On Semantic Underspecification

Manfred Pinkal

November 11, 1998

Abstract

Underspecification has become fashionable in computational semantics. In this chapter, I will try to give an idea of what semantic underspecification is, what it has been good for, and what the perspective for future application are. I will start with the inspection of the specific phenomena and techniques which are usually associated with the notion of underspecification (in Sections 1 and 2, respectively). In Sect. 3, I will try to indicate the main motivations for using underspecification techniques. Then, I will point out one important use of the concept in some more detail, i.e., direct reasoning with incomplete semantic information, discussing first the appropriate truth-conditional basis (Sect. 4), and second perspectives on efficient reasoning systems (Sect. 5). I will conclude with some remarks about the general status of the semantic underspecification concept, in Sect. 6.

1 Underspecification Phenomena

Scope, especially quantifier scope, is the first and best investigated phenomenon in the field of underspecification research. The use of logic representation languages in natural language semantics often requires a choice between alternative scopings in the interpretation process, where this choice seems to be premature or unnecessarily fine-grained. An example is (1), which can be assigned either (2) or (3) as logical representation, although in many cases the relative scope of universal and existential operator may be unclear, or irrelevant, or both.

(1) Two foreign languages are spoken by every linguist.

(2) $\exists^2 x(foreign\_language(x) \wedge \forall y(linguist(y) \rightarrow spoken\_by(x, y)))$

(3) $\forall y(linguist(y) \rightarrow \exists^2 x(foreign\_language(x) \wedge spoken\_by(x, y)))$

The basic idea behind the technique of underspecification is to represent the semantic information connected with a semantically underdetermined utterance A by a partial description of logical structure, rather than enumerating the possible completely specified interpretation alternatives for A. The pioneer for structural underspecification research was SRI Cambridge. The QLF formalism described in [AC92] provides an underspecified treatment of scope as well as certain aspects of referential ambiguity, ellipsis, and focus-background structure.

Another important source for the interest in underspecification is lexical semantics. Example (4) is a representative for a large field of ambiguity phenomena, which are conventionally classified as lexical ambiguities, but differ from trivial cases like the homonyms bank or pen in several important ways.

(4) John began the book

Rather than locating the source of ambiguity of Sentence (4) in the verb *"begin"*, which would require an infinite number of readings for the different kinds of complements the verb can take (*"begin to read"* , *"begin to write"* , *"begin to paint"* , *"begin to* *eat"*, ...), James Pustejovsky assumes that the different readings of *"begin"*+non-event-complement constructions are generated in a systematic way, in his theory of "Generative Lexicon" [Pus91, Pus95] . The process is triggered by the type conflict between the verb and its argument, which requires that the two objects be semantically related in an — initially underspecified — way. Specific information about this relation can be retrieved from special lexical information about the argument, e.g., that the typical way of using a book is reading its content (stored in the *qualia structure* of the lexical entry).

A third range of phenomena, which might turn out to be the most important application field for the technique of semantic underspecification, but just has started to attract the attention of semanticists, is spoken-language understanding. Spoken utterances are often incomplete, and they may be incompletely or wrongly recognized by the human hearer or by a speech recognition system. In (5), a standard example taken from the Verbmobil project is given (for Verbmobil see [Wah93]).

(5) (a) $[_S$ *Treffen wir uns*$]$ $[_{NPdat}$ *den nächsten zwei Wochen*$]$

(b) Shall we meet — the next two weeks

Since the speech recognition module was not able to identify the preposition in as part of the input utterance, the parser is not able to identify the last part of the input as an adjunct of the sentence. It passes a sequence of two unrelated analyses to the semantics module, which in turn is unable to combine the corresponding semantic representations directly. It can guess, however, that the two pieces stand in some kind of (underspecified) semantic relation, that the second is a PP modifier of the first (because emission of a preposition is probable), and that it is the special case of a temporal modifier (according to the semantic sort of the head noun). For a first approach to semantic processing with incomplete input see [Wor97].

2 Underspecification Techniques

A number of related formalisms for scope underspecification have been proposed during the last years (e.g., [Rey93, Bos96, Mus97, Pin96, CFS, EL96]). These proposals differ in the choice of the underlying semantic formalism, they differ to some extent in their expressiveness, and they differ in the theoretical status they assign to the respective formalisms. I will not discuss the approaches individually, but rather try to indicate the basic common ideas and intentions, by giving a simple semi-formal outline of a simple underspecification formalism.

I will use the familiar language of extensional type theory as underlying representation language, for the sake of simplicity (in fact, I will use only the FOL fragment, in my examples). In order to allow partial descriptions of logical formulas, the representation language is augmented by meta- variables X, Y, Z, X_1, X_2, ... (called *labels*, *holes*, or *handles* in the respective formalisms). Meta-variables range over expressions of the representation language, or, to be more precise, occurrences of such expressions. It is convenient to conceive of logical representations as trees, whose leaves are annotated with non-logical constants and variables, and whose non-terminal nodes are annotated with semantic operators (e.g., implication, abstraction with variable x, functional application) [1]. This allows to investigate underspecification formalisms in the well-understood framework of tree logic (see Vijai-Shanker et al.). In particular, meta-variables can be conceived of as referring to nodes, and constraints between meta-variables as partial tree descriptions. Underspecification formalisms usually provide a small inventory of basic constraint types, in addition to equality.

[1] To that purpose, we need an explicit operator symbol for function application (e.g., @). I employ the use the usual notation, here (i.e., $a(b)$ instead of $a@b$). For better readibility, I will in general use a notation which is as close as possible to logical standard notation (and also close to the notation of most underspecification frameworks).

These are dominance or subordination constraints $X \lhd^\star Y$, where $\lhd^\star$ is a two-place partial-ordering relation between occurrences of expressions (nodes), and constraints $X : A$ stating that part X of the representation under discussion is of form A, where A is a (partially specified) representation language expression. In (6), I give a scope- underspecified representation of (1)in terms of tree constraints, as an example. X_0 is the designated meta-variable referring to the top node, i.e., the described structure as a whole.

(6) $$\begin{aligned}&\{X_1 : \exists^2 x(foreign\_language(x) \wedge X_3),\\ &X_2 : \forall y(linguist(y) \rightarrow X_4), X_5 : spoken\_by(x, y),\\ &X_0 \lhd^\star X_1, X_0 \lhd^\star X_2, X_3 \lhd^\star X_5, X_4 \lhd^\star X_5\}\end{aligned}$$

The constraint set (6) is compatible with (the tree representations of) both (2) and (3). It specifies which quantifiers occur and which argument positions they bind, but underspecifies scope order. Actually, (6) is still compatible with an infinite number of logical terms, since it does not constrain the occurrence of additional material in the formula. Nevertheless, it provides a specification of the set consisting just of (2) and (3), since these formulae are the only minimal structures satisfying all constraints.

A mechanism for semantic construction which takes a syntactic analysis of Sentence (1) and generates the constraint set (6) from it, can be set up in a simple and straightforward way. [2] If later on context or subsequent discourse provide the additional information that *"two languages"* outscope *"every linguist"*, the corresponding constraint $X_3 \lhd^\star X_2$ can be added to (6). The resulting constraint set does not contain (2) as a minimal solution, which leaves (3) as the unique reading for (1). The example demonstrates that disambiguation can be postponed to the point where a maximum of relevant information is available.

Underspecification formalisms also provide a framework for the modeling of lexical reinterpretation cases: In example (4), the type

[2] Two related versions of the construction algorithm are given in [Mus97] and [Pin96], respectively.

conflict between *"begin"* and *"book"* blocks immediate application of the first to the second. We can read off the sentence, however, that the denotation of *"the book"* must occur somewhere inside the second argument of *"begin"*. This piece of safe knowledge can, in a first step, be expressed by the constraint set (7).

(7) $\{X_0 : begin(john, X_1), X_2 : the\_book, X_1 \lhd^{\star} X_2\}$

The second step consists in filling in the space between X_1 and X_2. This can be done by adding, e.g., $X_1 : reading\_of(X_2)$, or $X_1 : writing\_of(X_2)$, using additional semantic information contained in the qualia structure of the lexical entry of the argument, or world knowledge, and determining the contextually most appropriate alternative.

Let us look at the analysis of incomplete input next. In example (5), the input consists of a saturated sentence and a noun phrase, which cannot be directly related to each other. Given evidence (e.g., prosodic information, external conditions on the dialogue) that the semantic representations for both fragments are part of the same sentence representation, the constraint set (8) can be established.

(8) $\{X_1 : meet(we), X_2 : the\_next\_two\_weeks,$
$X_0 \lhd^{\star} X_1, X_0 \lhd^{\star} X_2\}$

In a second step, heuristic rules are applied, leading to the combination of X_2 with an underspecified temporal relation (a meta-variable which might be instantiated to *"during"*, *"after"*, *"immediately after"*, etc., given further context information), and applying the result to X_1. This can be formally expressed by adding the constraints $X_3 = T\_rel(X_2)$ and $X_0 = X_3(X_1)$ to (8), where T_rel is a meta-variable ranging over temporal relations[3].

[3]More correctly, possible instantiations of T_rel should have the form $\lambda t\ \lambda e.R(e,t)$, where a Davidsonian analysis of verbs and modifiers is assumed, and R ranges over relations of temporal location. Here, as throughout the

The examples should have illustrated one basic advantage of the underspecification technique: Whenever it is impossible to identify one unambiguous meaning representation at some stage in the interpretation process — due to insufficient disambiguation information, type inconsistencies between semantic objects, or incomplete input — we can encode the available safe information in a constraint set, thus specifying the range of possible readings. Later, we can try to obtain more information to narrow down the meaning potential, which is represented by a monotonic extension of the constraint set. To be sure, underspecification techniques by themselves do only a modest part of the job. The hard task of gaining the missing information needed for a satisfactory interpretation can only be done by problem-specific machinery, e.g., by modeling the reinterpretation process along the lines of the "generative lexicon", or by developing heuristic completion and combination for fragmentary input, as intended in the Robust Semantics module of Verbmobil. Underspecification formalisms do, however, provide a general framework to express semantic information with some degree of specificity, and to model its continuous increase in the interpretation process in a natural and efficient way.

I would like to conclude this section with two remarks about the underspecification formalism, one concerning its theoretical status, and the other one its expressive power. The underspecification formalism has been introduced above in a pretty informal way, as a mixture of logical representations and meta-elements (in fact, this is the way in which most underspecification formalisms present themselves). In principle, a tree-logical account of underspecification has to make a strict distinction between the meta-level description language, and the logical representation language, whose structures are described (henceforth *object language* or OL in short). Object-language expressions (conceived of as

chapter, I will ignore details of semantic analysis, unless they are relevant for the aims of the chapter.

trees) serve as models for the description language. Meta-language terms denote OL expressions, constraints are satisfied or dissatisfied by OL expressions. Since meta-language names for OL symbols can be the respective symbols themselves, the separation of levels does not require any major notational changes. The constraint sets given above can be understood as pure meta-language constructs.

Underspecification formalisms with dominance constraints are a rather expressive means to describe partial meaning information, but they are not expressive enough to describe all kinds of partial information that occur in natural language utterances. In particular, a problem arises when underspecification and discourse parallelism interact. Like underspecification, parallelism is a pervasive phenomenon in natural language. It occurs, e.g., in question-answer pairs, enumerations, contrastive constructions and corrections. There, it imposes strong semantic constraints requiring that the interpretations of the two utterances in a parallel construction are similar (ideally, identical), except for one pair (or several pairs) of contrasting constituents. This has been long recognized in lexical ambiguity and made use of in the well-known zeugma test.

(9) John owns a pen, and so does Bill.

Due to parallelism, the choice of a reading for *"pen"* in the first clause determines its reading in the second clause (mixed readings are impossible or at least pragmatically marked). A similar kind of interaction takes place in the case of scope ambiguities. The first detailed account of parallelism and scope interaction is given in Ivan Sags work on ellipsis ([Sag76]). The classical example was contributed by Hirschbuehler [Hir82], a variant of which is given in (10).

(10) Two European languages are spoken by every linguist, and two Asian ones are, too.

The first clause of Sentence (10) is two-ways ambiguous, and so is the second one. If we represent the underspecified semantics of both clauses by means of dominance constraints (as I did for Sentence (1) in (6)), the resulting constraint set will specify four (minimal) readings for (10). However, there are only two intuitively available readings: Due to parallelism *"every linguist"* must take either wide scope or narrow scope, in both clauses.

[Rey96] proposes a co-indexing technique for modeling covariance induced by parallelism. This technique works fine for the simple case of homonymy, but is difficult to handle in structurally complex cases of underspecification like (10). [Pin96] proposed context unification as a unified framework for scope underspecification and parallelism. In the Saarbrücken CHORUS project, we decided for another solution, which yields roughly the same expressive power, but is preferable in computational respects: The constraint language is extended by parallelism constraints, which come in two different forms. There is a two-place equality relation $X \sim Y$ stating structural equality between the trees dominated by X and Y. This equality relation is not sufficient to cover the standard cases of parallelism in natural language, since parallel constructions usually contain a pair of contrasting elements in corresponding positions of the respective representations, which are exempt from the parallelism requirement. Therefore, a second type of more complex constraints is employed, which we call "equality-upto constraints", and which we write as $X/U \sim Y/V$. $\sim$ is a four-place relation stating that the semantic representations of source and target clause are equal, except for the positions taken in by the contrasting elements (the representations of the NPs *"two European languages"* and *"two Asian languages"*, in the case of example (10)). Adding the appropriate equality-upto constraint to the scope-underspecified description of the first clause of (10) (which is structurally the same as (6)) yields a constraint set specifying just the two correct interpretations for (10))

[NPR97a, NPR97b]. I will not further comment on linguistic aspects of the interaction of underspecification and parallelism, but have to come back to the topic in connection with the question of the appropriate direct, truth-conditional interpretation of underspecified representations.

3 What is Underspecification Good for?

Underspecification improves efficiency, as the examples in the last section have illustrated: Working with fully specified logical representations requires either enumeration and separate processing of numerous interpretation alternatives, or early decision for one possible reading, which may necessitate subsequent revisions, and these tend to be expensive and difficult to control. Working with partial meaning descriptions allows to model semantic analysis (construction and resolution processes) in terms of monotonic extension of a constraint set, which in the optimal case results in the description of a unique logical formula, and in any case improves efficiency by allowing late disambiguation. This is the standard argument in favor of underspecification techniques, and it is a convincing one in view of the pervasive ambiguity and semantic openness of natural language.

However, underspecification opens interesting perspectives for efficient and natural semantic processing also in the case where all needed information is in principle available, and independent from the problem of ambiguity. A variety of different interaction modes between semantics and other linguistic levels can be realized, while the standards of a logic-based representation language are maintained. For underspecification enables semantic processing in a phase of the analysis, where the information is not yet completely there, but still in the process of being generated by other components. Thus, incremental architectures (spanning the levels of speech recognition, syntax, semantic analysis and reasoning), as well as concurrent processing modes on different levels of granularity become possible.

Underspecification moreover supports robust processing: Conventional syntax- and logic-based processing is reliable and precise where it succeeds, but is highly fragile, at the same time: If the conditions of complete and consistent input information are not met, this causes complete failure. "Flat" heuristic or statistical understanding systems always come up with a result. The results however are unreliable and horribly bad at times. A basic obstacle for combining deep and flat techniques in a sensible way has been the lack of a common semantic framework that allows logic- based, but partial and approximate meaning representation [4].

As already illustrated in Sect. 2, the underspecification framework allows to represent the information retrieved from fragmentary input by standard techniques of semantic construction, and to model the effect of heuristic rules employed to improve the fragmentary results. The *Robust Semantic Processing* module in the second Verbmobil phase is supposed to make use of these properties, in several respects: to combine heuristic and conventional processing methods, to combine the output of deep and flat parsers, and to process the parsers' output incrementally.

So far, I have pointed out the improvement of efficiency in the process of analysis gained by using underspecification techniques, as well as the improvement of robustness with respect to the variation and quality of the input for semantic analysis. In the remainder of the chapter I will deal with what possibly is the strongest motivation for an underspecification approach: The fact that humans are, and systems should be able to make use of un-

[4] The Verbmobil Research prototype is to my knowledge the first system which combines the advantages of 'deep' and 'flat' techniques, employing a deep linguistic analysis module and several flat parsers at the same time. If deep analysis is successful (within a given time limit), its result is preferred. Otherwise, one of the alternative flat analyses is selected. Since the different modules work in isolation, the method still leads to sub-optimal results.

derspecified output of the semantic analysis, without making reference to the single fully specified meaning alternatives. Ideally, the process of interpretation (semantic construction and resolution) reduces the set of possible readings to a singleton. In many cases, however, there is not enough information available and also no need to arrive at a unique reading. Humans seem to have no difficulties in making sense out of the utterance in these cases, and it seems that they exploit the underspecified information in a direct way, without taking the different possible readings into account. Take Sentence (11) as an example.

(11) Many linguists and computer scientists in three Saarbrücken institutes work on certain problems of computational semantics.

If you hear an utterance of (11) at an international workshop, it is highly improbable that you definitely decide on one of its numerous readings (Are there many researchers who are either linguists and computer scientists, or are there many linguists, as well as many computer scientists? Do they work on the same problems, or on different problems? Jointly or individually?). But nevertheless you might conclude that it is worth looking at the home page of the Saarbr"ucken's Computational Linguistics department or of the German Research Center for AI (DFKI). This means that you have drawn inferences from underspecified semantic information. Providing natural language systems with this kind of reasoning abilities would lead to another considerable gain in efficiency: Systems working on NL input would get around the alternative of either achieving a complete disambiguation (which may be very expensive and completely irrelevant, at the same time), or of performing classical deduction procedures on a large number of interpretation alternative, in parallel.

The question of how efficient methods for direct deduction might be realized will be addressed in Sect. 5. In advance, it must be discussed what kind of inferences should be drawn at all, from partial meaning information. What is the appropriate entailment concept, that direct deduction is supposed to meet? What do partial meaning descriptions really mean, after all?

4 What do underspecified representations mean?

Let us sort out a couple of simple but inappropriate answers right at the beginning. First, one could assume that the problem is already solved, since underspecification constraints as expressions of a tree-description language already have a perfect denotational interpretation, in terms of trees standing for logical structures. This is the wrong level of semantics however: (11) is an example of reasoning about the world, not about properties of logical structure. Accordingly, we have to care about a denotational semantics in terms of models that represent the state of the world, rather than object language terms. Entailments on the term-semantic level are "structure-preserving", e.g., of the general form: If a structure has property F, it has property G as well. Entailments on the level under consideration should be truth-preserving. The interesting instances will be those which do not preserve structure, i.e., which lead from (partial) descriptions of logical formulae to (partial) descriptions of different formulae, e.g., in the inference indicated in (12) (scope ambiguity being indicated by the prefixes in set notation).

(12)
$$\begin{array}{l} \{\forall x : \mathit{researcher}, \exists^2 y : \mathit{language}\}(\mathit{speak}(x,y)) \\ \forall x : \mathit{linguist}(\mathit{researcher}(x)) \\ \hline \{\forall x : \mathit{linguist}, \exists^2 y : \mathit{language}\}(\mathit{speak}(x,y)) \end{array}$$

Direct interpretation should refer to the level of the truth-conditional semantics used for the interpretation of the object language. The tricky question is how the object language interpretation concept and in particular the concept of truth can be adapted to descriptions which are compatible with more than one object-language expression.

There is a popular and intuitively appealing answer to this latter question: Just interpret ambiguity as object-level disjunction. Faced with an utterance of sentence (1), you are on the safe side if you take its meaning information to be the disjunction of its readings, which incidentally is identical with the narrow scope existential reading. Similarly, if you hear (13), and know already that Mary went to the savings-bank, it would be not constructive to question its truth on the basis of the falsehood of the other interpretation alternative.

(13) Mary left for the bank

According to Gricean maxims, the hearer should interpret an ambiguous utterance cooperatively, and this pragmatic requirement again is nicely met by the disjunctive interpretation, since it is the widest possible one. However, there is a problem with the argument, despite of its apparent plausibility. It works only in the case of plain, affirmative statements. If you are asked one of the questions (14) or (15), the answer "yes" might be misleading in spite of the fact that the corresponding assertions are true under one reading, namely in the case that the speaker had the other reading in mind.

(14) Is it true that two languages spoken by every linguist?

(15) Did Mary leave for the bank?

Similarly, the disjunctive interpretation is inappropriate, if the ambiguity is embedded in a negation, as in (16) and (17).

(16) It is simply wrong that two languages are spoken by every linguist

(17) Mary didn't leave for the bank

In these cases, disjunctive interpretation amounts to the conjunction of the negation of the two readings, by de Morgan's Law. This is the strongest possible interpretation, risky under the aspect of information acquisition, and maximally uncooperative as a policy of conversational behavior. Thus, the apparent semantic appropriateness of the disjunctive analysis seems to be due to some special pragmatic effects.

Before completely abandoning the option of a disjunctive interpretation, we have to inspect another way of relating underspecification to object-language disjunction. We could assume that an underspecified representation as a whole is semantically equivalent to the disjunction of the set of logical formulae which it describes. This amounts to giving disjunction always widest possible scope. De Morgan does not apply here: If A has the set of readings $\{A_1, A_2\}$, $\neg A$ has readings $\{\neg A_1, \neg A_2\}$, and accordingly, the interpretation of A and $\neg A$ would be taken to be equivalent to $A_1 \vee A_2$ and $\neg A_1 \vee \neg A2$ (rather than $\neg(A_1 \vee A_2)$), respectively. Unfortunately, there is a different problem with this approach. Take $\textsc{bank}_1$ and $\textsc{bank}_1$ as shortcuts for the *'savings-bank'* and the *'river bank'* reading of Sentence (18), respectively.

(18) This is a bank

(19) This isn't a bank

Then, (18) would be represented as $\textsc{bank}_1 \vee \textsc{bank}_2$, and (19) as $\neg\textsc{bank}_1 \vee \neg\textsc{bank}_2$. Since no object can fall under both readings of *"bank"*, the two readings of (18) are mutually exclusive, i.e., both $\textsc{bank}_1 \rightarrow \neg\textsc{bank}_2$ and $\textsc{bank}_2 \rightarrow \neg\textsc{bank}_1$ hold. On these premises, (18), which is equivalent to $\textsc{bank}_1 \vee \textsc{bank}_2$, entails as $\neg\textsc{bank}_1 \vee \neg\textsc{bank}_2$, which again is equivalent to (19). This means that any ambiguous expressions with mutually incompatible readings entails its negation — a conclusion which is hardly intuitively desirable. The problem is that the entailment concept resulting from the wide-scope disjunctive interpretation is much too weak. To prove $A \models B$, it would be sufficient to show that if A is true under one reading, there is one arbitrary true reading of B[5].

Interpretation of ambiguity via object-language disjunction is tempting, but definitely inappropriate. In the case of the *"bank "* example, the result of the wide-scope disjunctive interpretation is particularly unintuitive, since one has to switch the readings systematically, when proceeding from premise to conclusion. This seems to indicate the necessity to make the interpretation for ambiguous expressions context- sensitive, viz., dependent on previous choices of a disambiguation. One way of implementing this dependence is by means of a formal notion of context, defined as a function from ambiguous expressions to possible readings. Interpretation and entailment concept then can be defined in the spirit of the super-valuation technique, which has been proposed for the semantic analysis of vague expressions earlier already [Fin75, Pin95]. An underspecified formula A is true, if $k(A)$ is true for every possible context k, false, if $k(A)$ is false for every possible context k, and of undefined truth-value otherwise. Entailment is not defined on the basis of the global truth-value assignment, but by inspecting the semantic relation between the involved underspecified formulas context-wise. A entails B, if and only if for every context k: If $k(A)$ is (classically) true, then $k(A)$ is (classically) true. With respect to the problem case under discussion, the context-based definition of entailment clearly shows much better behavior. Since the choice of a reading in (18) fixes the context for the interpretation of (19), there is no possibility for (18) and (19) to become true at the same time. Thus, (19) not only does not follow from (18), but it turns out to be inconsistent with it — in nice accordance with classical logic.

[5]This is the $\exists\exists$-variant in Kees van Deemter's classification of alternative ambiguous entailment concepts [Dee96].

Again, there are basic problems concerning the intuitive adequacy of this proposal. Fixing a reading for an ambiguous expression once and for all in a sentence or discourse would not allow to provide sensible interpretations for sentences like (20) or (21).

(20) The bank is close to the bank (of the river)

(21) The pen is in the pen

As Reyle argues in [Rey96], even $A \models A$ is only justified under certain assumptions about the pragmatic relation between the two occurrences of A. If a speaker or system has added an underspecified representation to its knowledge base, based on an utterance of (15), without being able to resolve the ambiguity, and is asked the question (17), respectively, later, it cannot safely give an affirmative answer, since the reading intended by the informant may be a different one than the reading intended by the person asking the question. All this seems to suggest a much more cautious entailment concept which requires for $A \models B$ to hold that every possible reading of B be classically entailed by every reading of A: The truth of any disambiguation of the premise must guarantee truth of all disambiguations of the conclusion to cover all possible cases of resolution.

This cautious entailment (van Deemters $\exists\forall$-variant) can clearly serve as a safe semantic basis for direct deduction. However, a semantics which assumes free choice for all local disambiguation possibilities is not sufficient to account for the obvious interdependence of disambiguation decisions in (22) and (23).

(22) This is a bank, but it isn't a bank.

(23) This is a bank. So, this is a bank.

Sentence (22), to say the very least, has a strong tendency to be judged a contradiction, and likewise, there is a strong intuitive tendency to consider (23) a valid entailment, despite the ambiguity involved. Thus, it seems that the interpretation of underspecification has to take into account some kind or degree of covariation, though not in the obligatory way required by supervaluation semantics. [Dee96] proposes to employ a notion of coherence: According to his proposal, more uniform interpretations are preferred to less uniform ones among the set of consistent disambiguations of a discourse (everything else being equal). The degree of coherence or uniformity of an interpretation is computed, roughly speaking, by counting occurrences of ambiguous expressions which are assigned identical readings. However, van Deemter's proposal does not provide a real solution. In whichever way the coherence measure applies to the above examples, it will not explain the difference between (20)/(21) on the one hand, and (22)/(23) on the other hand[6].

The systematic difference between (20) and (21) on the one side, and (22) and (23) on the other side is that the latter are instances of the phenomenon of discourse parallelism. As I have pointed out already in connection with the zeugma test and the Hirschbuehler sentence in Sect. 2, parallelism imposes a constraint on semantic interpretation which forces parallel elements to have equal logical representations. Parallel constructions belong to the hard core of semantically relevant covariance cases. If we disregard possible other sources for obligatory covariance (it may well be that there aren't any), a straightforward solution to the direct interpretation problem is at hand. We assume that the cautious semantics considering all possible disambiguations is basically the correct one. Given equality constraints (in one or the other form,

[6]Pragmatics seems to be the appropriate level for coherence phenomena. Coherence-based decisions for interpretation are be guided by rather weak preferences (corresponding to the Gricean maxims of manner), which in turn are based on a variety of complex factors of textual and conceptual organization, rather than on a simple count criterion. Accordingly, van Deemter´s attempt to treat coherence on the level of semantics, leads to a highly non-standard logic for ambiguous expressions.

cf. Sect. 2), this does not automatically amount to independent choice of readings in all cases, but only within the limits imposed by the constraints. Let us look at the examples. There are no limits on the choice of readings for the occurrences of *"bank"* in Sentence (20), since the constraint set (24) describing its meaning potential covers all possible combinations of readings BANK_1 and BANK_2.

(24) $\{close\_to(X_1, X_2), X_1 \in \{\text{BANK}_1, \text{BANK}_2\},$
$X_2 \in \{\text{BANK}_1, \text{BANK}_2\}\}$

(25) $\{X_0 : X_2 \wedge \neg X_3, X_2 : X_4(a), X_3 : X_5(a),$
$X_4 \in \{\text{BANK}_1, \text{BANK}_2\}, X_5 \in \{\text{BANK}_1, \text{BANK}_2\}, X_2 \sim X_3\}$

On the other hand, the parallelism between the clauses of (22) introduces the equality constraint $X_2 \sim X_3$ into the description (25), which rules out different instantiations of X_4 and X_5, and therefore admits only the two inconsistent readings of (22)[7].

Let us consider next, how the approach can be extended to the entailment relation. If just the logical relation between a set of meaning descriptions σ and a meaning description A has to be judged upon, as in Reyles data-base query example, all different combinations of local disambiguations must be considered. In these cases, the cautious ($\exists\forall$) entailment concept seems to be the correct one. If we state soundness of (23), however, we do not just speak about the logical relation between its two clauses, but rather state that the discourse consisting of these sentences forms

[7]I have used the *'bank '* symbol for expressions of two different object language types, in the two examples, to keep the notation simple. In (24), BANK_1 and BANK_1 are in fact predicates occurring in the scope of the definite article, in a full, Montague-style quantificational analysis. Also, I have not said anything about the way the parallelism between the first and the second, negated clause of (22) is precisely to be stated, and how it maps to the equality constraint in (25). The solution is somewhat complex and not relevant for the argument.

a conclusive argument. Since the first and second clause stand in the relation of discourse parallelism which induces an equality constraint, the choice of the reading for *"bank"* in the conclusion of (23) must be the same as in the premise, and thus truth of the conclusion is guaranteed by truth of the premise. These considerations suggest that a distinction between two entailment concepts must be made, which is conceptually related to the distinction between static and dynamic entailment in dynamic semantics (cf. [GS91]). *'Dynamic entailment'* for underspecified representations takes constraints, in particular parallelism constraints, resulting from discourse semantic relations between premises and conclusion, into account, *'static entailment'* doesnt[8].

5 Reasoning with Underspecified Representations

Providing a direct semantics for underspecified representations is one problem, designing sound and complete calculi for direct deduction is a different one[9]. It is yet another question, however, how to build efficient software systems for reasoning with underspecified information. No answer to this question exists to date, nor even more or less concrete ideas of how such an answer could

[8]If we want to give a formal definition of the dynamic, discourse-related entailment concept, we face a technical difficulty: Since there may be covariance between premises and conclusion, we cannot define it by strengthening the cautious static entailment concept with additional conditions, but have to provide a possibility of context-wise evaluation again. We can do this by making technical use of the super-valuation approach mentioned above. Since the description language uses different meta-variables for different occurrences, there is no a priori covariance between different occurrences of an expression. If required, covariance can be explicitly brought about by equality constraints which reduce the set of admissible contexts.

[9]Solutions for several calculi have been proposed. I will not comment on them, however, here. [Rey93, Eij96]

look like. This is not too surprising because the problem of direct deduction on underspecified representations contains standard deduction for the corresponding object language as a sub-problem. Since the latter should have at least the expressiveness of FOL, a general and tractable method for direct deduction is a priori impossible. Unfortunately, it seems that we have to make use of the full machinery of heuristics and control mechanisms developed for theorem proving to make progress with the hard (in fact, AI-hard) problem of direct deduction for underspecified expressions. In this section, I would like to indicate, how a viable pragmatic alternative to this discouraging perspective might look like.

The inspection of practically relevant examples for direct deduction shows that many of those cases which appear as intuitive and natural belong to a special sub-class, viz., arguments which start from underspecified premises and yield completely specified conclusions. This sub-class of inferences is particularly useful, since it renders representations which may be further processed using standard reasoning techniques. Instead of working at a general solution of the problem of direct deduction, one could aim at dividing the problem into an easier task of providing special inference mechanisms for (hopefully) quick and easy extraction of safe knowledge, and the task of reasoning with the result of the extraction, which is a hard, but not problem-specific task, and thus can be left to standard theorem-provers.

Let us take Sentence (1) as an example, again. Since its narrow-scope existential reading is entailed by the stronger wide-scope one, the former can be inferred from the underspecified representation of (1) as safe information. The deduction rule could be roughly formulated in the following way: If a pattern like (26) occurs in a constraint set (stating that a universal and an existential quantifier apply to one nuclear expression without specifying a relative scope-ordering scope ordering), add constraint $X_4 \lhd^\star X_2$, (which amounts to a specification of the $\exists\forall$-reading).

(26) $\{X_1 = \exists x(X_3), X_2 = \forall y(X_4), X_0 \triangleleft^\star X_1, X_0 \triangleleft^\star X_2,$
$X_3 \triangleleft^\star X_5, X_4 \triangleleft^\star X_5\}$

Note that the rule does not make reference to fully specified OL readings, but rather works directly on the constraint set. This is not too impressive in the example under discussion, but becomes quite helpful in cases where dozens or hundreds of possible readings are around[10]. Another special, not too exciting, but useful type of direct deduction concerns cases of equivalent scope readings. Scope differences sometimes do and sometimes do not affect truth conditions. Examples (27) and (28) are cases where relative scope of the quantifiers does not matter.

(27) Some linguists speak a foreign language

(28) Every linguistics students knows every professor of linguistics.

A method that detects possible candidate pairs for scope equivalence by directly inspecting the semantic description would be obviously useful, since it allows to eliminate spurious scope ambiguities without enumerating and inspecting the disambiguations.

There is a very simple method to arrive at a fully specified object-language representation that denotes the maximum of safe information in an underspecified meaning description, given that there are finitely many readings. Just generate all readings, and connect them by object-language disjunction. The solution is theoretically incontestable: The disjunction denotes safe information, since its truth is guaranteed by the truth of any of the readings, and clearly it is also the maximum of safe information we can

[10] It should be mentioned that the solution is not as easy as I have sketched here, since a prerequisite for the applicability of the rule is that both quantifiers must occur in a positive context, and no different operators can possibly intervene between the two. I just want to give an idea of the special kind of rules I have in mind, and thus I will not go deeper into the problems, here.

get. With respect to practical application, it isn't a particularly attractive solution. It involves the schematic enumeration of all specific readings, again. This is unnecessary in many scope underspecification cases because they allow a compact treatment, as shown above, and it hardly leads to useful information in those cases which do not allow that compact treatment like *"bank"* and *"pen"* . The policy I have in mind is rather the following one: Do not aim at completeness, but pick out certain sub-problems of direct deduction (like the identification of weakest scope order, or equivalent scopings). Apply these special techniques of direct deduction whenever they are applicable. Otherwise, perform a (partial) disambiguation step, and try again to apply direct deduction rules to the result. If direct deduction works well for a given constraint set, a lot of processing time can be saved. Otherwise, there is a flexible transition to the conventional technique of working on disambiguations [11]. Apparently, also humans do not use one method of reasoning with underspecified information exclusively, but interleave direct reasoning, partial disambiguation, and maybe a number of other techniques. In earlier work ([Pin95]), I have argued for a classification of ambiguity phenomena, the main distinctive criterion being the need or tendency for disambiguation.

6 On the Status of Underspecified Representations

My comments on direct deduction may have left the impression that underspecification techniques have a rather limited range,

[11] To be more precise, this can either mean: make a case distinctions and work on the resulting cases in parallel, or: pick a preferred alternative, and if you run into a deadlock, revise your decision. Cf. [Poe96] who argues that the latter kind of reasoning is basic to human processing of underspecified information.

once the task of sentence-internal semantic composition is done: If an underspecified representation occurs, we either proceed to the specific kernel information shared by its different readings, or retract to disambiguations. Thus one could assume the following general pattern for semantic processing: First, sentence meaning is computed using partial descriptions (for which they are undoubtedly useful). Then, these descriptions are replaced by fully specified logical representations, in one or the other way, which are the only available and necessary data structures for further processing steps in the understanding process.

This view is not completely correct, for several reasons, however. On the one hand, the approach proposed in the last section was meant to provide an easy and viable method to exploit underspecified information. There was no claim of completeness involved. Examples like (12) indicate that there might well be useful and natural inference patterns leading from underspecified representations to underspecified representations. On the other hand, there is a specific argument in favor of a more global role of underspecification coming from discourse semantics. Look at (29), which is an example for the interaction between scope and anaphoric binding by virtue of accessibility constraints (cf. [KR93]).

(29) Two languages are spoken by every linguist. One of them is English.

If we had replaced the underspecified representation of the first clause of the text in (29) by the $\exists\forall$-reading, the wide-scope existential reading required by the anaphoric subject of the second clause would not be available any more. Another kind of evidence for the relevance of underspecification for discourse structure is provided by the discourse semantic phenomenon of parallelism. In cases of equivalent scope readings as (27) or (28) of Sect. 5, one can proceed from the scope neutral description to one arbitrary representation with fixed scope, without losing any truth-conditionally relevant information. However, even in this case the underspecified representation cannot be just dispensed with.

(30) At least one computer science student knows every professor of computer science, too.

Let an utterance of (29) be followed by (30), in which the scope order is crucial for truth conditions. As pointed out in Sect. 2, parallelism requires the corresponding quantifiers of the two sentences to be in the same scope relation. Taking one of the possible equivalent scope orders as the representation of (29), the truth conditions of (30) are fixed in a purely accidental way.

Apparently, there is a layer of information which may be truth-conditionally irrelevant, but is indispensable for discourse-semantics. This reminds of the situation in DRT: Translation of DRSes to formulas of predicate logic preserves the truth conditions, but the discourse-semantic semantic layer of the anaphoric potential is lost. In the same way, replacement of an underspecified representation by the safe kernel leads to loss semantic information which is relevant for discourse-semantic purposes. Rather than drawing the borderline between underspecification techniques and standard semantic frameworks at the sentence boundary, we may assume a global two-layer model for semantic processing. On the one hand, an underspecified description providing the potential for disambiguation or precisification must be available at any point of the discourse. On the other hand, we can at any point extract specific safe information, and try to do further inferencing on it by conventional means. Thus, the special techniques of direct reasoning proposed in the last section do not amount to replacing the latter, but mediating between the language-oriented level of semantic representation and the non-linguistic information processing level.

I think that this view of the role of underspecification in semantic information processing opens a perspective to tractable, and hopefully efficient solutions for some of the basic tasks in natural language understanding. Let me finally ask whether it could also serve as the basis of an adequate cognitive model of meaning processing. There is a very basic difficulty for giving a general answer to this question. We cannot be sure whether the methods we have available in semantics - be it on the level of logical representation languages or description language constraints - provide access to all dimensions, or even the most relevant ones of the mental encoding of meaning. In fact, there are strong intuitions that certain kinds of semantic information which can only be spelled out in a rather complex way using available techniques, have a rather simple and compact mental representation. One of these cases is ambiguity of quantifier scope. Maybe there is a completely different way of encoding the semantics of sentences with several scope-bearing elements, from which both kernel information and potential of precise readings follow. Another example are ontologically multi-layered common nouns (like *"newspaper"* and *"university"*), whose peculiar meaning structure James Pustejovsky tries to capture with the *dotted pairs* of his *Generative Lexicon.* I would refrain from claiming cognitive adequacy for any of the presently available approaches to semantic underspecification. Maybe, even the notion of underspecification itself is inappropriate because it suggests that there is a significant lack of information, where in certain cases at least, it might be the specific representational framework which disallows a direct unambiguous encoding.

However, I do not see any alternative framework which supports the basic kind of processes needed for understanding natural language - efficient semantic construction, context-sensitive meaning assignment, and inference. Thus, underspecification techniques are the best we have to date, although they might eventually turn out to be deeply inadequate.

7 Concluding Remarks

What I have said in this chapter, was mostly programmatic in character. Thus, I addressed things to be done and speculated about the general direction to take in order get these things done, but does not become too specific about details, as it is typically the case with programmatic papers. However, this one has the slight advantage that the program it presents is mainly intended as a program for a special project, the CHORUS project run at Saarbrcken University[12], rather than a program for the community in general. The principal aim of the project is a unified and tractable method for the representation and processing of all kinds of partial meaning information which have been exemplified in Sect. 1. Subgoals are the design of a constraint language for partial meaning representation (as indicated in Sect. 2); the investigation of methods for semantic analysis with partial input information (as indicated in Sect. 3); the development of direct deduction techniques, as indicated in Sect. 5, on the theoretical background outlined in Sect. 4. I hope we will be able to present results of this research in the near future.

Acknowledgements

The ideas and opinions presented in this chapter have grown in the environment of the CHORUS project and its immediate sister projects in the Sonderforschungsbereich 378. I am indebted to all colleagues working in this group and visitors to the group, in particular to Markus Egg, Jan van Eijck, Jan Jaspars, Michael Kohlhase, Alexander Koller, Reinhard Muskens, Joachim Niehren, and Peter Ruhrberg.

[12]CHORUS (Semantic Processing using Concurrent Constraints) is part of the SFB (Special Research Division) 378 "Ressource adaptive Cognitive Processes" funded by the German Science Foundation.

References

[AC92] H Alshawi and R. Crouch. Monotonic semantic interpretation. In *Proceedings of the 30th ACL*, pages 32–39, 1992.

[Bos96] Johan Bos. Predicate logic unplugged. In P. Dekker and M. Stokhof, editors, *Proceedings of the 10th Amsterdam Colloquium*, Amsterdam, 1996. ILLC.

[CFS] A. Copestake, D. Flickinger, and I. Sag. Minimal recursion semantics. An introduction. CSLI, Stanford University.

[Dee96] K. van Deemter. Towards a logic of ambiguous expressions. In K. van Deemter and S. Peters, editors, *Semantic ambiguity and underspecification*, pages 203–237. CSLI Publications, Stanford, 1996.

[Eij96] J. van Eijck. The logic of ambiguation. In The FraCas Consortium, editor, *Building the Framework, FraCas Deliverable 15*. 1996.

[EL96] M. Egg and K. Lebeth. Semantic interpretation in HPSG. Paper presented at the 3rd Conference on HPSG, Marseilles, May 20th-22th, 1996, 1996.

[Fin75] K. Fine. Vagueness, truth, and logic. *Synthese*, 30:265–300, 1975.

[GS91] Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics & Philosophy*, 14:39–100, 1991.

[Hir82] P. Hirschbühler. VP deletion and across the board quantifier scope. In *Proceedings of NELS 12*, 1982.

[KR93] Hans Kamp and Uwe Reyle. *From Discourse to Logic*. Kluwer, Dordrecht, 1993.

[Mus97] Reinhard Muskens. Order-independence and underspecification. Proceedings of the Workshops in Prague (Feb 1995) and Bad Teinach (May 1995), Volume 2: Comments and Replies, pages 155-166, Arbeitspapier des IMS, Universität Stuttgart, 1997. (appeared)., 1997.

[NPR97a] Joachim Niehren, Manfred Pinkal, and Peter Ruhrberg. On equality up-to constraints over finite trees, context unification and one-step rewriting. In *Proceedings of CADE'97*, 1997.

[NPR97b] Joachim Niehren, Manfred Pinkal, and Peter Ruhrberg. A uniform approach to underspecification and parallelism. In *Proceedings of ACL'97*, pages 410–417, Madrid, Spain, 1997.

[Pin95] Manfred Pinkal. *Logic and Lexicon*. Kluwer, Dordrecht, 1995.

[Pin96] Manfred Pinkal. Radical underspecification. In P. Dekker and M. Stokhof, editors, *Proceedings of the 10th Amsterdam Colloquium*, pages 587–606, Amsterdam, 1996. ILLC.

[Poe96] Massimo Poesio. Semantic ambiguity and perceived ambiguity. In *Semantic Ambiguity and Underspecification*. CSLI-Publications, Stanford, CA, 1996.

[Pus91] J. Pustejovsky. The generative lexicon. *Computational Linguistics*, 17:409–441, 1991.

[Pus95] J. Pustejovsky. *The generative lexicon*. MIT Press, Cambridge, 1995.

[Rey93] U. Reyle. Dealing with ambiguities by underspecification: construction, representation, and deduction. *Journal of Semantics*, 10:123–179, 1993.

[Rey96] U. Reyle. Co-indexing labeled DRSs to represent and reason with ambiguities. In K. Deemter and S. Peters, editors, *Semantic ambiguity and underspecification*, pages 239–268. CSLI Publications, Stanford, 1996.

[Sag76] I. Sag. *Deletion and logical form.* PhD thesis, MIT, Cambridge, 1976.

[Wah93] Wolfgang Wahlster. Verbmobil: Translation of face-to-face dialogues. In *Proceedings of the 3^{rd} European Conference on Speech Communication and Technology*, pages 29–38, Berlin, Germany, 1993.

[Wor97] Karsten L. Worm. Robuste semantische Verarbeitung. Verbmobil-Report 200, Universität des Saarlandes, Saarbrücken, December 1997. Available from http://www.dfki.de/verbmobil/.