Minimal Recursion Semantics as Dominance Constraints:
Translation, Evaluation, and Analysis

Ruth Fuchss Alexander Koller Joachim Niehren Stefan Thater

Abstract

We show that a practical translation of MRS descriptions into normal dominance constraints is feasible. We start from a recent theoretical translation, develop it into a practical system, and apply it to the output of the English Resource Grammar (ERG) on the Redwoods corpus. We validate the assumptions made by the theoretical translation for a large majority of cases; the MRS descriptions computed in all other cases seem to be systematically incomplete.

1 Introduction

Underspecification is the standard approach to dealing with scope ambiguity (Reyle, 1993; Pinkal, 1996). The readings of underspecified expressions are represented by compact and concise descriptions, instead of being enumerated explicitly. Underspecified descriptions are easier to derive in syntax-semantics interfaces (Egg et al., 2001; Copestake et al., 2001), useful in applications such as machine translation (Copestake et al., 1995), and can be resolved by need.

Two important underspecification formalisms in the recent literature are Minimal Recursion Semantics (MRS) (Copestake et al., 1999) and dominance constraints (Egg et al., 2001). MRS is the underspecification language which is used in large-scale HPSG grammars, such as the English Resource Grammar (ERG) (Copestake and Flickinger, 2000). The main advantage of dominance constraints is that they can be solved very efficiently (Althaus et al., 2003; Bodirsky et al., 2004).

Niehren and Thater (2003) defined, in a theoretical paper, a translation from MRS into normal dominance constraints. This translation clarified the precise relationship between these two related formalisms, and made the powerful meta-theory of dominance constraints accessible to MRS. Their goal was to also make the large grammars for MRS and the efficient constraint solvers for dominance constraints available to the other formalism.

However, Niehren and Thater made three technical assumptions which mean that it is not obvious whether their result can be immediately applied to the output of practical grammars like the ERG. Most importantly, they assumed that all linguistically correct and relevant MRS expressions belong to a certain class of constraints called nets.

In this paper, we evaluate the truth of these assumptions on the MRS expressions which the ERG computes for the sentences in the Redwoods Treebank (Oepen et al., 2002). The main result of our evaluation is that 83% of the Redwoods sentences are indeed nets, and 17% aren’t. A closer analysis of the non-nets reveals that they seem to be systematically incomplete i.e., they predict more readings than the sentence actually has; this supports the claim that all linguistically correct MRS expressions are indeed nets. We also verified the other two assumptions, one empirically and one by proof. Finally, we compared the runtimes of the LKB solver for MRS (Copestake, 2002) and the dominance solver from (Bodirsky et al., 2004), and obtained the result that the dominance solver is faster in most cases, and has a more predictable runtime.

Our results are practically relevant both to the HPSG/MRS and to the dominance constraint community. For the latter, we offer a way of obtaining dominance constraints for all sentences that the ERG can parse (and which are nets). To the former, we offer the possibility of resolving MRS much faster than before by using a dominance constraint solver. In addition, we can suggest nets as a verification tool to identify potentially problematic semantic outputs when designing a grammar.
Plan of the Paper. We first recall the definitions of MRS (§2) and dominance constraints (§3). We present our practical translation of MRS and prove its correctness for all nets (§4). We thereby show that the somehow surprising fact that the merging semantics of MRS can be freely abandoned for nets. This was one of two main assumptions of the theoretical translation. We evaluate our practical translation on the Redwood corpus, and discuss its last remaining assumption on net-hood (§5).

2 Minimal Recursion Semantics

We present a complete definition of MRS (Copestake et al., 1999). We do this in two steps. In the first step, we extend the idealized MRS version used in the theoretical translation (Niehren and Thater, 2003) by EP-conjunctions with a merging semantics. In the second step, we restrict the meaning of handle constraints beyond dominance and impose the existence of top handles.

Syntax of MRS. MRS constraints are formulas over the following vocabulary:

1. An infinite set of variables ranged over by h. Variables are also called handles.
2. An infinite set of constants denoting object language variables. We distinguish constants \( x, y, z \) denoting individual variables and constants \( e, e' \) denoting event variables.
3. A set of function symbols ranged over by \( P \), and a set of quantifier symbols ranged over by \( Q \). Pairs \( Q_q \) are further function symbols.
4. The symbol ‘\( =_q \)’

MRS formulas have three kinds of literals, the first two are called elementary predications (EPs) and the third handle constraints:

1. \( h : P(x_1, \ldots, x_n, h_1, \ldots, h_m) \), where \( n, m \geq 0 \)
2. \( h : Q_q(h_1, h_2) \)
3. \( h_1 =_q h_2 \)

In EPs, label positions are on the left of ‘:’ and argument positions on the right. Let \( M \) be a set of literals. The label set \( \text{lab}(M) \) contains all handles of \( M \) that occur in label but not in argument position, and the argument handle set \( \text{arg}(M) \) contains all handles of \( M \) that occur in argument but not in label position.

Figure 1: An MRS and its two configurations.

\[
P_1, P_2, \{ \ h_1 : P_1(h_2), h_1 : P_2(h_3), h_4 : P_3 \\
\ h_2 =_q h_4, h_3 =_q h_4 \ \}
\]

Figure 2: An unsolvable MRS with EP-conjunction

Definition 1 (MRS). An MRS is finite set \( M \) of MRS-literals such that:

M1 every handle occurs at most once in argument position in \( M \),

M2 handle constraints \( h =_q h' \) always relate argument handles \( h \) to labels \( h' \), and

M3 for every constant (individual variable) \( x \) in argument position in \( M \) there is a unique literal of the form \( h : Q_q(h_1, h_2) \) in \( M \).

An MRS \( M \) can contain different EPs with the same label \( h \). A set of such EPs constitutes an EP-conjunction. The intuition is that EP-conjunctions will be interpreted by object language conjunctions.

We usually draw MRSs as directed graphs. The nodes of these graphs are the handles of the MRS. EPs are represented as solid lines, and handle constraints are represented as dotted lines. For instance, the graph on the left of Fig. 1 represents the following MRS:

\[
\{ h_5 : \text{some}_y(h_6, h_8), h_7 : \text{book}(y), h_1 : \text{every}_x(h_2, h_4), \\
\ h_3 : \text{student}(x), h_9 : \text{read}(x, y), h_2 =_q h_3, h_6 =_q h_7 \ \}
\]

The relation between bound variables and their binders is made explicit by dotted edges (cf. C2 below), so called binding edges. Transitivity redundant dotted edges are omitted. For instance, there is no binding edge from some_y to book_y.

In Fig. 2 we give an MRS with EP-conjunctions and its graph. The function symbols of both EP’s are conjoined and their arguments are merged into a set.
**Configurations.** Let $M$ be an MRS and let $h, h'$ be handles in $M$. We say that $h$ immediately outscopes $h'$ in $M$ if there is an in $M$ with label $h$ and argument handle $h'$, and that $h$ outscopes $h'$ in $M$ if the pair $\langle h, h' \rangle$ belongs to the reflexive transitive closure of the immediate outscope relation of $M$.

A solution for an MRS is called a configuration, or scope-resolved MRS. Intuitively, a configuration is an MRS where all handle constraints have been resolved by plugging the “tree fragments” into each other.

**Definition 2 (Configuration).** An MRS $M$ is a configuration if it satisfies conditions $C1$ and $C2$:

$C1$ The graph of $M$ is a tree of solid edges whose handles are all labels: $\text{arg}(M) = \emptyset$ (such that $M$ contains no handle constraints), handles don’t properly outscope themselves, and all handles are pairwise connected by EPs in $M$.

$C2$ If $h : Q_x(h_1, h_2)$ and $h' : P(x, \ldots, x, \ldots)$ belong to $M$, then $h$ outscopes $h'$ in $M$ (so that all binding edges in $M$ are transitively redundant).

We say that $M$ is configuration of an MRS $M'$ if there is partial substitution $\sigma : \text{lab}(M') \rightsquigarrow \text{arg}(M')$ that states how to identify labels with argument handles of $M'$ so that:

$C3$ $M = \{ \sigma(E) \mid E \text{ is an EP in } M' \}$, and

$C4$ for all $h = q h' \in M'$, $h$ outscopes $\sigma(h')$ in $M$.

The value $\sigma(E)$ is obtained by substituting all labels in $\text{dom}(\sigma)$ in $E$ while leaving all other handles unchanged.

The MRS on the left of Fig. 1, for instance, has two configurations given to the right. The MRS of Fig. 2 does not have configurations since the argument handles of merged EPs cannot jointly outscope nodes of configurations.

We call a configuration merging if it contains an EP-conjunction, and simple otherwise. Merging configurations are needed to solve EP-conjunctions such as $\{ h : P_1, h : P_2 \}$. Unfortunately, they can also solve MRS that are free of EP-conjunctions, such as the MRS in Figure 3; the unique configuration of this MRS is a merging configuration. The labels $P_1$ and $P_2$ must be identified with the only available argument handle.

The admission of merging configurations may thus have important consequences for the solution space of arbitrary MRSs.

**Further Restrictions** Standard MRS interprets handle constraints as $\text{qeq}$-constraints. Let $M$ be an MRS and let $h, h'$ be handles in $M$. We say that $h$ is $\text{qeq}$ $h'$ in $M$ if either $h = h'$, or there is an EP $h : Q_x(h_0, h_1)$ in $M$ and $h_1$ is $\text{qeq}$ $h'$ in $M$. The $\text{qeq}$-relation is thus a specific form of dominance, thus every $\text{qeq}$-configuration is a configuration as defined above, but not necessarily vice versa. This restriction is relevant in theory but will turn out unproblematic in practice (see §5). Finally, the standard MRS requires the existence of top handles in all MRS descriptions. This condition matters only for MRSs with unconnected graphs (see (Bodirsky et al., 2004) for the proof idea), which clearly do not play any role in practical underspecified semantics.

**3 Dominance Constraints**

Dominance constraints are a general framework for describing trees. For scope underspecification, they are used to describe the syntax of object language formulas. Dominance constraints are the core language underlying CLLS (Egg et al., 2001) which adds parallelism and binding constraints.

**Syntax and semantics.** We assume a possibly infinite signature $\Sigma$ of function symbols with fixed arities and an infinite set of variables ranged over by $X, Y, Z$. We write $f, g$ for function symbols and $\text{ar}(f)$ for the arity of $f$.

A dominance constraint $\varphi$ is a conjunction of dominance, inequality, and labeling literals of the following form, where $\text{ar}(f) = n$:

$$\varphi ::= X \ll Y \mid X \not= Y \mid X : f(X_1, \ldots, X_n) \mid \varphi \land \varphi'$$

Dominance constraints are interpreted over finite constructor trees i.e., ground terms constructed from the function symbols in $\Sigma$. We identify ground terms with trees that are rooted, ranked, edge-ordered and labeled. A solution for a

![Figure 3: A solvable MRS without simple configuration](image-url)
dominance constraint $\varphi$ consists of a tree $\tau$ and an
assignment $\alpha$ that maps the variables in $\varphi$ to nodes
of $\tau$ such that all constraints are satisfied: labeling
literals $X : f(X_1, \ldots, X_n)$ are satisfied iff $\alpha(X)$ has
label $f$ and its daughters are $\alpha(X_1), \ldots, \alpha(X_n)$ in
this order; dominance literals $X \prec Y$ are satisfied iff $\alpha(X)$ dominates $\alpha(Y)$ in $\tau$; and inequality lit-
erals $X \neq Y$ are satisfied iff $\alpha(X)$ and $\alpha(Y)$ are
distinct nodes.

**Solved forms and configurations.** Every satis-
fiable dominance constraint has infinitely many
solutions. Constraint solvers for dominance con-
straints therefore do not enumerate solutions but
* solved forms i.e., “tree shaped” constraints. To
this end, we consider (weakly) normal dominance
constraint (Bodirsky et al., 2004).

We call a variable a hole of $\varphi$ if it occurs as in
argument position in $\varphi$ and a root of $\varphi$ otherwise.

**Definition 3.** A dominance constraint $\varphi$ is normal
if it satisfies the following conditions.

- **N1** (a) each variable of $\varphi$ occurs at most once
  in the labeling literals of $\varphi$.
  
  (b) each variable of $\varphi$ occurs at least once in
  the labeling literals of $\varphi$.

- **N2** for distinct roots $X$ and $Y$ of $\varphi$, $X \neq Y$ is in $\varphi$.

- **N3** (a) if $X \prec Y$ occurs in $\varphi$, $Y$ is a root in $\varphi$.

  (b) if $X \prec Y$ occurs in $\varphi$, $X$ is a hole in $\varphi$.

We call $\varphi$ is weakly normal if it satisfies the above
properties except for N1 (b) and N3 (b).

Note that Definition 3 imposes compactness:
the height of tree fragments is always one. This
is not a serious restriction, as weakly normal dom-
inance constraints can be compactified, provided
that dominance links relate either roots or holes
with roots.

Weakly normal dominance constraints $\varphi$ can be
represented by *dominance graphs*. The dominance
graph of $\varphi$ is a directed graph $G = (V, E_T \cup E_D)$
defined as follows. The nodes of $G$ are the vari-
ables of $\varphi$. Labeling literals $X : f(X_1, \ldots, X_k)$ con-
tribute tree edges $(X, X_i) \in E_T$, for $1 \leq i \leq k$, and
dominance literals $X \prec X'$ contribute dominance
edges $(X, X') \in E_D$. Inequality literals are not rep-
resented in the graph. In pictures, labeling literals
are drawn with solid lines and dominance edges
are drawn with dotted lines.

We say that a constraint $\varphi$ is in *solved form* if
its graph is in solved form. A graph $G$ is in solved
form iff it is a forest. The solved forms of $G$ are
solved forms $G'$ which are more specific than $G$
i.e., they differ only in their dominance edges and
the reachability relation of $G$ extends the reach-
ability of $G'$. A minimal solved form is a solved
form which is minimal with respect to specificity.

4 From MRS to Dominance Constraints

This section recalls Niehren and Thater’s trans-
lation from MRS-nets (to be defined below) into
normal dominance constraints. The translation has
been proven to be sound and complete for sim-
ple configurations i.e., configurations without EP-
conjunctions. We generalize the result and prove
that it is also sound and complete with respect to
arbitrary configurations, by proving that for nets,
all configurations are simple.

The bottom line is that EP-conjunctions are not
needed for MRS-nets. They can be resolved in
a preprocessing step, and a constraint solver is
never forced to merge different EPs into an EP-
conjunction.

In the next section, we then show empirically
that the two assumptions underlying the transla-
tion, the “net-assumption” that all relevant con-
straints are nets, and the “eqq-assumption” that
handle-constraints can be given a dominance se-
manics are met in practice.

4.1 MRS- and Dominance-Nets

A hypernormal path (Althaus et al., 2003) in a
constraint graph is a path in the undirected graph
that contains for every leaf $X$ at most one incident
dominance edge.

Let $\varphi$ be a compact weakly normal dominance
constraint and let $G = (V, E_T \cup E_D)$ be the graph
of $\varphi$ with all redundant dominance edges removed.
We say that $\varphi$ is a dominance net if $G$ is a net. $G$
is a net if every tree fragment with leaves $X_1, \ldots, X_k$
and root $X$ satisfies the following conditions:

- **N1** for all $Y \in \{X_1, \ldots, X_{k-1}\}$, there is a $Z \in V$
such that $(Y, Z) \in E_D$.

- **N2** for all $Y, Z \in V$, $(X, Y) \in E_D$ iff $(X_k, Z) \notin E_D$.

- **N3** for all $Y, Z \in V$ and for all $X' \in \{X_1, \ldots, X_k\}$,
  if $(X', Y) \in E_D$ and $(X', Z) \in E_D$, then $Y$ and
The literals of \( M \) are translated into a weakly normal dominance constraints \( \varphi_M \) as follows:

\[
h : P(x_1, \ldots, x_n, h_1, \ldots, h_k) \mapsto h : P_{x_1, \ldots, x_n}(h_1, \ldots, h_k)
\]

\[
h : Q_\varphi(h_1, h_2) \mapsto h : Q_\varphi(h_1, h_2)
\]

\[
h = q \ h' \mapsto h \vartriangleleft h'
\]

Additionally, dominance literals \( h \vartriangleleft h' \) are added for all \( h, h' \) s.t.: \( h : Q_\varphi(h_1, h_2) \) and \( h' : P(\ldots, x, \ldots) \) belong to \( M \) (cf. C2), and literals \( h \neq h' \) are added for all \( h, h' \) in distinct label position in \( M \).

**Normalization.** The constraint \( \varphi_M \) is normalized by replacing root-to-root dominance edges \( X \vartriangleleft X' \) with dominance edges \( Y \vartriangleleft X' \), where \( Y \) is a hole without outgoing dominance edges, which must exist by net-condition N2.

**Theorem 1 (Niehren and Thater, 2003).** If \( M \) is a connected MRS-net, then the simple configurations of \( M \) bijectively correspond to the minimal solved forms of its translation \( \varphi_M \).

**Lemma 1 (Niehren and Thater, 2003).** All minimal solved forms of a connected dominance net \( \varphi \) are simple.
grammar, in connection with the LKB system, a grammar development environment for typed feature grammars (Copestake and Flickinger, 2000). We use the system to parse sentences and output MRS constraints which we then translate into dominance constraints. As a test corpus, we use the Redwoods Treebank (Oepen et al., 2002) which contains 6612 sentences. We exclude the sentences that cannot be parsed due to memory capacities or words and grammatical structures that are not included in the ERG, or which produce ill-formed MRS expressions (typically violating M2) and thus base our evaluation on a corpus containing 6230 sentences. In case of syntactic ambiguity, we only use the first reading output by the LKB system.

To enumerate solutions of MRS expressions and their translations, we use the MRS solver built into the LKB system and a solver for weakly normal dominance constraints (Bodirsky et al., 2004).

5.1 Relevant Constraints are Nets
We check for 6230 constraints whether they constitute nets. It turns out that 5188 (83.27%) constitute nets and 1042 (16.73%) are no nets.

Non-nets. The evaluation shows that the hypothesis that all relevant constraints are nets seems to be falsified: there are constraints that are not nets. However, a closer analysis suggests that these constraints are incomplete and predict more readings than the sentence actually has.

Non-nets can be classified into two categories (cf. Fig. 5): The first class are violated “strong” fragments which have holes without outgoing dominance edge and without a corresponding root-to-root dominance edge. The second class are violated “island” fragments where several outgoing dominance edges from one hole lead to nodes which are not hypernormally connected. There are two more possibilities for violated “weak” fragments—having more than one weak dominance edge or having a weak dominance edge without empty hole—, but they occur infrequently (4.4%). If those weak fragments were normalized, they would constitute violated island fragments, so we count them as such.

124 (11.9%) of the non-nets contain empty holes, 762 (73.13%) contain violated island fragments, and 156 (14.97%) contain both. Those constraints that contain only empty holes and no violated island fragments cannot be configured, as in configurations, all holes must be filled.

Typical syntactic phenomena which lead non-nets are coordinations as in “it has a restaurant and a sauna,” and numerical expression as in “okay that would be one hundred and eighty six euros.”

The constraint on the left in Fig. 6 which the ERG computes for the sentence “the dog and the cat sleep” exemplifies the problem. The constraint is not an MRS-net as it violates the island-schema: the topmost fragment has outgoing dominance edges to otherwise unconnected subconstraints $\varphi_1$ and $\varphi_2$. Under the “strict plugging” semantics of the MRS dialect used in (Niehren and Thater, 2003) where every hole has to be filled exactly once, this constraint cannot be configured: there is no hole into which “sleep” could be plugged. However, standard MRS has a more general notion of configuration: holes can be filled more than once. The consequence is that these derived EP-conjunction can be merged in almost everywhere. In fact, the MRS constraint solver derives 14 configurations for the constraint, two of which are given to the right in Fig. 6, although the sentence has only two scope readings.

For the ERG, the problem can be illustrated with the average number of solutions: On average, nets have 1836 solutions, while non-nets have 14039 solutions, which is a factor of 7.7.

5.2 Qeq is dominance
For all nets, the dominance constraint solver calculates the same number of solutions as the MRS solver does, with 3 exceptions that hint at problems in the syntax-semantics interface. This result means that the additional expressivity of proper
qeq-constraints is not used in practice, which in turn means that in practice, the translation is sound and correct even for the standard MRS notion of solution, given the constraint is a net.

Note that we have not yet done a pairwise comparison of solutions, but the results strongly indicate that for nets, they do correspond.

5.3 Comparison of Runtimes

As a side-effect of comparing the solutions, we measure the runtimes of the two solvers. The measurements were performed in a multi-user environment, and should be taken to represent trends rather than strict comparisons. However, the tests were conducted pairwise, hence conditions were equal for every MRS constraint and corresponding dominance constraint.

The results are shown in Fig. 7. The figure on the left gives the runtimes for all constraints sorted by number of solutions. Where the dominance constraint solver exhibits a predictable runtime, the MRS solver performs well in some cases but badly in others. This suggests that the MRS solver has been optimized for certain kinds of constraints. The right-hand figure shows that for the dominance constraint solver, the runtime per solution is effectively constant, whereas it increases with the size of the constraint in the MRS solver and is generally much higher. Actually, the dominance constraint solver outperforms the MRS solver in over 85% of the cases.

We exclude cases in which one or both of the solvers do not return any results, which happens when we interrupt the dominance constraint solver (if the computation takes longer than 2 minutes) or the MRS solver runs out of memory. The MRS solver does not return any results in 134 cases (2.8% of all nets), the dominance constraint solver in only 29 cases (0.6%), all but 1 of which also cannot be computed by the MRS solver.

6 Conclusion

We developed Niehren and Thater’s (2003) theoretical translation into a practical system for translating MRS into dominance constraints, applied it systematically to MRSs produced by English Resource Grammar for the Redwoods treebank, and evaluated the results. We showed that:

1. most “real life” MRS expressions are MRS-nets, which means that the translation is correct in these cases;
2. for nets, merging is not necessary (and possible);
3. the practical translation works perfectly for all MRS-nets from the corpus; in particular, the $=_{q}$ relation can be taken as synonymous with dominance in practice.

As an interesting side-effect, we compared the runtimes of the constraint solvers. The evaluation shows that the dominance constraint solver outperforms the MRS solver and displays more predictable runtimes, which means that the translation is also interesting for researchers interested in MRS as it offers more efficient constraint solvers for MRS.

The evaluation also shows that a smaller, but still significant number of MRS constraints are not nets. We argued that these constraints may be linguistically problematic because they are structurally too weak and permit too many readings. In this sense, we believe that nets can serve as a correctness criterion for syntax-semantics interfaces.

In the future, it will be interesting to see whether it is possible to specify a “safety criterion” for semantics construction rules that guarantees that all
MRS expressions computed by grammars are nets. It might also be interesting to investigate handle constraints theoretically e. g., by providing a formal criterion that ensures that qeq is dominance for a certain class of constraints.

References


