Introduction

As we interact with the world and with each other we need to classify objects and events, that is, we need to make judgements about what types of objects and events we are confronted with. This is an important part of what is involved in planning the future actions we should carry out and how we should coordinate with other agents in carrying out collaborative actions. This is true of action in general, including linguistic action. The aim of this book is to characterize a notion of type which will cover both linguistic and non-linguistic action and to lay the foundations for a theory of action based on these types. We will argue that a theory of language based on action allows us to take a perspective on linguistic content which is centered on interaction in dialogue and that this is importantly different to the traditional view of natural languages as being essentially similar to formal languages such as logics developed by philosophers or mathematicians. At the same time we will argue that the tremendous technical advances made by the formal language view of semantics can be incorporated into the action-based view and that this can lead to important improvements both of intuitive understanding and empirical coverage.

Part I of the book (Chapters 1–3) deals with a theory of types related to perception and action and shows a way of presenting a theory of grammar within a theory of action. Part II (Chapters 4–7) then looks at a number of central issues in semantics from a dialogical perspective and argues that there are advantages to looking at some old puzzles from this perspective.

In Chapter 1 we introduce a notion of perception of an object or event as making a judgement that the object is of a type. In symbols, we write $a : T$ to indicate that object a is of type T. We shall talk interchangeably of an object being of a type or being a witness for a type. Our claim is that we can only perceive something as being of a type, even if that type is very general (like *PhysicalObject* or *Event*) – we cannot perceive it *simpliciter*. We present basic notions of the theory of types which will be developed in the book, TTR, a type theory with records, which builds to a great extent on ideas taken from the type theory of Per Martin-Löf although we have made significant changes both in the general design and aims of the theory and a number of details which appear to us to be motivated by cognitive and linguistic considerations. We also introduce some basic notions of a theory of action based on these types which will be developed further as the book progresses. The overall approach presented here owes much to the theory of situations and situation semantics presented by Barwise and Perry in the 1980's. One of the themes of this book is a working out of parts of the old situation theory using ideas taken from

Martin-Löf's type theory.

A central notion in TTR is that of *record*. The term "record" is used in computer science for what is often called an attribute-value matrix (AVM) or feature structure in linguistics. A record is a collection of fields consisting of a label (attribute or feature in the standard linguistic way of talking) and an object of some kind (which itself can be a record). An schematic example of a record is given in (1), where the ℓ_i are labels and the o_i are objects.

(1) $$\begin{bmatrix} \ell_0 & = & \begin{bmatrix} \ell_1 & = & o_0 \\ \ell_2 & = & o_1 \end{bmatrix} \\ \ell_3 & = & o_2 \end{bmatrix}$$

Records are witnesses for record types which are also collections of fields. Rather than objects, the fields in a record type contain types. In the schematic example in (2) the T_i are types.

(2) $$\begin{bmatrix} \ell_0 & : & \begin{bmatrix} \ell_1 & : & T_0 \\ \ell_2 & : & T_1 \end{bmatrix} \\ \ell_3 & : & T_2 \end{bmatrix}$$

The record (1) will be of the type (2) just in case the objects are of the types with the same labelling, that is, $o_0 : T_0$, $o_1 : T_1$ and $o_2 : T_2$. Martin-Löf's orginal type theory did not have records or record types though there have been many suggestions in the literature on how to add them. We have borrowed freely from some of these ideas in TTR although the way we have developed the notions differs essentially from previous proposals. We will use records and record types to model situations and situation types.

In Chapter 2 we apply the theory of types from Chapter 1 to basic notions of information update in dialogue. Here we build on seminal work on dialogue analysis by Jonathan Ginzburg and also related computational implementation by Staffan Larsson leading to the information state update approach to dialogue systems. We have adapted these ideas in a way that allows us to pursue the questions of grammar and semantics that we take up in the remainder of the book. A central notion here is that of the dialogue gameboard which we construe as a type of information state representing the current state of play in the dialogue from the perspective of a dialogue participant. It includes the dialogue participant's view of what has been committed to as being true in the dialogue so far and what questions are currently under discussion.

In Chapter 3 we show how syntax and semantics can be embedded in the theory of action characterized in Chapters 1 and 2. This is in contrast to a formal language view where language is seen as a set of analyzed strings of symbols associated with meanings of some kind. The philosophical ground of the action-based approach goes back to the relational theory of meaning

introduced in Barwise and Perry's situation semantics which focusses on the relation between utterance situations and described situations. This was perhaps the first attempt to generalize the Speech Act Theory developed by Austin and Searle to the concerns of compositional interpretation of syntactic structure. A more recent theory to which the ideas in this chapter are related is that of Dynamic Syntax (DS). While the particular formulations in our approach look rather different from those in DS the two theories have common aims relating to the analysis of language as action and an emphasis on the incremental nature of language which in this chapter we relate to the building of a chart type. There is also a common interest in the treatment of language as a system in flux where an act of speaking can create a new previously unavailable linguistic resource that can be reused in future speech events.

The theory of types that we employ gives us two important notions which will be important in the development of semantics in Part II. The first is the notion of *intensionality*. Types in TTR are intensional in that the identity of a type is not established in terms of the set of objects which are of that type. That is, types are not *extensional* in the way that sets are in a standard set theory. The axiom of extensionality in standard set theory requires that there cannot be two sets which have the same members. In contrast, there can be different types which have exactly the same set of witnesses. The second notion has to do with the facts that the types themselves are treated as objects that can enter into relations and be used to construct new types. We will call this *first class citizenship of types*, though it is related to notions of *intentionality* (with a "t") and *reflection* in programming languages, that is, the ability not only to carry out procedures but to reflect on and reason about them. In our terms, an important enabling factor for human language is that we not only can perceive objects and events in terms of types and act on these perceptions but that we can also reason about and act on the types themselves, for example, in ascribing them to other agents as beliefs or making a plan to achieve a goal by creating an event of a certain type. The types become cognitive *resources* which we can exploit in our communicative activity. In Part II we will look at a number of examples of this.

In Chapter 4 we examine reference by uses of proper names and occurrences of pronouns which are not bound by quantifiers. In order to account for this we need a notion of *parametric content*, which is to say that the content of an utterance depends on a context belonging to a certain type. For example, an utterance of the proper name *Sam* requires a context in which there is an individual named "Sam". But where in her resources should a dialogue participant look for such a context? One obvious place is the conversational gameboard that we introduced in Chapter 2. That is, the dialogue participant should determine whether there has been reference to somebody of that name already in the current dialogue according to her gameboard. Another place is the visual scene, or more generally the ambient situation which the agent can perceive by different sense modalities. This we also represent as a resource using a type – that is, the type for which the ambient situation would be a witness if the agent's perception is correct. Yet another place to look is the agent's long term memory (which we will equate with the agent's beliefs, although one may ultimately wish to make a distinction). This resource is also modelled as a type representing how the world would be if the agent's memory or beliefs are correct. The fact that we are reasoning about to what extent the context type associated with the utterance matches

the types modelling the agent's relevant resources enables us to talk about cases where there are names of non-existent objects (that is, the agent's resource types do not exactly match the world) or where a single object in the world corresponds to two objects in the resources or *vice versa* (another way in which there can be a mismatch between reality and an agent's resources).

In Chapter 5 we look at frames associated with common nouns. The idea of frames goes back to early work on frame semantics by Fillmore and also psychological work on frames by Barsalou. We will construe frames as situations (modelled as records in TTR). We will argue that frame types are an additional kind of resource which is exploited in natural language semantics. A common noun like *dog*, in addition to being associated with the property of being a dog, can also be associated with a type of situation (a frame type) which is common for dogs, for example, where the dog has a name, an age and various other attributes we commonly attribute to dogs. We will argue that such a frame can play an important role in interpreting utterances such as *the dog is nine* in the sense of "the dog is nine years old". Some nouns, such as *temperature*, seem to represent frame level predicates, following an analysis suggested by Sebastian Löbner in order to account for the analysis of utterances like *the temperature is rising* where it is not the case that some particular temperature is rising (say, 30°) but that different situations (frames) with different temperatures are being compared. Nouns which normally predicate of individuals can be coerced to predicate of frames. An example is the noun *ship* in an example originally discussed by Manfred Krifka: *four thousand ships passed through the lock* which can either mean that four thousand distinct ships passed through the lock or that there were four thousand ship-passing-through-the-lock events some of which may have involved the same ship. We argue that in order to interpret such examples you need to have as a resource an appropriate frame type associated with the noun *ship*.

In Chapter 6 we explore phenomena in natural language which are standardly referred to as *modality* and *intensionality*. We argue that types as we conceive them are better placed to deal with these phenomena than possible worlds that are used in standard formal semantics. In standard formal semantics propositions are regarded as sets of possible worlds. For example, the proposition corresponding to *a boy hugged a dog* is the set of all logically possible worlds in which a boy hugged a dog is true. What we substitute for this is the type of situations in which a boy hugged a dog. At an intuitive level these notions are quite similar. They both represent mathematical objects which allow for many different possibilities as long as the fact that a boy hugged a dog is held constant across them. One important difference is that sets of possible worlds are extensional sets whereas as our types are intensional. Thus it is possible for us to have two distinct types which have exactly the same witnesses. One pair of such examples we discuss is *Kim sold Syntactic Structures to Sam* and *Sam bought Syntactic Structures from Kim*. Intuitively we want these to represent different propositions and we argue that they can yield different truth conditions when embedded under a predicate like *legal*. (Under Swedish law, for example, it is illegal to buy sex but legal to sell sex.) Another pair involves so-called mathematical propositions which are true in all possible worlds but which nevertheless we would want to represent different propositions: *Two plus two equals four* and *Fermat's last theorem is true* (as proved by Andrew Wiles).

The chapter begins with a discussion of the problems associated with possible worlds analyses. We then continue with a discussion of modality and in particular of how Angelika Kratzer's notions of conversational background and ideals can be seen with advantage as resources based on types and the kind of topoi that Ellen Breitholtz has introduced in the TTR literature. In the third part of the chapter we discuss what are traditionally regarded as intensional constructions involving attitude verbs like *believe* and intensional verbs like *need* and *want*. We treat 'believe' as a relation between individuals and types (corresponding to the content of the embedded sentence). For an individual to believe a type it has to be the case that the type matches (in a way we make precise) the type which models the beliefs (or long term memory) of the individual, that is the same resource that was needed in Chapter 4 to get the dialogical analysis of proper names to work out.

In Chapter 7 we look at generalized quantifiers from the perspective of dialogic interaction. Traditionally generalized quantifiers are treated as sets of sets or sets of properties and the work of Barwise and Cooper on generalized quantifiers built on this idea. Barwise and Cooper also introduced the auxiliary notion of witness set for quantifiers under the heading "Processing quantified statements". In this chapter we turn things around and make the characterization of witness sets the primary notion in defining quantifiers. This makes it more straightforward to account for the anaphoric possibilities relating to quantified expressions in dialogue. We often use quantified statements in dialogue when we have inadequate information to determine their truth. This is particularly true of determiners like *every* and *most* when talking about large sets. We suggest that this phenomenon can be analyzed by estimating a probability based on the evidence presented in our cognitive resources (long-term memory or beliefs as discussed in Chapters 4 and 6). Finally, we give an account of how TTR types can be used to talk of content which is underspecified for quantifier scope. The idea is to compute content types which have the various available contents which can be associated with an utterance as witnesses.

There are a number of general themes which are woven together in this book and which have different emphasis at different points in the text:

- language as action
- linguistic content grounded in perception as type judgement
- language as interaction and coordination
- language as a system in flux
- types, not possible worlds
- types and reference to non-existent objects
- types and cognitive resources
- getting the balance right between language, the external world and mental states

- types as a way of doing underspecification
- avoiding an intermediate "semantic" language such as logical form or discourse representation language but rather giving a direct interpretation of linguistic events in terms of semantic a semantic universe containing structured objects

Behind all this is a desire to find a theory of types which can be used to talk about cognition in general as well as allow us to give a general account of language which includes many of the insights we have gained from separate linguistic theories, a foundation for a formal approach to cognition, if you like.

Why try to do all of this at once? Would it not have been better to write individual books and papers on each of these topics in turn? These are questions that I have asked myself at various points while writing this book. It worries me (and it will probably worry you) despite the fact that I know the answer: it is important to have a single approach to language in which all these issues can be addressed simultaneously. Taking the issues one at a time is not as convincing or ultimately as interesting as showing how these different aspects of language interact in a complex system, giving us a view of linguistic interpretation which both embraces an action oriented approach and preserves the insights we have gained from formal semantics as well as addressing some of the puzzles that it failed to solve adequately.

Others things that have been done and future possibilities