# NP Coordination in Underspecified Scope Representations

Alistair Willis
The Open University
Milton Keynes, UK.
`A.G.Willis@open.ac.uk`

### Abstract

Accurately capturing the quantifier scope behaviour of coordinated NPs can be problematic for some underspecified semantic representations. We present an extension to hole semantics that allows a natural representation of sentences containing NP coordination, and which correctly captures the quantifiers' scoping behaviour. We demonstrate that existing efficient algorithms developed to solve constraints on semantic construction also apply to the proposed extension. This allows NP coordination to be represented in practical systems for semantic underspecification.

## 1 Introduction

A common trend in computational semantics is to represent the meaning of ambiguous sentences with an underspecified representation. Ambiguous sentences often have too many possible meanings to be enumerated [13] and so underspecified representations are used, which provide a compact representation of the sentence's possible meanings [1, 9]. In general, an underspecified representation will also allow *partial* information about the sentence's meaning to be represented, so that contextual information can be incrementally used to determine the sentence's final meaning.

Many underspecified representations that deal with quantifier scope ambiguity represent meaning by defining constraints over possible compositions of semantic material, such as logical formulas or discourse representation structures [7, 14]. Hole semantics [3] and the constraint language for lambda structures (CLLS) [6] have both been developed as a means of representing and reasoning about the possible ways to compose the semantic parts into

1

the sentence's final meaning. In particular, dominance constraints [2] have been demonstrated to apply to a variety semantic formalisms [8, 11].

However, it can be difficult for representations based on semantic construction to correctly capture the behaviour of coordinated NPs. Consider sentence (1).

(1) Every student read a book and a paper.

Although the logical form of (1) contains three quantifiers, there appear to be only two readings of the sentence due to scope ambiguity. The two existential quantifiers take scope as a single NP, not independently of each other. So the reading in which the universal quantifier takes wide scope is:

(2) $\forall x.student'(x) \rightarrow (\exists y.book'(y) \wedge read'(x,y) \wedge \exists z.paper'(z) \wedge read'(x,y))$

and the reading in which the universal quantifier takes narrow scope is:

(3) $\exists y.book'(y) \wedge \forall x.student'(x) \rightarrow read'(x,y)$
    $\wedge \exists y.paper'(y) \wedge \forall x.student'(x) \rightarrow read'(x,y)$

These are the only two possible readings (due to quantifier scope ambiguity). This coordination behaviour, where coordinated NPs are not able to move out of the conjuction that they appear in, is known as the Coordinate Structure Constraint [15]. It is difficult to capture using representations such as dominance constraints, because the quantifiers, generally represented by different objects, must move together.

Also, notice in the logical forms (2) and (3), there are two occurrences of the predicate $read'$. Correctly dealing with this phenomenon can be difficult for mechanisms based on the nondeterministic composition of formulas, because it is not obvious from the initial sentence that multiple instances of the formula are required; the word *read* occurs only once in (1).

In this paper, we propose a meaning representation based on hole semantics that accounts for behaviour of coordinated NPs in sentences demonstrating quantifier scope ambiguity. We represent sentences with a *set* of underspecified forms, and then provide a mechanism for recomposing a final meaning from this set. Our proposal correctly accounts for the Coordinate Structure Constraint and multiple occurrences of formulas in the final logical form, as well as capturing the behaviour of quantifiers which arises from nested NPs.
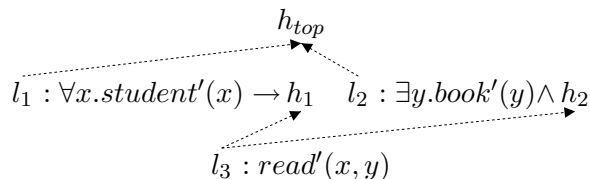
2

## 2    Hole Semantics for Quantifier Scope Ambiguity

For this paper, we use a form of hole semantics for an underspecified representation of scope ambiguity. We do not give formal definitions of the standard hole semantics here, for which Bos [3] should be consulted. In particular, we do not give a formal definition of the properties that a representation must have to lead to a well formed sentence meaning. However, we illustrate enough of the behaviour to motivate and justify the extensions proposed in section 3.

Hole semantics provides a mechanism for underspecified meaning representations by providing *holes* in formulas into which further formulas can be nondeterministically plugged (Bos [3] demonstrates how this can be applied to first order logic and (U)DRT). Different pluggings lead to different sentence meanings. For example, if first order logic were the language chosen to represent sentence meanings, then the holes appear in logical formulas where other formulas may appear. To illustrate this, consider sentence (4).

(4)  Every student read a book.

An underspecified meaning representation for (4), showing the different formulas and the possible ways of combining them, is illustrated in (5).

(5)

$$h_{top}$$

$$l_1 : \forall x.student'(x) \rightarrow h_1 \quad l_2 : \exists y.book'(y) \wedge h_2$$

$$l_3 : read'(x,y)$$

The underspecified representation (5) contains several formulas, each of which has a *label* taken from a set $\{l_1, l_2, \ldots\}$. The formulas themselves contain holes taken from a set $\{h_1, h_2, \ldots\}$, which may appear in a formula at any point where a further formula can be inserted. The holes are variables ranging over labelled formulas, and to construct the final logical form, each hole can be *plugged* by exactly one of the other formulas. A plugging is represented by an expression $h_i = l_j$ which represents the hole $h_i$ being plugged by the formula labelled $l_j$. For example, in (5), the equality $h_1 = l_3$ would represent the formula $read'(x,y)$ replacing the hole $h_1$ in the formula labelled $l_1$, to give $l_1 : \forall x.student'(x) \rightarrow read'(x,y)$. Note that the topmost

node is an unlabelled hole, $h_{top}$. The formula plugged into $h_{top}$ represents the final meaning of the sentence.

Restrictions upon which formulas can be plugged into which holes are expressed with *subordination constraints*, represented by the dotted lines in (5). A subordination constraint $l_j \leq h_i$ states that the hole $h_i$ is plugged either by the formula labelled $l_j$, or by a formula which itself contains $l_j$. So in (5), the dotted line from $l_3$ to $h_1$ states that $l_3$ is either used to plug $h_1$, or that $h_1$ must be plugged with a formula into which $l_3$ is plugged. The pluggings and associated logical forms (6) and (7) are the (only) two pluggings for (5) which respect the given constraints.

(6) a. $\forall x.student'(x) \rightarrow \exists y.book'(y) \wedge read'(x, y)$

b. $\{h_{top} = l_1, h_1 = l_2, h_2 = l_3\}$

(7) a. $\exists y.book'(y) \wedge \forall x.student'(x) \rightarrow read'(x, y)$

b. $\{h_{top} = l_2, h_1 = l_3, h_2 = l_1\}$

Subordination is a transitive property which can also used to represent partial scope between quantifiers. If (5) also contained the constraint $l_2 \leq h_1$, then only meaning (6) would be possible, in which the universal quantifier outscopes the existential quantifier.

Throughout the rest of this paper, we describe an *Unplugged Logical Form* (ULF) as a triple $\langle H, F, C \rangle$, where $H$ is a set of holes, $F$ is a set of labelled formulas, and $C$ is a set of subordination constraints of the form $l_j \leq h_i$. Bos [3] details the formal conditions under which such a ULF is well-formed ("*proper*") and consistent with a set of subordination constraints. We do not repeat the formal definitions here, except to note the important result that for a proper ULF, there exists a single hole that subordinates all the other holes and labels in that ULF. We call this the $TOP$ of the ULF. So for example, the ULF that describes the possible meanings of the sentence (4) is given in (8).

(8)

$$
\left\langle
\left\{
\begin{array}{l}
h_{top} \\
h_1 \\
h_2
\end{array}
\right\}
\left\{
\begin{array}{l}
l_1 : \forall x.student'(x) \rightarrow h_1 \\
l_2 : \exists y.book'(y) \wedge h_2 \\
l_3 : read'(x, y)
\end{array}
\right\},
\left\{
\begin{array}{l}
l_1 \leq h_{top} \\
l_2 \leq h_{top} \\
l_3 \leq h_1 \\
l_3 \leq h_2
\end{array}
\right\}
\right\rangle
$$

4

This is the underspecified representation illustrated in (5). The $TOP$ of this ULF is $h_{top}$.

A *possible plugging* for a ULF $\langle H, F, C \rangle$ is a bijective mapping from holes in $H$ onto formulas in $F$ which satisfies all the subordination constraints. The meaning of the sentence is the value of the hole $h_{top}$ in the solution.

We also follow Ebert [5] and Koller *et al* [8] in requiring that all the holes and labels that appear in a ULF must appear exactly once in that ULF (ie. all holes and labels in a ULF are distinct).

## 2.1 Algorithms for Hole Semantics and Underspecification

A practical underspecification system requires efficient algorithms to reason about the available scopes in an underspecified representation. In fact, Koller *et al* [8] have demonstrated that for linguistically valuable fragments of the two languages, a hole semantics representation can be converted into (and then back from) a dominance constraint representation. Bodirsky *et al* [2] have presented a polynomial time algorithm to determine the satisfiability of these classes of dominance constraints, demonstrating that representations based upon hole semantics can be solved efficiently.

In the next section, we demonstrate that sentences containing coordinated NPs can be represented using the structures from standard hole semantics. We conclude that hole semantics can be used for underspecified representations of sentences containing coordinated NPs without additional computational cost.

# 3 Sentences Containing Coordinated NPs

We now consider how to represent the meanings of sentences containing coordinated NPs. Consider again sentence (1).

(1) Every student read a book and a paper.

To capture the scope behaviour of this sentence correctly, we use a technique similar to the coordinated store element used in the Core Language Engine's quasi logical forms [1]. A single hole is used to represent the whole coordinated NP, and then individual structures represent the meanings of the constituent NPs. We introduce three extensions to the ULF language:

1. A sentence is represented with a *set* of ULFs, where coordinated NPs are represented by individual ULFs.

5

2. The set of labels is divided into two disjoint sets, $\{l_0, l_1, \ldots\}$ and $\{c_0, c_1, \ldots\}$. The labels $c_i$ are used to mark where a coordinated NP occurs in a ULF.

3. For a ULF which represents an NP, a special formula, $\mathcal{I}$, is used to mark the point into which another formula can be plugged (for a quantified NP, this generally represents the quantifier's scope).

Note that although the set of labels is split into two disjoint sets, the labels are not treated differently in the meta-language. That is, when reasoning about subordination constraints, each of the $c_i$ and the $l_i$ are just treated as labels in standard hole semantics. The distinction becomes relevant only when composing a final logical form from a ULF.

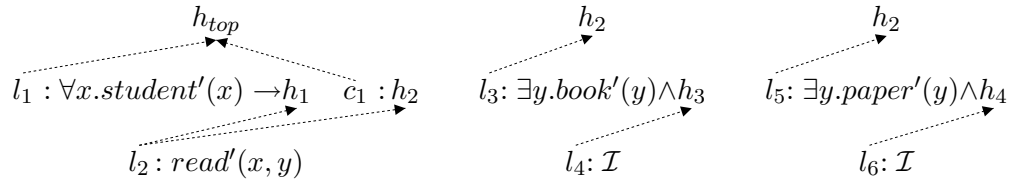## 3.1 Representing and Composing Coordinated NPs

To represent the meanings of the sentence (1), the labelled variable $c_1 : h_2$ is used in the ULF to represent the coordinated NP *a book and a paper*. This gives the ULF (9).

(9)

$$\left\langle \left\{ \begin{array}{c} h_{top} \\ h_1 \\ h_2 \end{array} \right\} \left\{ \begin{array}{l} l_1 : \forall x.student'(x) \to h_1 \\ c_1 : h_2 \\ l_2 : read'(x,y) \end{array} \right\}, \left\{ \begin{array}{l} l_1 \leq h_{top} \\ c_1 \leq h_{top} \\ l_2 \leq h_1 \\ l_2 \leq h_2 \end{array} \right\} \right\rangle$$

To represent the conjoined NPs, each is represented with a distinct ULF, which has $h_2$ as its $TOP$. In general, if the expression $c_i : h_j$ is used to represent coordinated NPs, then each of the conjoined NPs is represented as a ULF, with $h_j$ as its $TOP$, and $\mathcal{I}$ used to represent the point into which further formulas can be inserted (in a similar manner to the parameter which appears in a stored NP in Cooper storage [4]). The three ULFs needed to represent this sentence are illustrated in (10). Distinct ULFs are used to represent the NPs *a book* and *a paper*, and a hole with a coordination label is used to capture the coordinated NP.

(10)

$h_{top}$

$l_1 : \forall x.student'(x) \to h_1$    $c_1 : h_2$

$l_2 : read'(x,y)$

$h_2$

$l_3: \exists y.book'(y) \wedge h_3$

$l_4: \mathcal{I}$

$h_2$

$l_5: \exists y.paper'(y) \wedge h_4$

$l_6: \mathcal{I}$

The two ULFs representing NPs in (10) both have $h_2$ as their $TOP$, indicating that they are the conjuncts represented by the labelled variable $c_1 : h_2$ in the ULF representing the main sentence. Other than $h_2$, all the variables and holes used in the three ULFs are distinct. The meaning of the sentence is still given by the final value of the hole $h_{top}$. We require that all labels and holes in a set of ULFs be distinct throughout the set, unless they are the $TOP$ holes of ULFs used in coordinations (the conditions are stated fully in section 3.3). While the hole $h_2$ is plugged with different formulas across the *set* of ULFs, within each individual ULF, $h_2$ is plugged with only one formula, as is required for a proper ULF.

In general, where a ULF contains a node $c_i : h_j$, the value at that node is the (logical) conjunction of each tree whose topmost node is $h_j$. In those conjoined trees, the placeholder $\mathcal{I}$ is replaced by the value of $h_j$ in the main ULF. To demonstrate this, we consider how to compose the solutions to the ULFs in (10) give the meanings of the sentence (1). Clearly, the ULFs representing the NPs *a book* and *a paper* have only one solution each, but there are two possible solutions to the first ULF. The first solution and subsequent meaning construction is illustrated in figure 1. The left hand side shows the first solution of the ULF (with bold arrows indicating assignment of labels to holes). The node labelled $c_1 : h_2$ is then replaced with the conjunction of the solutions of the remaining two ULFs.
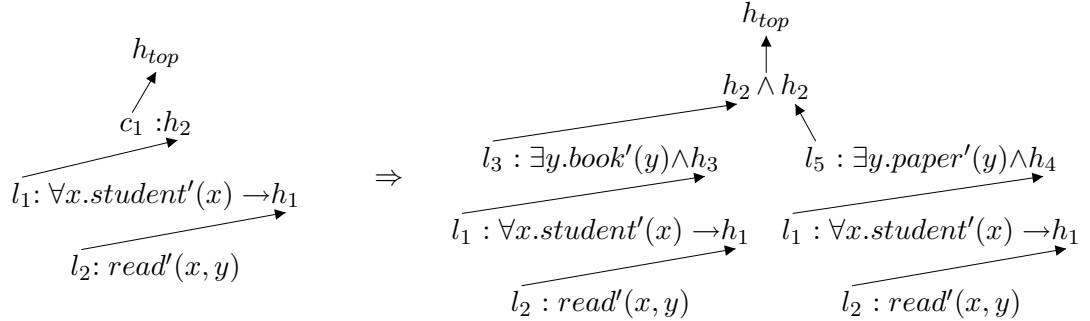


Figure 1: Replacing $c_1 : h_2$ with Coordinated NPs ($l_1 \leq c_1$)

The construction in figure 1 results in (3), the meaning of the sentence where the universal quantifier takes narrow scope.
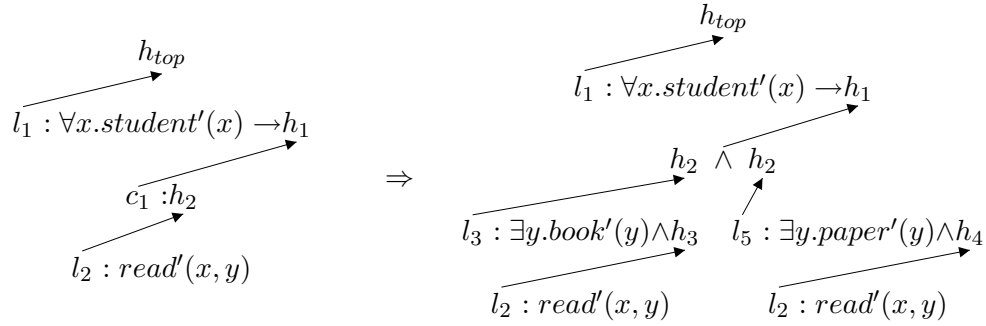
7

Figure 2: Replacing $c_1 : h_2$ with Coordinated NPs ($c_1 \leq l_1$)

(3) $\exists y.book'(y) \wedge \forall x.student'(x) \rightarrow read'(x,y)$
$\wedge \exists y.paper'(y) \wedge \forall x.student'(x) \rightarrow read'(x,y)$

Note that the repeated occurrence of the predicate $read'$ has been captured without requiring any complex preprocessing of the structure, because the same formula is plugged into $\mathcal{I}$ in both the conjoined ULFs.

The construction for the reading where the universal quantifier takes wide scope, (2), is illustrated in figure 2.

(2) $\forall x.student'(x) \rightarrow (\exists y.book'(y) \wedge read'(x,y) \wedge \exists z.paper'(z) \wedge read'(x,y))$

The ULFs have generated the two readings of the sentence required.

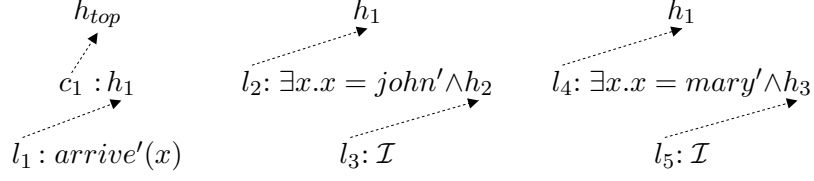How should we treat sentences such as (11), in which the coordinated NPs are proper names?

(11) John and Mary arrived.

This sentence would traditionally be represented in logic as:

(12) $arrive'(john') \wedge arrive'(mary')$

with $john'$ and $mary'$ being constants in the object language denoting individuals in some domain. As illustrated in figures 1 and 2, variables which occur in the repeated formulas are not renamed, and so this method will not generate the logical form (12). We therefore treat proper names by using an existentially quantified variable and equality over constants (a representation also used in the logical forms generated by DRT). By representing (11) with the ULFs illustrated in (13), the logical form (14) is generated.

8

(13)

$$h_{top} \qquad\qquad h_1 \qquad\qquad\qquad h_1$$

$$c_1 : h_1 \qquad l_2\colon \exists x.x = john' \wedge h_2 \qquad l_4\colon \exists x.x = mary' \wedge h_3$$

$$l_1 : arrive'(x) \qquad\qquad l_3\colon \mathcal{I} \qquad\qquad\qquad l_5\colon \mathcal{I}$$

(14) $\exists x.x = john' \wedge arrive'(x) \wedge \exists x.x = mary' \wedge arrive'(x)$
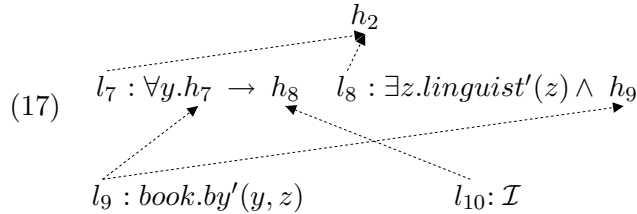
## 3.2 Coordination with NP Nesting

The technique described in section 3.1 extends to NPs which contain further nested NPs, while respecting the Coordinate Structure Constraint. Consider sentence (15)

(15) A student read every book by a linguist and every paper.

As discussed in section 1, the NPs *every book by a linguist* and *every paper* cannot move out of the coordination that they appear in. Because a distinct ULF is used for each of these NPs, it is not possible for the existential quantifier from *a linguist* to take scope over the universal quantifier from *every paper*, as predicted by the constraint.

However, different scopes *within* the nested NP can be captured, without the constituent parts moving out of the coordination. To represent the meanings of the NP *every book by a linguist*, the ULF (16) is used, illustrated as (17). Again, we have used $h_2$ as $TOP$ for this NP, as though we intended to use it as a conjunct in (10).

$$
(16)\ \left\langle \left\{ \begin{array}{c} h_2 \\ h_7 \\ h_8 \\ h_9 \end{array} \right\} \left\{ \begin{array}{l} l_7 : \forall y.h_7 \rightarrow h_8 \\ l_8 : \exists z.linguist'(z) \wedge h_9 \\ l_9 : book.by'(y,z) \\ l_{10} : \mathcal{I} \end{array} \right\}, \left\{ \begin{array}{l} l_7 \leq h_2 \\ l_8 \leq h_2 \\ l_9 \leq h_7 \\ l_{10} \leq h_8 \\ l_9 \leq h_9 \end{array} \right\} \right\rangle
$$

$$h_2$$

(17) $\quad l_7 : \forall y.h_7 \rightarrow h_8 \qquad l_8 : \exists z.linguist'(z) \wedge\ h_9$

$$l_9 : book.by'(y,z) \qquad\qquad l_{10}\colon \mathcal{I}$$

9

The ULF (16) has two solutions, (18) and (19), which can be reintroduced into (9) without NPs moving out of the coordination.

(18) a. $\forall y. \exists z. linguist'(z) \wedge book.by'(y, z) \rightarrow \mathcal{I}$

b. $\{h_2 = l_7, h_7 = l_8, h_9 = l_9, h_8 = l_{10}\}$

(19) a. $\exists z. linguist'(z) \wedge \forall y. book.by'(y, z) \rightarrow \mathcal{I}$

b. $\{h_2 = l_8, h_7 = l_9, h_8 = l_{10}, h_9 = l_7\}$

Note that $\mathcal{I}$ falls within the scope of the existential quantifier in (19) but not (18).

Because the mechanism described here prevents quantifiers from taking scope outside the NP that they appear in, this method can also be used to implement *scope islands* for the NPs. In particular, there is not the redundancy in the mechanism proposed by Lev [10], although Lev's approach does not require reasoning over multiple ULFs. In fact, if *all* quantified NPs are represented with a separate ULF, rather than only the coordinated ones, then the model of quantification posited by Park [12] results. Park argues that quantifiers may not "intercalate" between (that is, take scope between) different quantifiers in the same NP.

## 3.3 Properties of Coordinated Unplugged Logical Forms

Based on the discussion in the previous sections, we can give the properties for a *coordinated ULF* (CULF), where a CULF is a set of distinct ULFs. A CULF $U$ is a set of proper ULFs, with the following additional properties:

1. $h_{top}$ appears in exactly one ULF in $U$, and is $TOP$ for that ULF.

2. No label $\{l_1, l_2, \ldots, c_1, c_2, \ldots\}$ appears in more than one ULF in $U$.

3. A hole $h_i \in \{h_1, h_2, \ldots\}$ only appears in more than one ULF in $U$ if:

    - $c_j : h_i$ is a labelled formula in (exactly) one ULF in $U$ for some label $c_j$, and

    - $h_i$ is $TOP$ for at least one other ULF in $U$.

4. Each ULF which does not have $h_{top}$ as its $TOP$ contains exactly one labelled formula $l_i : \mathcal{I}$ for some label $l_i$.

As before, the meaning of the sentence is given by the final formula plugged into $h_{top}$. Because each component of the CULF is a ULF in its own right, an algorithm such as that of Bodirsky *et al* can be used to compute all scopings at no greater computational price than for sentences which do not contain NP coordination.

## 4  Conclusions

We have proposed a method of representing coordinated NPs within a framework such as hole semantics. By representing the meaning of the sentence as a set of ULFs, existing efficient algorithms can be used to find the possible meanings for sentences displaying quantifier scope ambiguity. The representation also accounts naturally for the multiple occurrences of formulas in the sentence's final logical form and the Coordinate Structure Constraint.

## References

[1] Hiyan Alshawi, editor. *The Core Language Engine*. MIT press, 1992.

[2] Manuel Bodirsky, Denys Duchier, Sebastian Miele, and Joachim Niehren. A new algorithm for normal dominance constraints. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 54–78, 2003.

[3] Johan Bos. Predicate logic unplugged. In *Proceedings of the 10th Amsterdam Colloquium*, pages 133–143, 1996.

[4] Robin Cooper. *Quantification and Syntactic Theory*. Reidel, 1983.

[5] Christian Ebert. *Formal Investigations of Underspecified Representations*. PhD thesis, Kings College, London, 2005.

[6] Markus Egg, Alexander Koller, and Joachim Niehren. The constraint language for lambda structures. *Journal of Logic, Language and Information*, 10(4), December 2001.

[7] Hans Kamp and Uwe Reyle. *From Discourse to Logic*, volume 42 of *Studies in Linguistics and Philosophy*. Kluwer Academic Publishers, 1993.

[8] Alexander Koller, Joachim Niehren, and Stefan Thater. Bridging the gap between underspecification formalisms: Hole semantics as dominance constraints. In *Meeting of the European Chapter of the Association of Computational Linguistics*, pages 195–202, April 2003.

[9] Esther König and Uwe Reyle. A general reasoning scheme for underspecified representations. In Hans Jürgen Ohlbach and Uwe Reyle, editors, *Logic, Language and Reasoning: Essays in Honour of Dov Gabbay*. Kluwer Academic Publishers, 1999.

[10] Iddo Lev. Decoupling scope resolution from semantic composition. In *Proceedings of the International Workshop on Computational Semantics (IWCS-6)*, pages 139–150, 2005.

[11] Joachim Niehren and Stefan Thater. Bridging the gap between underspecification formalisms: Minimal recursion semantics as dominance constraints. In *41st Meeting of the Association of Computational Linguistics*, pages 367–374, July 2003.

[12] Jong C. Park. Quantifier scope and constituency. In *Proceedings of the Association for Computational Linguistics*, pages 205–212, June 1995.

[13] Massimo Poesio. Semantic ambiguity and perceived ambiguity. In Kees van Deemter and Stanley Peters, editors, *Semantic Ambiguity and Underspecification*. CSLI Publications, 1996.

[14] Uwe Reyle. On reasoning with ambiguities. In *Proceedings of the EACL*, Dublin, 1995.

[15] John Robert Ross. *Constraints on Variables in Syntax*. PhD thesis, Massachusetts Institute of Technology, 1967.