are partitions. A partition is a set (here, a partially ordered set/lattice) made of non-overlapping sub-sets (cells), with the following properties: given a set \( C \) and a set of sub-sets \( S \), \( \bigcup S = C \) holds, \( A \neq B, A \cap B = \emptyset \), with \( A, B \leq S \). In words, a partition is such if the sum of cells make up (cover) the partitioned set, and if no cells overlap, provided that they are distinct. The two generator sets are the sets of atomic states/events in \( S \) and \( E \). Atomic states (referents) are
referents which have no proper parts except themselves: if the referent \( c \) is an atom, then there will be no referent \( d \) such that \( d \leq c \). For instance, a state \( \{s\} \) has no proper parts, whereas a sum state \( \{a,b\} \) has two proper parts, \( \{a\} \) and \( \{b\} \), its constituting states\(^8\). The (generator) set of \( S \) will be defined as \( AT(S) = \{s,\ldots,s'\} \), whereas the generator of \( E \) will be \( AT(E) = \{e\} \), as per discussion, with \( S \cap E = \emptyset \) and \( S \cup E = R \). In words, the set of spatio-temporal situations is split between two mutually exclusive sets (i.e. partitions), that of states and that of events.

Let us now focus on the lattices generated by these sets. I shall assume that \( S \) is formed by the all the possible sums of atomic states. Intuitively, if a state represents a “spatial” situation, then more complex situations can be defined by combining these basic situations together, in a recursive fashion. Informally, the “up” situation (i.e. \( up \)) can be merged with the “front” situation (i.e. \( f \)), to form the “upfront” situation (i.e. \( \{up,f\} \)). Since \( S \) is a lattice, the relation \( s \leq S \) holds: the set of states can contain an infinity of sum states and is ordered by the part-of relation. It has a unique maximal sum, or ideal, \( S \) itself. The set \( E \) is also a lattice, which is obtained by merging (“unifying”) the elements of \( S \) with those of \( E \), in a point-wise fashion. Since \( E \) coincides with \( e \), then the only way for this set to “expand” is to combine with the situations from the complementary set of states. The result will be that the relation \( e' \leq E \) holds, with \( e' : \{s,e\} \). In words, an event is defined as the sum of a possibly complex state and the event of change, and the lattice of events can contain an infinity of events ordered by the part-of relation; it also has an ideal member, \( E \) itself. Intuitively, these relations respectively capture the difference in interpretation between more “static” verbs and adpositions, e.g. Ps such as in front of, which do not involve any change of position; and adpositions which denote some change of location (and thus an event of change), such as through.

Once we have defined the structure of the two partitions, we need to define how the two partitions are merged together in a single domain, and how this domain is ordered by the part-of relation. This process is obtained via the so-called transitive closure principle, which defines

---

\( ^8 \)The union of two cells has the effect of forming a “new” cell, \( \{a,b\} \), in which the two atoms are in a sense “hidden” in the structure. See Landman (1991:100-110); Schwarzschild (1996: appendix); Link (1998); for discussion. “Modal” treatments of partitions, their properties and applications can also be found in e.g. Groenendijk & Stockhof (1984:ch.2); Landman (1991:ch.2, 2004:ch.1); Schwarzschild (1996:ch.2); Aloni (2001:ch.1); inter alia
an isomorphism \( h \) between elements of the two partitions (cf. Landman 1991:ch.2, Landman 2004:18-24). For each \( s \leq S \), there will be a corresponding \( e' \leq E \), such that the equivalence \( h(s) = e' \) holds: for each state, we have a corresponding event. Since the inequality \( s \neq e \) holds for each \( s \leq S \), a relation between states and events/processes will be irreflexive, although transitive and antisymmetric. The transitive closure principle will define a strict partial order (or total order) between events and states, e.g. \( e < s \), derived in a straightforward manner from the general partial orders defined over situations. Informally, the transitive closure principle says that states and events are part of distinct domains, but they still be ordered via the (strict) part-of relation. Intuitively, Ps such as to and from capture this kind of “hybrid” relation, but others could be defined as well.

If we can define an isomorphism \( h \) from \( S \) to \( E \), then we can define an equivalent isomorphism \( k \) from \( E \) to \( S \), which maps an event onto its “originating” state. Consequently, for each strict order of the type \( e < s \), there will be a reverse or complementary order of events \( s < e \). Again, two examples are to and from. A simple way to capture this duality is via the use of negation. In a lattice, negation is defined as set complementation, the set including all elements not in a given set, e.g. \( \neg b = \cup \{c \leq W : b \cap c = \emptyset \} \) (cf. Landman 2004:30). Consequently, the relation \( s < e \) can be seen as the complementary relation to \( e < s \), i.e. we have \( \neg(e < s) = s < e \).

States and events represent one the complement set of the other, something we can represent as \( S = \neg(E) \) (and thus \( AT(S) = \neg AT(E) \)). We can now formally capture that as partitions, States and Events are complementary domains. Also, the event of change is defined as a situation in which no other states hold, hence representing the intuition that change can be conceived as any interval in which a transition occurs (e.g. going from one location to another).

One further fact stemming from these definitions is that situations (states and events), according to these definitions, can be represented in two slightly different ways, so they correspond to two slightly different types of interpretation for PPs. If a situation can be defined as a relation between “smaller” situations (e.g. \( s : (s' \leq s'') \)), then the equation \( s : (s' \cup s'' = s'') \), or just \( s : \{s', s''\} \) holds. This equation says that a situation can also be defined as sum of smaller
situations. Intuitively, if behind denotes a relation between figure and ground along a certain axis (i.e. a set of “ordered” positions: the vase behind the box), then it can also denote a non-oriented “region” occupied by the figure, conceived as a distinct portion of “Space” (e.g. Mario sits behind). Furthermore, since our DRT fragment anaphoric relations as identity relations between referents, it is possible that a situation can be identified with its constituting parts: from \( s : \{s', s''\} \), we can derive the identity \( s = (\{s', s''\}) \). Intuitively, this “anaphoric” identity says that pro-forms such as the indexicals here and there can be interpreted complex situations, with part of this identity being context-dependent.

These assumptions regarding our the structure of our lattice of situations\(^9\) make specific predictions regarding the interpretation of Ps, since they spell out specific properties of the relations that Ps denote, if they find their denotation on the domain of situations. The predictions, presented in compact form, are the following.

First, if Ps denote relations over situations, then they will either denote a “static” part-of relation over the states’ domain, of the form \( s : (s' \leq s'') \), or a relation over events, of the form \( e' : (s \leq e) \), with \( e' = \{s, e\} \).

Second, if they do not denote one of these two relations, then they will denote a (strict) part-of relation including events and states, which may either be “positive” or “negative” in order (i.e. we either have \( s < e \) or \( e < s \)).

Third, since both restricted part-of relations may be defined as positive or negative, each (restricted) relation over states/processes will have a “mirror” relation, which expresses the inverse order of situations (i.e. if we have \( (s \leq s') \), then we have \( (s \leq s') \)).

Fourth, for each situation defined as a relation between situations, there will be a corresponding situation in which this order is “removed”. For each situation represented as a relation (i.e. \( s : (s' \leq s'') \)), we will have a situation represented as a sum (i.e. \( s : \{s', s''\} \)), the same holding for events. The type of situation will also determine which parts of speech can combine with a P: “ordered” situations combine with other “ordered” situations, “unordered” situations with

\(^9\)I shall leave open the question on how a similar sortal distinction can be “backtracked” to our logic of Vision, although it should be in theory consistent with the assumptions behind this treatment.
“unordered” situations.

Fifth, if we merge syntactic objects apparently denoting “different” complex situations, we can merge these situations as a single complex situation (roughly, we have $s' : (x) \cup s'' : (y) = s : (x, y)$). We can “compress” situations and their constituting parts into a single, more complex situation. In the case of complex, relational situations, we can merge these relations in complex relation (i.e. transitivity, we have $s' : (x \leq y) \cup s'' : (y \leq z) = s : (x \leq y \leq z)$). We can also establish relations between a “compressed” situation and its constituting situations, via this process of “compression”. Furthermore, if a situation can be represented as an identity statement between the parts and their sum/relation (e.g. $s : \{s', s''\}$ equivalent to $s = \{s', s''\}$), then a situation can also be interpreted anaphorically, something that we can represent in the standard format “$s = ?$”. The practical import of these predictions, present in a non-compact and informal way, is the following.

The first prediction respectively establishes the existence of Ps such as in and in front of, Ps which denote “location”, as a relation between states; and the existence of Ps such as through and along, which denote “motion”, as a relation between different states making up the “trajectory” of a moving figure, and which involves an event of change, an interval of time in which the figure is moving rather than in a given position in space.

The second prediction establishes the existence of Ps such as to and from, Ps denoting motion that results in a change of position, as a relation between an event and a state occupied by a moving figure. This state can be the “result” state, the position reached by the moving figure after the event of change (to), or it can be the figure’s initial position (from). In either case, event of change and state(s) are distinct situations.

The third prediction establishes the existence of antonym-like pairs of Ps, such as in front of and behind, above and below, on and off and so on, which denote “orientation”, a relation between states taken in a certain order. For relation that denotes an “oriented” set of states, then a relation involving the complementary set of states can be defined, or: for a P like in front of, we have a P like behind.
The fourth prediction establishes the existence of two possible readings for any P. If *behind* denotes an oriented set of regions with respect to a ground, then it may also denote a certain region of space corresponding to the sum of these regions. Oriented regions can then be “measured out”, i.e. they can be combined with expressions denoting e.g. their length or orientation; or they combined with other regions. This prediction has some interesting consequences: it predicts that *behind* can combine with measure phrases such as *ten meters*, or that we can have coordinated PPs such as *above and in front of*, denoting either complex regions or “vectors”.

The fifth prediction establishes the existence of complex regions obtained in a recursive manner, such as those denoted by “P-within-P” structures (e.g. *in front of, on top of*). It establishes that the situation(s) denoted by *in front* includes the situation(s) denoted by *in front of*, and more in general to recursively combine the denotation of several Ps into a complex situation (e.g. *to in front of, from behind*, etc.: see e.g. Svenonius 2010). Consequently, the fifth prediction also establishes the existence of logical (entailment and sub-set) relations between Ps, which mirror their syntactic relations in the semantics, such as the relations between *on top of* and *on*, or *into* and *to*, and so on. Furthermore, it establishes the existence of anaphoric-like elements such as indexicals *here* and *there*, pro-forms that denote an “atomic” situation s. The anaphoric nature stems from their ability to be eventually identified with some specific situation in context (e.g. “here” being “in the garden”).

It should be obvious that these predictions are tightly connected to the data that were beyond the reach of our proposal at the beginning of this chapter. However, before exploring in full detail these predictions and how they are related to these data, we must define the relation between semantic relations and syntactic categories, and thus address the *syntax-semantic interface problem*. Let us start from type matters. The semantic type associated with the domain of situations is $s$, the type of situations/possible worlds. Aside this basic (or “lexical”) type, we can recursively define more complex “functional” types, in the common format: if $a$ is $\alpha$ type and $b$ is a type, then $\langle \alpha, b \rangle$ is a type.

Since we are only reasoning with situations and relations between situations, we will mostly
need the following types: \(\langle s \rangle\) (i.e. arguments), \(\langle s, s \rangle\) (i.e. properties), and \(\langle s, \langle s, s \rangle \rangle\) (i.e. relations). The “intermediate” type \(\langle s, s \rangle\) corresponds to the type of X’ constituents, as we will see in a few paragraphs. Our types do not “end in \(t\)”, i.e. they do not represent functions (unary and binary) that take possible worlds and return truth-values, unlike standard Montagovian approaches (e.g. Gamut 1991; Heim & Kratzer 1998; Chierchia & McConnell-Ginet 2000). The reason for this difference is justified by one simple empirical consideration. If we aim to assign the same semantic type to Vs and Ps (since they are heads), then the use of the “classical” type \(\langle w, \langle w, t \rangle \rangle\) for Ps (our \(\langle s, \langle s, t \rangle \rangle\)) would yield the type \(\langle t \rangle\) for PPs, which would not be a proper input for a V. A V would also receive the semantic type \(\langle s, \langle s, t \rangle \rangle\), since it is a head. Consequently, it would require that its argument, a P, would be of type \(\langle s \rangle\), not \(\langle t \rangle\). We would have a type-mismatch problem.

One solution to this problem is to assume a system of covert type-shifting principles, which would “correct” the type so that the interpretive process can proceed (e.g. Partee & Rooth 1983; Partee 1987; Chierchia 1998; or categorial approaches such as Jacobson 1999,2004). Rather than this cumbersome and empirically unjustified solution, I shall instead use the very regular syntactic structure emerging from X-bar schema, and assume that each lexical item is typed as a simple or complex situation, i.e. a situation, or function/relation over situations. This is also suggested in Situations Semantics approaches (e.g. Kratzer 2007); works such as Morzycki (2005)\(^{10}\), or earlier “sortal” approaches to type theory (e.g. Keenan & Faltz 1985; Chierchia & Turner 1988; see Partee 2009 for recent discussion). The intuition is simple: if there is only one domain in the universe of discourse, then syntactic objects must denote objects within this domain, whether they are simple or complex in their semantic structure.

Since we shall mostly be concerned with heads and phrases, we will only need to explicitly represent \(\text{part-of}\) relations and arguments. A \(\text{part-of}\) relation is translated as \(\lambda x, \lambda y, x, s_y : (x \leq y)\), of type \(\langle s, \langle s, s \rangle \rangle\). An argument is translated as \(x_s\) or as \(s_s : (x \leq y)\) of type \(\langle s \rangle\), respectively translating the denotations of DPs and PPs in “P-within-P” cases. Subscripts represent the type

\(^{10}\)Note that Morzycki (2005) uses the type of properties \(\langle e, t \rangle\) for my situation type \(\langle s \rangle\). This difference is here immaterial.
of the corresponding objects. In words, a relation merges with two arguments and returns what we could call a “structured” situation; an argument is a situation, which can have its own internal structure as possibly the result of some previous operations of “conversion”\(^\text{11}\). I shall omit subscripts in the remainder of the chapter, as they will not be crucial for the exposition of the arguments. The distinction between the two denotational variants for syntactic phrases is a reflection of our choice to treat DPs as instances of Hale & Keyser (2002)’s (d)-type, and thus avoid a more fine-grained analysis of this category (but see section 4.5.4 for a brief discussion). The exact interpretation of a complex situation can be best understood via an example.

A PP such as in front of the car is the result of merging the PP in front with the P of and the DP the car, and denotes a certain (spatial) situation. We must then type the specifier PP in front and the complement DP the car as \(\langle s \rangle\). The DP the car denotes a simple situation \(c\) (say, the referent associated to a car). The PP in front denotes a complex situation, in which the figure is defined as being “oriented” with respect to the ground. This notion of orientation can be captured via the part-of relation: in front denotes a situation/region \(s\) which corresponds to the sum of the external and frontal region of an object, in turn representable as the (singleton) situations \(in\) and \(f\). The result will be the complex situation \(s : (in \leq f)\). The PP in front denotes a situation in which the “smaller” situations that make it up are taken in a certain “order”, as the “axial” interpretation of this PP suggests (cf. also Nam 1995; Zwarts & Winter 2000).

Simplifying matters somewhat, once this PP and the DP the car are merged with of, of type \(\langle s, (s, s) \rangle\), we obtain the complex situation \(s : ((in \leq f) \leq cr)\), of type \(\langle s \rangle\), as the denotation for the PP in front of the car. In words, there is a (specific) situation in which a figure occupies, it is (temporarily) part of a frontal, external position/situation defined with respect to a given car, itself defined as another position/situation. Although we maintain the basic intuitions behind several geometric treatments of Ps, we do so by using an enriched and yet simplified notion of “Space”, since this notion is best suited to capture the fine-grained semantic details of Ps and their interpretation.

\(^{11}\)The type of structured situations we will hence define are roughly equivalent to the “structured meanings” of Cresswell (1985), inter alia. This technical aspect is not crucial, here.
I will spell out the full details of this treatment in the remainder of the chapter. A few words must be spent on the introduction of the \(\lambda\)-operator in our formalism. In standard and dynamic Montagovian approaches, \(\lambda\)-operators are a way to represent unsaturated functions and relations, while in DRT they can be seen as \(\text{DRSs}\) without a “fixed” referent. Hence, \(\lambda\)-operators can be seen as a functional way to represent (generalized forms of) entailment, roughly representable in the format \((a \land b \land \ldots \land z) \rightarrow f\). In words, \(\lambda\)-terms allow to represent in a compact manner characteristic functions, lists of values which are part of a function \(f\) (e.g. Gamut 1991; Heim & Kratzer 1998; Chierchia & McConnell-Ginet 2000). Functions can be then combined with arguments and, via function application, compute a point-wise as their output: from applying \(\lambda x. f'(x)\) to \(d\), we obtain \(f'(d)\).

In our DRT framework, function application (also known as \(\text{Apply}\)) is the semantic counterpart of merge (e.g. Muskens 1996; van Eijk & Kamp 1997; Blackburn & Bos 1999,2005). When a function and an argument are merged, the resulting \(\text{DRS}\) can be roughly represented as: \(\lambda x.[f'(x)] + [d] = \lambda x.[f'(x),d]\), a \(\text{DRS}\) including both function and argument as distinct conditions. Given the conjunctive interpretation of conditions in \(\text{DRSs}\), we have \(\lambda\)-conversion (i.e. \(\lambda x.[f'(x),d] = [f'(d)]\)) as a form of \(\text{modus ponens}\), i.e. \(\lambda x.f'(x)(d) = f'(d)\) is equivalent to \((f(d) \land (f(d)\ldots f(z))) \rightarrow f(d)\). Function application has the effect of “reducing” the type of expressions: a relation of type \(\langle s,\langle s,s\rangle \rangle\) takes an argument of type \(\langle s\rangle\) as an input and returns a property of type \(\langle s,s\rangle\) as an output. A property takes an input of type \(\langle s,s\rangle\) and returns a “structured” situation of type \(\langle s\rangle\) as an output. Hence, “basic” situations have the type of complex situations, because of this process.

The \(\lambda\)-operator, via function application, allows us to reduce the complexity of \(\text{DRSs}\) by substituting an open-value referent with a specific referent, in a manner similar to anaphora res-

---

12 A technical note: the entailment (binary) operator “\(\rightarrow\)” can be defined via \(\text{merge}\) and negation, as usually assumed in standard logic (see e.g. Landman 1991:1-7).

13 Blackburn & Bos (1999,2005) use the symbol “@” to denote application, i.e. we have \(f@x\) or \(f@d\), rather than \(f(x)\) or \(f(d)\). Also, as operators (e.g. negation), \(\lambda\)-terms are prefixed to \(\text{DRSs}\), which in turn have an empty universe, to be filled after \(\beta\)-conversion with the “captured” referent. I represent empty universes (for \(\lambda\)-terms) by simply omitting the relevant brackets. Note that I represent arguments as special \(\text{DRSs}\), since they lack conditions (cf. Kamp et al. 2005:772, def. 0.2.).
olution, and thus to have a tight correspondence between syntactic and semantic objects. This correspondence is known as the Curry-Howard isomorphism: for each step in a syntactic derivation, we will have a corresponding step in a semantic operation, representable via a combination of (specific) $\lambda$-terms (e.g. Blackburn et al. 2006:50-54). In our specific case, heads will correspond to (part-of) relations, bar elements to properties/partially saturated relations, phrases to arguments, either simple or complex.

For Ps, the following equations represent this isomorphism: $pp = \langle s \rangle$, $p' = \langle s, s \rangle$, $p = \langle s, \langle s, s \rangle \rangle$. DPs, which are typed as $n$, will also be semantically typed as $\langle s \rangle$, as per discussion. As syntactic arguments of a P (e.g. in front and the car), both PPs and DPs will denote semantic arguments of the relation denoted by a P (e.g. of). Once we can define the relation between syntactic and semantic objects, we can define a (partially ordered) set of semantic sorts. This set is represented as $SORT = \langle LEX, \leq \rangle$, which will be defined as the set of interpreted DRSs, in turn the set of lexical items (i.e. $LEX$) which will have an interpretation in the model, given by the part-of relation “$\leq$”. In words, the partially ordered set $SORT$ defines all the set of all possible sub-sets (sorts, hence the name) of possible worlds within our universe of discourse (cf. again Asher & Pustjovsky 2004; Asher 2011) and the relations between these sets, including the ones denoting spatio-temporal regions. The isomorphism between syntactic and semantic objects can be roughly captured as follows. For each $d \leq LEX, d \leq TYPE$, the following holds: $[d] := d$, i.e. for each type of syntactic object in our lexicon, there will be a corresponding semantic object in the universe of discourse. In words, a P always denotes a relation (e.g. in front of), a PP and DP an argument (e.g. respectively in front of the table, the boy), and so on. The function $[.]$ is the interpretation function over our universe of discourse. Here onwards, I shall use the symbol $d$ to represent an interpreted term (i.e. a DRS, and not a discourse referent), when this should not give rise to confusion.

At this point, we can give a basic template for the interpretation of our Ps. The following facts about interpretation hold: $[p] = \lambda x_s. \lambda y_s. s : (x \leq y)$, $[p'] = \lambda y_s. s : (k \leq y)$, $[[pp]] = s : (k \leq n)$, and $[[n]] = s$. In words, a P (i.e. a head) denotes a relation, a bar-constituent $P'$ denotes
a relation in which one element has been merged (i.e. a property). So, both verbs (to be, come, go, etc.) and Ps (e.g. to, in front of, from, behind, etc.) denote relations between situations. Both PPs and “simple” constituents (DPs) denote a situation, but do so in rather different ways: DPs denote single referents/situations (e.g. the boy), PPs denote complex situations, e.g. the location of an figure as being outside and aligned with respect to the frontal position of the ground (e.g. in front). This is consistent with the Curry-Howard isomorphism: for each syntactic category we have a corresponding semantic object, although different sorts of semantic objects may be “born” at the same semantic type. In words, both a DP and PP denote a situation, an object of type ⟨s⟩, but a DP denotes a “simple” situation (e.g. a referent representing “Mario”), whereas a PP may denote a situation which is constituted of different parts, or “smaller” situations (e.g. a situation in which a boy is in the garden).

One consequence of this approach is that, regardless of the complexity of a sentence, it will always denote a situation, once the syntactic process is over: we cannot access directly its truth-conditions. I shall assume that, in order to retrieve its truth-conditions, a truth- or assertion-operator may merge with a sentence and map the corresponding denotation onto a truth-value. This operator is represented as T, and is of type ⟨s,t⟩. It may be seen as a general instance of more specific operators specifying the assertoric content of a sentence, as suggested in fine-grained treatments of the “C” category (e.g. Rizzi 1997)14.

Following DRT and other dynamic approaches, I shall assume that a true sentence or proposition is obtained via existential closure on situations15 (e.g. Heim 1982; Chierchia 1995b:86; Zeevat 1999:2; SDRT treatments such as Asher & Lascarides 2003; Partee 2009; etc.). Informally, when a derivation ends, this operator evaluates whether the situation denoted by the resulting sentence is true or false, i.e. if there is at least one such situation in the model of discourse that corresponds to the evaluated situation (true) or not (false). For instance, once a

14In our approach to syntactic structure, this element would however be merged at the rightmost part of clausal structure, rather than in its leftmost position, as in Rizzi (1997). Since I have rejected the FSEQ assumption of cartography, this issue is immaterial here.

15In DRT, a proposition is a set of possible worlds that verify a DRS, as per our definition (cf. Kamp et al. 2005:771).
sentence such as *The man sleeps in the garden* is obtained and interpreted as a situation $s \leq s'$. then this operator tests whether there is at least one situation with that structure in the model of discourse. If there is, then the sentence will considered true (or false).

Syntax-wise, this “special” constituent belongs to the category $\text{d}$, with $\text{d} = \{\text{p}, \text{n}, \text{v}\}$. Intuitively, this category adds no new syntactic units to the derivation, hence it acts as a “halt” command to the syntactic process, which in turn licenses the evaluation of situation derived from this process. The $\text{d} + \text{pp}_v = \text{pp}_v$ holds, so the assertion-operator states that a sentence is part of the set of sentences (complex lexical items) of a grammar. From a semantic perspective, in our Boolean format existential closure is represented as $\lambda s . \bigcup(s)$, i.e. an operator that takes the set (union) of situations which describe a certain complex relation holding between “smaller” situations, hence binding an otherwise free referent.

If a sentence denotes a relation, we will have $\lambda s . \bigcup(s : R(x,y)) = \bigcup s : R(x,y)$, a relation which is true for at least one situation in a model. Since “$\bigcup$” is the Boolean counterpart of the existential quantifier “$\exists$”, the equivalence between this notation and standard first-order notation should be obvious (e.g. $\exists s[R(s,x,y)]$ is equivalent to $\bigcup s : R(x,y)$), in our DRT format. Also, since existential closure defines a function that “embeds” a set of situations in a model, it is (roughly) equivalent to the definition of truth found in more “standard” theories (e.g. Landman 1991:8). I shall only consider cases in which a sentence is evaluated as true, for the remainder of the chapter. In chapter 5 I shall discuss other possible values that this operator assigns to sentences.

From this point onwards, I shall drop type subscripts in the formulae, as they will not be crucial. The treatment I outlined in this section permits us to offer the following answer to the third research question, which is as follows:

**Q-3:** *Ps can be distinguished in different types because they denote different part-of relations defined over the domain of situations;*

This answer tells us that the difference between static Ps such as *in front of* and dynamic ones such as *to* is not arbitrary, but reflects a general ontological property of how semantic domains
are organized. If the semantics of natural language expresses a distinction between states and events, and thus between relations over states and relations over events, then this distinction will be expressed by those parts of speech that denote relations. We can predict that English Ps can (and should) be distinguished along these lines, and thus we can predict that they will have the properties outlined in our five predictions. We are now in a position of not only being able to account the problematic data in (102)-(112), but to actually predict their existence as a consequence of our unified theory of Ps. The remainder of this chapter is thus based on testing whether these predictions are correct or not, and thus to test the empirical adequacy of our extended DRT fragment.

4.5 The Semantics of Ps: Analysis of the Data

In this section I shall offer an analysis of the P data, testing whether these data support the predictions offered in the previous section or not. Before I start, I shall sketch a standard semantic classification of (spatial) Ps that will help us in dividing the data into easily analysable classes.

Spatial Ps, at least since Bennett (1975) and Cresswell (1978), are usually distinguished into locative Ps or just locatives, such as at, in, on; and directional Ps or just directionals, such as to, towards, along. Locatives denote the static or unchanging location of a figure with respect to a ground (e.g. in or in front of), whereas directionals spell out a temporal or causal dimension of meaning, e.g. that the location of a given figure is the result or the beginning of a previous an event of motion (e.g. to or from, respectively). Locatives can be divided into two different classes, non-projective locatives (e.g. at) and projective locatives (e.g. above). Projective Ps express the location of a figure with respect to the axes of a ground, while non-projective locatives mostly express more basic spatial and non-spatial (topological) notions, such as inclusion, support or overlapping. Most, if not all Ps appear to be ambiguous in their interpretation between a locative and directional meaning (e.g. along: Parsons 1990; Zwarts 2006), and virtually all Ps can be anaphorically related with indexical PPs, such as here and there (e.g. Cresswell 1978).
I shall follow this classification of the data, by first analyzing the semantics of locative PPs (section 3.5.1); the semantics of directional Ps (section 4.5.2); the semantics of the still understudied ambiguous cases (4.5.3). I shall then sketch some rather basic considerations on other categories (Vs, DPs) in order to define the place of Ps in the clausal architecture, in section 4.5.4.

4.5.1 Locative Ps

Locative Ps, as we have seen, can be distinguished between projective and non-projective Ps. Some non-projective locative Ps are the following:

(140) Mario is at the cinema

(141) The pen is in the box

(142) The ring is on the finger

In (140), Mario is somewhere close to the cinema, if not inside it: *at* seems to express a rather general (spatial) relation holding between Mario and the cinema, or perhaps between Mario and one of the cinema’s “parts” (e.g. Herskovits 1986; Nam 1995; Zwarts & Winter 2000). In (141) the pen is occupying the same location of the box as if the two entities were points in space (e.g. Jackendoff 1983; Landau & Jackendoff 1993). In (142), the ring is occupying the external space immediately adjacent to the finger, and being prevented from falling because of its (implicit) connection to the finger (e.g. Herskovits 1986; Wunderlich 1991, 1993). Non-projective locatives convey basic spatial relations (in our extended meaning), but do not include more fine-grained information regarding the orientation of the figure. As observed by Parsons (1990), these locatives do not entail any other states or events than themselves, as illustrated in (142): they express inherently “static” relations over time. I shall focus on projectives first, in order to introduce some basic notions.

Some projective locatives are instead the following:

(143) The pen is in front of/behind the box
(144) The pen is to the left of/to the right of the box
(145) The pen is above/behind the box
(146) The pen is inside/outside the box
(147) The stadium is North/South of the river
(148) The city is at the edge of the southern border
(149) The engine is in the front of the car

In (143)-(146), the pen’s position is defined via one axis (or projection) going from the box to one part of the box, respectively its front, left side, positive vertical position or “internal” side. The figure is generally understood to be in an external position defined with respect to the ground, at a variable distance, with a case limit being adjacent to the relevant portion of space (e.g. the front’s box: Zwarts 1997; Zwarts & Winter 2000). Projective locatives may also express orientation with respect to a certain environmental position, as in (147), but their syntax and semantics appears to pattern in line with other projectives (e.g. Kracht 2004, 2008). The projectives in (143)-(147) all have an antonym-like term, juxtaposed to the relevant P in the examples (cf. in front of/behind). Certain complex Ps, such as at the edge of or in the front of in (148) and (149), are sometimes considered as non-projective locatives, derived from projective ones, since they express the location of the figure as coinciding with the “end-point” of a certain orientation (e.g. Cresswell 1978; Zwarts & Winter 2000; Levison & Meira 2003; Svenonius 2006). Projective locatives are also “static”, given the entailment pattern(s) displayed in (149).

These two classes of Ps are apparently different in the specific type of information they express, but their basic syntactic and semantic properties turn out to be homogeneous. All of these Ps constitute cases of “P-within-P” structures, which take a DP and a PP as their argument phrases. I shall repeat the basic structure of PPs offered in (82) as (150) in three of their instantiations, and offer some sample structures for the Ps we have listed so far:

(150) \([pp]([PP]([P]DP))\) / \([pp][PP][P]DP\) / \([pp][PP](P)DP\)
(151) \([pp](at)(P)[the box]\)
The minimal differences across examples (151)-(155) is that, in some projectives, *of* may spell out an overt realization of the P head, in some cases. For non-projective Ps, this does not occur, as (151) shows, while in some cases other morphemes may be assumed to realize this relation (e.g. *-side* as in *inside*, (152)). These examples instantiate the “simpler” structure on the left, in (150), while example (155) instantiates the “P-within-P” structure on the right in (150), with *the* being a particular realization of the “internal” P head, as we discussed in 3.

From a semantic perspective, both types of locatives offer simple evidence that the first and third prediction are borne out, since they respectively express location as a relation between states/regions (first prediction), which can be interpreted in either “direction” of a virtual axis (e.g. *inside* vs. *outside*, third prediction). The P head *of* (or possibly the morpheme *-side*) will be of type \(<s, (s, s)>\) and denote the relation \(\lambda x. \lambda y. s : (x \leq y)\). Since all locative Ps denote inherently “static” relations, this relation is restricted to the domain of states, i.e. that we have \(x, y, s \leq S\).

The PPs and DPs in argument position will be of type \(<s>\), with DPs invariably denoting a simple situation (e.g. *b* for *the box*). I shall then assume that PPs in specifier position may denote a complex situation. A PP such as *in front* denotes the complex situation \((in \leq f)\). The relation between the two underlying states *in* and *f*, standing for the arguments PP *in* and *front*)\(^{16}\) is expressed by the phonologically null head. Its antonym P *behind* can be represented as \(s : (b \leq h) = s : (\neg(in \leq f))\), the complex situation involving the complementary positions/situations of *in front*. This treatment can be extended to all such pairs, so that we can easily capture their antonym-like nature, as well as the intuition that these PPs combine with *of* to denote “axial”

\(^{16}\)From here onwards, I shall assume that the morpheme *in*, when occurs as part of other Ps, does not denote “inclusion”, but a position definable as “external position”. This is based on etymological grounds (cf. also *on*). See Svenonius (2006:79-84) for further discussion.
spatial relations, i.e. projective Ps, ultimately situations in which the constituting parts are taken in a certain “order”.

In order to illustrate how the semantics of these Ps emerges from the combination of the underlying syntactic units and their interpretation, I shall offer a sample derivation to show the interpretation of a PP, e.g. in front of the box in (153). I shall leave aside syntactic matters, and only represent the semantic component of derivations. I shall “enrich” derivations by adding the semantic type for each interpreted object, to the right of the corresponding interpretation.

The derivation the following:

(156)  
\[ t. \langle \text{pp}, \text{in}, \langle s \rangle \rangle, \quad \text{(assumption)} \]
\[ t + 1. \langle p, \lambda x. \lambda y. s : (x \leq y), \langle s, \langle s, s \rangle \rangle \rangle, \quad \text{(assumption)} \]
\[ t + 2. \langle \text{pp} + p = p', (\text{in}) \lambda x. \lambda y. s : (x \leq y) = \lambda y. s' : (in \leq y), \langle s, s \rangle \rangle \quad \text{(m. intr., f. application)} \]
\[ t + 3. \langle \text{pp}, \text{front}, \langle s \rangle \rangle, \quad \text{(assumption)} \]
\[ t + 4. \langle (p') + \text{pp} = \text{pp}, \lambda y. s' : (x \leq y)(f) = \lambda y. s' : (in \leq f), \langle s, s \rangle \rangle \quad \text{(m. intr., f. application)} \]
\[ t + 5. \langle p, \lambda x. \lambda y. s : (x \leq y), \langle s, s, s \rangle \rangle \quad \text{(assumption)} \]
\[ t + 6. \langle \text{pp} + p = p', (\text{in} : (in \leq f)) \lambda x. \lambda y. s : (x \leq y) = \lambda y. s : (s' : (in \leq f) \leq y) = \lambda y. s : (in \leq f) \leq y), \langle s, s \rangle \rangle \quad \text{(m. intr., f. application)} \]
\[ t + 7. \langle n, b, \langle s \rangle \rangle, \quad \text{(assumption)} \]
\[ t + 8. \langle p' + n = \text{pp}, \lambda y. s : ((in \leq f) \leq y)(b) = \lambda y. s : ((in \leq f) \leq b), \langle s \rangle \rangle \] \quad \text{(m. intr., function application)}

In words, the PP in front of the box is interpreted as denoting a relation between a box, taken as a state/region of space, and the two ordered states/regions denoted by in front. The “projective” interpretation of this P emerges as the result of taking the specifier PP as denoting a complex situation “embedded” within the situation denoted by of, and the same consideration can be extended to virtually any other projective PP. The derivational steps from \( t \) to \( t + 4 \) display how
the interpretation of this PP is derived, according to the “P-within-P” hypothesis. The remaining steps show how this complex situation becomes part of the “larger” PP, in front of the box. Importantly, the situation denoted by in front becomes part of the one denoted by the “whole” PP, once it is merged with the relation denoted by of. Informally, the “oriented” region/situation denoted by in front is defined as ordered with respect to some other region/situation, which turns out to be the ground, the situation denoted by the box, via transitivity. One aspect of the fifth prediction is borne out: we obtain a single, “structured” situation denoted by in front of out of two structured situations, respectively denoted by in front and of.

A virtually identical treatment can be offered for non-projective PPs. In this perspective, a minimal syntactic difference between e.g. in and at the edge of is that the first P is assumed to combine with a phonologically null head, realized by of in the second P (cf. also inside). Semantically speaking, this type of locative Ps appear to denote “regions”, states that apparently lack an internal structure (“orientation”). If in denotes the ground as a single position, then at the edge of denotes the most extreme, external point of a certain ground (in (148), the southern border).

Intuitively, this non-projective interpretation for at the edge of the border can be roughly represented as the situation s: {a, ed} ≤ sb: the combination (sum) of an external and extreme position, which is part of the extended space corresponding to the southern border sb17. If projective Ps denote ordered regions/situations, then non-projective Ps denote unordered situations (or sums thereof). Both denote structured situations, but differ with respect to the “type” of structure they denote: a sum of situations (non-projective Ps), a part-of relation between situations (projective Ps). Indirectly, this difference reconstructs the contrast between “regions” and “vectors” found in the reviewed literature, although it does so by highlighting the compositional processes underlying Ps.

17In order to give a more adequate explanation of this interpretation, we would need a more precise semantics of Ds, which is beyond the scope of this thesis: this coarse-grained analysis is precise enough for the data at hand. See section 4.5.4 for discussion.
The treatment for “complex” non-projective Ps can be extended to “simple” ones straightforwardly. I shall focus on two special cases, \textit{in} and \textit{at}, since they seem to denote a special type of “region”. Regarding \textit{at}, some works assume that it denotes external, close relation to a ground (e.g. Herskovits 1986; Zwarts & Winter 2000), others that it denotes a general locative P, without a specific interpretation (e.g. Kracht 2002; Hale & Keyser 2002; Levinson & Meira 2003). Regarding \textit{in}, if this P denotes “spatial identity”, one can expect certain complex entailment patterns to occur (cf. Zwarts & Winter 2000:170-175).

The relevant examples are the following:

(157) Mario is at the cinema\{=Mario may be in front of the cinema OR Mario may be in the cinema OR Mario may behind the cinema\}

(158) Testaccio is in Rome, Rome is in Italy, and Italy is in Europe\{= Testaccio is in Europe\}

In (157), \textit{at the cinema} entails that several, alternative-like positions for Mario are possible, including one in which he is in the cinema. In (158), we can conclude that the quarter “Testaccio”, being in Rome (and Italy), is also in Europe. The first entailment pattern suggests that \textit{at} denotes the sum of all possible states (i.e. \(S\)) that are part of the reference state denoted by the ground. We can represent this relation as \(s: (S \leq y)\), which is equivalent to \(s: (S = y)\): if the ground is the maximal situation in this relation, then it will be identical to the set composed of all situations that are part of it (i.e. the front, or the inside, or behind, etc).

The second entailment pattern can be captured by assuming that \textit{in} denotes a situation which only includes the empty state, e.g. \(s: (0 \leq y)\). In this case, the figure will identified with the “center” of the ground, the region/situation which is part of all other situations/regions making up the ground (i.e. the empty set). Consequently, if Testaccio is identified with the center of Rome and Rome is identified with the center of Italy, then Testaccio is identified with the center of Italy. I shall discuss more in detail these relations and their role in a general theory of Ps in section 4.5.3, however, since the precise interpretation of these Ps becomes clear once we also discuss the semantic contribution of the copula.
This analysis of locative Ps also allows us to test our fourth prediction in some thorough detail. First, under this approach we predict that projective Ps have a derived region/coordinated interpretation, which allows the combination of Ps via coordination/disjunction, to form complex Ps. Second, projective Ps can be modified by measure phrases, whereas non-projective ones cannot. However, before testing these predictions, we need to spend a few words on coordinated Ps and measure phrase modification.

The coordination of heads of phrases has been a topic of several syntactic semantic works in the literature (e.g. Kayne 1994; den Dikken 2006 on their Syntax; Partee & Rooth 1983; Keenan & Faltz 1985; Link 1998; Winter 2001; on their Semantics). The “P-within-P” assumption predicts that e.g. two coordinated PPs, such as above or below the table, will correspond to the structure $\left[PP[PP\text{above}]or[PP\text{below}]](P)\text{the table}\right]$. The two lexical items or or and, or other coordinating elements, are syntactically Ps (e.g. Emonds 1985; Kayne 1994), and instantiate the “inner” P head, in this construction, similarly to the D the occurring in the PP at the edge. Consequently, the interpretation of a coordinated P will amount to the “logical” combination of the two (or more) combined PPs, as per fifth prediction.

Interestingly, the fourth prediction states that complex Ps may also have a non-projective reading, i.e. they may denote the position of a figure as being in a certain region, rather than at the end of a certain set of oriented regions. The subtle difference between the two readings is that, on the projective reading, above or below the table denotes the figure’s position as the end of one sequence of oriented regions, and thus can only be either in the upper vertical region or in the lower vertical region; on the non-projective reading, any vertical position is acceptable, as a position occupied by the figure.

Measure phrases denote specific values of a “scale”, a partial order over a given domain, and must combine with other elements denoting partial orders (e.g. Kennedy 1999, 2007). A measure phrase such as ten meters denotes a set of situations ordered with respect to their spatial extension. When it combines with a PP such as in front of the house, it will “select” the situation in which the frontal position of the figure will also be at ten meters of distance from the house.
Importantly, measure phrases can only combine with PPs denoting a “closed” scale, i.e. PPs that do not denote situations which potentially have an infinitesimal or infinite extension (Winter 2005, 2006).

Some relevant examples are the following:

(159) The stone is above or below (P) the table
(160) The stone is (above and in front) of the table
(161) The stone is below the box and on top of the table
(162) The car is ten meters behind or in front of the house
(163) #Mario is ten meters at the desk
(164) # Mario is ten meters at the edge of the desk

Examples (159)-(161) show that projective Ps can be combined to intuitively denote certain positions in which the stone is located. For instance, the PP *above or below the table* denotes a situation in which the stone is either in some position along the upper or the lower vertical axis. In each case, aside a standard projective interpretation, a non-projective one is also possible. For instance, two projective Ps such as *above* and *in front of*, are interpreted as non-projective manner (i.e. we have $s':\{a,v\}$ and $s'':\{i,fr\}$, the unordered regions/situations denoted by these Ps) and then combined together via coordination\(^{18}\) (i.e. we have $s'':\{a,v\} \cap s':\{i,fr\}$), which in this case denotes the “product” of the two coordinated referents, i.e. the region that is at the same time “above” and “in front” the table, as per intuitions (i.e. $s:\{a,v,i,fr\}$, with $s = s' \cap s''$ cf. Winter 2001:ch.2; *inter alia*).

In this case, the inclusive reading is possible, since it denotes a region which includes both portions of the vertical axis. Furthermore, we can have the coordination of “full” PPs, each combining with a different DP, as in (161). In this case, however, the sentence is interpreted only in a non-projective way: the stone is in some position that is in the lower vertical region of the

\(^{18}\)Syntax-wise, I assume that *and* is a P head that combines with two PPs, *above* and *in front of*, in a manner entirely equivalent to *or* (cf. Kayne 1994; Hale & Keyser 2002:ch.6).
box, and which is also in the upper, adjacent vertical region of the table. Our proposal can easily account facts which are somewhat problematic in other proposals, in a relatively straightforward way (e.g. Zwarts & Winter 2000:197-200). Note that in (159) represent the phonologically null head (P), which takes the two coordinated Ps in the specifier position: in (160), this element is overtly realized as of, and thus combines with the two coordinated PPs above and in front, as per predictions.

Examples (162)-(164) show that only projective Ps (e.g. in front of, but not at) can combine with measure phrases, according to predictions and consistently with previous literature, but also that this pattern can be extended to Ps that are “born” at the level of a non-projective interpretation (e.g. at the edge of), a fact that has apparently not been noticed in the literature. Importantly, when measure phrases combine with coordinated Ps, the only possible reading is the projective one: in (162), we understand that the car is either ten meters behind or ten meters in front of the house but, for obvious reasons, it cannot be in both positions. This fact is also straightforward on our analysis, but problematic in other approaches (e.g. Nam 1995: ch.4). Our theory can not only predict and account several “core” facts about Ps, but it can also predict and account subtler, more “peripheral” facts about this category and its interaction with other categories.

Before moving to directional Ps, we can observe that some data which were outside our reach at the beginning of this chapter have now become evidence supporting our predictions. I shall briefly discuss this aspect by repeating (102), (103), (110)-(112) respectively as (165)-(169) to illustrate the point:

(165) Mario is in front of the blonde girl
(166) Mario is behind the blonde girl
(167) Mario sits ten meters in front of the desk
(168) *Mario sits ten meters at the desk
(169) Mario has gone to the pub and is sitting at the counter
These data support our predictions in the following ways. They support the prediction that some Ps denote static relations (first prediction); that Ps can stand in an antonym relation while having the same underlying syntactic structure (third prediction, (165) and (166)); that projective (i.e. relational) Ps can combine with measure phrases, also denoting relations, while locative Ps can’t (fourth prediction, (167) and (168)); and that Ps can be combined via coordination, to denote complex locations/directions, possibly related via entailment (fifth prediction: (169)). The second prediction, that certain Ps denote “dynamic” relations between situation, must still be tested. The relevant data for this and the other predictions will be discussed in the following sections.

4.5.2 Directional Ps: the Facts

Directional Ps consist of Ps that cannot combine with the copula to be and other “static” Vs, since they inherently denote change of location and the direction of this change (e.g. Zwarts 1997; Zwarts & Winter 2000). A non-exhaustive list includes to, from, into, through, across, along, towards, as in the following examples:

(170) Mario is going to the cinema
(171) *Mario is to the cinema
(172) The bird has flown up to the sky
(173) *The bird is up to the sky
(174) Mario is coming from the shop
(175) *Mario is from the shop
(176) The car has gone through the mountains
(177) *The car is through the mountains
(178) Mario is going across the river
(179) *Mario goes across the river
The examples in (170)-(183) show that directional Ps cannot combine with the copula and thus license a locative interpretation. Although these Ps can combine with other “static” verbs, often denoting orientation and “posture” (e.g. *stretch, lie*) and may combine with the copula in certain idiomatic readings (e.g. *to be through with problems*), in their inherent “spatial” reading they strongly convey some relation between situations involving a change from one state to another (Fong 1997, 2001).

Syntax-wise, the structure of these Ps is the same of locative Ps, e.g. *up to* being a case of “P-within-P” structure. Their interpretation, as illustrated by Parsons (1990:84-90), consists of a figure ending up in a ground, as a consequence of an event of motion. I shall repeat these entailments in (184) and (185) to make the point precise:

(184) Mario has went to the cinema $\models$ Mario was at the cinema

(185) Mario has gone across the road $\models$ Mario is across the road

In words, if Mario went to the cinema, this entails that Mario was at the cinema as a consequence of this event of motion. For other directionals such as *across*, this entailment seems to hold when the verb is in present perfect form, as well (e.g. Parsons 1990:ch.13). If Mario has gone across the road, he is across the road, i.e. at the ending point of an “across path”, as a consequence of this event of motion. Directionals denote the same underlying sequence or of situations involving an event of change, but may differ with respect to the precise states that are involved in this “sequence” or “shape” of states (e.g. Zwarts 2004, 2006). Directionals thus support the second prediction, as they denote transitive closure relations, i.e. relations involving total orders between events and states, over the domain of situations. I shall discuss the subtler “static”
cases and some consequences regarding the entailment patterns in section 4.5.3, focusing in this section on the basic dynamic interpretation.

Following Parsons (1990); Fong (1997), I shall assume that a P such as to includes an implicit irreflexive relation between an event of change and a “resulting” state, which can be represented as \( \lambda x.\lambda y.e < s : (x \leq y) \), of type \( \langle s, \langle s, s \rangle \rangle \). In words, we explicitly represent that dynamic Ps include an event of change which is distinct from the certain (resulting) position for the moving figure, and that Ps such as to and from explicitly express this total order. In particular, according to the third prediction, from will denote the complementary order of events, i.e. \( \lambda x.\lambda y.\neg(e < s : (x \leq y)) \). Complex Ps such as up to show that this relation can be restricted to denote a certain resulting (or originating) state, with a complementary example being e.g. away from (e.g. Jackendoff 1983; Zwarts 2005).

The structure for these Ps is:

(186) \([pP](P)[to][\text{the cinema}]\]

(187) \([pP][up](to)[\text{the sky}]\]

(188) \([pP][through](to)[\text{the mountains}]\]

(189) \([pP][across](to)[\text{the street}]\]

Examples (187) and (188) should be straightforward. Examples (189) and (190) represent an interpretation of through, or similar other Ps (here, across) in which the car may still be still “inside” the mountains after moving (but in a different position than the starting one), but at some point it will be involved in an event of motion. Different interpretations can arise if the ground is understood to have a finite shape: going through the tunnel may either be understood as going into and out of the tunnel, or as staying inside the tunnel for the whole “journey” (e.g. Zwarts 2005). As lexical matters regarding the interpretation of DPs and pragmatic factors play a role in these cases, we can safely leave them aside without the risk of offering a too coarse-grained analysis.
The specific interpretation of the phonologically null head depends to a good extent from the related verb: one may go or come through some ground, in a sense realizing the two complementary orders of events. The ability of these Ps to interact with Vs having a “weaker” ability to denote change of position (e.g. walk) suggest that this phonologically null head may denote various types of orders over situations. For commodity, we may just assume that this head denotes a pre-order in its most basic meaning, which may be under-specified for polarity (i.e. a null to or from), however. Informally, while directional Ps denote relations over situations involving a distinct event of change, the exact “direction” of change may be expressed by other parts of speech than the P.

The interpretation of these Ps is thus the following. I will offer two sample PP derivations, one for up to the sky and the other for through the mountains, to highlight the slight interpretive differences. The derivations are:

(190) \[ t. \quad \text{[up]} := up, \langle s \rangle \] (assumption)
\[ t + 1.\text{[[to]]} := \lambda x.\lambda y.(e < s : (x \leq y)), \langle s, \langle s, s \rangle \rangle \] (assumption)
\[ t + 2.\text{[[up]]} \text{[[to]]} := (up)\lambda x.\lambda y.(e < s : (x \leq y)) = \]
\[ \lambda y.(e < s : (up \leq y)), \langle s, s \rangle \] (m. intr., f. appl.)
\[ t + 3.\text{[[sky]]} := sk, \langle s \rangle \] (assumption)
\[ t + 4.\text{[[up to]] ([[the sky]])} := \lambda y.(e < s : (up \leq y))(sk) = \]
\[ e < s : (up \leq sk), \langle s \rangle \] (m. intr., f. appl.)

(191) \[ t. \quad \text{[through]} := tr, \langle s \rangle, \] (assumption)
\[ t + 1.\text{[[to]]} := \lambda x.\lambda y.(e < s : (x \leq y)), \langle s, \langle s, s \rangle \rangle \] (assumption)
\[ t + 2.\text{[[through]]} [[\text{[to]]]} := (up)\lambda x.\lambda y.(e < s : (tr \leq y)) = \]
\[ \lambda y.(e < s : (tr \leq y)), \langle s, s \rangle \] (m. intr., f. appl.)
\[ t + 3.\text{[the tunnel]} := tn, \langle s \rangle \] (assumption)
\[ t + 4.\text{[[through(to)]]} [[\text{[the tunnel]]]} := \lambda y.(e < s : (tr \leq y))(tn) = \]
\[ e < s : (tr \leq tn), \langle s \rangle \] (m. intr., f. appl.)
In words, the derivations in (190) and (191) introduce a relation between states which also explicitly states the final state (up, through) as being the resulting position of an event of motion. Intuitively, the interpretation of up to denotes that a certain “up” region, coinciding with the sky, is reached as a consequence of an event of motion. In case to occurs without any PPs in the specifier, the “result” state can be understood as a phonological null at, as the Parsonian entailments (in (117)) suggest. So, going to some location may entail that one ends in any state defined with respect to a ground.

The interpretation of through the tunnel also follows the Parsonian assumptions, and says that an event of motion results in a state in which the figure has covered the whole inner section of the tunnel, i.e. he has been through the tunnel. In this and other cases, the verb’s contribution also plays an important role in interpretation in making fully explicit that the moving figure may have reached the end (beginning) of a certain “path” expressed by a P, as the example in (186) shows. I shall defer this discussion to section 4.5.4, however.

The analysis I have offered so far for directional Ps is also consistent with the fourth prediction, as in the case of locative Ps. Some examples supporting this prediction are the following:

(192) Mario has gone into and out of the town
(193) The bird is flying above or below the cloud
(194) Mario has gone from the store to the shop
(195) *Mario is going one kilometer to the store
(196) The car has gone fifty kilometers across the mountains and through the woods

Examples (192)-(196) show that directional Ps may receive a coordinated interpretation and thus be coordinated via to form complex Ps. In the case of into and out of in (192), Mario first enters the town, then he exits it. In the case of above or below in (193), we may just don’t know in which region he is flying, with respect to the cloud, and we may even assume that he may fly in
both regions, but at different intervals of time (as per non-projective reading). Similar accounts and predictions for “dynamic” Boolean constituents, although via complex systems of “type-lifting”, are offered in Partee & Rooth (1983); Winter (2001). In the case of from the store to the shop in (166), the apparent juxtaposition of PPs can be treated as a particular case of “P-within-P” configuration, with from the store being in the specifier position of to the store. The resulting interpretation will roughly be $e < s : ((\neg s < e) \leq st)$, in turn equivalent to $\neg s < e < s : (s \leq sh)$, which intuitively says Mario goes from store to shop.

Examples (194) and (195) show that modification patterns appear to be the same as in the case of locative Ps: while directional Ps denoting an underlying “region” (e.g. to) cannot combine with measure phrases, the ones with a projective flavor can freely combine with measure phrases (e.g. across). Note that, in cases such as (196), we understand that the total distance covered by the car is distributed, somehow, between the “journey” across the mountains and the one through the woods. For instance, the car has covered 25 kilometers while going across the mountains, and then another 25 while going through the woods, for a total of 50 kilometers.

By this point of our discussion, the following data have also become evidence in support of our theory, rather than recalcitrant data. I shall repeat sentences (104)-(106), (109), (112) and (113) as (197)-(202) to briefly illustrate the point:

(197) Mario has gone to the shop
(198) Mario has arrived from the shop
(199) Mario is at the shop
(200) Mario has gone ten meters towards the desk
(201) *Mario has gone ten meters to the desk
(202) Mario has gone to the pub and is sitting at the counter

The predictions supported by these data can be briefly stated as follows. Some Ps denote “dynamic” relations, relations that denote a distinct change of position, such as to and from (second prediction, (197) and (198)). For a P such as a to, which expresses directed motion in one
“verse”, the complementary P from exists (third prediction, (197)-(198)). Directional Ps such as towards can be modified by measure phrases, in a certain principled way (fourth prediction, (200) and (201)); Ps such as to and at are logically related via entailment relation, since the corresponding sentences are also connected via this relation (fifth prediction, (197) and (199)). Also, directional Ps can be coordinated with other Ps, as a consequence of the “P-within-P” hypothesis (also fifth prediction, (202)). In words, directional Ps amply support the predictions I have offered so far regarding Ps, much like locative Ps. Other cases still in need of a precise treatment will be discussed in the next two sections.

4.5.3 “Intermediate” Ps, and other Phenomena

In this section I shall discuss how our theory can predict and thus account the recalcitrant data that are still in need of an explanation. I shall start by discussing the logical relations holding between sentences such (106) and (107) as an instance of the sub-set relation, relations that are predicted by the fifth prediction. These sentences are repeated here as (203) and (204)\(^{19}\):

\[ (203) \quad \text{Mario is in the shop} \]
\[ (204) \quad \text{Mario is at the shop} \]

Under our approach, these two Ps receive a “special” type of semantics, since they respectively denote the most specific and the most general (static) relation holding between two referents. While in captures a relation in which the figure is related to no specific situation except itself (i.e. \(\emptyset\)), at denotes a relation in which the figure is related to all the possible situations related to the ground, taken as a sum situation (i.e. \(S\)). If we represent Ps as their corresponding “complex” situations, then we can organize these situations in a lattice-based model of Ps. In such a case, in can be seen as the filter or bottom element of the lattice of Ps (i.e. the P denoting the “basic” situation \(\emptyset\)), whereas at represents the ideal or top element, the “complex” situation \(S\). Obviously, the relation \(\emptyset \leq S\) holds, or: the denotation of in is a part (a sub-set) of the

\[^{19}\text{This sentence is repeated here a third time just for the sake of clarity: I hope that the reader will not mind this further reiteration.}\]

156
denotation of *at*. Consequently, if a sentence such as (203) can be seen as denoting a *sub-set* of the situations denoted by a sentence such as (204), then *in* can be seen as denoting a sub-set relation of the relations denoted by *at*.

A similar consideration can be made for other Ps: intuitively, *in front of* or *behind* can also be seen as denoting a part of the situations that are denoted by *at*, at least in their non-ordered (“sum”) readings. The generalization stemming from this pattern is that Ps are related not only via entailment relations (ultimately, super-set relations), but also via sub-set relations. Intuitively, if Ps denote situations of varying complexity, then Ps involving more situations will in a sense include certain Ps denoting less complex situations. One example is *at* that, in its proposed denotation, acts as a “general” situation including all other (locative) situations.

Other cases, discussed for instance by Levinson (2000:ch.4), are instead the following:

(205) The box is on the table
(206) The box is on top of the table
(207) The box is under the table
(208) The box is underneath the table

According to Levinson (2000), the sentence in (205) can be understood as denoting a relation between box and table that is true in a sub-set of the cases in which the sentence in (206) is also true. If the box is on top of the table, it will be supported by the table (i.e. on the table), and also in contact with the table’s top surface. Similarly, the sentence in (207) is true in a sub-set of cases in which (208) is true, i.e. when the box is under the table, but not it is adjacent as well. Other similar cases can be discussed, and several cross-linguistic examples are discussed in Levinson & Meira (2003). The discussion in section 2.4 of chapter 2 offers related, experimental evidence too (e.g. the “overlap” of *in* and *on, above* and *on*; etc.). More in general, then, not only our original recalcitrant data become evidence in favor of our theory, but several other examples now fall within the range of our proposal’s predictions.
Let us now move to other “P topics”. The fourth prediction states that a P may have a “non-relational” interpretation coexisting with a relational one. The first and second prediction state that a P may denote either a “static” relation, or more “dynamic” relation involving a relation between an event of change and other states. These predictions, combined together, state that Ps may be ambiguous between a more “static” reading, in which the relevant situations are not ordered, and the more “dynamic” reading in which the order of situations implies a change of position for the figure. We predict that Ps may be ambiguous between a locative and a directional interpretation, as known in the literature (e.g. Fong 1997, 2001; De Swart 1998; Zwarts 2005, 2006).

Examples supporting this prediction are the following:

(209) Mario is at the shop
(210) Mario has stopped at the shop
(211) Mario is in the room
(212) Mario is jumping in the room
(213) Mario sits on top of the hill
(214) Mario is running on top of the hill
(215) Mario waits behind the chair
(216) Mario is walking behind the chair
(217) The mountains decline towards the valley
(218) The river lies across the valley
(219) The man sits along the river
(220) The man stretches around the wall
(221) Mario is in front of the girl
(222) Mario has gone in front of the blonde girl
The paired examples going from the pair (209)-(210) to the pair (221)-(ref216), show that Ps can denote both a purely "static" or "dynamic" sort of relation, hence being able to combine with different sorts of verbs. The two examples in (221) and (222) are our original recalcitrant data, (102) and (108). In this second case, a P intuitively denotes a slightly different type of relation than the "default" one, since in the resulting situation the undergoing event is an undistinguished part of the resulting situation. This fact can be represented via the relation \( \lambda x.\lambda y.e': \{\{e,x\} \leq y\} \), which says that the situation denoted by a PP includes an event of change. The event of change, however, lacks a clear position with respect to the undergoing event of motion, since we have the identity \( e' = \{e,s\} \) (with \( e' \leq E, s \leq S \)).

A P such as across, under this interpretation, can be represented as \( \lambda y.e': \{e,ac\} \leq y \), with \( ac \) denoting the "across" region/situation that is "occupied" during a certain event. This is the correct interpretation for PPs which combine with "located motion" events (e.g. Fong 1997), such as jumping and running (across something), and it is also compatible with dynamic verbs such as has stopped, which may involve a more fine-grained perspective on the order of events. Intuitively, if a P can alternate between a locative and a directional reading, then then relevant reading will be compositionally disambiguated by the verb, after the two constituents are merged together.

One general observation is that most if not all Ps may be treated as being under-specified, in the sense that they may enter a syntactic derivation without having a specific interpretation (informally, "relation" vs. "sum"). In our "left-to-right" model of syntax, this is not a problem, since the exact interpretation may emerge via the merge of V and P, as observed in the closely related literature on Ps and Lexical Aspect (e.g. Fong 1997; De Swart 1998; Kratzer 2003; Rothstein 2004; Zwarts 2005, 2006; Ramchand 2008).

Examples (217)-(220) show that directional Ps can also combine with "static" verbs, verbs that denote the orientation or "posture" of a subject with respect to a ground (e.g. Gawron 2005; Zwarts 2006; den Dikken 2008). In this case, these Ps express the position of a subject as if changing with respect to the ground, an interpretation known as "fictive motion" (e.g. Cresswell
intuitively, if we walk across the valley and follow the river, its parts will not be “across” the valley, but once we have completed this “Cresswellian journey”, we can say that the river, in its entirety, is across the valley. For these examples, the relevant interpretation is also $\lambda x.\lambda y.e' : (\{e,x\} \leq y)$, which can be interpreted as denoting some real motion, as opposed to the fictive one.

Under this treatment, across denotes an event of going across a certain location, which thus includes a change of location (captured by $e$). A static reading will exclude the event of change. For instance, (219) may either denote a situation in which a man is currently sitting parallel to a river, or a situation in which the man lowers and consequently lies parallel to a river. A similar reasoning can be applied to (220), and more in general to Ps that can combine with verbs with an inherently ambiguous interpretation (e.g. Fong 1997; Gawron 2005). This rather subtle distinction supports the fourth prediction: if Ps denote events, they may do by either denoting such events in a static (i.e. sum) perspective, or in a more dynamic (relational) perspective. In the first case, the “fictive motion” reading of (218) arises, in the second the “actual” motion arises (e.g. (222).

Another example I shall discuss is that of PPs acting as “nominal modifiers”, amply discussed in the literature (e.g. Jackendoff 1983, 1990; Fong 1997; Hale & Keyser 2002, inter alia). Ps can combine with two DPs and express a form of modification, i.e. they can express that a certain referent is located/oriented with respect to a another referent:

(223) The apple inside the box
(224) A road to the mountains
(225) Mario kicks (the box under the table)

Examples (223)-(225) show that both locative and directional Ps can act as nominal modifiers, possibly in the syntactic context of a full sentence. According to Hale & Keyser (2002) (and contra e.g. Jackendoff 1983), however, Ps in these cases have a predicative function, and the two DPs are actually their specifier and argument phrases of a P. The resulting PP may then be
the complement of a relevant verb, as in the case of forms of complex predication such as (225) (again, Hale & Keyser 2002:ch.2-4).

Syntax-wise, the minimal assumption we need to account these examples is that the “extra” DP is merged in the specifier position of the specifier PP. Semantics-wise, we can roughly represent the denotation of this PP as \( s : ((ap \leq (in \leq sd) \leq bx)) \), a slightly more complex relation between situations involving a given apple and its “inside” location in a given box. A proviso for directional Ps is the following: in cases such as (224), to denotes the direction of a road, rather than a sequence of events (of motion). In this and similar other cases, the accessed reading is one involving “fictive motion”. Informally, the road is such that, if we cover it in its entirety, we will be at the mountains (e.g. Cresswell 1978; Jackendoff 1983, 1990). The fourth prediction appears to be borne out for directional Ps as well.

Another phenomenon I shall discuss regards Ps (or, more accurately, PPs), and their ability to be identified with anaphoric pro-forms, indexicals such as here. Our fifth prediction is that, since PPs denote complex situations, they may do so by “hiding” the situations making up this complex referent, which may be eventually retrieved from the implicit or explicit context. Two examples are the following:

(226) Mario is here (=in front of the car)

(227) I am in front of the car and Mario is here, too

(228) Mario is sitting here

In (226), here is interpreted as denoting some implicit information, that Mario is located in front of the car. In (227) this information can be retrieved from the PP in the first conjunct of the sentence, to which the PP here is in an anaphoric relation. Note that (228) is our original problematic example, (112).

As an anaphoric element, here denotes the anaphoric relation “? = s : (loc\*\(s\))”, with “? = s” being the anaphoric component, and loc\* a condition that grants that the bound situation is spatial in nature (cf. Zeevat 1999:2-3). In (226), this condition grants that the anaphoric relation
introduced by *here* is resolved as \((in \leq fr) \leq cr) = s\), equivalent to \(s = (in \leq fr) \leq cr\). In words, since (P)Ps can be also interpreted as anaphoric (identity) relations, they may be realized as pro-forms (indexicals), and thus depend for their interpretation from the context, as per fifth prediction. For more complex related cases we have mentioned, such as ellipsis and answers, we will need to discuss clausal matters in the next section, before being able to have an answer.

### 4.5.4 Basic Facts of Clause Architecture and Sentence Interpretation

In this sub-section I will discuss the semantics of DPs in minimal detail, as well as the import of our predictions with respect to Vs as the “other” category typed as \(p\), and the remaining loose topics (P answers, ellipsis). None of my proposals is meant to be exhaustive, but should be thought as offering the minimal assumptions that allow us to treat the semantics of other parts of speech and their (compositional) interaction with Ps, within a broader perspective of sentence structure (and processes).

The first topic is that of DPs. I have assumed that DPs are syntactically typed as \(n\) and semantically as \(<s>\), so they can combine with Ps as arguments. This semantic type is justified to some extent if we look at “who” questions types, which offer evidence that discourse referents are a specific sort within the domain of situations (cf. discussion in section 4.4; von Fintel 1994; Kratzer 2007). It is also consistent with standard treatments of *definite descriptions*. Informally, a DP such as *the boy* is the merge of the definite article *the* and the common noun *boy*. If we assume that *the* denotes an operator that selects a single/maximal referent out of the set denoted by a common noun (e.g. *boy*), then *the boy* will denote a unique (atomic or sum) situation, as per standard assumptions (e.g. \(t\)-operator: Gamut 1991; Heim & Kratzer 1998; Chierchia 1998, 2010; \(\sigma\)-operator, Link 1998; Landman 2000, 2004; \(\varepsilon\)-operator, von Heusinger 2003; *inter alia*).

A different picture emerges if one treats DPs as denoting generalized quantifiers, and thus inherently as relational structures. The theory sketched here might be integrated with a theory of *generalized quantifiers*, informed by recent results in dynamic semantics regarding their syntactic properties and their interpretation in discourse contexts (e.g. Krifka 1999; Nouwen 2003;
Brasoveanu 2008; and references therein for a recent and thorough overview). This approach would also be consistent with the idea that D heads may instantiate (b)-types syntactic types, or “abstract” Ps. Exploring such possibilities is, however, well beyond my current goals, so I shall leave such option for future research. For all relevant purposes, I will maintain that DPs are interpreted as “special” situations, non-logical constants that denote individual referents represented as e.g. $d$, as I have assumed so far.

Let us move to Vs. According to our syntactic theory, Vs are assigned the abstract type $p$, which is interpreted as a relation (of type $\langle s, \langle s, s \rangle \rangle$). The predictions I have laid out for Ps should thus extend to Vs, although they should take a somewhat different form: Vs can express temporal and modal information, either via their suffixal morphology or in combination with auxiliary verbs/copulae of various types. I shall leave aside these dimensions of meaning, and analyze our predictions as predictions regarding the lexical aspect of Vs, since they are about the situation structure denoted by Vs (cf. also e.g. Kratzer 2003; Rothstein 2004; Ramchand 2008; inter alia). Therefore, much of what I will say in this section on this topic should be consistent with general assumptions in the literature, although there should be a non-trivial differences between the following proposal and more fine-grained proposals, to which the reader is deferred (e.g. Dowty 1979, 1989; Moens & Steedman 1988; Verkuyl 1993, 2008; De Swart 1998; Fong 1997; Krifka 1998; Kratzer 2003; Rothstein 2004; Borik 2008; Ramchand 2008; Portner 2010; inter alia). My discussion in this section can thus only be seen as being inherently speculative in nature, and will be heavily geared towards the discussion of verbs that have already been mentioned so far, so that it will be possible to offer at least examples of some full sentence interpretation we have seen so far.

I shall start from the first prediction: there should be Vs which either denote relations between states or relation between events. The copula be is one example of the first class, and several “dynamic” verbs are examples of the second class. Some examples are the following:

(229) Mario is in the kitchen

(230) Mario sits with Luigi
Example (229) shows that the copula can denote a relation between Mario and a situation in which he is located in the kitchen. Examples (230)-(232) show that verbs such as *sits*, *waits*, *stretches* can also denote a relation between Mario and some event in which he is sitting (stretching, waiting). The first prediction seems to be easily borne out.

The second prediction is that there should be verbs which denote relations involving the event of change as distinct from the final or initial state. The third prediction is that verbs may come in pairs, and thus express two complementary orderings of their respective domains. Some obvious examples supporting these predictions are:

(233) Mario goes to the room
(234) Mario comes from the room
(235) Mario enters the room
(236) Mario exits the room
(237) Mario sits on the sofa
(238) Mario stands on the sofa

Examples (233) and (234) are typical examples of verbs involving an explicit change of position, *come* and *go*, and several others could be proposed, e.g. *enters* and *exits* (235) and (236). Examples (237) and (238) may be seen as involving two antonym-like verbs denoting a static relation (i.e. *sit* vs. *stand*), which can also be interpreted as denoting (complementary) processes. The second and third prediction also appear to be borne out.

The fourth prediction is that verbs may either have a “relational”, more dynamic interpretation, or they may have a static interpretation. The verbs *stretches* and *sits*, among others, support this prediction, as the examples show. Two tests that reveals this ambiguity are the classical
modification by temporal adverb test and the “almost” test (e.g. Dowty 1979; Winter 2005, 2006; and references therein, for recent discussion):

(239) Mario sits on the sofa for a few seconds
(240) Mario sits on the sofa in a few seconds
(241) Mario almost sits on the sofa

Examples (239) and (240) respectively show that sits can be understood as a verb involving a situation in which Mario only sits till on the sofa for a few seconds (“static” reading), or lies down on the sofa in a blink of an eye (“process” reading). Example (241) can only be understood as Mario changing his mind about sitting, and thus not performing this action. In this case, the accessed interpretation is only the dynamic one, since it is the basic (“unmarked”) interpretation of these verbs.

The fifth prediction is somewhat subtler. In one incarnation, it states that the complex situation denoted by a verb may be identified with the constituting situations, i.e. that we have \( (a \leq b) \), from \( s : (a \leq b) \). For verbs, we observe that the opposite process is also possible. For instance, if we treat verbs as denoting relations over situations, verbs such as stretch intuitively denote specific types of relations, in which the combination of constituting situations are interpreted as being of the “stretch” type. This can be roughly represented as \( str : (a \leq b) \), or: a stretching situation (type) \( str \) involves a relation between two “smaller” situations. Since we have the identity \( s = (a \leq b) \), we can state this format for verbs as \( s = str : (a \leq b) \), in turn equivalent to \( str : (a \leq b) \). In words, a stretching situation is a situation that includes certain constituting situations making up this complex event (e.g. moving arms and legs, yawning, etc.). In this way, we reconstruct the notion of “Condition” in DRT, or “situation type” in Situation Semantics, as the sum of all such situations that instantiate this sort.

These predictions allow us to offer coarse-grained denotations for verbs, as also denoting part-of relations over situations, of type \( \langle s, \langle s, s \rangle \rangle \), which have a general “denotational template” representable as \( \lambda x.\lambda y. s : \text{Con}'(x \leq y) \) (examples (196)-(198), and thus to offer sample deriva-
tions for full sentences. I shall assume that PPs are merged in “one fell swoop” to simplify derivations, and that our only present perfect verb, has flown, includes a total order between an event (of motion) and the resulting state, spelt out by the P, as per standard assumptions (e.g. Parsons 1990; Kratzer 2003, 2004; Ramchand 2008; inter alia).

I shall repeat (143), (213) and (172) respectively as (245), (247) and (249)), followed by their corresponding derivations:

(242) \[ [[is]] := \lambda x. \lambda y. s : (x \leq y) \]

(243) \[ [[sits]] := \lambda x. \lambda y. s : sit : (x \leq y) \]

(244) \[ [[has flown]] := \lambda x. \lambda y. e < s : fly : (x \leq y) \]

(245) The pen is in front of the box

(246) t. \[ [[the pen]] := p, \langle s \rangle, \]

\( t + 1. [[is]] := \lambda x. \lambda y. s : (x \leq y), \langle s, \langle s, s \rangle \rangle \) (assumption)

\( t + 2. ([[the pen]] [[is]] := (p) \lambda x. \lambda y. s : (x \leq y) = \lambda y.s : (p \leq y), \langle s, s \rangle \) (m. intr., f. appl.)

\( t + 3. [[in front of the box]] := ((i \leq fr) \leq b), \langle s \rangle \) (assumption)

\( t + 4. [[the pen is]] ([[in front of the box]]) := \lambda y.s : (p \leq y) ((i \leq fr) \leq b) = s : (p \leq (i \leq fr) \leq b), \langle s \rangle \) (m. intr., f. appl.)

\( t + 5. [[T]] := \lambda x. \bigcup(s), \langle s, t \rangle \) (assumption)

\( t + 6. [[T]] ([[the pen is in front of the box]])) := \lambda s. \bigcup(s : (p \leq (i \leq fr) \leq b)) = \bigcup(s : (p \leq (i \leq fr) \leq b), \langle t \rangle \) (m. intr., f. appl.)

(247) Mario sits on top of the hill

(248) t. \[ [[Mario]] := m, \langle s \rangle \] (assumption)

\( t + 1. [[sits]] := \lambda x. \lambda y. s : sit : (x \leq y), \langle s, \langle s, s \rangle \rangle \) (assumption)

\( t + 2. ([[Mario]] [[sits]] := (m) \lambda x. \lambda y. s : sit : (x \leq y) = \lambda y.s : (m \leq y), \langle s, s \rangle \) (m. intr., f. appl.)
\[ t + 3. [\text{on top of the hill}] := ((on \leq t) \leq h), \langle s \rangle \]  
(assumption)

\[ t + 4. [\text{Mario sits}][\text{on top of the hill}] := \lambda y.s : sit : (m \leq y)((on \leq t) \leq h)) = \lambda y.s : sit : (m \leq (on \leq t) \leq h), \langle s \rangle \]  
(m. intr., f. appl.)

\[ t + 5. [T] := \lambda x. \mathbb{U}(s), \langle s, t \rangle \]  
(assumption)

\[ t + 6. [T]([\text{Mario sits on top of the hill}]) := \lambda s. \mathbb{U}(s)(s : (m \leq (i \leq fr) \leq b)) = \mathbb{U}(s : (m \leq (i \leq fr) \leq b)), \langle t \rangle \]  
(m. intr., f. appl.)

(249) The bird has flown up to the sky

(250) \[
\begin{align*}
t &. \quad \text{[The bird]} := b, \langle s \rangle \quad \text{(assumption)} \\
t + 1. [\text{has flown}] := \lambda x. \lambda y.e < s : fly : (x \leq y), \langle s, \langle s, s \rangle \rangle \quad \text{(assumption)} \\
t + 2. ([\text{the bird}])[\text{has flown}] := (b)\lambda x. \lambda y.e < s : fly : (x \leq y) = \lambda y.e < s : fly : (b \leq y), \langle s, s \rangle \quad \text{(m. intr., f. appl.)} \\
t + 3. [\text{up to the sky}] := e < s : (up \leq sk), \langle s \rangle \quad \text{(assumption)} \\
t + 4. [\text{The bird has flown}][\text{up to the sky}] := \lambda y.e < s : fly : (b \leq y)((e < s : fly : (up \leq sk)) = e < s : (m \leq (i \leq fr) \leq b))(s) \quad \text{(m. intr., f. appl.)} \\
t + 5. [T] := \lambda x. \mathbb{U}(s), \langle s, t \rangle \quad \text{(assumption)} \\
t + 6. [T]([\text{The bird has flown up to the sky}]) := \lambda s. \mathbb{U}(s)(e < s : (m \leq (i \leq fr) \leq b)) = \mathbb{U}(e < s : (m \leq (i \leq fr) \leq b)), \langle t \rangle \quad \text{(m. intr., f. appl.)}
\end{align*}
\]
somewhat different from the standard translation found in DRT, but basically identical to Montagovian approaches (e.g. Chierchia 1998; Landman 2000, 2004). The translation for the other verbs can be “converted” in the standard DRT format (i.e. \( s : \text{Con}'(x,y) \)), although our format is better suited to highlight the intuition that all Ps denote some form of “spatial overlap”.

We can now make the contribution of \( \text{in} \) and \( \text{at} \) to a sentence more precise. By following the translations offered in these examples, sentences including \( \text{in} \) and \( \text{at} \) (e.g. \( \text{Mario is in/at the shop} \), 203 and 204) can be roughly translated as \( m \leq \emptyset \leq sh \) (equivalent to \( f \leq sh \)) and \( m \leq S = sh \), respectively. Intuitively, \( m \) stands for the denotation of \( \text{Mario} \), \( sh \) for the denotation of \( \text{shop} \). In words, these relations respectively denote that Mario is identified with the core of the shop, and Mario may be in any position making up the “extended space” of the shop. These appear to be the correct interpretations for sentences involving these to Ps, according to our previous discussion.

Let us focus back on our derivations. The coarse-grainedness of these derivations stems from at least three layers of missing information. First, these derivations lack temporal information; second, if we take a more fine-grained approach to syntactic derivations, the “specifier” PP would first be merged with the verb, and then re-interpreted as part of the following P; Third, the merged DPs lack their \( \theta \)-role. It is, however, possible to cancel this coarse-grainedness: I shall sketch one method per dimension, for illustrative purposes.

The first piece of missing information may be added via the inclusion of a standard treatment of temporal information and anaphora: DRT offers a very thorough theory on this topic, so its exact implementation is only a technical matter.

The second piece has no direct consequences of interpretation, only on temporary syntactic structure. Semi-formally, it can be represented as follows. If syntactic objects are merged in a piece-meal fashion, at some point we will have an object such as \( \text{the pen is in front} \), representable as \( s : (p \leq (in \leq fr)) \) (type \( \langle s \rangle \)). The merge of \( of \) can be roughly represented as: \([s : (p \leq (in \leq f))] + [s : (x,y)] = [s : (p \leq (in \leq f \leq y))]\); we merge the new syntactic object, form the unit in front of via associativity (and its semantic reflex, transitivity), and obtain \( \text{the pen is in front} \)
of, an object of type \((s,s)\), which can then be merged with the ground DP, in this case *the box*. The syntactic and semantic result is the same, but the derivation is somewhat longer, although equivalent to the ones offered here.

The third piece may be added by assuming that each situation merging with a verb must first satisfy a sortal constraint in interpretation: rather than representing a “locative” verb roughly as \(s : (x \leq y)\), we may represent it as \(s : (\text{th}'(x) \leq y : \text{loc}'(y))\), i.e. as a verb that assigns thematic roles to the arguments it combines with. The implementation of this aspect is straightforward, and both DRT- and lexical-based accounts abound in the literature (e.g. respectively Zeevat 1999; Bos 2010; Landman 2004; *inter alia*). In theory, this layer of information may also be introduced to account how lexical aspect is computed in a compositional manner, and thus how verbs may determine (or not, if they are themselves ambiguous) the interpretation of Ps. Fong (1997); Zwarts (2005, 2006); Ramchand (2008) offer insights on this process that are compatible with the theory sketched so far. I leave the exact implementation of these aspects for future research, however.

To conclude, I shall very briefly discuss the remaining topics on ellipsis and questions, again in a rather speculative way. A general consensus across semantic theories of questions is that questions are interpreted as requests for a certain specific sort of information: informally, if we ask where Mario is, we are requesting information about the situation expressing his location. At a very coarse-grained level, the interpretation of a question and an answer exchange such as (251):

(251) Q: Where is Mario?
    A: In the garden

(252) I tried to find the missing check, but couldn’t (find the missing check)

Involves a step \(t\) at which the question is interpreted as a “partial” sentence, roughly representable as \(\lambda y. s : (m \leq y)\), whereas the answer offered at a step \(t + 1\) is interpreted as \(s : (g)\) (“in the garden”). The merge of question and answer, representable as \(s : (m \leq y)\), represents
the situation that is taken by both participants to be true in discourse, much like the complete answer *Mario is in the garden*, and which is in a sense inferred, rather than explicitly uttered in discourse.

A similar approach can be extended to fragments, with a specific *proviso*. If a sentence denotes a complex situation, then an uttered fragment denotes a part of this situation (in turn, a situation), which is explicitly spelt out. If the answer in (252) is offered as a stand-alone comment to someone looking for Mario, the fragment *in the garden*, i.e. the situation \( s : (g) \), might be interpreted as presupposing the background information \( \lambda y.s : (m \leq y) \) and which might be accommodated in discourse, as assumed in various incarnations of DRT (e.g. Geurts 1999; Geurts & van der Sandt 2004; Zeevat 1999; SDRT, e.g. Asher 2003; *inter alia*). In this case, VP ellipsis, represented in (252) appears to support the fifth prediction. Very informally, the elided part of sentence in (252) should be interpreted as anaphorically being linked with the previous VP: what I could not do is to find the missing check.

Similar considerations can be extended to other discourse-oriented phenomena, such as stress and focus, complex entailment patterns in discourse (e.g. “sequence-of-tense” phenomena), cross-sentential anaphora, as well as the explicit representation of the agents uttering the different “steps” of a discourse. In these cases, both DRT/SDRT frameworks offer more than adequate tools to properly treat these data, so the remaining problems are purely a matter of implementation.

While the current embryonal explanations are nothing more than a mere sketch, there appears to be no particular theoretical problem, or empirical evidence that suggests for these explanations to be on the wrong track. This is intuitively a further argument in favor of the proposal offered so far, along with the much stronger argument that all of our predictions are apparently supported. We can now come to the conclusions of this chapter.
4.6 Conclusions

In this chapter I offered a novel proposal regarding the semantics of Ps. I have done so by offering a quite thorough analysis of English adpositions (Ps) and semantic properties, and their contribution to the overall interpretation of sentences they occur in. This theory “merges” several strands of analysis found in previous semantic literature on Ps, and offers a unified and improved theory regarding the linguistic properties of this category. At the same time, it updates the analysis offered in chapter 1 under one key aspect, and allows offering an answer to the global research question of part II, introduced at the beginning of chapter 3 and repeated in this chapter.

This answer is as follows:

A-A: *Ps combine with other Ps and with other parts of speech in a compositional way, and express the possible types of relations defined over the domain of (spatio-temporal) situations in a principled way;*

This answer tells us that the contribution of Ps can be defined in a rather fine-grained way, so that each syntactic unit making up a P offers its semantic contribution in a uniform way. Thanks to the Curry-Howard isomorphism, we can easily capture the regularity of Ps and their interpretation as a consequence of their syntactic regularity, hence offering a fully compositional treatment of this category.

Other important results of this chapter, and more in general of part II, stem from the fact that we now have a systematic way to make predictions regarding the interpretations of the members this category, as well as other categories as well; and that these predictions are tightly connected to the correct predictions that we can make about the syntactic structure of Ps and their emergence over time, as well as the predictions about other categories that combine with Ps. We started this chapter by observing that several data regarding Ps were in need of an explanation in our DRT fragment, whether they were already treated or still recalcitrant in previous proposals. By developing a more thorough theory on the semantics of Ps based on the syntax presented in the previous chapter, we have been able to offer a semantic theory of Ps that not only can cover
these examples, and thus be on par with previous approaches, but can also successfully cover the recalcitrant data that were in need of an explanation in previous work (e.g. measure phrases and Ps; indexicals; entailment and sub-set patterns; coordinated Ps; under-specified interpretation; et cetera).

This unified theory stems from an approach to Ps that is able to predict the types of possible relations that these elements denote, via a simple and yet elegant principle, the Curry-Howard isomorphism. By assuming that the domain of (spatio-temporal) situations has a certain partitioned structure, the basic and less basic properties of Ps follow in a natural way, and as a consequence of a unified perspective. Our improved theory of Ps is now fully compositional, offers a transparent relation between syntax and semantics, and has a quite wider empirical adequacy than previous proposals. In particular, it predicts and correctly accounts, under a unified theoretical perspective, data that were problematic or even against the predictions of previous proposals, regardless of their various theoretical declinations.

Once that we have a solution to our second outstanding problem, we can concentrate on addressing the third problem: what is the psychological reality of such the fully compositional approach to Ps outlined in this part. For this purpose, I shall leave aside the more fine-grained details of our theory of Ps, such as the predictions regarding measure phrases, indexicals and multiple interpretations of Ps. I shall concentrate on an experimental investigation regarding the on-line status of Ps by offering an investigation of the basic interpretation of Ps and their logical relations, the topic of Part III of this thesis.
Part III

The Psychological Reality of Adpositions
Chapter 5

The Psychological Reality of Adpositions, Part I: Theoretical Preliminaries

5.1 Introduction

In this chapter and the next I will address the problem of the psychological reality of our theory of adpositions, by offering experimental evidence regarding their on-line interpretation in adults and children, as well as their acquisition in children. The goal of the third part of the thesis is to offer a solution to the third outstanding problem, which can be captured via the following global research question:

Q-A: What is the interpretation of Ps and their logical relations by English speakers, and how this interpretation emerges over time;

Anticipating matters a bit, I shall offer an answer that confirms the five empirical predictions on the interpretation of Ps as being predictions about speakers’ interpretation of this category; and that confirms two further predictions that I shall offer at the end of this chapter, regarding how
Ps are acquired by children, over the development of their budding grammars. This chapter will act as a short theoretical introduction to chapter 6, while 6 will present three experiments that will offer empirical, on-line evidence for our predictions.

In this chapter I shall adopt a slightly different divide et impera strategy than the one adopted in the previous chapters, in order to solve this problem. I shall first address two problems that are still in need of a solution in our proposal: what we know about sentence processing and acquisition matters, and how we can integrate this knowledge within our DRT fragment. The empirical import of solving these problems can be observed via the following examples:

(253) “Mario is in front of the car”

(254) Mario has gone to the desk

(255) Mario is sitting !at the desk

In (253), we have an English utterance, a string of phonological words. Intuitively, we know that once a speaker parses this string, he will access the corresponding sentence and its interpretation, via the processes outlined in part II. However, our DRT fragment still lacks a precise way to represent this process. In (255), we have an English sentence involving a P, at. English adult speakers may straightforwardly access the interpretation of this sentence. Children, however, may be able to access the intended meaning of (254) and the P to only at a certain developmental phase, but they may not be able to access its intended meaning of (255), as represented by the pre-theoretical symbol “!”’. Consequently, they may still not be able to access the entailment relation holding between the two sentences (254) and (255). As it stands, our DRT fragment cannot treat these data, and thus cannot yet be used to offer an answer to our global research question.

The specific goal of this chapter is thus to integrate this knowledge within our DRT fragment, and thus have a principled way to account how the experimental evidence is in favor (or against) our predictions regarding the interpretation of Ps. This goal can be met by offering an answer to the following research questions:
Q-1: What do we know so far about the mapping from processed utterances to their underlying sentences;

Q-2: What do we know so far about Language Acquisition as an incremental process, and what predictions we can make about Ps and their emergence;

In order to answer these questions, I will first review some DRT models of parsing that specifically include the processing of utterances (e.g. Poesio 2003), and I will then review some theories of Language Acquisition (e.g. Crain & Thornton 1999; Tomasello 2003), offering an argument on integrating Crain & Thornton’s *Modularity Matching Model* within our approach.

This chapter is organized as follows. Section 5.2 presents an analysis of theories of language processing and their relation with our DRT fragment. Section 5.3 presents an analysis of theories of language acquisition, and a way to integrate one proposal in DRT. Section 5.4 offers some conclusions.

## 5.2 DRT as a Theory of Parsing

A central goal of DRT is to model the interpretation of sentences as an on-line, incremental process, in which the syntactic/semantic and pragmatic components of sentence processing interact in a seamless way (Kamp *et al.* 2005:750-752). In this regard, DRT offers a more thorough perspective to several of the processing theories found in the language processing literature, which mostly focus on syntactic matters (e.g. our PIG, Phillips 1996; Altmann 2001; see again chapter 3, section 3.3 for references). Differently from these theories, however, DRT has not been principally driven by experimental data, in its formulation.

There are, however, two processing theories with a firm base on empirical findings and which share the same underlying assumptions of DRT: *Referential Theory* (e.g. Crain & Steedman 1985; Altmann & Steedman 1988), and one variant of *Centering Theory* based on Massimo Poesio’s work, which implements (a version of) DRT as an underlying theory of grammar (e.g. Poesio *et al.* 2000, 2004; Poesio 2003). The core claims in both theories are quite similar.
The two theories differ, however, with respect to their pragmatic assumptions, in particular their approach to the resolution of anaphoric relations for NPs (our DPs) at a cross-sentential level, but not with respect to other anaphoric elements of interpretation, including our Ps.

Both theories assume that sentences are processed piece-meal, with syntactic structure driving semantic interpretation. The parsing process starts with an agent parsing *phonological words*, which act as “external” inputs signalling which lexical items must be retrieved and composed together. The underlying structure of a sentence is then accessed in a “left-to-right” fashion, which is roughly equivalent to our approach to syntactic production. When an agent parses a sentence, his parser will derive a syntactic structure according to the derivational processes outlined in the previous two chapters. As soon as syntactic structure is built, the interpretation of the single lexical items is accessed, and with it the compositional interpretation of the merged constituents. The result will be the interpretation of a sentence or fragment thereof, which may still need further extra-linguistic processing to be disambiguated (e.g. anaphora resolution).

I shall capture these assumptions in our DRT fragment as follows. I shall assume that phonological words represent “external” instantiations of lexical items, our labeled DRSs. Other modalities of “externalization” (e.g. Sign Language) are possible, but not discussed here. Although not necessary, I shall assume that phonological words form a partially ordered set $N$, so for each phonological word $n$ the following holds: $n \leq N$, i.e. a phonological word is part of a Phonological Lexicon. This process is defined as an isomorphism $j$, as in the case of the process mapping VRSs (Visual Representation Structures) to DRSs. For each phonological word $n$, the parser will retrieve a labeled DRS $\langle d, d \rangle$, via the function $j$. For instance, the phonological word /ka't/ will be mapped onto the labeled DRS $\langle n, \{x\} : \text{cat'}(x) \rangle$ via the function $j$, so that the following relation holds: $j(/ka't/) = \langle n, \{x\} : \text{cat'}(x) \rangle$, the word “cat” is mapped onto its corresponding syntactic category (a noun) and semantic interpretation (a condition). *Language Processing*, in this formulation, corresponds to the process that maps combined\(^1\) phonological words into "concatenated" larger units, such as utterances; phonemes can also be concatenated to form syllables and more complex units (cf. Samuels 2009:ch.1). For our purposes, this operation is equivalent to *merge*, as defined over phonological words. Differences in definition

\(^1\)In several phonological theories, phonological words are assumed to be “concatenated” into larger units, such as utterances; phonemes can also be concatenated to form syllables and more complex units (cf. Samuels 2009:ch.1). For our purposes, this operation is equivalent to *merge*, as defined over phonological words.
words (i.e. utterances) to the (interpreted) sentences that correspond to these utterances, and can be represented as a triple \( \langle n, \langle d, d \rangle \rangle \), or in a more compact format as a pair \( \langle n, DRS \rangle \).

It would be possible to define more in detail both the structure of the phonological vocabulary and the nature of this isomorphism, but this process would lead us too far afield\(^2\). The literature contains more thorough proposals that are much in line with the basic assumptions I offer here (e.g. Reiss & Hale 2008; Samuels 2009). It would also be possible to reconstruct several, if not all, the predictions regarding the interpretation of syntactically and semantically ambiguous sentences (e.g. garden-path constructions, categorial ambiguity, polysemy, quantifiers and scope, etc.), but this process would also take us too far afield. Again, the literature offers a far more thorough analysis of these topics, and proposals which are in line with the proposal I sketch here (e.g. Altmann 2001, 2004; Phillips 1996, 2003; Poesio 2003; Poesio et al. 2000, 2004; \textit{inter alia}). Even with these partial definitions, however, we are now able to capture the parsing process in a more precise way, since we can establish a one-to-one mapping between phonological forms parsed by agents and the corresponding \textit{DRSs} that make up an interpreted sentence.

I shall thus assume that our DRT fragment represents how agents parse utterances, and are able to process the sentences corresponding to them, in a compositional way. An example of how this process comes about is the following. I shall represent a derivation corresponding to the utterance in (201) (in quotation marks) by representing for each derivational step a triple of phonological word, syntactic category and semantic interpretation. I will omit semantic types and compress passages, to make the derivation more concise:

\begin{align*}
(256) & \quad \text{“The ball is in front of the table”} \\
(257) & \quad t. \quad \langle \text{the ball}, \mathbf{n}, b \rangle \quad \text{(assumption)} \\
& \quad t + 1. \langle \text{is}, \langle \mathbf{p}, \lambda x. \lambda y. s : (x \leq y) \rangle \rangle \quad \text{(assumption)}
\end{align*}

\(\text{are here immaterial.}\)

\(\text{\(^2\)It should be obvious that the inverse function of } j, \text{ call it } z, \text{ defines the process by which } \textit{DRSs} \text{ are “externalized” during an instance of } \textit{Language production}, \text{i.e. how an agent maps a } \textit{DRSs} \text{ representing a produced sentence onto a corresponding word: we have } z(\langle d, d \rangle) = n. \text{ See again Levelt (1993) and 2 for discussion.}
\[ t + 2. \text{the ball is, } ((n + p) = p_v', (b)\lambda x.\lambda y.s : (x \leq y) = \]
\[ \lambda y.s : (b \leq y) \quad \text{(m. intr., f. appl.)} \]
\[ t + 3. \text{in front of the table, } pp, ((in \leq f) \leq b) \quad \text{(assumption)} \]
\[ t + 4. \text{the ball is in front of the table, } p_v', + pp = pp_v, \lambda y.s : (b \leq (in \leq f) \leq t)) \quad \text{(m. intr., f. appl.)} \]
\[ t + 5. ((T), d, \lambda s. \bigcup(s)) \quad \text{(assumption)} \]
\[ t + 6. \text{the ball is in front of the table). } pp_v + d = pp_v, \]
\[ \lambda s. \bigcup(s)(s : (b \leq (in \leq f) \leq b)) = \bigcup(s : (b \leq (i \leq f) \leq t)) \quad \text{(m. intr., f. appl.)} \]

In words, the derivation in (257) says that an utterance is interpreted as a sentence which is true in a model of discourse. For each step of the processing task, an agent parses a phonological word and retrieves the corresponding DRS, merging DRSs as the derivation unfolds. The derivation offered in (257) can be seen as theory-neutral, since both Poesio and Crain & Steedman’s theories treat syntactic and semantic details in a similar way (and phonological, in Poesio’s case). The differences between the two theories ultimately lie in pragmatic matters, e.g. how and when, in a derivational time, inter-sentential aspects of interpretation are resolved (e.g. presuppositions, implicatures, inter-sentential anaphora resolution). As these aspects are not crucial to our discussion, I shall remain agnostic to which pragmatic theory to choose.

At this point, we can offer an answer to the first research question of this chapter:

**Q-1:** The incremental, derivational processing of utterances allows the access to the corresponding derivational processing of sentences;

This answer captures the rather simple intuition that the on-line processing of utterances allows the consequent on-line processing of sentences. As soon as speakers can parse words and combine them together into utterances, the corresponding sentences will also be “formed” and interpreted. We can now represent in our DRT fragment the intuition that there is a tight connection between an utterance such as “Mario is in front of the car” in (253) and the sentence *Mario*.
is in front of the car. Example (253), and with it any other utterance and corresponding sentence, can now be properly accounted. The same holds for the data that will presented in chapter 6. Before looking at those data, though, I shall propose a theory of Language Acquisition within DRT.

5.3 DRT as a Theory of Acquisition: Proposal and Predictions

In this section I will suggest how to integrate our fragment of DRT with a theory of Language Acquisition, and thus offer an account on how DRSs are processed for the first time and eventually stored (in long-term memory), with the result of “expanding” an agent’s model of discourse, in particular children.

The problem of Language acquisition in children has received an enormous deal of attention in the literature. However, among the several theories regarding language acquisition, few are compatible with the DRT approach outlined so far. For instance, our DRT fragment is basically incompatible with probability-based models of acquisition, since a cornerstone of DRT is the (deterministic) treatment of Language as an interpreted set of well-formed syntactic structures (e.g. Hirsh-Pasek & Golinkoff 1996). It is also incompatible with theories that reduce Language principles to Cognition principles, e.g. van Hoek (1995); Tomasello (2003). Although the “Logic” behind Language is based on principles that are domain-general, basic syntactic and semantic notions such as “head” (and thus the syntax of Ps) or “entailment” (and thus the semantics of Ps) find their reason of existence precisely as language-based information units, which represent how different “types” of knowledge are integrated together in one universe of discourse. Without language-specific rules, these phenomena would be unexplainable. It is also incompatible with constructionist approaches to grammar and grammar acquisition, which do not entertain the possibility that Language is definable via a precise set of rules, but focus instead of sets of fixed, non-compositional “constructions” (Culicover & Jackendoff 2005; Goldberg 2006). Given the rule-based, dynamic and compositional approach behind our fragment of
DRT, the incompatibility of these approaches should be obvious.

Most of the remaining models of acquisition are those approximately falling in the generative tradition, e.g. Guasti (2002); Culicover & Nowak (2003); Crain & Thornton (1999). The first two theories can be said to be syntactic-bound, since they take a representationalist approach to syntactic matters. They assume hierarchies of fixed positions, as in the syntactic theories reviewed in chapter 3. Thus, they are also incompatible with my derivational stance, since I assume that there are no fixed positions with a fixed interpretation, and thus that the derivational process allows a given interpretation to “emerge”. The third proposal, Crain & Thornton (1999) and their *Modularity Matching Model* (henceforth: MMM) is not so distant from my assumptions, so I shall review it in some detail. This proposal has a more logically-oriented approach, with four core assumptions playing a particularly important role.

First, “Language” is seen as a modular faculty (part of our general information-processing faculties), which can be decomposed in a *syntactic module* and a *semantic/pragmatic module*. The syntactic module represents the “central” part of the faculty, and has the role of merging “lexical items” into more complex units, sentences and discourses. Although the underlying model of (generative) syntax is “representationalist” (as in Guasti 2002), its definition is general enough to be compatible with ours. The semantic/pragmatic module interprets sentences with respect to a model of discourse, with the output possibly modifiable by subsequent computation of pragmatic-based phenomena (e.g. presupposition, implicatures). The semantic/pragmatic module is based on dynamic theories of semantics (e.g. Chierchia 1995b; DRT). The parser, which is modeled according to Reference Theory, has the task of parsing utterances and accessing the intended sentences, which are then processed accordingly. The treatment of sentence parsing offered in the previous section is a good approximation of how this theory treats this phenomenon.

Second, it is assumed that children and adults share the same underlying language faculty. Although children may seem to have problems in producing/understanding adult-like linguistic expressions, their problems stem from more basic limits in “processing power”: children may
access the basic rules to produce a correct English sentence, but their working memory resources may be too limited for the process to be correctly carried out. Children’s apparent non-adult production and interpretation of sentences may stem from their lower short-term memory resources, and thus from a more limited performance in linguistic tasks. Their competence, their innate knowledge and ability to access linguistic processes, is assumed to be adult-like. Children can produce and interpret sentences much like adults, insofar as their short-term memory limitations allow them to do so, and insofar as they are able to find a precise interpretation in their own still developing model of discourse, their still growing long-term memory.

Third, children’s Language acquisition process is governed by the Language Acquisition Device (henceforth: LAD), a component of the Language Faculty that updates new sentences and meanings in a child’s grammar, or retrieves the intended structure for a sentence and its meaning of a sentence in cases of ambiguity/underspecification. While the parser may generate multiple (interpreted) sentences for an utterance, the LAD selects which sentence represents a proper sentence of a given grammar. Since the child may be constrained by processing limits, he may be able to access only a part of the intended meanings of a sentence, a phenomenon known as the sub-set problem. Cases amply discussed by Crain & Thornton (1999) include scope relations, scalar implicatures, the interpretation of inclusive-or and many others. These cases seem to suggest that children may interpret sentences in a non-adult way when only part of the intended meaning can be accessed, against the full set of meanings accessible to adults (see also e.g. Crain & Khlentzos 2008, 2009).

Fourth, the Language faculty emerges piece-meal in children. When the acquisition process starts, a child is able to access only minimal information regarding syntactic and semantic/pragmatic rule. Via the continuous process of parsing new sentences and meanings, and thus expanding one’s Lexicon (set of well-formed sentences), the child slowly but progressively acquires a grammar of a language (e.g. English), and with it he obtains access to which syntactic structures and corresponding semantic interpretations are possible. During this process, the child may also “assume” incorrect sentence structures, insofar as they are (still) possible syntac-
tic structures in other languages (e.g. German), what is known as the continuity hypothesis (e.g. Pinker 1984; Crain 1991).

These four assumptions are entirely compatible with our DRT fragment, as well as our general assumptions about cognitive architecture. I shall adopt them, and discuss a way to formalize the third assumption, or how to capture the notion of LAD within our DRT fragment. We have defined the recursive merge of two DRSs as the progressive update of an explicit context, a DRS, with novel information, another DRS (see 2, section 2.3.5). Under this dynamic perspective, sentence processing is inherently an update process in which new information (i.e. interpretation of upcoming constituents) is added piece-meal, and which results in a final, “static” step: evaluating whether a sentence is part of a model of discourse or not (true or false). A third possibility is that a sentence may not be directly evaluated as true or false, but treated as having a “truth-value gap”, i.e. it is undefined since its conventional meaning, its interpretation in a model, is not (completely) known, once the parsing process is over. Typical examples are not only sentences containing various types of syntactic and semantic vague predicates (e.g. scalar adjectives such as tall, mass nouns such as water, etc.), but also sentences containing novel words, which offer an obvious source of undefinedness to agents (e.g. Reyle 1993; Poesio 2003; Poesio & Rieser 2010).

In order to model this fact, I shall expand one of my assumptions from chapter 4, and assume that the assertion-operator can compute three truth-values for a sentence: true, false or undefined, represented as “1”, “0” and “*” (the “Kleene star”). A sentence is undefined when its interpretation could be either true nor false, in a model of discourse. This property which can be very roughly represented as \( *T(s) = (s \cap \neg s) \), a sentence which is interpreted as both being possibly true (i.e. s) or possibly false (i.e. \( \neg s \)), when evaluated (cf. also Stenning & van Lambalgen 2008:ch.4). I shall then assume that, when a sentence is evaluated as undefined, then an agent is not in a position to decide whether it is true or false, but it is in a position to judge that the sentence is undefined.

\footnote{Undefinedness in DRT is a reflection of the intuitionistic nature of DRT’s logic. In certain forms of Intuitionistic logic, differently from classical ones, it is possible to have predicates which are undefined, i.e. they are neither true nor false in a given “state” of a model, but may become true (or false) when the model is updated with more information (e.g. new propositions). In this case, the interpretation function is said to be partial, since it does not always assign a “true” or “false” truth-value to a situation. See Landman (1991:150-170) for discussion on intuitionistic logic and vagueness models.}
may retrieve the “original” situation and test whether he may obtain a true or false sentence by updating (part of) the model of discourse. In words, an agent that cannot evaluate whether a sentence such as (256) (i.e. the ball is in front of the table) is true or false in discourse, will attempt to “test” whether adding the interpretation of this sentence is consistent (e.g. true) in a model that includes this sentence as well. If this process is successful, the agent will add this bit of information to his model of discourse (i.e. that the ball is front of the table), as a true (or false) fact about the world.

I shall then assume that a test-operator $A$ of type $\langle t, s \rangle$ will then retrieve the original situation, i.e. that we have $A(T(s)) = s$. In words, a test-operator acts as the inverse function of the assertion-operator, and retrieves the original situation denoted by a sentence. The use of the “test” label is theory-neutral, but the definition of this operator is very similar to many similar “locutionary” operators found both in DRT and “extended” variants of Situations Semantics (e.g. Blackburn & Bos 2005:ch.2; Egg, Koller & Niehren 2001; Asher & Lascarides 2003; Poesio 2003; Poesio & Rieser 2010; Ginzburg 2010). The interpretation of this operator can be represented as $\lambda t.t : (s)$, which reads: a test-operator takes a truth-value and returns the situation that is evaluated as true, false or undefined by this operator. Intuitively, if a sentence is undefined, then an agent may first retrieve the original situation, and eventually update it in a model of discourse.

I shall then assume that this situation is then updated via an update relation, of type $\langle s, \langle s, s \rangle \rangle$. An update relation $U$ takes an undefined situation and its corresponding sort, and merges the two situations to create an “updated” sort. It is thus defined as $\lambda x.\lambda .sort.(x + sort) = sort'$, with “sort” being the new set including the missing situation. If a sentence contains a novel P, for instance, the updated sort will be that of spatio-temporal situations $R$. The explicit context (previous and subsequent sentences) and possibly the implicit context as well (e.g. gestures, visual inputs) can offer further, crucial disambiguating information that allows establishing the “bit” of information that must be updated in the model of discourse. For instance, a “novel” noun such as cat can be defined as the noun corresponding to a cluster of previously mentioned features (e.g. Bloom
2000:ch.1; Carey 2010); or it can be ostensively identified with an object to which the noun refers (Poesio 2003:18-20; Ginzburg 2010:10-20 and references therein for discussion).

Intuitively, this update relation adds the interpretation of a novel P (e.g. *at*) to one’s model of Ps and their interpretation. If *sort* corresponds to a set/lattice of currently accessible Ps to a speaker (e.g. *in*, *to*, *from*), then the update function will “extend” this set to include *at*. Once our DRT fragment includes a way to update a model, it becomes at the same time a theory of Language processing, Language production and Language Acquisition, since it can represent via the same set of basic operations how sentences are processed, produced and acquired.

I shall now offer a detailed example involving Ps. Remember from chapter 4, section 4.5.1, that *in* denotes a spatial relation which involves no other specific regions/situations than the one corresponding to figure and ground, and thus expresses their “spatial identity”. If a figure is “in” a ground, then no other spatial regions are involved in this relation. In our lattice-oriented approach, in represents the filter (bottom) element in the domain of spatial Ps, and thus the “first” element that offers information regarding this sort, from a developmental perspective. Suppose, then, that a child parses (258), and interprets a novel word, the P *in*. In doing so, he may be able to update his model of discourse to include not only this novel word, but also to access this domain for the first time.

The derivation is offered in (259):

(258) The ball is in the container

(259) \[ t.(^{\ast}T(\text{the ball is in the container}), d + \text{pp}_v = \text{pp}_v, \] (assumption)
\[ \lambda s.^{\ast}(s_r : (b \leq (0 \leq t))) = ^{\ast}(s_r : (b \leq (0 \leq t))) \] (m. intr., f. appl.)
\[ t + 1.(A, d, \lambda t.t : (s)) \] (assumption)
\[ t + 2.(A, d + \text{pp}_v, \lambda s.(a : (s))(^{\ast}(s_r : (b \leq (0 \leq ct))) = s_r : (b \leq (0 \leq t))) \] (m.intr., f. appl.)
\[ t + 3.(U, d + \text{pp}_v = \text{pp}_v, \lambda x.\lambda \text{sort}.(x + \text{sort}) = \text{sort}^\prime) \] (assumption)
\[ t + 4.(U, d + \text{pp}_v = \text{pp}_v, ((c)\lambda x.\lambda \text{sort}.(x + \text{sort}))(P) = P^\prime) \] (assumption)
The partial derivation in (259) says the following. When a child parses *in* for the first time, he will consider the sentence undefined: there is no situation in his model of discourse that includes the interpretation of *in*, as denoting a relation of total identity in “Space” between two entities. Given the denotation of *in* as a “basic” spatial relation, the child will update a the most basic region/situation belonging to the $R$ domain, in his model of Ps (as the “0” indicates). He will also update a subtler type of information: that certain classes of syntactic objects denote a specific sort of information, the spatial one, as the derivation shows (steps $t$ to $t+5$), use of “$r$” subscript). The child will become able to access information regarding the sort of spatial Ps, and thus he will become able to learn more complex Ps in a piece-meal fashion, as evidenced in the literature, to an extent converging to a closed system once he will be able to access at as well. Furthermore, when this sentence is evaluated as true, it can be (correctly) matched with a visual input, which signals that a ball is in the same position as the container (i.e. $g(\text{the ball is in the table}) = v$ holds). The child may also access extra-linguistic (“conceptual”) information regarding this P, in the opportune implicit context, which acts as an indirect confirmation that the type of linguistic information he is accessing is e.g. a (novel) P.

The approach I am proposing here gives a rather precise formalization of the MMM within DRT. It is consistent not only with the ample experimental literature supporting this model, but also with the vast literature on word-learning, which shares almost the same assumptions regarding the acquisition of novel words (e.g. Bloom 2000; Carey 2010; and references therein on concept acquisition). Under this approach, Language Acquisition can be seen as a procedure that allows an agent’s model of discourse to recursively expand, via progressive and cyclic steps. A child may often access only a part of the intended meaning of a sentence, and may thus have problems in interpreting a sentence in an adult-like way. Once the child is able to update his model, though, the adult-like interpretation becomes accessible. An obvious question
that emerges from this discussion is whether we are able to make predictions regarding the emergence of Ps in children, and thus make predictions about which Ps shall emerge once in is successfully acquired, and why. I shall thus suggest a sixth and seventh prediction regarding the emergence of Ps based on how logical relations (entailment, sub-set) can guide the acquisition process, which can be stated as follows.

The sixth prediction is that Ps appear in (partial) order of complexity, from the most specific to the most general one. This predicts that in, a P that denotes “spatial identity”, is the first locative P that should emerge in children grammars, and that other Ps should subsequently emerge piece-meal. It also predicts that other Ps of various types should emerge once a child can access the interpretation of in. For instance, both directional Ps (e.g. to, through, from) and projective Ps (e.g. in front of, above) should emerge piece-meal, once this basic P is accessed; at, as possibly the most complex P in the grammar, should be accessed at a later time, after to and from are accessed, because of the entailment relation(s) holding between these Ps.

The seventh prediction is that, as a result of this emergent process, the entailment and sub-set relations holding between these Ps will emerge as a (logical) consequence, and will thus represent how the process of updating the domain of (spatial) Ps comes about and is made consistent, as well as other domains. This prediction has some specific consequences: one is that the entailment relations between Ps, e.g. the one holding between to and at, should define the order by which these Ps emerge in children’s grammar: for instance, that at is accessed once to is also consistently accessed. Another is that, if the interpretation of one P is a sub-set of the interpretation of a second P (i.e. it emerged at an earlier stage), then the two Ps may be interpreted as equally acceptable, in the opportune context. For instance, if in is acceptable in a given context, at will also be acceptable. A third aspect is that these two relations can interact: if to entails at and at can overlap with in, in the opportune contexts to and from will indirectly entail in. By acquiring different Ps, the child will at the same time be able to access the logical relations holding between these Ps, which are a reflection of a more general process of acquisition defined as model update.
At this point, we can offer an answer to the second research question, which is the following:

Q-2: *Language Acquisition occurs as a process incremental update, i.e. by children being able to interpret more and more well-formed sentences, and thus being able to access their logical relations as well;*

This answer also captures why we can make certain predictions about the “steps” that the acquisition process goes through. By acquiring novel sentences and corresponding interpretations, children also acquire the logical structure underlying these representations, in the form of the entailment and sub-set relations holding between sentences and adpositions within these sentences. Children do not just acquire meanings, but they also acquire the “logical role” of these meanings within the system of a grammar such as English, being guided by this supplementary information regarding how these meaning are related (i.e. entailment and sub-set relations) in the process. Consequently, we predict that there will be a period of time in which children will be able to interpret our example (254) (i.e. *Mario has gone to the desk*), but will not yet be able to interpret our example (255) (i.e. *Mario is sitting at the desk*), since this latter meaning needs to be implemented in their budding model of discourse.

The next chapter has the task of verifying the seven predictions we now have regarding the interpretation and acquisition of Ps, and thus to offer an answer to our third global research. Before coming to conclusions, I shall spend a few words on the notion of Universal Grammar. As a theory of the initial state of grammar adopted in MMM, Universal grammar represents a theory of which combinatorial (syntactic), interpretive (semantic) and interface properties must be present in humans’ genetic endowment. There is a vast literature on this topic, so I shall defer the reader to the literature for discussion (e.g. Chomsky 2004; 2007; Sauerland & Gärtner 2007; and references therein). Here I shall simply clarify the relation between the MMM assumptions about Universal Grammar, our definition of a DRT fragment as a theory of Language Acquisition and, given the assumptions about grammar in MMM, as a possible theory of the initial state of Universal Grammar. The brief proposal is as follows.
Universal Grammar is defined as $UG = L_0 = \langle D, +, \leq, \cap \rangle$, the tuple of syntactic and semantic operations and objects that are accessible to a child at the initial state, when a Language/Logic must still compute any values, as the subscript “0” suggests. The generative space that can be derived from this initial state is infinite, since it can represent any Language, or more accurately all the possible sentences that can be generated across different Languages, and their corresponding interpretations. Within this space, a Language-specific grammar can be represented as an indexed grammar, e.g. the grammar of English can be represented as $L_{EN} = \langle D, +, \leq, \cap \rangle$, the set of all possible interpreted sentences included in a(n idealized) grammar of English, from which the syntactic and semantic rules of this grammar can be “read off”, as respectively elements of the $LEX$ and $TYPE$ set. I shall assume that as an initial state, Universal Grammar contains information about Ps in the form of their syntactic and semantic representation, i.e. that each P belongs to a certain syntactic category (i.e. pp) and semantic domain (i.e. $R \leq W$). Consequently, the sixth and seventh predictions may be seen as predictions regarding how a child moves from an initial state of the grammar and slowly accesses knowledge regarding the properties of Ps, not necessarily restricted to English. In our case, we will focus our attention on one specific case, that of Ps and their logical properties, leaving aside the more thorny issue on whether these predictions can be extended to other parts of speech (and Languages) as well, and thus whether our theory gives evidence in support of the Universal Grammar hypothesis or not.

A more thorough discussion of this approach to grammars and its universal properties, beyond the mere sketch proposed here, is discussed in Keenan & Stabler (2003). I shall assume, like in MMM, that UG can be seen as the initial state of our DRT fragment, and can generate a grammar of English as one of the many possible grammars. In particular, I shall assume that one particular sub-set of this grammar, the grammar of P, can be successfully generated via the basic rules we have defined so far. I shall however spell out the predictions about acquisition in the next chapter.
5.4 Conclusions

In this chapter I have analyzed more in detail how two psychological processes, Language Processing and Language Acquisition, can be modeled within our DRT fragment. I have suggested, in line with much reviewed literature, that our DRT fragment easily allows us to integrate processing aspects of phonology, and permits with some minimal new assumptions to model language acquisition as a process of “model update”, hence offering a formal counterpart to the Modularity Matching Model. The next chapter will offer empirical evidence supporting the predictions offered in this and the previous chapter, and thus will offer empirical evidence in support of our theory of Ps, as well as answer to our third global research question.
Chapter 6

The Psychological Reality of
Adpositions, Part II: Experimental
Data

6.1 Introduction

In this chapter I will offer experimental evidence in support of the theory of Ps I outlined in the second part of thesis, and thus offer an answer to the third global research question offered in chapter 5. Informally, I shall offer evidence supporting the view that Ps denote different types of “spatial” relations, such as the “dynamic” ones denoted by directional Ps (e.g. to, from), and the “static” ones denoted by locative Ps (e.g. in, at); and that the logical relations of entailment (e.g. to entailing at), and sub-set (e.g. in being “part” of at) are accessed by speakers, in the opportune explicit contexts. The obvious empirical import of offering this answer is that our DRT treatment, which so far can only be seen as an off-line theory of how the syntactic-semantic processing and acquisition of Ps comes about, will turn out to be an on-line theory supported by the data collected from native speakers of English, adults and children alike.

Given the experimental nature of this chapter, the central goal is to present experimental
evidence that supports our theory. However, there are three research questions that underpin these data and will be addressed before the experimental studies are presented in detail. These three research questions are:

Q-1: What do we know so far regarding the interpretation of Ps in adults, and what Ps are still in need of a more thorough investigation;
Q-2: What do we know so far regarding the interpretation and acquisition of Ps in children, and what Ps are still in need of a more thorough investigation;
Q-3: Why the acquisition process of Ps in children emerging from the literature follows certain developmental patterns, and how it is connected to our empirical predictions;

The answers to these three research questions will form the platform from which I will present the three experiments that will offer support in favor of our theory of Ps. Anticipating matters a bit, I will offer evidence that focuses on a certain group of Ps which are still understudied in the experimental literature, as well as the entailment and sub-set relations holding between these Ps and the sentences they are part of.

The four target Ps (and relations) are to, from, in and at, offered again in the sentences below:

(260) All the tank engines have gone to the station
(261) All the tank engines have arrived from the station
(262) All the tank engines are sleeping at the station
(263) All the tank engines are sleeping in the station

The sentences in (264)-(266) offer an example of the test sentences that I used in the three experiments investigating the on-line interpretation of Ps, which will offer an experimental validation to the theory of Ps outlined in Part II.

The three experiments were the following. One study covered the interpretation of Ps in (Australian) English speakers, and tested whether the predictions regarding their interpretation
accurately describe this interpretation. A second and a third study covered the acquisition of the interpretation of Ps in (Australian) English children, as it unfolds over time. The first study acted as a control study for the second and third one. It was first tested whether our theory correctly predicts adults’ interpretation of Ps, then the results are compared against children’s results, so that it is also possible to assess whether the continuity and the subset hypothesis hold, when restricted to Ps. The evidence offered is positive, and strongly suggests that our predictions regarding the interpretation of Ps have a psychological reality, i.e. they correctly account the way adults and children interpret Ps.

At the end of the chapter, once we will have presented the experimental data, we will also be able to answer the third global research question, repeated below:

Q-A: What is the interpretation of Ps and their logical relations by English speakers, and how does this interpretation emerge over time;

The answer to this global research question will thus offer a solution to our third outstanding problem, what is the on-line interpretation of Ps in English speakers, and thus what is the psychological reality of the theory of Ps offered so far. This chapter is organized as follows. In section 6.2, I review previous literature on Ps. In section 6.3, I review experimental methods and analysis, and present the studies. In section 6.4 I offer some conclusions.

6.2 Previous Literature on the Processing of Ps

In this section I review previous research regarding the interpretation of Ps, and evaluate whether these works support, and to what extent, the predictions made by my theory, and more in general the “relational” approach to the semantics of Ps outlined in part II. I shall first analyze previous work on adults’ interpretation of Ps (section 6.2.1), and then review experimental studies regarding children’s acquisition of Ps (section 6.2.2). I then offer an explicit proposal regarding how Ps emerge in children’s grammars (section 6.2.3).
6.2.1 The Processing of Ps in Adult English Speakers

The literature regarding Ps and their processing is quite rich, although it tends to be focused on certain groups of Ps. As seen in chapter 2, section 2.2.2, recent research in psychology has studied in detail how sentences including Ps are processed and compared against (matching) visual scenarios. In this section I shall discuss in some more detail the relevant theories and works, rehearsing and expanding the relevant details of the discussion from chapter 2, and which types of Ps have been investigated. I shall also discuss how these data compare to our theoretical proposal offered in part II, highlighting which empirical voids emerge as still in need to be filled.

The *Functional Geometric Framework (FGF)* approach of Coventry & Garrod (2004) finds its empirical justification on a series of studies on the interpretation of several locative Ps (e.g. *above, under, on, in*) against appropriate visual contexts. The basic assumption is that each P is represented in long-term memory as a combination of mechanical, geometrical and affordance-based properties, and that when agents interpret Ps and the sentences including them against a visual context, they will find a sentence more or less felicitous depending on “how” much of this information matches the visual input. Typical experiments involved picture-matching tasks, in which the experimenters displayed a series of pictures to participants and asked them whether certain sentences were a correct description of facts or not, or were asked to judge which sentence was more appropriate to describe a certain scenario, by scoring the sentence on a 1 – 5 scale, with “1” being the worst and “5” being the best degree of appropriateness, respectively (“Lickert scale”: e.g. Coventry, Carmichael & Garrod 1994; Coventry & Prat-Sala 1998, 2001; Coventry 1998, 1999, 2003; Coventry, Prat-Sala & Richards 2001).

One example is the following. In a standard experimental set-up, participants were showed a picture of a man holding an umbrella to protect himself from the rain, with the umbrella at different angles of inclination, with respect to the body. Participants judged the sentence:

(264) The man is under the umbrella

To be maximally appropriate (e.g. judged as a “5”) when matched against a picture of a man
holding an umbrella right above his head, but less appropriate when the umbrella did not protect
the man completely, e.g. when the umbrella’s main axis was not aligned with the man’s. These
and further works within this framework investigated several projective English Ps, e.g. above, under, below, but non-projective Ps such as in, on, near, far (e.g. Coventry & Garrod 2004,
2005; Coventry & Frias-Lindqvist 2005; van der Zee & Slack 2003; Carlson & van der Zee
2005; Coventry & Guijarro-Fuentes 2008; Coventry et al. 2009; Mix et al. 2010). Other Ps
have been left out, however, including all directional Ps (e.g. through, to) and several non-
topological Ps (e.g. on, at, beside). Consequently, the picture offered by this framework is to an
extent incomplete.

Although these works do not follow a model-theoretic approach, their findings suggest that
the interpretation of many projective Ps (e.g. in front of), but also some non-projective ones (e.g.
in) can be based on a richer model of “Space” which includes non-geometrical information, as
the one I advocated so far. More importantly, the data stemming from these studies suggest, at
least indirectly, that Ps are interpreted as relations between two entities, for instance a man and
an umbrella; and that these relations are in a sense “restricted” in their inherent interpretation.
For instance, when participants accepted a sentence such as (264) against the appropriate type of
picture, they indirectly accepted that under captures a relation between man and a certain part of
the umbrella, its lower vertical “region”, and they preferred cases in which this relation is clearly
defined, i.e. it does not overlap with other relations.

Other theories about adults’ interpretation of Ps suggest similar conclusions. For instance,
the Vector Grammar theory of O’Keefe (1996, 2003) offers a purely geometric-oriented ap-
proach to (English) Ps. It suggests that Ps convey information about vector fields, sequences of
vectors that represent the shortest connection (“path”) from ground to figure, as it is represented
in the Boundary Vector Cells (BVC), cells which have the function to represent spatial config-
urations in the Hippocampus. Consequently, when one interprets under in (264), the BVC cells
will fire in such a way that they will represent a vector field going from the representation of the
umbrella to the one of the man.

195
The Attentional Vector Sum model (AVS) offers a model regarding the interpretation of Ps in which Ps denote clusters of “visual” aspects (mechanical, functional, geometric) that act as vectors, pointing to a region of space that should be the attentional focus of an agent (e.g. Regier & Carlson 2001; Regier & Zheng 2003, 2007; Carlson et al. 2003, 2006; Regier et al. 2005). According to this theory, the P under in (264) is interpreted as a set of visual properties, modeled as vectors, that direct the agent’s attentional focus from the umbrella to the vertical, negative version of this object, where the man is found. Similarly to FGF, it is assumed that agents will focus their attention to a visual region which maximally instantiates the underlying meaning of a P: in our case, agents will focus on the region immediately the lower space of the umbrella rather than side regions.

Both frameworks offer evidence to support their claimed based mostly on projective Ps as in the case of the FGF framework. The evidence they offer, setting aside different assumptions and theoretical motivations, is quite similar to the one offered in FGF, also in missing any evidence regarding directional Ps (e.g. to and through). English speakers appear to interpret both types of locative Ps as relations between entities with a definite, but not exclusive, “geometric” component, which are modeled in both theories as “vectors”. Both theories thus support, although in a different and indirect format, our relational approach to the semantics of both in and in front of, much like the FGF framework. If agents interpret sentences involving Ps as denoting relations between “spatial” regions, then they should accept these sentences against corresponding visual scenarios, as these studies show.

A series of works focusing on a more detailed analysis of non-projective Ps can be found in Feist (2000, 2002, 2004, 2006, 2010); Feist & Gentner (2002, 2003). These studies contain a detailed analysis of how different languages vary in capturing the notions of “containment” vs. “attachment” or “support”, what is roughly associated (in cognitive semantics literature) to English interpretations of in and on. For instance, they show that while in English the most “appropriate” way to express the relation between a medicating bandage and a hurt leg is via on (e.g. the bandage is on the leg), whereas in Spanish, the most appropriate expression is en
Similar findings are offered in Feist (2006), which shows that *at* is used when agents do not know the precise location of a figure, unlike *in* (e.g. *Mario is in the cinema* vs. *Mario is at the cinema*). As we can see, these works are also indirectly consistent with our approach to Ps, since they offer evidence suggesting that non-projective Ps, such as *at* and *in*, are interpreted as more “general” relations between two entities.

The literature reviewed so far mostly focuses on locative Ps, both non-projective and projective ones. Recent works shed some light on some directional Ps. A recent study, Papafrogou (2010), offers production data on certain directional Ps, e.g. *towards* and *out of* (i.e. “Goal” and “Source” Ps of Jackendovian tradition), and briefly analyzes how children and adults may interpret these Ps in context as relation between a moving figure, reaching (or leaving) a ground during an event of motion. The works in Language Acquisition of Stringer (2005, 2006a, 2006b) also include adults’ data, taken as a control group sample, that shed light on directional Ps such as *through*, *across* and similar others. While the focus is mostly on production, these studies also include comprehension tasks, which suggest that adults interpret a P such as *through* as denoting a “dynamic” relation between a moving figure and a ground, in which the figure goes in and out of the ground. These works also support the relational view on Ps advocated so far, shedding some light on some otherwise poorly studied directional Ps.

The reviewed works offer a rich and interesting set of data regarding the interaction between implicit context and Ps, although they give a relatively indirect picture of the interpretation of mostly projective Ps. Once we leave aside non-linguistic aspects, such as the degree of accuracy by which sentences match pictures\(^1\), these works suggest that the interpretation of Ps such as *in*, *under* or *above* consists in expressing a relation between certain regions of space (or a specific region), which can also include more abstract information (as I argued in chapter 2), often modeled in terms of vectors (Vector Grammar, AVS). However, since much attention has been paid to projective Ps, we still now little about the interpretation of some directional and some non-projective locative Ps: for instance, we do not really know much about the interpretation

\(^1\)See again 2, section 2.4.1 on this topic, however.
of to, from and at, among others. Consequently, previous literature on the interpretation of English Ps in adults is consistent with our theoretical assumptions, but does not cover a number of non-projective Ps, and says little about directional ones.

At this point, we are able to offer an answer to the first research question:

Q-1: *Adults interpret Ps as expressing information regarding the spatial relation holding between figure and ground; however, little evidence is known regarding most (non-projective) locative Ps and directional Ps, and their logical relations;*

The first research questions expands and clarifies some of the topics mentioned as early as chapter 2 (section 2.2), and tells us that Ps such as in, at, to and from, as well as the logical relations (entailment and sub-set relations) holding between these Ps are still in need of a more solid empirical verification.

### 6.2.2 The Emergence of Ps in English-Speaking Children

The literature on the acquisition of Ps, both on adpositions and other members of this category, is rich and quite varied, although the study of production (in children) takes the centre stage, as opposed to comprehension. For instance, Clark (1973) reports the early emergence of in in young children (as young as 1;3 years), who overproduced in as a sort of “general” P. A more detailed study is Johnston & Slobin (1979), which focuses on English, Turkish, Italian and Serbo-Croatian children between 2;3 and 3;5 years of age, and their production of projective Ps (e.g. front, back) in their everyday language, offering some evidence that the “horizontal” and “vertical” terms (e.g. front, up) emerge before the “lateral” (e.g. left) ones\(^2\). Similar evidence can be found in Durkin (1981), who observed that children in their second year of age spontaneously produce several Ps, often starting from in or other Ps denoting “basic” relations (e.g. on, under).

A rich program of research on Ps, also studying their categorization and influence on memorization, is that of Dan Slobin and associates (e.g. Slobin 1985, 1992, 1997; and references

\(^2\)The authors compare languages which have different members of Ps to realize spatial concepts (e.g. case markers in Turkish, Serbo-Croatian), and observe that case markers tend to appear somewhat earlier than adpositions. This aspect is not crucial, here.
therein). These studies offer data from a cross-linguistic perspective, and focus in particular on Talmy’s “verb-framed” vs. “satellite-framed” distinction (e.g. Talmy 1978, 1983, 2000). These works use a specific type of elicitation task, known in the literature as the “frog task”\(^3\). In this kind of task, the experimenters use an illustrated and text-less book in which a little frog escapes from its jar, and the frog’s owner has several adventures in his attempt to find the little frog. Children are invited to describe a picture and the illustrated scenario (e.g. the frog going under a tree trunk), and thus to produce a P corresponding to this scenario. These answers are then analyzed against adults’ data, to study how children acquire Ps and underlying concepts over developmental time. An early example is Slobin & Bocaz (1988), which studies the development of narrative abilities in English and Spanish children, aged 4 to 8 years. Other important examples including English data are e.g. Berman & Slobin (1994); Slobin (1996, 2004); Ragnarsdóttir et al. (1997); Naigles & Terrazas (1998); Brown (2004); Slobin (2000, 2003, 2005).

Most of these works offer evidence suggesting that children, when able to access Ps, can do so in an adult-like way; and that possible mistakes often consist in producing syntactic structures which are not observed in one Language, but acceptable in another (as per continuity hypothesis: Pinker 1984; Crain 1991; Crain & Thornton 1999). For instance, Stringer (2005) offers cross-linguistic evidence (English, French, Japanese) from children in the 3;0-7;0 years age range. He shows that English children may produce ungrammatical forms such as *he went crossing the river* (Stringer 2005:196), which are acceptable syntactic structures in Japanese, however. Similarly, children of 3 years of age tend e.g. to choose *crosses* instead of *goes across* roughly 10% of the time, whereas adults in the respective control groups do so roughly 4% of the time (Stringer 2005:178-182).

Other works also focus on production and offer similar findings, but suggest that the different cross-linguistic patterns support a different pre-linguistic partition of spatial concepts. For instance, several studies focus on how children learning different languages (e.g. Korean, English, Dutch, German, Tzeltal) develop spatial concepts such as “tight attachment” and “loose

---

\(^3\)The “frog, where are you?” stories originate from an old children’s book from the 60’s, Meyer (1969).
attachment”, as respectively expressed by the PPs in the box and inside the box, in English. They suggest that since English children learn express these concepts via the Ps in vs. inside, they will express a different non-linguistic concept than Korean children, who learn to express these concepts via the verbs keitta vs. kkitta (lit. “(be) in” vs. “be inside”: e.g. Choi 1997, 2006; Choi et al. 1999; Bowerman & Choi 2001, 2003; Bowerman 1991, 1995, 2001; on German, Dutch; Brown 2001, 2004 on Tzeltal; inter alia4).

Works which do not follow a cognitive semantics tradition and focus on comprehension are somewhat less frequent, but certainly important. A classic study in Miller & Johnson-Laird (1976) reports anecdotal evidence on how children interpret non-projective Ps such as in, on, at. Such evidence, however, is quite coarse-grained, although methodology and children’s judgements are not reported in detail. Several subsequent works have investigated how children can acquire various projective Ps when the implicit context give enough cues to their meaning, often investigating quite young children (e.g. top or bottom in 2;0 years old children, Clark 1980; above, between, in front of, in range 2;0-4;0 years old children, Johnston 1981, 1984; from, with and by in 3;0-4;0 children, Clark & Carpenter (1989); Vandeloise 1994, 2005; on French en in 2;0-4;0 children).

More recent research has both investigated cross-linguistic data, including Sign Language, and has offered a more precise picture regarding the emergence of the interpretation of Ps (e.g. Choi 1991, 1995; Vandeloise 1994, 2005; Bowerman & Choi 1995; Thorseng 1997; Sinha et al. 1999; Clibbens & Coventry 1996 on Sign Language; Richards, Coventry & Clibbens 2004; Richards & Coventry 2005; Coventry & Guijarro-Fuentes 2008). They offer evidence that comprehension of basic locative Ps (e.g. in and on) tend to emerge quite early in child language (approximately 1;7 years), usually followed or concomitant with emergence of projective Ps’ comprehension (e.g. under, over, above, etc.), and that can proceed until later years (e.g. 7;0 years, Richards & Coventry 2005). Some works (e.g. Choi’s) also suggest that comprehension may be different across languages, mirroring how non-linguistic concepts can differ from one

4A problem with this claim is that Korean also has a rather sophisticated case system to express spatial relations e.g. (Son 2006; Svenonius 2006). I will leave aside whether this fact goes in favor or against Choi’s hypothesis.
language to another, and are mirrored (early) on in children’s acquisition.

Very recent works have uncovered evidence showing that children as young as 1;3 years may understand basic spatial relations and correctly associate them to pictures (e.g. inclusion, support, lower position) even if they cannot yet produce the corresponding Ps (e.g. *in*, *on*, *under*: Rohlfing 2001, 2004, 2005; Choi & Rohlfing 2010). The studies in Stringer (2005, 2006a, 2006b) reported a set of experiments that tested the production of Ps in children ranging from 3;0-7;0 years of age, and which investigated the comprehension of several Ps, including many directional Ps (e.g. *across*, *along*, *through*). These tests involved both “frog story” tasks and complex picture-matching tasks, in which children were asked whether they accepted sentences involving e.g. *through* (e.g. *the car has gone through the tunnel*) as correctly describing a sequence of pictures in which a car first went into and then out of a tunnel. Consequently, they study both production and comprehension of Ps in children, with a focus on the former. They also include data on adults’ production and comprehension, as we have seen in the previous section, which offer some evidence regarding directional Ps. For the most part, children offered adult-like answers, by e.g. considering as wrong a *through*-sentence when matched with a sequence of pictures in which a car did not exit the tunnel, by the end of the sequence.

These works invite the conclusion that children understand Ps in an adult-like manner, once they can access their interpretation. For instance, children were able to access an adult-like interpretation of *through* as involving an figure moving in and out of a ground, roughly during the latter part of their fourth year; once this interpretation became accessible to them, they were able to correctly accept true sentences involving through, and reject false sentences involving this Ps. Summing up, children’s interpretation of locative Ps, the ones which have received the most attention in the literature, is consistent with our theoretical assumptions. Children, much like adults, interpret Ps as expressing “spatial” relations, relations involving a figure and a ground, which can involve more than “just” geometric information, but which are inherently restricted to a specific type. The data discussed so far suggest that children interpret locative Ps such as *on* and *in front of*, but also (some) directional Ps such as *through*, as spatio-temporal
and causal relations involving two entities; hence, that their interpretation is adult-like, once it becomes accessible to them.

We thus have a wealth of evidence regarding children’s interpretation of most Ps as well as indirect evidence of how adults interpret some directional Ps, but we have little evidence on how children interpret certain “core” non-projective and directional Ps, namely at, from, to, in, at least with respect to its exact interpretation. As in the case of adults’ data, then, we are in need of experimental findings that shed light on these Ps, as well as testing whether our theoretical predictions correctly account their interpretation in adults and children.

At this point, we are able to offer an answer to the second research question:

Q-2: Children interpret Ps as expressing information regarding the spatial relation holding between figure and ground; however, little evidence is known regarding some core Ps which are poorly studied in adults as well, as well as the logical relations of entailment and sub-set holding between Ps;

The second research question tells us that we know little about children’s interpretation of the four Ps in, at, to and from, and their logical relations, as in adults’ case. Consequently, we now have a clear picture of which Ps must investigated further, and can thus offer a solid base for our experimental study. The next section will address the third research question, before presenting the studies in detail.

6.2.3 The Emergence of Ps: a Proposal

The literature reviewed so far offers a rich set of data, which is intuitively consistent with our theory of Ps. However, in order to make more precise the relation between this literature and the theory of Ps I have presented in the previous chapter, I shall analyze which of the seven predictions are supported by these works, and which appear to be still in need of empirical validation.

The first prediction states that Ps can denote a “static” relation or a “dynamic” one. The data
on projective Ps for adults and children support part of this prediction, as we have discussed (e.g. Coventry & Garrod 2004; Stringer 2005). Some relevant data are however missing: we have no exact data regarding the interpretation of *in* and *at* and the entailment relations they participate in, for both adults and children.

The second prediction states that Ps can also denote a “change” relation, such as *to* and *from*. Indirect support for other directional Ps (e.g. *through*) confirms that this prediction is overall correct (e.g. Stringer 2005). However, the data on both populations (adults and children) show that this prediction must still be confirmed for these two Ps, in particular with respect to their entailment relations.

The third prediction states that Ps will either express a certain order of situations or its reverse order. Both sets of data support this prediction, with sets of data such as Johnston & Slobin (1979) suggesting that the “positive” antonyms (e.g. *front*) tend to emerge in children’s production before the “negative” antonyms (e.g. *behind*).

The fourth prediction states that projective Ps can be interpreted in a non-projective way. This prediction is indirectly supported by both sets of data, in particular those data that show how adults focus their attention on certain zones of space, or Stringer’s data regarding how children produced e.g. *across* in both “static” and “dynamic” scenarios (e.g. Stringer 2006a).

The fifth prediction states that Ps may denote the location of a figure as occupying a “complex” location. Data stemming from e.g. Coventry & Garrod (2004) and their experiments indirectly support this view, since they suggest that agents may focus their attention on complex regions of space. One P which is still in need of a more thorough investigation is *at*, which respectively expresses general location, as discussed in chapter 4. With *at*, its related entailments are also in need of empirical evidence. Both predictions are in need of explanation, consequently.

The sixth prediction states that Ps emerge after the most specific P *in* emerges as the first P. This prediction is supported by various works supported so far, e.g. Vandeloise (2005); Rohlfing (2001). It also predicts that the understudied Ps *to* and *from* should emerge once *in* has emerged, and that *at* as a “general” P is possibly the last P to emerge (e.g. *to, from*: see discussion in
chapter 3; cf. also Levinson & Meira 2003). This prediction is still in need of testing, as well.

The seventh prediction states that once Ps emerge, the entailment relations they are involved in should emerge as well, and become part of their interpretation. Once children can properly interpret *to* and *from*, they will also accept the entailment relation holding between the sentences involving these Ps, and thus be able to access *at* as a P which is entailed by other Ps, hence the (possibly) last element to emerge in this sub-domain, or at least it should emerge once directionals that participate in entailment relations with *at* should also emerge. It also predicts that the sub-set relations holding between Ps should emerge, once a child can access both Ps. For instance, a child should accept both *at* and *in* as true in the opportune context, since their interpretation “overlaps”.

A rather idealized model of this emergent process, ascribed to our understudied Ps, is the following:

![Diagram of Ps]

**Figure N.1**
The lattice/diagram of Ps. Red lines denote the partial order of emergence of Ps and entailment relations, blue lines the sub-set relation holding, via transitivity, between *at* and *in*. 

204
The intuition behind the highly idealized model in (261) is that a child, once he can access the sub-domain of Ps, will slowly update the model as soon as he is able to successfully interpret Ps. A child will start from \textit{in}, become able to access \textit{to} and \textit{from} (and several other Ps) at some further point, and then converge towards the acquisition of \textit{at}, as the process unfolds. This prediction must still be tested, since few, if none of the works mentioned so far contemplate the theoretical hypothesis that Ps can be “connected” via logical relations, let alone attempt to test its psychological reality.

An important consideration regarding predictions is the following. While previous literature offers (indirect) support that the third, fourth and (part of the) sixth of our seven predictions are basically correct, it says little about the first, second, fifth, sixth and seventh predictions with respect to adults and children. Previous literature offers consistent support in favor of a relational approach to the semantics of Ps both for adults and children, although it does not cover certain Ps, such as \textit{at, to, from}. It does not cover, or at least it does so in a very indirect way, the logical relations that hold between Ps, such as entailment and sub-set relations. This is not surprising, since virtually all models of Ps discussed so far do not attempt to offer a “model” of Ps as a “structured” category.

Our lattice-based approach, on the other hand, strongly suggests that precise logical relations hold between Ps, which can also represent how these relations emerge over time. In our theory Ps such as \textit{in} and \textit{at} both represent “spatial” relations; however, since \textit{in} denotes a very specific relation between two objects (e.g. inclusion), the domain of this relation will be a part of the more general domain of \textit{at}, which denotes a very general relation between two objects (e.g. inclusion, or proximity, or partial inclusion, etc.). Consequently, our theory has now two broad empirical goals to be met. First, we must find evidence regarding our predictions on the interpretation of our core Ps, and whether this evidence is consistent with previous findings. Second, we must find evidence regarding our predictions on the logical relations between these core Ps (entailment and sub-set), which doubles as evidence in favor of our lattice-based approach to the
emergence of Ps.

The remainder of this chapter is aimed at offering experimental evidence suggesting that these predictions are borne out, in the following way. I shall present data regarding three experiments, one on adults’ interpretation of Ps, and two on children’s interpretation and acquisition of Ps. These data will present evidence on how participants interpreted the four target Ps: in, to, from and at. They will also present evidence on how participants interpreted the relevant entailment and sub-set relations: in particular, they will present evidence regarding the entailment relation holding between to and at, the sub-set relation holding between at and in, and the “mixed” relation holding between from and in.

At this point we can offer an answer to the third and final research question:

Q-3: The acquisition process of Ps in children appears to definable as the consequence of children acquiring the logical relations between Ps, i.e. as accessing Ps from least complex to more complex;

This answer tells us that, if children acquire Ps starting from in and progressing through the Gamut of Ps without a very precise order (e.g. acquiring in front of and through roughly at the same time), then this process appears to realize, over developmental time, a general process of “model update”, as we have suggested in the previous section. Intuitively, children first acquire in as expressing the relation of spatial identity (the most specific relation), and then acquire less and less specific Ps, possibly acquiring at as the last (and most general) P. In doing so, then, they should also acquire to and from before at, since these Ps represent preliminary “logical steps” in this process of acquisition. In this case, logical relations also represent patterns of development, as we suggested in chapter 5.

The remainder of this chapter offers experimental evidence in support of this answer and the other research answers and related predictions. It will offer an answer to the global research question of Part III, repeated below:

Q-A: What is the interpretation of adpositions and their logical relations by English
The data answering this question will be presented in the next section.

6.3 The Interpretation and Acquisition of Ps: the Experiments

In this section I present three experiments on the interpretation of Ps. I introduce the task used in the experiments in the remainder of this introduction. In section 6.3.1, I present an experiment regarding adults’ interpretation of our target Ps; in section 6.3.2 and 6.3.3, I will respectively present two experiments regarding children’s interpretation of Ps, by presenting data on two Australian English-speaking children. In section 6.3.4, I will offer a general discussion. Before moving to the experiments, I shall briefly outline the type of experimental task used in each experiment.

In each experiment I used a simplified variant of the *Truth Value Judgement Task* (henceforth: TVJ task, Crain & Thornton 1999). I will briefly motivate this choice and summarize how the standard format of the task works, referring the reader to Crain & Thornton (1999) for a more thorough introduction. I shall start from its motivation for children’s experiments. Most of the reviewed literature on Ps and their production in children is centered on the “frog task” and on Ps’ production. Although this task is particularly useful for the elicitation of production data, it makes the testing of interpretation of Ps and their entailments problematic. Since children are “free” to express a possibly small but non-unique set of Ps to describe a story, it is not possible to test how children may not only interpret any of our target Ps, but most importantly whether they accept the entailment patterns holding between Ps, since an experimenter would need to elicit two sentences from the same picture.

The TVJ task allows testing how children interpret sentences, and more specifically Ps, by eliciting a yes-no answer from a child who has observed a brief story, hence eliciting whether a statement accurately describes a particular situation alluded to in some context or preamble. A standard TVJ yes-no task involves two experimenters, possibly a third experimenter acting as a
cameraman. One experimenter acts out a certain story involving a number of toys and props in front of a child, while a second experimenter plays the role of a puppet (e.g. Kermit the Frog) who observes the story with the child. At the end of the story, the puppet offers a brief comment about he thinks that happened in the story, followed by a test yes-no sentence to the child. For instance, if the story involved a group of horses who went to one lake, Kermit can ask:

(265) Has every horse gone to the lake?

The child’s answer offers evidence on whether he can correctly interpret the sentence or not. In this case, if the child offers a “no” as an answer, he will offer evidence that he can correctly interpret the sentence and evaluate which of the two possible answers correctly represent the actual outcome of the story. This is roughly the basic format of the task; the specific versions used in the experiments will be presented in the relevant sections.

The TVJ Task can be used with a condition regarding the plausibility of both answers (positive and negative), by having both the equivalent scenarios to be possible outcomes in the story. This is the Condition of Plausible Dissent (henceforth: CPD, Crain & Thornton 1999:225-226).

The nature of this condition is relatively intuitive to grasp: in order for a sentence to be considered true, its counterpart must have been true at some earlier moment in discourse. Importantly, the possibility that a false answer at some point was acceptable is made explicitly during the acting out of the story. For instance, in a story that can be followed by a question such as (265), each horse involves in the story makes explicit his intention to reach the lake; however, one horse is not able to reach the lake (for instance, he sprains his ankle), so he can’t reach the lake. Hence, the explicit context (i.e. the story) makes clear that both answers are equally acceptable, although only one of them is actually true (in this case, the one corresponding to the answer “no”), and the other false (in this case, the one corresponding to the answer “yes”), since not all horses have gone to the lake.

Although this condition is far from controversial (see Meroni et al. 2006 for a review), it offers a good testing criterion for Ps, since it predicts that e.g. to in (265) will be false when some of the horses will not have completed an event of going to the lake, but true once the missing
horse(s) have completed this event. Consequently, one can offer questions which test the experimental hypotheses at different “phases” of the story, and obtain equally valid, although slightly distinct, types of data. The specific use of this condition will be spelt out in each experiment, however.

6.3.1 Experiment 1: the Adult Group

The goal of this study was to test how adults interpret the four target Ps (to, in, from, at), and whether they accept the entailment and sub-set relations defined among these Ps (i.e. respectively between to and at, between in and from; and between at and in). These data were then used as the baseline for the data regarding children’s experiments (i.e. second and third experiment).

Participants

22 undergraduate participants from the department of Psychology were tested, who received course credit as a reward for attendance.

Materials

The experiment involved a simplified version of the TVJ yes-no task, adapted for adult participants. This variant was chosen to test participants in a less “entertaining” and faster way, since adult participants could only attend one session, but as adults could offer a more focused participation to the experiment. The changes from the standard TVJ task were the following. Each experimental session involved a power-point presentation which depicted 7 stories (one per experimental hypothesis) involving a number of fictional characters, mostly taken from “Thomas and friends” fictional universe. Each slide was accompanied by text consisting of two or three sentences, depicting the events occurring in the story. The text was read aloud by one experimenter, myself, to ensure that all participants knew the nature of the events depicted in the slides.
No second experimenter was involved, since this version of the task could be carried out by a single experimenter.

Before the story, a brief introduction explaining the answering procedure was offered. This introduction also presented the main characters in the story, and the character who was going to offer the questions, called “Mr. Little Bears”. “Mr. Little Bears” was described as an amnesiac teddy bear, who watched the stories with the participants but, because of his memory problem, he had to ask a question regarding the story, at the end of the narration. The power-point slides were shown on a 25” Mac, at the Language Acquisition Lab. Each participant was given a pen and an answer sheet on which to sign his/her answers during the experiment. The following target questions were used in the experiment, with a brief lead-in, i.e. a comment by “Mr. Little Bears” in the story offering the questions to the participants. A typical lead-in was the comment “our tank engines look very hungry today. I don’t remember one thing, though.”; other lead-ins followed a similar structure.

In order to use the CPD in the aforementioned way, the administered test sentences (questions) contained DPs in subject position that were universally quantified, i.e. they included the universal quantifier all (e.g. all the horses). In this way, each sentence was false at some earlier moment in the experiment, and true once each of the relevant props on display performed a given action. A note on interpretation before we move to the predictions. The proper treatment of all and other quantifiers in our DRT fragment would require some changes in our assumptions (e.g. introduction of duplex conditions). Here I shall just assume that all acts as a form of “distributive” operator: when it combines with plural definite DPs (e.g. the horses), it says that a given predicate must be applied to each of the referents in the denotation of the DP (all the horses e.g. Keenan & Faltz 1985; Link 1998; Brisson 1998, 2003). The specific predictions are the following.

The first four questions were offered in the following order (see discussion below):

(266) “[Lead-in] Have all the tank engines gone to the farm?”

(267) “[Lead-in] Are all the tank engines showering at the water tower?”
The entailment questions were offered as pairs of questions after a given scenario, and in the given order:

(270) a. “[Lead-in] Have all the tank engines gone to the farm?”
    b. “[Lead-in] Are all the tank engines eating at the farm?”

(271) a. “[Lead-in] Have all the tank engines started from the farm?”
    b. “[Lead-in] Have all the tank engines eaten breakfast in the farm?”

(272) a. “[Lead-in] Have all the tank engines arrived at the hotel?”
    b. “[Lead-in] Are all the tank engines sleeping in the hotel?”

A note on the entailment relations. Recall from chapter 4 that Parsons (1990:78-84) discusses these entailments as being licensed when both paired sentences involve the simple past or the future. For sentences involving verbs in the present perfect tense, he suggests that the entailments are licensed when the explicit context (i.e. previous discourse) confirms the completion of the event of motion. Thus, the entailment relations we discussed so far can also be licensed when pairs of sentences such as (270) and (271) are involved. Our experiments satisfied this contextual requisite, making the testing of these entailment relations valid, and intuitively more accurate, since the stories described events concluded a few moments before the question was presented.

The exact use of these questions in the story is described below.

Procedure

Participants were shown the seven stories in a sequential way. As the main experimenter, I offered a brief introduction, explaining that participants had to closely watch the story and follow the narration, and then circle their answer on the answer sheet. At the end of each story, the target question was offered, A fifth character, “Mr. Little Bears” (a plushie bear) also appeared as a witness to the stories but, since he had bad memory, he had to ask the participants about what
happened in the stories. This happened at the end of each story, with a final slide showing Mr. Little Bears and one of the target questions. Participants had to circle their preferred answer (“yes”, “no”, “not sure”) after each story and, if they changed their mind, they could cross the first answer and choose a second answer. In such a case, only the second answer was considered valid.

The order of the tested Ps was the following: first, the four basic Ps were tested (i.e. to, at, from, in, in this order), then, the two entailment relations and the sub-set relation were tested (i.e. “to|=at”, “in|=from”, “in≤at”, in this order). In the “basic” stories, one of the four s changed his mind during the story so he did not end up at a given location. The correct answer to these stories was “no”. In the entailment stories, the missing tank engine joined the other tank engines at the end of the story, making the target question true. Participants were asked to answer two questions in a row, after each story. For instance, participants were shown a story in which each train went to the farm, and ended up being at the farm afterwards; Mr. Little Bears asked a to-type question and at-type question, one slide after the other. The correct answer to both questions was “yes”. In both types of stories, the CPD was respected because both outcomes were possible at some possible moment in the story, but only one outcome was realized at the end.

Let us make the predictions precise. Adult participants should have interpreted to as denoting that each tank engine reached the farm after changing position; they should have interpreted in as denoting total inclusion, i.e. that each tank engine should have been located in e.g. the hotel at the end of the story; they should have interpreted at as denoting that each tank engine was located somewhere around the water tower; they should have interpreted from as denoting that each tank engine started its movement in the same location, and then went somewhere else. Hence, they should have answered “no” to each question regarding the four stories, since the “false” outcome was the correct one. In the entailment/sub-set scenarios, participants should have maintained the same interpretation, and accessed both pairs of sentences as related (i.e. true in the same explicit context). So they should have answered “yes” to each pair of questions, since the “true”
outcome was the correct one. Consequently, we expect that the answer percentages should have been 5%/90%/5% for the first four stories, and 90%/5%/5% for the fifth to seventh story.

Results and Discussion

The main findings of this experiment were that the hypotheses were borne out. The observed answer percentages were the following. Given N=22, for the to-type answers, the percentages were 0%/100%/0%; for the at-type answers, the percentages were 0%/100%/0%; for the from-type, the percentages were 4,5%/95,5%/0; for the in-type, the percentages were 4,5%/95,5%/0. Consequently, the results strongly suggest that the predictions regarding the interpretation of these Ps are borne out. Note that, since the experiments presented idealized scenarios, the speakers’ “not sure” (undefined) answers stemmed from the possibility that adults genuinely did not know the answer, rather than from some accidental factor interfering in their interpretation. As the data suggest, this was never the case; furthermore, only two participants, one each for the from and in hypotheses, gave an incorrect answer. This was in line with our predictions.

Let us look at the entailment/sub-set data. I shall make a précis first. For entailments, I have considered each pair of “yes” answers as in favor of the hypothesis, whereas answer pairs in which either the first or the second answer were “no”, were considered as against the hypothesis. Undefined answers are those involving an undefined answer in any of the two “conditions” (cf. Landman 1991:150-55 for discussion). Recall that, in this case, participants should have answered “yes” for the most part, with “no” and “not sure” at chance, since the “true” answer was the correct one, as all the tank engines reached some location. Consequently, we would expect the answer percentages to be 90%/5%/5%, for “yes”/“no”/“not sure” answers. The results were the following. Given N=22, for the to|at-type of entailment, the answer percentages were 95,5%/0%/4,5%; for the in|from-type of entailment, the percentages were 95,5%/0%/4,5%; for the in≤at-type of sub-set relation, the percentages were 95,5%/4,5%/0.

These data very strongly suggest that our adult participants interpreted Ps according to our predictions, since the results were near-ideal, for the most part. Interestingly, the only par-
participant that offered a “not sure” answer on the fifth scenario (the “to = at” scenario) did so because she did not remember a small detail about the scenario, so she answered “not sure” to the at-question. Consequently, the entailment relation was considered “undefined”, as per assumptions. Other participants offered negative answers in similar cases, e.g. when they remembered incorrectly some detail regarding the story. These data are perfectly in line with our hypotheses: adult participants interpreted Ps according to our predictions, unless for some accidental, performance-based reasons, they offered an incorrect response, or chose not to commit themselves to an answer. It may be argued that the scenarios were clear enough that participants could easily avoid such problems, as the data suggest.

Let us look in detail at how the data support the specific predictions. The first prediction was supported by adults’ interpretation of at and in as denoting static relations (second, fourth story), and by their ability to be equally acceptable in the correct implicit contexts (seventh story). The second prediction was supported by adults’ interpretation of to and from as denoting “dynamic” relations (first, third story), which crucially involved an event of change. Both Ps were also accepted as being part of the corresponding logical relations (fifth, sixth story), hence the second prediction was also supported in their entailment-based aspect. The fifth prediction was supported by participants’ interpretation of at, but also of the entailment stories. For instance, since both Ps were interpreted as true in the last story, the possibility that in is interpreted as “part of” of at was supported, and with it the entailment component of these Ps (hence, also supporting the first prediction). These experiments strongly suggest that our predictions about Ps’ interpretation in adults’ are correct, and are hence consistent with previous literature on this topic. Let us move to the second experiment.

6.3.2 Experiment N.2: Terence P.

This experiment aimed to test the interpretation of experimental hypotheses in children, in particular how the different Ps and their interpretations became accessible over developmental time. For this purpose, one child was interviewed on a regular basis, and his data regarding Ps’ inte-
pretation collected via a simplified of the TVJ task, explained below.

Participants

The child involved in this experiment was Terence P. Terence P. was interviewed while in the age range 3;1-3;11 for a total of 17 sessions, during an age range on which there few data regarding our experimental hypotheses. All of the testing sessions were carried out at our Language Acquisition laboratory. Here, the child was interviewed on roughly a fortnight basis, for sessions ranging from 20 minutes to one hour. In most of the sessions, I was the only experimenter, since the set-up in the lab allowed to record sessions without the need of an experimenter acting as the camera-man.

Materials

Each session was recorded via the use of a portable video-camera mounted on a tripod allowing complete coverage of the experimental display, with a set of connected microphones. In all of the sessions, the child’s mother was present, but she was not aware on the exact nature of the tasks, and no mention was made regarding the experimental hypotheses in any extra- session chat. Each of the sessions was video-taped, then converted into Idmovie format and segmented into clips of variable length (3 to 10 minutes).

Each session involved the testing of several mini-stories in which the 7 hypotheses were tested on roughly a cyclic basis (e.g. first to, then at, then an entailment scenario, etc.). In order to avoid memorization in answer patterns, I used a wide choice of different “tank engines” (for a total of 41 different engines) and possible locations (for a total of 18 different locations), so that each test sentence was (often) tested against a different implicit context, as well as using different “motion” (e.g. going, jumping, running, etc.) and “location” verbs (e.g. sitting, sleeping, eating, etc.). Typical test sentences included, but were not limited to, the ones offered in examples.

5The pseudonym is an obvious homage to Terence Parsons, as the child’s true name was omitted for privacy matters.
When more opportune in context, questions involving resultative constructions were also used, in particular for to and from. Examples are the following:

(273) have you taken all the tank engines to the farm?

(274) have you taken all the tank engines from the oil factory?

Regardless of the specific syntactic construction, all sentences were true (or false) in exactly the same type of story, as per hypothesis. The role of the puppet offering questions was “taken” by a Godzilla puppet, which I used to narrate unfolding scenarios. The importance of using this puppet during the experiments will be explained in the next section.

**Procedure**

The use of a “live” experimental set-up was motivated by the need to actively involve the child in the stories, and test children before scholar age (hence, unable to fill in an answer sheet). Since the goal was to collect data over a prolonged period of time via regular interviews, the use of a dynamic set-up allowed the child to be actively participate in the experiments, and thus be able to participate and answer in several yes-no TVJ tasks over long periods of time, without the risk of external factors (boredom, distractions) influencing the outcome of the tasks. For instance, the child was often invited in setting up the target scenario (i.e. bringing a relevant group of tank engines to a given location), with the experimenter completing the task in several cases in which the child did not complete the requested task. This involvement granted that the child was for the most part focused on the events, and reduced the risk of distractions from extra-experimental factors.

In each session, the first two or three minutes were warm-up unstructured play time, so the child could get acquainted with the new toys. Similarly, the last five minutes were also unstructured play, so that the child could relax before leaving the lab. The variant of the TVJ task used in the experiment consisted in having the child to answers posed by Godzilla after a scenario was complete. For instance, after all the relevant tank engines went to a given farm,
Godzilla said that he was not sure of what happened and asked if the tank engines went to the farm. The experiment also differed from the adults’ experiment in two aspects.

The first aspect was the use of a simplified version of the TVJ task. In this simplified version, I acted as the only experimenter that participated in the experiments, acting out the tasks and controlling the puppet at the same time. The Godzilla puppet was introduced at the beginning of the experimental session, as a witness that had to make questions regarding the tank engines and their adventures: since he had poor eyesight, he could not properly see what events occurred during the stories. As the main experimenter, I first acted out a story involving some tank engines, then at the end of the stories acted as Godzilla, when the target question was offered to the child. Since the puppet was provided with the trademark long tail, it was possible to manipulate Godzilla while at the same time manipulating other toys on display, thus “collapsing” the roles of the two experimenters into one.

The second aspect was the use of both “true” and “false” outcomes as possible target answers. In a typical story, I narrated events similar to those depicted in the adults experiments, for instance by showing a group of tank engines attempting to reach a given farm. One of the tank engines, for various reasons, had problems in trying to reach this farm. The child, then, was either asked the target question before the target scenario was completed, or after the scenario was completed. Consequently, in the first case the child had to answer “no” to the question, whereas in the second case he had to answer “yes” to the question, in order to offer an appropriate answer. The CPD condition granted that both set-ups offered valid data, since both outcomes were equally possible at some point in the story, but only one outcome was actually true. The ratio during the experiments was roughly three to one (positive to negative). While I tried to maintain an even ratio in the first sessions, the use of positive questions was often easier, since it was often necessary to move the props on display and make them easily visible to the child. The use of a puppet granted that the child’s answers were offered without a bias towards the experimenters’ authority (see e.g. Crain & Thornton 1999:ch.5 for discussion).

For each session, I transcribed the dialogue of experimenter, puppet and child onto a .doc file,
also segmented in “clips”. Each clip was segmented in chunks, each chunk representing a mini-
story testing one of the four Ps and related type-sentences (to, at, in, from), or one of the three
entailment types. Each chunk starts when the experimenter introduced a new mini-scenario, and
ends when a child gave an answer of any sort. In certain cases, it was not possible to transcribe
some words, so for those cases I reported this “noise” as “[xxxx]”. Extra-linguistic comments
were also marked with square brackets, e.g. “[right answer]”. Each chunk was scored according
to the child’s answer: “correct” if the child offered a correct answer, “wrong” if he made a
mistake, “undefined” if the child did not offer an answer. Answers were scored as “undefined”
in two cases: when the child did not offer an explicit answer (e.g. “yes” or “no”), although
he may had shown non-verbal understanding of the task; and when he produced an utterance
after the question, but the utterance was not pertinent to the question at hand. For instance,
if the child answered only “look mommy!” to a question such as (273), then the answer was
considered undefined. Other cases in which external factors prevented the child from answering
properly (e.g. the child tripped on display) were discarded from the final count.

Finally, once the test sentences were transcribed and organized into chunks, they were ana-
yzed in detail and organized in a data-base (i.e. excel sheets). These sheets reported data on
session, type of sentence, occurrence of token in session (1st, 2nd, etc.), entities corresponding
to the figure/ground on display (i.e. referents), the actual utterance by the child, the scoring as-
tioned to the sentence and general remarks. For entailment-based sentences, each sentence was
reported separately, then its connection (“temporal contiguity”) with the relevant session was
reported.

For consistency reasons, in the coding of the data I left aside a number of cases (20/257
tokens, 7,7% of the total) for which I could not come to a final evaluation for the sentences (i.e.
whether they were true, false or undefined). These sentences have been aggregated to another
group of properly scored sentences, which amounted to 15,9% of the data (41 tokens). This
aggregate cluster, consisting in 23,7% of the total sentences, has been subject to an inter-coder
integrity check. A second member of the Language acquisition laboratory, well acquainted
with the experimental hypotheses and design, has provided a second score for both set of data\(^6\). For sentences lacking a final judgement, 16 tokens/80% of the total sentences “on hold” has been adjudicated as “undefined”, 4 tokens/20% has been adjudicated as “false”, after reaching a shared consensus. For the sentences involving a second judgement, the shared consensus has covered 98.5% of the cases, with the remaining (1 token) adjudicated as “true” rather than “undefined”.

Before turning to the results, let us rehearse the predictions. I shall assume that, if Terence P. was able to access the interpretation of each of the four tested Ps, and each of the three entailment relations, he should have done so in an adult-like fashion. However, since as a child he was likely to be affected by stronger memory limitations, a non adult-like performance could have been a signal that he may still have lacked the necessary resources to interpret a given question and offer an answer. I shall assume that this lack of resources should take form in a higher but not critical percentage of undefined answers. I shall also assume that his might be the only relevant divergence from adults’ performance: hence, that Terence P. was expected to produce an adult-like amount of errors. The predicted results are the following. For stories in which the child had to answer “yes”, I shall assume a baseline of 90%/5%/5%, respectively for “yes”/“no”/“not sure” answers. For stories in which the child had to answer “no”, I shall assume a baseline of 5%/90%/5%, respectively for “yes”/“no”/“not sure” answers. I shall then discuss whether a result higher than 5%, for not sure answers, would eventually be statistically significant or not.

Let us turn to results, then.

**Results and Discussion**

The main findings of this experiment were that the main hypotheses were borne out, as in the adults’ group case. I shall start by presenting the basic data involving stories that tested the four main Ps (to, at, from and in), in which the child had to answer “yes”, because Godzilla offered the question after the story was complete, for instance when all tank engines went to a

---

\(^6\)Thanks to a colleague for the support and time spent in re-coding these data.
certain location. For each type of answer, I report number of tokens and results. For the to-type, given N=61, the answer percentages were 81.9%/7.4%/10.7%; For the at-type, given N=57, the answer percentages were 85.8%/7.1%/7.1%; For the from-type, given N=34, the answer percentages were 79.4%/5.8%/14.8%; For the in-type, given N=36, the answer percentages were 83.3%/5.5%/11.2%. As the data show, the results are intuitively in support of our experimental hypotheses, since the correct (“yes”) answers prevail, while the error rate is overall within predictions, according to the hypothesis that children produce errors in an adult-like fashion, i.e. because of performance factors. The “not sure” (undefined) answers appear to be higher than chance rate, but consistent with our predictions. A $\chi^2$ test on these apparently divergent data revealed that the results were statistically non-significant. For instance, the apparently problematic “not sure” answers for the from-type data yield a $\chi^2 = 3.34$, $p = .20$ value range, above statistical significance (i.e. above $p < .05$). Note that, although there is a higher number of to- and at-type tokens stemming from contingent factors, this difference in size sample is statistically irrelevant.

The next group of data involve stories in which the child had to answer “no”, because Godzilla offered the question before the story was complete, for instance when one tank engine could not go to a certain location, while others could reach this location. For each type of answer, I report number of tokens and results. Recall that, for this type of sentences, the “no” answer is predicted to be the correct one, hence the one observed at 90%, while the other two answers should be observed at chance, i.e. 5%. Informally, the “central” value in the percentages is the relevant one. For the to-type, given N=21, the answer percentages were 11.2%/77.6%/11.2%; For the at-type, given N=18, the answer percentages were 5.5%/83.5%/11%; For the from-type, given N=9, the answer percentages were 0%/100/0%; For the in-type, given N=12, the answer

---

7The $\chi^2$ distribution test is computed via the following formula: $(O_i - E_i)^2 / E_i$, with $O_i$ being the observed frequency, $E_i$ the estimated frequency. Values of $p$ below 5% (i.e. $p < .05$) are statistically significant, i.e. they tell us that the hypothesis is not borne out. Since we have three possible outcomes, the degree of freedom is $df = 2$. See e.g. Woods, Fletcher & Hughes (1986:ch.2).

8While the four types were presented in regular cycles, it was often the case that to and at were tested one more time per session, given the overall length of the tasks. The “asymmetry” in the data mostly stems from this factor.
percentages were 8.3%/83.4%/8.3%. As the data show, the results are intuitively in support of our experimental hypotheses, since the correct (“no”) answers prevail. The “yes” and “not sure” answer rates are overall low, suggesting that children performed better at this type of task, and overall according to hypotheses.

Let us now look at the entailment relations and their testing. Remember that, whenever possible, I offered questions involving “connected” Ps in a sequence, and checked whether the child correctly answered to both. As in the adults’ case, all entailment sentences were tested when “true”, i.e. the child had to answer “yes” to both questions. I tested *to* and subsequently *at*, as well as *in* followed by *from* and *in* followed by *at*, whenever it was possible to test both types of P within the same (sub-)scenario. For the *to*=*at*-type of entailment, given N=34, the answer percentages were 84.9%/3%/12.1%; for the *in*=*from*-type of entailment, given N=16, the answer percentages were 93.75%/0%/6.25%; for the *in*≤*at*-type of sub-set relation, the answer percentages were 78.4%/7.2%/14.4%. As we can observe, the error rate is well within predictions, whereas the relatively high percentage of undefined answers seem to suggest that performance factors and limitations on resources of various kinds played a small role in Terence P’s answers. For instance, in the case of the sub-set relation, a $\chi^2$ test on the “not sure” answer yields the results $\chi^2 = 2.4, p < .30$, thus confirming that the data are within the experimental hypotheses.

Some more specific observations are the following. For the *in*-, *to*- and *from*- types, the first correct answers appeared within the first three sessions, whereas the first errors appeared at a later time (session 10th, 10th and 8th session, respectively). One crucial aspect regards the *at*-type: for the first five sessions, Terence P. either did not answer or answered erroneously to *at*-type sentences, whereas by the sixth session errors became rarer, so they could be attributed to performance factors. This suggests that *at* became part of his grammar after 3;3 years of age, whereas the other Ps already emerged beforehand, and thus offer support to the prediction that *at* should emerge at a later time, than other Ps (sixth prediction). In general, Terence P. made mistakes at random points of the study, as he was able to correctly interpret *to*, *in*, *from* since
the earliest phases of the study. For *to*, a number of errors are actually caused by matters of real world knowledge: in the 14\textsuperscript{th} session, Terence P. rejected two otherwise correct sentences because he considered a bad tank engine, Diesel 10, as “extraneous” to the set of tank engines under discussion. Overall, errors appeared later than correct responses, and in a rather sporadic way, hence we can attribute them to simple performance problems, or non-linguistic factors. For *at*, the picture is slightly different. For the first five sessions (age range: 3;1-3;4 years), Terence P. either did not answer or answered erroneously to *at*-types. Half of the errors came from these sessions. Afterwards, his errors became rarer, and in general he showed a firmer interpretation and shorter answering times to this type of sentences. This suggests that Terence P. developed access to the interpretation of the more complex P in our “hierarchy” at a later time than the other Ps, as predicted by the theory.

For illustrative purposes, I shall present some of Terence P.’s data, starting from a typical case in which Terence P. had to offer a negative answer, according to the experimental hypothesis:

(275) Exp.:“Hi Terence P., Hello, have you taken all the tank engines to the farm?”
T.P.: Not yet, they are a bit late [true, still hasn’t complete task, taking time?]
Exp.:“Oh yeah ok, rar, have you taken all the tank engines to the farm now?”
T.P.: “yes” [true now, he has completed task]

In this example from the 14\textsuperscript{th} session, Terence P. answered “not yet” since the queried state of affairs did not correspond to the actual story. At the beginning of the story, he was asked to take the tank engines to the farm. Since he was taking his time to effectively take all the tank engines to the farm, he pointed it out when asked a first time. When he was done, he instead confirmed that the task was over, and thus that the tank engines have all gone to the farm.

The typical entailment stories also can be useful to illustrate the specific type of task and Terence P.’s answers. Look at this example regarding *to* and *at*:

(276) Exp.:“Oh guys, have all the tank engines gone to the farm T.P.?”
T.P.: yes [true]
Exp.: “but are all the tank engines at the farm now?”
T.P.: “yes” [Freddie is out, others in]

In this case, taken from Terence P.’s 14th session, Godzilla (i.e. the experimenter acting as the puppet) asked an “at”-question right after the child answered to a “to”-question. In this specific case, the child accepted a case in which one tank engine was outside, and all tank engines were in the shed.

The entailment relation between *at* and *in* is also tested via these sentences:

(277) Exp.: “are all the tank engines at the farm now?” T.P.: yes they are [waits a second before answering, “in” is true too]
Exp.: “Oh wait let me be sure, now all the tank engines are in the farm, is that right, let me count”

In (277) the tank engines were lying inside a toy farm with surrounding fences, as a result of a prior event of motion, and Terence P. accepted both *at* and *in* sentences as being appropriate descriptions (i.e. true) of the facts. While *in* is in a sense more appropriate, in this context, the child also accepted *at*, as predicted by the theory. Note also that here Terence P. agreed with Godzilla’s statement, rather than answering to a question.

The relation of entailment between *in* and *from* is illustrated by the following example:

(278) Exp. “but”
T.P.: Thomas is sleeping in the bed and Percy is sleeping... Arthur is sleeping in here
Exp. “but wait a second T.P. Not all tank engines are sleeping in the farm”
T.P.: but look here!
Exp.: “Oh? Where is Thomas, oh sorry, sorry, maybe I should sleep more”, “you know”
T.P. Godzilla this time went to sleep very late, lots of tank engines to stomp yesterday...
rar, wait a second T.P. Are you taking all the tank engines from the farm?”
T.P.: yeah, but thomas is sleeping, one moment [slow, finally takes thomas out] Exp: “arar but now have you taken all the tank engines from the farm?”

223
In (278), T.P. first confirmed that the tank engines (here, tank engines\textsuperscript{9}) were in the farm, then he started taking them and moving them (away) from the farm, confirming that this action was complete once he was done taking the tank engines from this location (and also confirming that he is doing that, beforehand). Before moving to the third experiment, I shall offer a more general discussion regarding the data, and how they support the predictions that stem from our theory of Ps.

First, Terence P. accepted sentences that correctly described the tank engines as being in/at a certain location, or having gone to/come from a certain place, and rejected sentences which did not corresponded to the depicted scenarios, and accepted entailment relations in an adult-like fashion. The first, second and fifth predictions are overall supported, since Terence P. accepted Ps as denoting static relations (first prediction, \textit{in} and \textit{at} data), relations denoting change (second prediction, \textit{to} and \textit{from} data). Terence P. accepted at as denoting a relation between the tank engines and a “complex” location they occupied (e.g. two in front, two behind, one inside); hence, the fifth prediction was also borne out.

Second, the late emergence of a correct interpretation of \textit{at}, compared with other target Ps, suggests that our sixth and seventh predictions are on the right track. Terence P. accessed Ps in a piece-meal fashion (\textit{in}, \textit{to} and \textit{from} before \textit{at}: sixth prediction), and accepted entailment and sub-set predictions as soon as he was able to access the interpretation of these Ps (seventh prediction). The difference in results between Terence P. and the adult participants appears to stem for the most part from performance factors, possibly including the different experimental set-up. Since no result diverges from the experimental hypothesis in a significant way, I conclude that Terence P.’s interpretation of Ps was adult-like, as per hypothesis.

\textsuperscript{9}Although I mostly used the term “tank engines” during the sessions, both children also accepted the occasional “tank engine(s)” to refer to Thomas and friends, with no visible problems.
6.3.3 Experiment N.3: Fred L.

This Experiment aimed to test the interpretation of experimental hypotheses in children. Differently from the first experiment, this experiment targeted the emergence of Ps in a younger child. For this purpose, one child was interviewed on a regular basis, and his data regarding Ps’ interpretation collected via a variant of the TVJ task.

Participants

Fred L. is the second of the two children who participated in the study at age range 2;2-3;0 years. Fred L. was interviewed on a fortnight basis, for sessions ranging from 28 to 56 minutes, for a total of 18 sessions.

Materials

Each session was recorded according to the procedure described in the previous section. The child’s mother was also present, but was unaware of the nature of the experiment. Each session involved the testing of several mini-scenarios in which the 7 hypotheses were tested on roughly a cyclic basis (e.g. first to, then at, then an entailment scenario, etc.). Differently from the previous experiment, I used a relatively wide array of props in his sessions. A consistent minority of cases revolved around rugby balls (6 in total), which strongly captivated Fred L.’s interest. Whenever possible, I also used different types of “locations” (boxes, houses, etc.) acting as containers, sleeping resorts for the balls, etc. for a total of 9 different locations. Regardless of the props, the tasks and questions were virtually identical to the ones used in the first experiment. The same combination of sentences used in Terence P.’s experiment was also used in the experiment, i.e. resultative sentences were used when more feasible.

A special mention pertains the use of puppets. Fred L. had often showed a strong “intolerance” to the presence of any marionettes on display, invariably getting distracted or actively trying to take the experimenter’s role in using the puppet. For this reason, I narrated the stories and directly offered the questions to the child. For this reason, answers inviting a “yes” answer,
i.e. answers offered after a given story was completed, only met the CPD condition, but note the “experimenter’s positive bias” condition. However, previous research on children in this age range suggest that positive bias towards adult experimenters are quite limited, if not absent (e.g. Unsworth 2005; Notley et al. 2008). Consequently, the data based on this type of answer can be also seen as offering evidence regarding Fred L.’s interpretation of Ps. Nevertheless, this aspect will be discussed in the results section, when relevant.

Procedure

In each session, a variant of the TVJ task was administered without resorting to Godzilla. I narrated the stories and, at some point either before or after the end of the story, I offered a question to the child. In each session, the first 2 or 3 minutes were warm-up practice, and the last five or so were based around play, to let the child relax after the session. Differently from Terence P.’s experiment, I was not able to consistently test logical relations, as two of the three relations were not accessible to the child. This aspect will be discussed in the “results” section.

For consistency reasons, in the coding of the data I left out all those cases in which I could not come with a final evaluation, as for Terence P.’s case, for a total of 12 tokens. This group of sentences has been aggregated to a second group, for which I proposed an integrity checking, for a total of 16 cases (16/125, 12.8% of the total). The first set has been scored in the following manner: we have judged 2 sentences as “true” (18.1%), 3 as “false” (27.2%) 5 as “undefined” (54.7%). We scored the second set unanimously (i.e. no sentence required a second evaluation).

As in the previous experiments, I shall assume that Fred L. was able to offer correct answers in an adult-like manner, if he was able to accept a given interpretation, and that wrong and undefined answers were at chance level. Consequently, I shall assume that any deviation from chance for undefined answers signaled Fred L.’s memory limits, as in the case of Terence P. and the second experiment. The predictions are as follows. The child was expected to answer “yes” to sentences offered after a story, so the pattern 90%/5%/5% is expected (i.e. “yes”/“no”/“not

---

10 Thanks again to my colleague.
sure”); he was also expected to answer “no” to sentences offered before a story, so the pattern 5%/90%/5% is expected.

Results and Discussion

The main findings of this experiment were that the some, but not all, of the hypotheses were borne out. In particular, it was found that Fred L. was still not able to access the interpretation of at, and with it one entailment relation, the one holding between to and at, and the sub-set relation holding between in and at. These findings are not surprising, if we consider that Terence P. was not able to access at before 3;3 years of age, and Fred L. stopped attending the sessions at a younger age than that. Let us discuss the first group of relevant data: the stories in which Fred L. had to offer “yes” answers. For the to-type, given N=29, the answer percentages were 82.7%/3.4%/13.9%; For the at-type, given N=13, the answer percentages were 30.7%/7.6%/61.7%; For the from-type, given N=18, the answer percentages were 81.8%/4.5%/13.7%; For the in-type, given N=22, the answer percentages were 81.8%/4.5%/13.7%. Before discussing further these results, I shall present the data regarding the stories in which Fred L. had to offer “no” answers. For the to-type, given N=14, the answer percentages were 7.4%/84.2%/7.4%; For the at-type, given N=0, the answer percentages were 0%/0%/0%; For the from-type, given N=10, the answer percentages were 4.5%/81.8%/13.7%; For the in-type, given N=9, the answer percentages were 0%/100%/0%.

The testing of the entailment and sub-set relations was severely limited, since Fred L. was not able to access at during the study. These data suggest that Fred L.’s interpretation of the four target Ps was overall at an earlier stage than the one observed in Terence P. The data regarding positive answers require some careful considerations. A general pattern was that Fred L. became easily distracted when he was invited to participate in at-type scenarios, and thus made the testing of this hypothesis somewhat hard. Although it was not possible to set up negative types of scenarios, the statistically significant “undefined” answers suggest the child was still not able to access the interpretation of this P.
This fact was probably influenced by the nature of the task, but given that the child was able to offer appropriate answers in other positive types of scenarios, we may assume that his inability to properly interpret at was evidence that this P was still inaccessible to him, as the still not perfect interpretation of to seems to suggest, as well. This is not only consistent with our hypotheses, but also with the findings from the second experiment: if Terence P. had problems in accessing at before 3.3 years of age, we would expect that Fred L. would have been unable to access this P, given that he participated while in an earlier age range. Overall, although these data suggest, although in an indirect manner, that Fred L.’s access to Ps was according to the predictions we outlined so far, and they invite the following conservative conclusions. As per the first and second experiment, the appendix contains a more thorough, but overall optional, statistical analysis.

Overall, Fred’s data support most of our predictions still in need of an empirical confirmation, although they do so from a different perspective than the one offered by adults’ and Terence P.’s data. First, Fred L.’s comprehension of in as a static P (first prediction) and to and from as Ps denoting “change” (second prediction) in stories inviting “no” answers, the ones offering the most accurate evidence, was according to our predictions, although the testing of the entailment and sub-set relations, and thus part of the fifth prediction, was beyond our reach. Second, Fred L. was able to access three of the four target Ps, but was still in the process of “updating” at in his model of Ps, and was probably quite far from this objective: the sixth prediction was supported, since Fred L. was slowly converging towards a “new” model of Ps, informally speaking. Since the entailment relations between Ps were still not accessible because of the child’s inability to access the interpretation of at, evidence supporting the seventh prediction was not found, unlike in the case of Terence P. At this point, it is possible to offer a general discussion regarding the three experiments.


6.3.4 General Discussion

In this chapter I presented three experiments that investigated whether the predictions made by the “extended” theory of Ps presented so far were borne out or not. Although the evidence offered by these experiments stemmed from different experimental conditions and experimental set-ups, all three experiments contributed in elucidating the empirical validity of the theory, and thus the psychological reality of our DRT fragment focused on Ps. The following points can be made about this evidence.

The first experiment offered evidence regarding the interpretation of some still poorly understood Ps in adults (i.e. *in, at, to* and *from*). It also offered a study on how the logical relations of entailment and sub-set guide adults’ interpretation of sentences involving Ps. The evidence found in this experiment strongly supported the first, second and fifth predictions of our theory, offering quite novel evidence regarding these still understudied Ps and these predictions still in need of testing. In detail, it offered evidence supporting the prediction that *in* and *at* are interpreted as static relations (first prediction), that *to* and *from* denote dynamic relations, relations involving change (second prediction); and that *at* can denote the (static) location of several “scattered” entities as being a “single” location (fifth prediction).

The second experiment offered evidence regarding the interpretation of these Ps in an Australian English-speaking child, Terence P., during his “intermediate” steps in the acquisition of Ps (and other parts of speech), i.e. age range 3;1-3;11 years. This evidence offered support for most of the seven predictions, and certainly for those which were still in need of (at least) partial investigation: first, second, fifth, sixth and seventh prediction. In particular, Terence P.’s adult-like interpretation of *in, at, to* and *from* in a progressive way supports not only the first, second and fifth discussion as per discussion in the previous paragraph; it also supports the assumption that these Ps emerge in a certain logical order, with *in* emerging before *to* and *from*, and *at* emerging after these three Ps (sixth prediction), and with entailment and sub-set relations holding between these Ps emerging in a piece-meal fashion as well (seventh prediction).

The third experiment offered evidence regarding the interpretation of these Ps in another
Australian English-speaking child, Fred L. Since the data were collected during Fred’s third year (i.e. age range 2;2-3;0 years), the data offered more preliminary evidence than the one found in Terence P.’s study. These data nevertheless offer support regarding the interpretation of in, to and from; they offer evidence that Fred L. was still unable to access the interpretation of at during the time of the study, and with it the relevant entailment and sub-set relations in which this P participates. Since our predictions state that children should acquire these relations piece-meal, these data offer indirect evidence of the dynamicity of this process. While Terence P. showed that he was able to reach this linguistic maturity during the early phases of the experiment, Fred L. was still not at the relevant phase that allowed him to properly interpret at. A revealing fact, in each of the three studies, was that participants did not make mistakes, i.e. deviations from the predicted interpretations, in any statistically significant quantities. When they were able to offer an answer, they offered it according to the predictions. When they were not, they possibly chose not to answer at all, as per predictions.

Each experiment offered evidence for some of the seven predictions that were still need of a more thorough empirical support, especially regarding the still poorly covered in, to, at and from, as well as the entailment and sub-set relations holding between these Ps. The first and second experiment supported all of our five predictions still in need of verification, while the third experiment fully supported almost four of them, since it offered partially incomplete support for the sixth, and no support for the seventh. The empirical contribution of each experiment certainly offered solid evidence that the tested predictions were overall borne out, hence shedding light on adults interpret Ps and the logical relations holding between Ps, and how these interpretations emerge over developmental time in children, in line with previous literature and our predictions. These experiments offer enough evidence to invite the conclusion that our DRT fragment can predict the interpretation and emergence of Ps in a rather accurate (although not perfect) way. Let us move to the general conclusions, then.
6.4 Conclusions

In this chapter I offered the data of three experiments regarding the interpretation and acquisition of Ps. I first outlined which Ps were still poorly studied in the experimental literature, and isolated the four mono-morphemic in, to, at and from as four Ps still in need of more thorough empirical coverage. I then suggested that the study of their interpretation, as well as the novel study of the logical relations holding between these Ps, would offer us quite novel and solid empirical evidence regarding the psychological reality of our DRT-bound theory of Ps.

The three experiments I presented support the predictions made by the theory to various positive degrees. The first experiment with adult participants very strongly supported our predictions about adults’ interpretation of in, to, at and from, as well as their ability to correctly access the logical relations holding between these Ps. The second experiment with Terence P. strongly supported our predictions about children’s interpretation being adult-like in nature, once the “correct” developmental phase is reached. The third experiment with the younger Fred L. supported our predictions to a lesser, but still quite good extent, since Fred L. was able to offer adult-like responses to those Ps which were already accessible to him, but he was not able to evaluate the logical relations holding between these Ps, to an extent. Overall, these different shades of evidence offer good support in favor of our theory of Ps, since five of the seven predictions of the theory, which were still in need of testing, were overall supported by seemingly different types of participants.

These experiments offer enough on-line evidence that supports the following answer to the global research question:

A-A: English speakers interpret Ps as relations over situations, which convey information about the “spatial” relations between a ground and one or more figures; they acquire the interpretation of Ps by acquiring the most specific Ps first, and then other Ps in order of increasing complexity, being guided by the logical relations holding between Ps in this process;
This answer tells us that English speakers interpret Ps according to the predictions made by our DRT fragment, and its treatment of Ps. The experimental data we have discussed overall offer ample support on the hypothesis that our fragment offers a psychologically plausible model of Ps, and thus can be seen as a model of the syntactic and semantic processes that underpin the processing of these parts of speech in speakers *in vivo*. The general conclusions regarding this empirical study, however, will be discussed in the next chapter, the conclusions to the thesis.
Chapter 7

Conclusions

This thesis has presented a novel study on English spatial adpositions, which we have called “P” as an easily identifiable syntactic label (e.g. in, at, to, from, in front of, ahead of, etc.). The core goal of this study has been that of offering three integrated solutions to three outstanding problems regarding spatial adpositions: first, what is the relation between adpositions and the non-linguistic, visual content they may capture; second, what are the syntactic and semantic principles that describe and explain their linguistic properties; third, what is the on-line, psychological reality of these principles, i.e. whether our theory correctly accounts the interpretation of adpositions in adults, and their “growth” in children.

At the beginning of this thesis, the three outlined problems represented aspects of a general theory of Ps that appeared to be without a clear and easily accessible solution. The examples (1)-(6) in the introduction, repeated here as (279)-(284), offered a basic but accurate picture of the nature of these problems. Look at the examples again:

(279) The boy is sitting in front of the desk
(280) The boy is sitting on top of the desk
(281) The boy is sleeping in the bed
(282) The boy went to the desk
The boy was sitting at the desk

The boy has arrived from the room

Before we addressed each problem in detail, we still could not offer a clear answer regarding the principles at work behind these three basic intuitions. First, we could not explain in an exhaustive and precise way that Ps can capture our non-linguistic (visual) understanding of spatio-temporal relations. Second, we could not explain that the subtle differences in interpretation between Ps such as in front of vs. on top of rested on the “changing” parts of these Ps, namely in front vs. on top, and what grammar principles can capture these differences in a principled way. Third, we could not explain that the interpretation of Ps in English speakers, whether they are adults or children, can be accurately accounted by the principles accounting for off-line data regarding Ps, and whether these principles can also account how these interpretations emerge in children’s Language.

In order to offer a solution for each of these problems, we started our “theoretical journey” by discussing in detail the first problem, the relation between Ps and the non-linguistic content. The solution to this first problem was conceived as an answer to the following research question:

Q-A: What is the relation between Vision and Language;

This question sums up a basic problem that we have found in the literature: the lack of a precise theory of how non-linguistic spatial information, in particular visual information, is related with spatial Language, or more accurately with adpositions and their content.

The first part and chapter offered an answer to this question, offered at the end of chapter 2/part I, which is as follows:

A-A: The relation between spatial Vision and spatial Language is an isomorphism, as both models represent the same “amount” of information via different types of information;

This answer aimed to capture one basic intuition: that “what we see” can be captured by “what we say” at the same level of precision, insofar as we decide to do so, and that both modes of
information-processing differ in the type of information they process, but not in the underlying principles. This answer allowed us to give a fairly precise and unified account of the relation between a sentence and a possible scenario that this sentence can represent. Suppose that we observe a scenario in which a certain boy (e.g. Mario) is sitting on a blue chair, and this chair happens to be aligned with the frontal part of a given desk, at 2 meters of distance. In this scenario, a sentence such as (279) expresses only a part of the visual information that is accessible to us, but nothing prevents that more precise details regarding this spatio-temporal relation, a relation between situations, could be added by further sentences. Importantly, though, this sentence conveys information regarding the boy’s position by capturing at a linguistic level of comprehension a relation between two entities, which has the same formal properties of its visual counterpart.

The solution of this problem allowed us to have a clearer picture of the relation between Vision and Language as distinct but interacting parts of Cognition. At the same time, though, it left open the solution to the second problem: we could still not give a precise and principled compositional account of Ps, both from a syntactic and semantic point of view. The solution to this problem was offered in part II of the thesis, which focused on both the syntactic and semantic properties of Ps (respectively, chapter 3 and 4). This solution was conceived as an answer to the following research question:

Q-A: *How adpositions can combine together and with other parts of Speech, and express distinct but related types of spatio-temporal relations;*

This answer aimed to capture one basic intuition: that Ps express spatio-temporal relations in a compositional way, so that each distinct morpho-syntactic unit making up a P (e.g. *in, front* and *of*) could be specified with respect to the interpretation of the “whole” P, *in front of*, and the contribution to sentence interpretation of this P. It also aimed to capture the intuition that the interpretive differences between Ps stemmed from general principles regarding the structure of our model of discourse. If Ps denote relations between situations and these relations are structured in different sorts, then Ps can (and should) receive different interpretations, also mirroring
these relations.

It also allowed to offer the following answer to the second research question, which is as follows:

\textbf{A-A:} \textit{Ps combine with other Ps and with other parts of speech in a compositional way, and express the possible types of relations defined over the domain of (spatio-temporal) situations in a principled way;}

This answer attempted to capture the logical relations holding between e.g. \textit{to} and \textit{from}, in (282) and (284), or \textit{to} and \textit{at in} (282) and (283) as a consequence of the distinction between states and events in the domain of situations. From a technical point of view, this result was achieved by offering an extension of our framework of choice, DRT, which aimed to include results and explanations from other frameworks (syntactic Minimalism, Lattice theory of situations). Since the DRT fragment outlined in the first part was not powerful enough to correctly describe and explain all the linguistic properties of adpositions, the second part improved this fragment.

The solution to this problem paved the way for the solution of the third problem, analyzed in Part III. Since the second solution ultimately offered an off-line theory of Ps and their interpretation, it also offered a natural starting point to address the third and final problem analyzed in the thesis, the on-line status of this theory: whether the theory could correctly predict the interpretation of Ps in children and adult speakers, and their emergence in children. This third solution was conceived as an answer to the third global research question:

\textbf{Q-A:} \textit{What is the interpretation of Ps and their logical relations by English speakers, and how this interpretation emerges over time;}

This question aimed to make precise the relation between off-line analysis and on-line data. The answer to this question was articulated in two chapters, which respectively introduced some preliminary background regarding Language processing and Acquisition, and a detailed presentation and discussion of the three experiments offering empirical evidence (chapters 5 and 6).
The answer I offered to this question was the following:

A-A: English speakers interpret Ps as relations over situations, which convey information about the “spatial” relations between a ground and one or more figures; they acquire the interpretation of Ps by acquiring the most specific Ps first, and then other Ps in order of increasing complexity, being guided by the logical relations holding between Ps in this process;

This answer aimed to capture one basic intuition: that the theory of adpositions I offered in the first and second part could (and should) be seen as a psychological theory of adpositions, since the answer is based on the empirical support of the theory offered by the three experiments offered in the fifth chapter. This answer captures the intuition that the interpretation assigned to sentences (279)-(284) by English speakers are those predicted by the theory; and that the logical relations holding between these sentences are also confirmed by speakers’ intuitions, as the data showed. Speakers accepted the entailment relation holding between (282) and (283), and thus the logical relation between to and at; they also accepted the sub-set relation holding between (281) and (283), and thus the logical relation between in and at. Hence, our theory made correct predictions about the psychological reality of these relations.

The solutions offered to these three problems, if correct, appear to offer a substantial improvement on our understanding on what are the properties of Ps. As we have seen, they can offer a principled account on three apparently disparate phenomena regarding this part of speech under a unified theoretical and empirical perspective. The theoretical perspective is the one offered by the common formal Language of our extended DRT fragment: by the end of this thesis, we can offer an explicit and precise account of visual, linguistic and developmental processes, as well as the fine-grained relations holding between these processes. The empirical perspective consists of our new-found ability to account apparently very different sets of data as “data about Ps”: we have seen that the interpretation of Ps in English speakers, both from an off-line an on-line perspective, is a unitary phenomenon involving how we convey and comprehend the spatio-temporal relations between objects, that we can observe in various scenarios.
This thesis does not exhaust the discussion of empirical problems regarding this part of speech. Several other topics could be discussed, in relation to conceptual holes left in the thesis.

A first problem is the physiological or biological plausibility of our theory. This problem represents a specific instance of the binding problem, since it is based on the empirical question of how different types of information are connected together (e.g. Marcus 2001). As we have seen in chapter 2, certain theories such as HCM and Vector Grammar make specific claims regarding the neurological processes that occur when we process spatial information. Although these claims appear not to be specific to linguistic processes, it is still an open question on what is the exact import of our theory for a study of the biological bases of Language. If the Vector Grammar hypothesis is correct, then Ps can be said to have a biological basis, since they express at a linguistic level a certain pattern (a vector) of Boundary Vector Cells (BVCs) in the Hippocampus, that fire when we observe a visual scenario (O’Keefe 2003; Burgess 2006). Language production and processing, however, occurs in the temporal lobe. It is usually assumed that Broca’s area is involved with speech production and the ability to produce syntactic well-formed sentences (e.g. Caplan 2006; and references therein). It also been suggested that different portions of this lobe are involved in the production and retrieval of syntactic categories, such as nouns and verbs (respectively, superior and middle anterior temporal gyrus: e.g. Caramazza & Finocchiaro 2002; and references therein), or for semantic information (e.g. Bi et al. 2010; and references therein). Under our current assumptions, it is an open problem how these two different parts of the brain can “combine” different types of information, visual vs. linguistic, into a possibly unified representation. Although we may have a clear and precise formalism to represent different processes and a way to combine these processes together, we do not know whether this formalism represent processes that are biologically real, i.e. whether they occur in the brain or not.

A second problem that arises from this initial problem is the following. If hypotheses such as Caramazza & Finocchiaro (2002) are correct, then an open question is “where” Ps may be located, on a map of the temporal lobe. Recent work has suggested that Ps may be associated with
specific regions of the temporal lobe, since the processing and production of locative Ps such as *over* and *under* appears to involve at least the superior and middle posterior gyrus, in the temporal lobe (e.g. Amorapanth, Widick & Chatterjee 2009; and references therein). It is an open question on whether other Ps should be associated to these “regions” of the temporal lobe, as we would expect, or if other parts are involved as well. At the same time, it is also an open question on how these different regions of the brains are “connected” together, when sentence processing and production occurs. It is also an open question on how the compositional processes we have discussed so far involving Ps are actually instantiated in the brain, and thus how different portions of the temporal lobe including Broca’s area, interact during the production/comprehension of sentences (279)-(284).

A third problem that is tightly connected with these two problems pertains Language disorders. It is also an open question on whether a theory of how these processes occur can double as a theory of language disorders, i.e. whether our theory is consistent with the several works on aphasia and the omission of Ps we have discussed in chapter 1 (e.g. Trofimova 2009 and references therein). The works cited in the previous two paragraphs include evidence including various forms of brain lesions, and thus appear to support this basic intuition, that the lesion of dedicated “regions” for Ps would critically affect their production and comprehension. However, it is still an open question on what is the exact relation between these two levels of information-processing.

These three problems represent only the tip of a much bigger conceptual Iceberg, which is based on one single aspect of Ps we have completely left aside. Several other problems, closer to the topics we have discussed in this thesis, are awaiting a solution. For instance, we are still in need of a more thorough empirical coverage of the Vision-Language interface offered in Part I: we still do not have evidence that such theory can successfully account how we may match the interpretation of a sentence involving *to*, such as (282), with a corresponding visual scenario, and what kind of visual relation might correspond to the entailment relation holding between (282) and (283), *inter alia*. We are also not aware on whether our theory of Ps can be safely extended.
to other Languages. As Levinson & Meira (2003) observe, Ps are a highly variable category across Languages, with the richness of P members found in English being an exception: several languages appear to have only a handful of Ps which express rather “general” concepts, as the concept of general location associated to at. We still need to test whether the theory offered in Part II can account these data in a straightforward manner, or further changes and corrections to the theory would be necessary, once we broaden our view beyond English. This off-line problem doubles as an on-line problem, since we have discussed in Part III the psychological reality of our theory. A particularly stringent question is how we can use our theory to account acquisition data in languages that have few Ps, and what changes to the theory may be necessary to account these data. In a language such as Italian, for instance, the two relations denoted by to and at in English correspond to one P, a: the PP alla scrivania can be both interpreted expressing the English equivalents to the desk and at the desk. It is still necessary to test whether our theory can straightforwardly account these more challenging data, or if we need to make our assumptions more precise, in order to account them.

These are only some of the many possible problems that are still in need of a solution, once we focus on Ps. Several other could be listed. However, for the time being, these problems can be safely left aside, since the goal of this thesis has been met, and further problems may become the goal for works to come.
Bibliography


258


260


Partee, Barbara H. Do We Need Two Basic Types?. In Sigfrid Beck and H. Gärtner (eds.), *Snippets: Special Issue in Honor of Manfred Krifka*, Berlin: Kluwer.


266


Richards, Lyn, & Kenny Coventry. (2005). Childrens production of locative prepositions in English: the influence of geometric and extra-geometric factors. In Laura Carlson & Emile van
der Zee (Eds.), *Functional features in language and space: Insights from perception, categorization and development*, 163-173. Oxford University Press.


270


272


Winter, Yoad (2008). *Between logic and common sense, or Dikkertje Dap meets Dr Halfbaked*. Talk presented at Nijmegen Institute for Cognition and Information.


