Semantic underspecification

Markus Egg

Abstract

Semantic underspecification is a technique to capture several readings of an ambiguous expression in one single representation by deliberately omitting the differences between the readings in the representation. First, underspecification formalisms will be presented to introduce underspecification in general and to outline important properties of these formalisms that allow their classification into subgroups. After expounding the kinds of ambiguity to which underspecification can be applied, the article then presents various motivations for the use of underspecification, and shows how underspecified semantic representations can be further processed.

1 Introduction

Underspecification can be defined as the deliberate omission of information from linguistic descriptions to capture several alternative realisations of a linguistic phenomenon in one single representation. Underspecification emerged in phonology (Steriade 1995; Harris 2007) and was later adopted by semanticists to model *ambiguity*. Underspecified semantic representations capture whole sets of different meanings (one for each reading of an ambiguous expression) in one representation. Semantic underspecification focusses on expressions with systematically related sets of readings, in particular, on scope ambiguity.

In natural language processing, underspecification is endorsed to keep semantic representations of ambiguous expressions tractable and to avoid unnecessary disambiguation steps; a new use of underspecification is its use in *hybrid processing*, where it is used as a common format for the results of deep and shallow processing. Underspecification is used also in syntax and discourse analysis (Marcus et al. 1983; Rambow et al. 2001; Muskens 2001; Duchier and Gardent 2001; Schilder 2002; Egg and Redeker 2008; Regneri et al. 2008).

The next section outlines underspecification formalisms in general and presents important properties of underspecification formalisms which distinguish different subgroups of these formalisms. Then the range of semantic phenomena to which underspecification can be applied will be sketched in section 3. Various motivations for using underspecification in semantics are outlined in section 4. The last section describes how these representations can be further processed.

2 Approaches to semantic underspecification

This section introduces the technique of underspecification and outlines general properties of underspecification formalisms that allow their classification into several subgroups. First and foremost, these formalisms handle ambiguity by either *describing* the different readings of an ambiguous expression or by providing a procedure to *derive* these readings. But these formalisms also differ with respect to other properties, in particular, their *expressivity* (can they specify not only the set of readings of an ambiguous expression but also arbitrary subsets of this set).

The ambiguous examples whose underspecified treatment is expounded in this section are so-called *quantifier scope* ambiguities. (The word ‘quantifier’ refers to DP meanings, formally, sets of properties, except in expressions such as 'universal quantifier'.) Consider e.g. the well-worn (1) with its two readings (2a) 'for every woman, her own man' ($\forall > \exists$; '$>$' indicates wide scope of its left argument over the right one) and (2b) 'one man for all women' ($\exists > \forall$):

(1) Every woman loves a man

(2) (a) $\forall x(\mathbf{woman'}(x) \rightarrow \exists y(\mathbf{man'}(y) \wedge \mathbf{love'}(x, y)))$ (b) $\exists y(\mathbf{man'}(y) \wedge \forall x(\mathbf{woman'}(x) \rightarrow \mathbf{love'}(x, y)))$

The formulae in (2) consist of the same three parts (roughly, the semantic contributions of the verb and its two arguments), and the relation of loving as introduced by the verb always gets lowest scope. The formulae only differ in the arrangement of the semantic contributions of the arguments of the verb.

Since quantifier scope ambiguity is the prototypical domain for the application of underspecification, involved cases of quantifier scope ambiguity like (3) have developed into benchmark cases for underspecification formalisms:

(3) Every researcher of a company saw most samples

(3) is a case of *nested quantification* in that the subject DP introduces a quantifier and comprises another DP that introduces one more quantifier. The challenge of nested quantification is fact that the number of readings is less than the number of the possible permutations of its quantifiers w.r.t. their scope ordering. E.g., in (3), there are $3! = 6$ possible permutations but at least one scope ordering is not attested ($\forall > \mathbf{most'} > \exists$; Hobbs and Shieber 1987).

Appropriate underspecification formalisms must be able to represent the exact range of readings of an ambiguous expression and may not overgenerate by predicting unattested readings. This is accomplished in two ways.

First, ambiguity can be *described*: Expressions of a formalism describe the set of readings of an ambiguous expression so closely that this suffices to determine the range of its readings. Procedures that derive the individual readings then merely enumerate the readings, they do not restrict them in any way.

Second, ambiguity can be *derived*: Some formalisms provide an initial, more general characterisation of the readings; the exact range of readings is then only determined by specifying a procedure (an algorithm) to derive fully specified readings from the general characterisation.

2.1 Describing ambiguity

The first way of implementing semantic underspecification are *partial descriptions* for the sets of semantic representations for the readings of ambiguous expressions. These descriptions by themselves delimit the range of readings of the ambiguous expression and specify them.

This strategy is based on the fact that sets can be characterised by a property that exclusively holds for their elements. For ambiguous expressions, sets of semantic representations for their readings are defined by describing the common ground between these representations only. Since this deliberately omits the differences between them, the description can only be partial.

Most underspecification formalisms that follow this strategy distinguish an object level (semantic representations) and a meta-level (descriptions of these representations, called *constraints*). The formalisms define the expressions of the meta-level and their relation to the described object-level representations. As a simple example, consider (1) and its readings (2a-b) and the description of the common ground between (2a-b) in the constraint (4):

$$\Box$$

(4) $$\forall x\ (\mathbf{woman'}(x) \rightarrow \Box) \qquad \exists y\ (\mathbf{man'}(y) \wedge \Box)$$

$$\mathbf{love'}(x, y)$$

(4) comprises four *fragments* of semantic representations (here, λ-terms) which may comprise *holes* (parts of fragments that are not yet determined, indicated by '$\Box$'). Holes and fragments are related by a relation R (depicted as dotted lines), if R holds for a hole h and a fragment F, F must be part of the material that determines h.

R determines a partial scope ordering between fragments: A fragment F_1 outscopes another fragment F_2 iff F_1 comprises a hole h such that $R(h, F_2)$ or $R(h, F_3)$, where F_3 is a third fragment that outscopes F_2 (Copestake et al. 2005). We assume that the description explicates all the fragments that show up in the described object-level representations and that variable binding operators in a fragment F bind occurrences of the respective variables in all fragments outscoped by F (which simplifies matters somewhat, see Egg et al. 2001).

(4) can be paraphrased as follows: The fragment at the top is just a hole, i.e., the described representations are not yet known. But since the relation R relates this hole and the right and the left fragment, they are both part of these representations - only their order must be fixed. Finally, the holes in both the right and the left fragment are related to the bottom fragment in terms of R, i.e., the bottom fragment is in the scope of either quantifier. The only semantic representations compatible with this description are (2a-b), as desired.

To derive the described readings from such a constraint (its *solutions*), R is extended until the scope of all fragments is fixed. For the solution (2a), the tuple consisting of the hole in the left fragment and the right fragment is added:

$$\Box$$

$$\forall x(\mathbf{woman'}(x) \rightarrow \Box)$$

$$\exists y(\mathbf{man'}(y) \wedge \Box)$$

(5) $$\mathbf{love'}(x, y)$$

From the viewpoint of scope, we can minimise (5) by omitting all tuples that express a scope ordering that already follows from the transitivity of scope (in

(5), the tuple consisting of the top hole and the existential fragment, and the tuple consisting of the hole of the universal fragment and the bottom fragment, respectively).[1] Then all the holes are related to a specific fragment, and all the fragments except the one at the top are related to a hole:

(6)

$$\Box$$

$$\forall x\ (\mathbf{woman}'(x) \rightarrow \Box)$$

$$\exists y\ (\mathbf{man}'(y) \wedge \Box)$$

$$\mathbf{love}'(x, y)$$

Pairwise identification of the hole-fragment tuples in (6) (the 'plugging' of Bos 2004) finally yields (2a), the first solution of (4). For the other solution (2b), start the procedure by adding to R the tuple consisting of the hole in the right fragment and the left fragment.

Underspecification formalisms that implement scope in this way comprise Underspecified Discourse Representation Theory (Reyle 1993, 1996; Frank and Reyle 1995), Minimal Recursion Semantics (MRS, Copestake et al. 2005), the Constraint Language for Lambda Structures (CLLS; Egg et al. 2001), the language of Dominance Constraints (subsumed by CLLS; Althaus et al. 2001), Hole Semantics (Bos 1996, 2004; Kallmeyer and Romero 2008), and Logical Description Grammar (Muskens 2001).

Scope relations can also be expressed by variables (whose instantiation determines a specific reading), e.g., in the Underspecified Semantic Description Language (Pinkal 1996, Niehren et al. 1997, Egg and Kohlhase 1997), the Quasi-Logical Form in Alshawi and Crouch (1992), or Glue Language Semantics (Dalrymple et al. 1997; Crouch and van Genabith 1999; Dalrymple 2001).

After this expository account of the underspecified account of the simple (1), consider the nested quantification in (7) [= (3)] and its constraint (8).

(7) Every researcher of a company saw most samples

(8)

$$\Box$$

$$\exists y(\mathbf{company}'(y) \wedge \Box) \qquad \forall x((\mathbf{researcher}'(x) \wedge \Box) \rightarrow \Box) \qquad \mathbf{most}'(\mathbf{sample}', \lambda z.\Box)$$

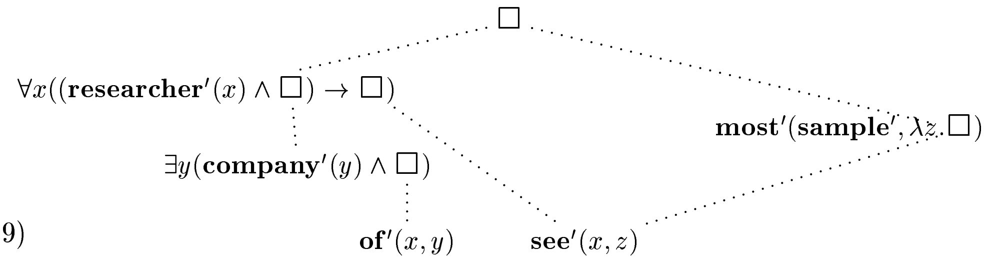
$$\mathbf{of}'(x, y) \qquad \mathbf{see}'(x, z)$$

The challenge for underspecification lies in the fact that expressions with nested quantification have less readings than the factorial of the number of the involved DPs, since some scoping options are ruled out. E.g., (7) has five readings, the impossible one with the scope ordering $\forall > \mathbf{most}' > \exists$ must be excluded in a suitable underspecified representation of (7).

To show that (8) indeed describes exactly five readings, I will now derive these readings from (8). As a first step of disambiguation, the existential and

[1] Recall that relations S are transitive iff $\forall x, y, z(S(x, y) \wedge S(y, z) \rightarrow S(x, z))$.

the universal fragment are ordered. Giving the existential fragment narrow scope yields (9):



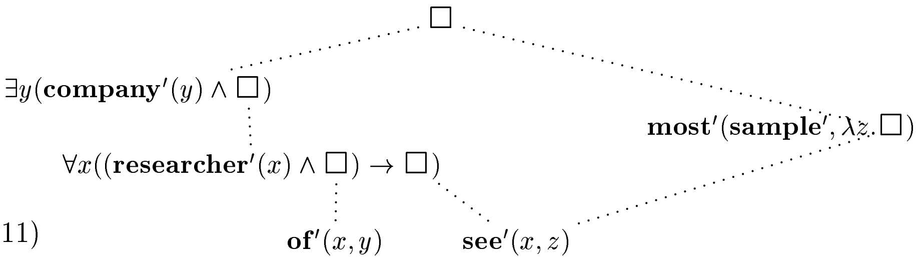
(9)

But now the existential fragment can no longer interact scopally with the *most*- and the *see*-fragment, because it is part of the *restriction* of the universal quantifier. Therefore (9) encompasses only two readings, with the *most*-fragment or the universal fragment taking widest scope. This rules out a reading in which *most* scopes below the universal, but above the existential quantifier:

(10) (a) $\forall x((\mathbf{researcher'}(x) \wedge \exists y(\mathbf{company'}(y) \wedge \mathbf{of'}(x, y))) \to \mathbf{most'}(\mathbf{sample'}, \lambda z.\mathbf{see'}(x, z)))$

(b) $\mathbf{most'}(\mathbf{sample'}, \lambda z \forall x((\mathbf{researcher'}(x) \wedge \exists y(\mathbf{company'}(y) \wedge \mathbf{of'}(x, y))) \to \mathbf{see'}(x, z)))$

Giving the existential scope over the universal one in (8) returns (11). This constraint describes the three readings in (12), whose difference is whether the *most*-fragment takes scope over, between, or below the other two quantifiers.



(11)

(12) (a) $\mathbf{most'}(\mathbf{sample'}, \lambda z \exists y(\mathbf{company'}(y) \wedge \forall x((\mathbf{researcher'}(x) \wedge \mathbf{of'}(x, y)) \to \mathbf{see'}(x, z))))$

(b) $\exists y(\mathbf{company'}(y) \wedge \mathbf{most'}(\mathbf{sample'}, \lambda z \forall x((\mathbf{researcher'}(x) \wedge \mathbf{of'}(x, y)) \to \mathbf{see'}(x, z))))$

(c) $\exists y(\mathbf{company'}(y) \wedge \forall x((\mathbf{researcher'}(x) \wedge \mathbf{of'}(x, y)) \to \mathbf{most'}(\mathbf{sample'}, \lambda z.\mathbf{see'}(x, z))))$

Constraint (8) thus encompasses five readings altogether, as desired. But (7) is only a simple case of nested quantification, e.g., the more complex (13) has no less than 42 readings (Hobbs and Shieber 1987). Appropriate underspecification formalisms must be able to handle nested quantification in general.

(13) Some representative of every department in most companies saw a few samples of each product

Nested quantification highlights the two main characteristics of this approach to semantic underspecification: Underspecified expressions *describe* a set of semantic representations and at the same time *delimit* and *fully specify* the range of this set. The derivation of solutions from such expressions does thus not add information in that it restricts the number of solutions in any way.

2.2 Deriving ambiguity

Other approaches to semantic underspecification describe sets of semantic representations in two steps. First, there is an *initial description* of these sets, e.g., (15) for (14) [= (1)] in Schubert and Pelletier (1982). They render the semantics of DPs as *terms*, scope-bearing expressions with a not yet determined scope. Terms are triples of a quantifier, a bound variable, and a restriction:

(14) Every woman loves a man

(15) $\mathbf{love}'(\langle \text{forall} \;\; x \;\; \mathbf{woman}'(x)\rangle, \langle \text{exists} \;\; y \;\; \mathbf{man}'(y)\rangle)$

To derive a set of fully specified representations from such a description, a *resolution algorithm* integrates terms into descriptions by 'discharging' them (i.e., applying them to suitable parts of the description and thereby determining their scope). E.g., to obtain (2a) for the reading 'for every woman her own man' of (14), the existential term is integrated first: The term is replaced by the bound variable and the quantifier with the term's bound variable and restriction is prefixed to the resulting expression, which yields (16):

(16) $\exists y(\mathbf{man}'(y) \wedge \mathbf{love}'(\langle \text{forall} \;\; x \;\; \mathbf{woman}'(x)\rangle, y))$

Integrating the universal term then yields (2a); to derive (2b) from (15), one would integrate the universal term first. Such an approach is adopted e.g. in the Core Language Engine version of Moran (1988) and Alshawi (1992).

Hobbs and Shieber (1987) present an algorithm for more complicated cases, in particular, nested quantification. Initial semantic descriptions for nested quantification comprise nested terms, as e.g. in the description (17) for (7):

(17) $\mathbf{see}'(\langle \text{forall} \;\; x \;\; \mathbf{researcher}'(x) \wedge \mathbf{of}'(x, \langle \text{exists} \;\; y \;\; \mathbf{company}'(y)\rangle)\rangle,$
$\langle \text{most} \;\; z \;\; \mathbf{sample}'(z)\rangle)$

Resolution of nested terms requires that the inner quantifier may never be integrated before the outer one. For (7), this rules out the unwanted sixth possible permutation of the quantifiers, which otherwise could have been generated by integrating the terms in the order '$\exists, \mathbf{most}', \forall$'.

Another formalism that belongs to this group of algorithms is Ambiguous Predicate Logic (Jaspars and van Eijck 1996).

In sum, the underspecification formalisms expounded in this subsection give initial underspecified descriptions for ambiguous expressions that do not by themselves delimit the range of intended representations fully, this delimitation is the joint effect of the initial descriptions and the resolution algorithm.

The difference between underspecification formalisms that describe the readings of an ambiguous expression and those that derive these readings is thus not the existence of an algorithm to enumerate the readings, but the question of whether such an algorithm is essential in determining the set of solutions.

2.3 Expressivity

Underspecification formalisms are *expressive* if they can represent not only the set of readings of an ambiguous expression but also any of its subsets (König and Reyle 1999; Ebert 2005). E.g., suppose that in the case of (7) the reading (12b) with the scope ordering $\exists > \mathbf{most'} > \forall$ is ruled out contextually.[2] Underspecification approaches that model scope in terms of partial order between fragments of semantic representations are not expressive in this sense: In these approaches, any constraint that covers the four other readings of (7) also includes reading (12b) (König and Reyle 1999; Ebert 2005).

Expressivity is gradable; approaches that express quantifier scope by lists (e.g., Alshawi 1992) are less expressive than those that use dominance relations, or scope lists together with an explicit ordering of list elements as in Fox and Lappin (2005). The approach of Koller et al. (2008), which uses *Regular Tree Grammars* (Comon et al. 2007) for scope underspecification, is fully expressive.

3 The domains of semantic underspecification

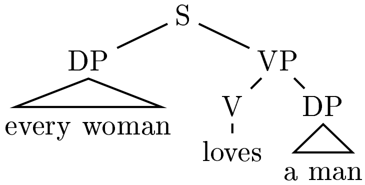
This section offers a classification of ambiguity to identify those kinds of ambiguity that are in the focus of work on underspecification formalisms. Ambiguous expressions will be grouped into four classes according to two criteria. The criteria compare the readings of these expressions from a semantic and a syntactic point of view and are called *semantic* and *syntactic homogeneity*, respectively:

- Do the readings comprise the *same semantic material*?
- Is it possible to give a *single syntactic analysis* for all the readings?

3.1 Semantically and syntactically homogeneous ambiguity

Classic representatives of ambiguous expressions that fulfil the two homogeneity conditions are quantifier scope ambiguities. Reconsider e.g. (14) with the simplistic syntactic analysis (18) and its two readings in (2). In (18) and in (24) below, unary branching nodes are omitted to enhance readability.

(18)



[2]Kallmeyer and Romero (2008) claim that (7) lacks this reading right from the start.

Other scope-bearing items enter into the same kind of ambiguity, e.g., negation and modal expressions, as in (19) and (20).

(19) Everyone didn't come ($\forall > \neg$ or $\neg > \forall$)

(20) A unicorn seems to be in the garden ($\exists >$ *seem* or *seem* $> \exists$)

Scope ambiguity may also occur below the word level. The scope-bearing elements in these cases may but need not correspond to morphemes.

(21) beautiful dancer

(22) John almost died

In (21), the adjective may pertain to the noun as a whole or to the stem only, which yields two readings that can roughly be glossed as 'beautiful person characterised by dancing' and 'person characterised by beautiful dancing', respectively (Larson 1998). This can be modelled as scope ambiguity between the adjective and the nominal affix *-er* (Egg 2004).

The two readings of (22), viz., 'John was close to undergoing a change from being alive to being dead' (i.e., in the end, nothing happened) and 'John underwent a change from being alive to being close to death' (i.e., something did happen) can be modelled as scope ambiguity between a change-of state operator like BECOME in the verb semantics and the adverbial (Dowty 1979).

Most of the work on underspecification focusses on semantically and syntactically homogeneous ambiguity, which is sometimes called *structural ambiguity* in the literature. But this term is itself ambiguous in that it is sometimes used in the broader sense of 'semantically homogeneous' (i.e., syntactically homogeneous or not). But then it would also encompass the group of semantically but not syntactically homogeneous ambiguities discussed in subsection 3.2.

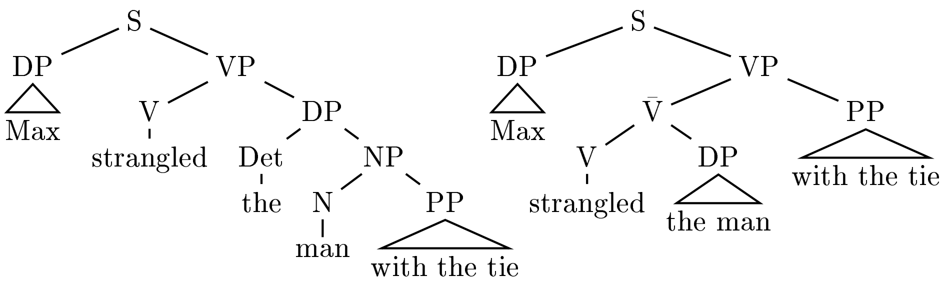
3.2 Semantically but not syntactically homogeneous ambiguity

In semantically but not syntactically homogeneous ambiguities, the same syntactic material is arranged in different ways in the readings of an expression. Consequently, the meanings of the readings all consist of the same semantic material (though differently ordered, depending on the respective syntactic structure), but the readings do not share a common syntactic structure. The notorious modifier attachment ambiguities as in (23) are a prime example of this kind of ambiguity:

(23) Max strangled the man with the tie

The two readings of (23) have different syntactic structures. In the reading that the man is wearing the tie, the constituent *with the tie* is part of the DP *the man with the tie*. In the other reading, in which the tie is the instrument of Max' deed, *with the tie* enters a verbal projection (as the syntactic sister of *strangled the man*) as a constituent of its own. Neither tree would be suitable as the syntactic analysis for *both* readings.

(24) (a) ‘tie worn by victim’ (b) ‘tie as instrument of crime’



Semantically but not syntactically homogeneous ambiguity is typically not accounted for in terms of semantic underspecification in the same fashion as semantically and syntactically homogeneous ambiguity (exceptions include Muskens 2001 and Richter and Sailer 1996).

3.3 Syntactically but not semantically homogeneous ambiguity

The third kind of ambiguity is instantiated by expressions whose readings share a single syntactic analysis but do not comprise the same semantic material.

These expressions can be classified in four subgroups. First, there are expressions with *lexically* ambiguous words, whose ambiguity is inherited by the whole expression. E.g., the ambiguity of the noun *school* with readings like ‘building’, ‘institution’, or ‘teaching activity’ makes expressions like (25) ambiguous, too.

(25) Max abhors school

Polysemy but not homonymy belongs to this group, because the different readings of a polysemous item belong to the same lexeme (no syntactic differences), whereas different readings in the case of homonymy are based on different lexemes, i.e., homonymy is a case of syntactic heterogeneity.

The range of readings of a polysemous lexeme can be captured in terms of an underspecified core meaning common to all readings. This is worked out in the so-called *two-level* semantics (Bierwisch 1983; Bierwisch and Lang 1987; Bierwisch 1988), which distinguishes a level of semantics (where the core meanings reside) and a conceptual level (where these meanigs are specified into the individual readings). E.g., the core meaning of *school* is the property of being related to processes of teaching and learning; this meaning is specified on the conceptual level by operators that enrich the core meaning with properties like ‘building’ or ‘institution’.

Cases of *reinterpretation* (metonymy and aspectual coercion) belong to this group (Hobbs et al. 1993, Dölling 1995; Pulman 1997; de Swart 1998; Egg 2005), if they are modelled in terms of underspecified operators that are inserted during semantic construction to avoid impending clashes between otherwise incompatible semantic material. E.g., in (26) a coercion operator is inserted between *play the Moonlight Sonata* and its modifier *for some time*, which cannot be combined directly:[3]

(26) Amélie played the Moonlight Sonata for some time

Different specifications of a coercion operator are possible (for (26), to a progressive or an iterative operator, i.e., she played a part of the sonata or played it repetitively), which lead to different readings of (26). But then the readings of such expressions no longer comprise the same semantic material.

Minor cases of syntactically but not semantically homogeneous ambiguities emerge by *referential ambiguity* (not yet fixed reference of deictic expressions; cf. Asher and Lascarides 2003 and Poesio et al. 2006) and *missing information* (parts of a message could not be decoded due to problems in production, transmission, or reception; Pinkal 1999).

Most underspecification formalisms do not address syntactically but not semantically homogeneous ambiguities. Those that do focus on polysemy, among them the semantic representation language in the PHLIQA question-answering system (Bronnenberg et al. 1979), Poesio's (1996) Lexically Underspecified Language, and Cimiano and Reyle's (2005) extension of Muskens's (2001) Logical Description Grammar.

3.4 Neither syntactically nor semantically homogeneous ambiguity

Finally, homonyms are ambiguous expressions that are neither syntactically nor semantically homogeneous. They are generally ignored in underspecification formalisms, because there is not enough common ground between the readings. Consequently, underspecified semantic representations of homonyms would be too general to be distinctive (different from the underspecified representations of other homonyms). E.g., a semantic representation for *plant* that includes the readings 'organism' and 'factory' could only be 'concrete object' and would thus not be different from one for *temple* (including the readings 'flattened region on the sides of the head' and 'building for divine worship').

4 Motivation

Motivations for semantic underspecification formalisms range from more theoretical considerations like the relation between syntactic and semantic structures to more practical issues of Natural Language Processing.

4.1 Functionality of the syntax-semantics interface

With semantic underspecification formalisms the syntax-semantics interface (the mapping from syntax to semantics) can be kept *functional* (see Westerståhl 1998 and Hodges 2001) despite semantically and syntactically homogeneous ambiguities like (14). Even for these expressions the functionality of

[3]The VP is a *bounded* expression (it denotes states of affairs with inherent boundaries), but the adverbial selects for unbounded expressions where there are no such boundaries.

semantic interpretation is preserved by mapping their single syntactic structure onto an underspecified semantic structure that encompasses all their readings.

Short of relinquishing the functionality of the syntax-semantics interface altogether and relating one syntactic structure with several semantic ones (like e.g. Cooper 1983 or Steedman 2007), the only other alternative would be to multiply syntactic structures for semantically and syntactically homogeneous ambiguities (one for each reading): Each reading corresponds to a unique syntactic structure (on a semantically relevant syntactic level). In other words, these ambiguities are modelled as syntactically *heterogeneous*.

This is the strategy implemented in particular in *Generative Grammar*. It postulates a syntactic level of *Logical Form* (LF) that hosts syntactic structures unique to specific readings. For (14), (27) and (28) are grossly simplified LF-structures for its readings that are merely meant to illustrate the general point:[4]

(27) $[_S$ [a man]$_j$ $[_S$ [every woman]$_i$ $[_S\ t_i\ [_{VP}$ loves $t_j]]]]$ $(\exists > \forall)$

(28) $[_S$ [every woman]$_i$ $[_S$ [a man]$_j$ $[_S\ t_i\ [_{VP}$ loves $t_j]]]]$ $(\forall > \exists)$

Here the DPs are moved (leaving behind a coindexed trace) and adjoined to an S node, starting with the subject DP in (27) and the object DP in (28). The semantic scope between the quantifiers is then put down to the relative syntactic position of the DPs in the syntactic trees (the relation of *c-command*; see Heim and Kratzer 1998). For (27) and (28), the respective leftmost DP, which is adjoined last, gets widest scope, which yields the two readings of (14).

4.2 Ambiguity and negation

Semantic underspecification offers a more adequate account of negated ambiguous expressions than disjunctive representations of ambiguity. A negated ambiguous expression is interpreted as the denial of *one* of its readings (a disjunction of the negated readings), e.g., in (29), the negation of an example of Bierwisch, as a denial of the acoustic or of the intellectual reading:

(29) Faulkner is not hard to understand

But a disjunctive representation of ambiguity would have to model negated ambiguous expressions in terms of negating the disjunction of the readings (or the conjunction of the negated readings), e.g., for (29), as a denial of *all* readings. For (14), a disjunctive representation can be abbreviated as (30), which turns into (31) after negation:

(30) $\forall\exists \vee \exists\forall$

(31) $\neg(\forall\exists \vee \exists\forall) = \neg\forall\exists \wedge \neg\exists\forall$

[4]See e.g. Hornstein (1995) for a much more sophisticated and independently motivated version of this approach.

To avoid this problem, the meaning of an ambiguous expression can be modelled as the set of its fully specified readings, assuming that its assertion is interpreted as the disjunction of this set. Then, e.g., the meaning of (14) is $\{\forall\exists, \exists\forall\}$; its assertion is interpreted as (30), its negation, as the disjunction of its readings $\{\neg\forall\exists, \neg\exists\forall\}$ (van Eijck and Pinkal 1996). To define such sets of readings efficiently, underspecification is called for.

4.3 Underspecification in Natural Langugage Processing

Semantic underspecification is widely endorsed in Natural Language Processing (NLP) as an answer to the problem of *combinatorial explosion* (Poesio 1996; Ebert 2005): In many cases, there are too many readings of an ambiguous expression to be generated and enumerated, let alone to be handled efficiently in successive modules of an NLP system (e.g., for Machine Translation).

These high numbers are especially due to *spurious ambiguities* that come in during the analysis of the expressions, but are also due to the high frequency of scope-bearing constituents per expression (apart from DPs, also negation, modal verbs, quantifying adverbials like *twice*, etc.). E.g., Koller et al. (2008) report a median number of 56 scope readings per sentence in the Rondane Treebank (Copestake and Flickinger 2000), most of which are spurious. Record holder is (32) with its about 4.5×10^{12} scope readings:

(32) Myrdal is the mountain terminus of the Flåm rail line (or Flåmsbana) which makes its way down the lovely Flåm Valley (Flåmsdalen) to its sea-level terminus at Flåm.

Deriving an underspecified representation of these readings and expanding it to a full representation of one of the readings only by need is less costly than generating all possible interpretations and then selecting the relevant one or processing all not yet excluded readings in parallel. Often a complete disambiguation is not even necessary. E.g., many scope ambiguities are irrelevant for translation, which was the reason why the Verbmobil project (machine translation of spontaneous spoken dialogue) used a scopally underspecified semantic representation (Schiehlen 2000).

Semantic underspecification is furthermore used in NLP for *hybrid processing* as an *interface* between deep and shallow processing. Hybrid NLP applications combine the results of deep and shallow processing on the semantic level, e.g., in the 'Heart of Gold' architecture developed in the project 'DeepThought' (Callmeier et al. 2004). Underspecification is particularly apt to model the partial semantic representation to be gained on the basis of a shallow syntactic analysis (e.g., by a part-of-speech tagger or an NP chunker): Such analyses ignore part of the syntactic structure of expressions, therefore they are bound to miss the semantic information that could be derived from this syntactic information.

Finally, semantic underspecification is useful for *semantic construction* in general, as the construction of underspecified semantic representations is pretty independent of the syntactic structures on which it is based. This makes semantic underspecification formalisms very *portable*, which is shown by the wide range of possible couplings of (surface-oriented) syntactic analyses with underspecification formalisms. E.g., Head-Driven Phrase Structure Grammar (Pollard and Sag 1994) can be coupled to Minimal Recursion Semantics (Copestake et al. 2005), Glue Language Semantics (Asudeh and Crouch 2002), Underspecified DRT (Frank and Reyle 1995), and Hole Semantics (Chaves 2002).

This independence can also be exploited to derive semantic representations for unambiguous expressions whose (surface) syntactic and semantic structures do not match in an obvious way and which are hence considered problems for semantic construction. Such expressions include cases of negative concord (Richter and Sailer 2006) and the modification of indefinite modifiers like *something* (semantically, the set of properties such that some entity has them). Their modifiers seem to pertain exclusively to the restriction of the quantification in their semantics, e.g., *something blue* is semantically the set of properties such that *some blue entity* has them.

Egg (2004, 2006) exploits the independence of underspecification formalisms from underlying syntactic structures for the semantic construction of modified indefinite pronouns and other problematic cases of semantic construction. Simply treating 'thing' as a noun that happens not to be separated from its determiner 'some' but is modifiable just like any other noun would not do even for syntax alone because of the postnominal AP position in cases like *something blue* (as opposed to *some blue thing*). To get this ordering and semantic construction right, Abney (1987) assumes an enclitic (but modifiable) noun *thing*, which eventually attaches to *some* by head-to-head movement; semantic interpretation precedes this movement. The challenge for interfaces in surface-oriented syntactic analyses is to get the interpretation right without assuming such a movement, and this is where underspecification proves helpful.

5 Processing underspecified semantic representations

Underspecified semantic representations can be further processed in order to derive fully specified (or at least less ambiguous) semantic representations.

First, one can enumerate the readings by *resolving* the constraints with the help of so-called solvers. Such solvers are available e.g. for MRS representations (Copestake and Flickinger 2000) and the language of dominance constraints (Koller and Thater 2005; Koller et al. 2008). Related to the enumeration of solutions is work on *redundancy elimination*, which weeds out spurious ambiguities either during the resolution process (Moran 1988; Alshawi 1992) or directly on the underspecified representations (Koller et al. 2008).

Some underspecified semantic representations allow the *deduction* of fully specified information. E.g., if Amélie is a woman, then it follows from (14) that she loves a man, no matter which reading of (14) is at stake (Reyle 1993, 1996; Jaspars and van Eijck 1996).

Finally, specific readings can be chosen (or the number of potential readings be reduced) if one strengthens underspecified representations by *preferences* for specific kinds of readings.

The first group of preferences are syntax-based, in particular, those derived from surface linear order (Johnson-Laird 1969; Lakoff 1971) or c-command (e.g., VanLehn 1978). Consider e.g. (33a), which has a clear preference for the $\exists\forall$-reading over the $\forall\exists$-reading, to (33b) [= (14)], where the $\forall\exists$-reading is preferred:

(33) (a) A woman loves every man

(b) Every woman loves a man

However, claims that preceding or c-commanding constituents get wider scope are not universally valid, in particular not for nested quantification like in (7) (Kurtzman and MacDonald 1993).

Other preferences are based on grammatical functions and thematic roles. E.g., VanLehn (1978) suggests a scope preference hierarchy stretching from topic (strongest preference for wide scope) over subject and PP complement down to object. Syntactically and functionally based hierarchies overlap to a certain extent (at least in English), because DPs higher on the functional hierarchy also tend to c-command DPs lower on the hierarchy.

The determiners themselves have also been suggested as indications for scope preferences, with the hierarchy ranging from *each* and *every* (strongest preference for wide scope) down to *each* and *a few*, e.g., by Ioup (1975) or VanLehn (1978). CLE incorporates some of these preferences (Moran 1988; Alshawi 1992).[5]

Many researchers argue that the whole range of quantifier scope effects can only be accounted for in terms of an interaction of different principles (Ioup 1975; Fodor 1982; Hurum 1988; Pafel 2005). Kurtzman and MacDonald (1993) adduce the contrast between a sentence like (33b) and its passive version (34) as evidence for such an interaction. The passive variant no longer shows a preference for the $\forall\exists$-reading:

(34) A man is loved by every woman

If a single principle determined preferences, the passive version should exhibit a preference, too, either for its (new) subject, or for the *by*-PP (the former demoted subject, which keeps its thematic role). But as soon as one assumes an interaction of syntax-oriented principles with the thematic role principle, the results are to be expected: Most subjects have higher thematic roles, hence, the principles introduce the same scope preference for the subject in the active sentence. In contrast, the principles yield conflicting preferences for the passive sentence, where the new subjects typically do not carry the highest thematic role in the sentence. Passive sentences like (34) therefore no longer exhibit a scope preference (Kurtzman and MacDonald 1993).

References

Abney, S. (1987). *The English noun phrase in its sentential aspect.* Ph. D. thesis, MIT.

[5]See Pafel (2005) for further kinds of preferences.

Alshawi, H. (ed.) (1992). *The Core Language Engine.* Cambridge: MIT Press.

Alshawi, H. and R. Crouch (1992). Monotonic semantic interpretation. In *Proceedings of the 30th ACL*, 32–39.

Althaus, E., D. Duchier, A. Koller, K. Mehlhorn, J. Niehren, and S. Thiel (2001). An efficient algorithm for the configuration problem of dominance graphs. In *Proceedings of the 12th ACM-SIAM Symposium on Discrete Algorithms*, Washington, 815–824.

Asher, N. and A. Lascarides (2003). *Logics of conversation.* Cambridge: Cambridge University Press.

Asudeh, A. and R. Crouch (2002). Glue Semantics for HPSG. In F. van Eynde, L. Hellan, and D. Beermann (eds), *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, 1–19. Stanford: CSLI Publications.

Bierwisch, M. (1983). Semantische und konzeptionelle Repräsentation lexikalischer Einheiten. In R. Růžička and W. Motsch (eds), *Untersuchungen zur Semantik*, 61–99. Berlin: Akademie-Verlag.

Bierwisch, M. (1988). On the grammar of local prepositions. In M. Bierwisch, W. Motsch, and I. Zimmermann (eds), *Syntax, Semantik und das Lexikon*, 1–65. Berlin: Akademie-Verlag.

Bierwisch, M. and E. Lang (eds) (1987). *Grammatische und konzeptuelle Aspekte von Dimensionsadjektiven*, Volume 26/27 of *studia grammatica.* Berlin: Akademie-Verlag.

Bos, J. (1996). Predicate logic unplugged. In P. Dekker and M. Stokhof (eds), *Proceedings of the 10th Amsterdam Colloquium*, 133–142. ILLC.

Bos, J. (2004). Computational semantics in discourse: Underspecification, resolution, and inference. *Journal of Logic, Language and Information 13*, 139–157.

Bronnenberg, W., H. Bunt, J. Landsbergen, R. Scha, W. Schoenmakers, and E. van Utteren (1979). The question answering system PHLIQA1. In L. Bolc (ed.), *Natural language question answering systems*, 217–305. London: McMillan.

Callmeier, U., A. Eisele, U. Schäfer, and M. Siegel (2004). The DeepThought core architecture framework. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, and R. Silva (eds), *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC) 2004*, Lisbon, 1205–1208. ELRA: European Language Resources Association.

Chaves, R. (2002). Principle-based DRTU for HPSG: a case study. In *First International Workshop on Scalable Natural Language Understanding*, EML, Heidelberg.

Cimiano, P. and U. Reyle (2005). Talking about trees, scope and concepts. In *Proceedings of the 6th International Workshop in Computational Semantics*, Tilburg, 90–102.

Comon, H., M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi (2007). Tree automata techniques and applications. Available on: `http://www.grappa.univ-lille3.fr/tata`.

Cooper, R. (1983). *Quantification and syntactic theory*. Dordrecht: Reidel.

Copestake, A. and D. Flickinger (2000). An open-source grammar development environment and broad-coverage English grammar using HPSG. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation*, Athens, 591–600.

Copestake, A., D. Flickinger, C. Pollard, and I. Sag (2005). Minimal Recursion Semantics. An introduction. *Research on Language and Computation 3*, 281–332.

Crouch, R. and J. van Genabith (1999). Context change, underspecification and the structure of Glue Language derivations. In M. Dalrymple (ed.), *Semantics and Syntax in Lexical Functional Grammar: The Resource Logic Approach*, 117–189. Cambridge: MIT Press.

Dalrymple, M. (2001). *Lexical Functional Grammar*. Number 34 in Syntax and Semantics. New York: Academic Press.

Dalrymple, M., J. Lamping, F. Pereira, and V. Saraswat (1997). Quantifiers, anaphora, and intensionality. *Journal of Logic, Language, and Information 6*, 219–273.

Dölling, J. (1995). Ontological domains, semantic sorts and systematic ambiguity. *International Journal of Human-Computer Studies 43*, 785–807.

Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel.

Duchier, D. and C. Gardent (2001). Tree descriptions, constraints and incrementality. In H. Bunt, R. Muskens, and E. Thijsse (eds), *Computing Meaning, Volume 2*, 205–227. Dordrecht: Kluwer.

Ebert, C. (2005). *Formal investigations of underspecified representations*. Ph. D. thesis, King's College, London.

Egg, M. (2004). Mismatches at the syntax-semantics interface. In S. Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*, 119–139. Stanford: CSLI Publications.

Egg, M. (2005). *Flexible semantics for reinterpretation phenomena*. Stanford: CSLI Publications.

Egg, M. (2006). Anti-Ikonizität an der Syntax-Semantik-Schnittstelle. *Zeitschrift für Sprachwissenschaft 25*, 1–38.

Egg, M. and M. Kohlhase (1997). Dynamic control of quantifier scope. In P. Dekker, M. Stokhof, and Y. Venema (eds), *Proceedings of the 11th Amsterdam Colloquium*, 109–114. ILLC.

Egg, M., A. Koller, and J. Niehren (2001). The Constraint Language for Lambda-Structures. *Journal of Logic, Language, and Information 10*, 457–485.

Egg, M. and G. Redeker (2008). Underspecified discourse representation. In A. Benz and P. Kühnlein (eds), *Constraints in Discourse*, 117–138. Amsterdam: Benjamins.

van Eijck, J. and M. Pinkal (1996). What do underspecified expressions mean? In R. Cooper et al. (eds), *Building the framework*, FraCaS Deliverable D 15, section 10.1, 264–266.

Fodor, J. (1982). The mental representation of quantifiers. In S. Peters and E. Saarinen (eds), *Processes, beliefs and questions*, 129–164. Dordrecht: Reidel.

Fox, C. and S. Lappin (2005). Underspecified interpretations in a Curry-typed representation language. *Journal of Logic and Computation 15*, 131–143.

Frank, A. and U. Reyle (1995). Principle-based semantics for HPSG. In *Proceedings of the 7th conference on European chapter of the Association for Computational Linguistics*, Dublin, 9–16.

Harris, J. (2007). Representation. In P. de Lacey (ed.), *The Cambridge handbook of phonology*, 119–137. Cambridge: Cambridge University Press.

Heim, I. and A. Kratzer (1998). *Semantics in Generative Grammar.* Malden: Blackwell.

Hobbs, J. and S. Shieber (1987). An algorithm for generating quantifier scoping. *Computational Linguistics 13*, 47–63.

Hobbs, J., M. Stickel, D. Appelt, and P. Martin (1993). Interpretation as abduction. *Artificial Intelligence 63*, 69–142.

Hodges, W. (2001). Formal features of compositionality. *Journal of Logic, Language, and Information 10*, 7–28.

Hornstein, N. (1995). *Logical Form. From GB to Minimalism.* Cambridge: Blackwell.

Hurum, S. (1988). Handling scope ambiguities in English. In *Proceedings of the second conference on Applied natural language processing*, Morristown, 58–65. Association for Computational Linguistics.

Ioup, G. (1975). Some universals for quantifier scope. In J. Kimball (ed.), *Syntax and Semantics*, Volume 4, 37–58. New York: Academic Press.

Jaspars, J. and J. van Eijck (1996). Underspecification and reasoning. In R. Cooper et al. (eds), *Building the framework*, FraCaS Deliverable D 15, section 10.2, 266–287.

Johnson-Laird, P. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology 21*, 1–13.

Kallmeyer, L. and M. Romero (2008). Scope and situation binding in LTAG using semantic unification. *Research on language and computation 6*, 3–52.

Koller, A., M. Regneri, and S. Thater (2008). Regular tree grammars as a formalism for scope underspecification. In *Proceedings of ACL-08: HLT*, Columbus, 218–226. Association for Computational Linguistics.

Koller, A. and S. Thater (2005). The evolution of dominance constraint solvers. In *Proceedings of the ACL-05 Workshop on Software*, Ann Arbor.

König, E. and U. Reyle (1999). A general reasoning scheme for underspecified representations. In H. J. Ohlbach and U. Reyle (eds), *Logic, Language and Reasoning. Essays in Honour of Dov Gabbay*, 1–28. Dordrecht: Kluwer.

Kurtzman, H. and M. MacDonald (1993). Resolution of quantifier scope ambiguities. *Cognition 48*, 243–279.

Lakoff, G. (1971). On generative semantics. In D. Steinberg and L. Jakobovits (eds), *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, 232–296. Cambridge: Cambridge University Press.

Larson, R. (1998). Events and modification in nominals. In D. Strolovitch and A. Lawson (eds), *Proceedings from SALT VIII*, Ithaca, 145–168. CLC Publications.

Marcus, M., D. Hindle, and M. Fleck (1983). D-theory: talking about talking about trees. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA, 129–136. Association for Computational Linguistics.

Moran, D. (1988). Quantifier scoping in the SRI core language engine. In *Proceedings of the 26th ACL*, Buffalo, 33–40.

Muskens, R. (2001). Talking about trees and truth-conditions. *Journal of Logic, Language, and Information 10*, 417–455.

Niehren, J., M. Pinkal, and P. Ruhrberg (1997). A uniform approach to underspecification and parallelism. In *Proceedings of ACL'97*, Madrid, Spain, 410–417.

Pafel, J. (2005). *Quantifier scope in German.* Amsterdam: Benjamins.

Pinkal, M. (1996). Radical underspecification. In P. Dekker and M. Stokhof (eds), *Proceedings of the 10th Amsterdam Colloquium*, Amsterdam, 587–606. ILLC.

Pinkal, M. (1999). On semantic underspecification. In H. Bunt and R. Muskens (eds), *Computing Meaning, Vol. 1*, 33–55. Dordrecht: Kluwer.

Poesio, M. (1996). Semantic ambiguity and perceived ambiguity. In K. van Deemter and S. Peters (eds), *Ambiguity and underspecification*, 159–201. CSLI Publications.

Poesio, M., P. Sturt, R. Artstein, and R. Filik (2006). Underspecification and anaphora: Theoretical issues and preliminary evidence. *Discourse Processes 42*, 157–175.

Pollard, C. and I. Sag (1994). *Head-driven Phrase Structure Grammar.* CSLI and University of Chicago Press.

Pulman, S. (1997). Aspectual shift as type coercion. *Transactions of the Philological Society 95*, 279–317.

Rambow, O., D. Weir, and V. Shanker (2001). D-tree substitution grammars. *Computational Linguistics 27*, 89–121.

Regneri, M., M. Egg, and A. Koller (2008). Efficient processing of underspecified discourse representations. In *Proceedings of ACL-08: HLT, Short Papers*, Columbus, 245–248. Association for Computational Linguistics.

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

Reyle, U. (1996). Co-indexing labeled DRSs to represent and reason with ambiguities. In K. Deemter and S. Peters (eds), *Semantic ambiguity and underspecification*, 239–268. Stanford: CSLI.

Richter, F. and M. Sailer (1996). Syntax für eine unterspezifizierte Semantik: PP-Anbindung in einem deutschen HPSG-Fragment. In S. Mehl, A. Mertens, and M. Schulz (eds), *Präpositionalsemantik und PP-Anbindung*, 39–47. Gerhard-Mercator-Universität Duisburg.

Richter, F. and M. Sailer (2006). Modeling typological markedness in semantics: the case of negative concord. In S. Müller (ed.), *Proceedings of the HPSG 06 Conference*, Stanford, 305–325. CSLI Publications.

Schiehlen, M. (2000). Semantic construction. In W. Wahlster (ed.), *Verbmobil: Foundations of speech-to-speech translation*, 200–216. Berlin: Springer.

Schilder, F. (2002). Robust discourse parsing via discourse markers, topicality and position. *Natural Language Engineering 8*, 235–255.

Schubert, L. and F. Pelletier (1982). From English to logic: Context-free computation of 'conventional' logical translation. *American Journal of Computational Linguistics 8*, 26–44.

Steedman, M. (2007). Surface-compositional scope-alternation without existential quantifiers. Draft 6.0, August 2007. Available from `http://www.iccs.informatics.ed.ac.uk/~steedman/papers.html`, last checked 25 August 2009.

Steriade, D. (1995). Underspecification and markedness. In J. Goldsmith (ed.), *The handbook of phonological theory*, 114–174. Cambridge: Blackwell.

de Swart, H. (1998). Aspect shift and coercion. *Natural Language and Linguistic Theory 16*, 347–385.

VanLehn, K. (1978). Determining the scope of English quantifiers. Master's thesis, MIT, Cambridge, MA. MIT Technical Report AI-TR-483.

Westerståhl, D. (1998). On mathematical proofs of the vacuity of compositionality. *Linguistics & Philosophy 21*, 635–643.