A new well-formedness criterion for semantics debugging

Dan Flickinger[1] and **Alexander Koller**[2] and **Stefan Thater**[2]
[1]CSLI, Stanford University, Stanford
[2] Department of Computational Linguistics, Universität des Saarlandes, Saarbrücken

Abstract

We present a novel well-formedness condition for underspecified semantic representations which requires that every correct MRS representation must be a *net*. We apply this condition to identify a set of eleven rules in the English Resource Grammar (ERG) with bugs in their semantics component, and thus demonstrate that the net test is useful in grammar debugging. In addition, we show that a partly corrected ERG derives 3 % less non-nets on the Rondane treebank and we expect that after completing the correction of the ERG, only 5.5 % non-nets are derived, which we take as support for our initial hypothesis.

1 Introduction

A very exciting recent development in (computational) linguistics is that large-scale grammars which compute semantic representations for their input sentences are becoming available. For instance, the English Resource Grammar (4) is a large-scale HPSG grammar for English which computes underspecified semantic representations in the MRS formalism (4). It is standard to use underspecification to deal with scope ambiguities; apart from MRS, there is a number of other underspecification formalisms, such as dominance constraints (4) and Hole Semantics (4).

However, the increased power of the new grammars comes with a new challenge for grammar engineering: How can we be sure that all semantic outputs the grammar computes (through any combination of semantic construction rules) are correct, and how can we find and fix bugs? This problem of *semantics debugging* is an important factor in the 90% of grammar development time that is spent on the syntax-semantics interface (4).

Grammar development systems such as the LKB implement some semantic sanity checks, which are practically useful, but rather shallow, and therefore limited in their power. On the theoretical side, there are attempts to formalise "best practices" of grammar development in a *semantic algebra* (4), but this is quite a far-reaching project that is not yet fully implemented.

One potential alternative method for semantics debugging comes from Fuchss et al.'s recent work on *nets* (4). They claim that every underspecified description (written in MRS or as a dominance constraint) that is actually used in practice is a *net*, i.e. it belongs to a restricted class of descriptions with certain useful structural properties, and they substantiate their claim through an empirical evaluation on a treebank. If this "Net Hypothesis" is true, we can recognise a grammar rule (or combination of rules) as problematic if it produces only non-nets on a variety of inputs.

In this paper, we show that such a use of nets is indeed possible. We use the ERG to derive MRS representations for all sentences in the Rondane treebank (distributed with the ERG) and the Verbmobil sections of the Redwoods treebank (4). Our first result is a small set of eleven rules which systematically cause the MRS representations to be non-nets for every sentence in which they are used. These rules all have faulty semantics components, i.e. we have identified semantically buggy rules. We are currently correcting the grammar by hand. The partly corrected grammar produces 89.5 % nets and only 8 % non-nets for the syntactic analyses in the Rondane corpus, and we expect that after completing the correction of the problematic rules, only 5.5 % non-nets are derived, which we take as further support of the Net Hypothesis.

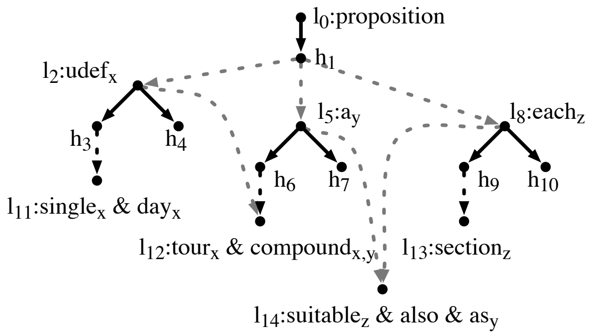


Figure 1: Graphical representation of the MRS for "Each section is also suitable as s single day tour."

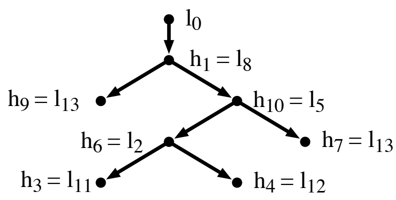


Figure 2: Configuration of the MRS in Fig. 1.

2 Minimal Recursion Semantics

We start with a brief overview of Minimal Recursion Semantics (MRS). MRS (4) is the standard scope underspecification formalism used in current HPSG grammars, such as the English Resource Grammar (ERG; (4)) or grammars derived from the Grammar Matrix (4). Its purpose is to separate the problem of resolving scope ambiguities from semantics construction.

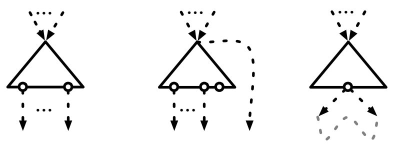
Fig. 1 shows a graphical representation of the (slightly simplified) MRS which the ERG derives for the sentence "Each section is also suitable as a single day tour" from the Rondane treebank. It consists of *elementary predications* (EPs) such as l_2:udef(x,h_3,h_4), l_5:a(y,h_6,h_7), l_{12}:tour(x,y), and l_{12}:compound(x,y), and of *handle constraints* such as $h_6 =_q l_{12}$. Elementary predications specify the parts that a semantic representation must be made up of, and handle constraints $h =_q l$ specify, approximately, that h must outscope l. Terms l_i on the left-hand side of EPs are called *labels*, terms h_i are called *(argument) handles*, and terms x, y, etc. are ordinary first-order variables. Notice that there are two EPs for the label l_{12}; this is called an *EP conjunction*, and interpreted as conjunction of the two formulas labelled by l_{12}.

The graph in Fig. 1 can be given an explicit interpretation as a representation of an MRS structure (4). The nodes correspond to the labels and handles in the MRS, and the solid edges correspond to the EPs. We call the subgraphs that are connected by solid edges the *fragments* of the graph. The dashed *dominance edges* are used to represent handle constraints, the outscoping requirement between a variable and its binder (such as between the quantifier at l_2 and the variable in l_{12}), and the implicit constraint that the "top" label l_0 must outscope all other EPs. Note that we assume that the graph does not contain transitively redundant edges; for instance there is no binding edge between l_2 and l_{11}. EP conjunctions are represented by explicit conjunction at the graph nodes.

An underspecified MRS structure describes a set of configurations, or *scope-resolved* MRS structures. The scope-resolved MRS structures can be computed by arranging all the fragments of an MRS structure into a tree, in such a way that every label except for the one at the root is identified with a handle, and all the outscoping requirements are respected. One of the five scope-resolved MRSs for the MRS in Fig. 1 is shown in Fig. 2 (we omit EPs for clarity). Note that in general it is possible that more than one label is assigned to the same handle, and that the scope-resolved MRS structure can contain more EP conjunctions than the original MRS structure. In such a case, we call the scope-resolved MRS structure a *merging configuration*.

3 MRS-Nets

We say that an MRS structure is a *net* if all the fragments in its graph are of one of the three forms shown in Fig. 3. In a *strong* fragment, every leaf (argument handle) and no other node has exactly one outgoing dominance edge. For example, the nuclear fragments l_{11} and l_{14} in Fig. 1 are strong fragments. In a *weak* fragment, every leaf but one has exactly one outgoing dominance edge, and the root of the fragment has one outgoing dominance edge too. Weak fragments correspond to quantifiers (such as l_2 and l_8 in Fig. 1) where the dominance edge from the root represents the implicit variable binding. Fi-



(a) strong (b) weak (c) island

Figure 3: Fragment Schemata of Nets

nally, *island* fragments may have leaves with multiple outgoing dominance edges, but then all dominance children must be connected by *hypernormal paths*. A hypernormal path is an undirected path which doesn't use two dominance edges that come out of the same node. An example for an island fragment is the topmost fragment in Fig. 1. Its dominance children are the three quantifier fragments, and there is a hypernormal path between each pair of these fragments – for instance, l_2, l_{12}, h_6, l_5 and l_5, l_{14}, l_8.

Because all fragments in Fig. 1 are strong, weak, or island, the MRS it represents is a net. By contrast, Fig. 5 shows two MRS structures which are not nets. Both structures violate the *island* condition because the topmost fragment has outgoing edges to quantifier fragments (e.g. in the left-hand graph, the fragments for "a bit" and "two young Norwegians") which are only connected via the top fragment itself, and not by an additional hypernormal path. The left-hand graph also contains a quantifier fragment ("a bit") which violates the *weak* condition, as there is an open argument handle without a corresponding dominance edge out of the root of the fragment.

Nets were introduced in (4; 4) as a technical restriction; the key theorem about nets is that they can be translated into normal dominance constraints (4) and Hole Semantics (4). This means that nets can be solved efficiently using the solvers for normal dominance constraints (4). Nets have other useful properties: For example, nets have no merging configurations, so all EP conjunctions can be resolved to true conjunctions in a preprocessing step.

The crucial restriction that nets impose is that the dominance children of island fragments must be hypernormally connected. Intuitively, hypernormal connectedness means that nets must be "downwards" connected: In the example in Fig. 1, l_2 and l_8 are "tied together" by the zig-zag path through l_{12} and l_{14}, whereas "a bit" and "two young Norwegians" in Fig. 5 have no such connection. A linguistic intuition for this is that quantifiers that are syntactic arguments of the same verb remain hypernormally connected because their variables occur as arguments of this verb.

4 Nets in Semantics Debugging

Now we show that nets can indeed be used to identify grammar rules with incomplete semantics components, and that non-nets are so infrequent in practice that it is reasonable to assume that all correct MRS structures are indeed nets.

4.1 Previous Work

Recently, Fuchss et al. (4) presented a first evaluation of whether the MRS structures that can be derived using the ERG are nets or not. They found that about 83% of the MRS structures derived for all syntactic readings of all the sentences in the Redwoods treebank (4) are in fact nets. Their impression from inspecting some non-nets was that non-nets seemed to be systematically incomplete. They took this as suggestive of what they call the *Net Hypothesis*: that all MRS structures needed in practice (i.e. for the parses of a treebank according to a large-scale grammar) are nets.

4.2 Experiment

If the Net Hypothesis is true, the 17% non-nets must be the results of errors in the annotation or the grammar rules, and every MRS that is not a net can be taken as an indicator that the grammar rules used in producing it might be candidates for debugging.

In order to analyse this in more detail, we reran Fuchss et al.'s evaluation, using the October 2004 version of the ERG. As test corpora, we used the Verbmobil sections of the Redwoods 5 Treebank (Jan. 2005) which contains 10503 sentences, and the Rondane Treebank (1034 sentences) distributed with the ERG. Both corpora are annotated with HPSG syntactic structures, for each of which a unique MRS structure can be extracted.

The table in Fig. 4 shows the results of the experiment. Each sentence in the treebanks was classified into one of three categories: (1) sentences whose MRS structure was not well-formed according to the shallow tests in the LKB system (e.g., structures

Treebank	#Sents.	Ill-formed	Non-Nets	Nets
Verbmobil	10503	11 %	17 %	72 %
Rondane	1034	8 %	11 %	81 %

Figure 4: Classification of the sentences in the treebanks.

containing variables that aren't bound by any quantifier, or structures with cycles); (2) sentences whose MRS structures were okay according to the LKB checks, but were not nets, and (3) sentences whose MRS structures were nets. Of all the MRSs that are well-formed according to the test in the LKB, 81 % (Redwoods) and 88 % (Rondane) are nets, and 19 % (Redwoods) and 12 % (Rondane) aren't. Interestingly, the ratio of nets to non-nets is much smaller if we look not only at the annotated syntactic analyses, but at *all* possible analyses (as Fuchss et al. did).

4.3 Semantic Debugging

Then we checked which rules were "responsible" for the introduction of non-net structures. We found that there is a group of eleven rules which systematically derive only non-nets for all syntactic analyses of all sentences in the treebanks; these rules account for approx. 55% of the non-nets:

1. Measure noun phrases like "2–3 hours" (MEASURE_NP, BARE_MEASURE_NP)
2. Coordinations of more than two conjuncts like "train, bus or car" (P_COORD_MID, N_COORD_MID)
3. Sentence fragmens like "Delicious!" (rules FRAG_PP_S, FRAG_R_MOD_I_PP, FRAG_ADJ, and FRAG_R_MOD_AP)
4. Other rules: VPELLIPSIS_EXPL_LR, NUM_SEQ, TAGLR.

Indeed, the semantics components of all eleven rules are buggy, in that the MRS graphs that they compute have too few dominance edges or unconnected fragments that should constitute an single fragment (e.g., by forming an EP-conjunction). This is illustrated by the structures shown in Fig. 5. The structure on the left is derived by the ERG for the sentence "a bit further on we meet two young Norwegians". In this structure, the quantifier "a bit" (whose analysis uses the MEASURE_NP rule) introduces a bound variable *x* that is used only in its restriction, but in none of the predicates in its scope ("meet further on"). This is obviously not intended. Because the missing variable binding also relaxes the constraints on how fragments can be plugged together, the underspecified description admits structurally wrong readings, e.g. by plugging "young Norwegian" into the scope of "a bit" (see Fig. 6). If we fix the structure by using *x* in the EPs for "further on", this introduces an additional dominance edge in the graph which makes the structure a net.

A similar bug occurs in the right-hand MRS structure. The EPs "and" and "implicit_conj" are two different components of the same collective "tea, milk, and coffee", and should therefore be connected. Because they aren't, the structure has meaningless scopings such as the one shown in Fig. 6 (and almost 1000 further scopings) where "and" and "drink" have been merged into the same argument handle. If we connect "and" and "drink" either by collecting them into a single EP-conjunction, or by introducing additional material (e.g., an qauntifier fragment) that connects the two nodes, the MRS structure again becomes a net.

Treebank	#Sents.	Ill-formed	Non-Nets	Nets
Rondane	961	2.5 %	8 %	89.5 %

Figure 7: Classification of the sentences in the Rondane treebank for the partly corrected version of the ERG

4.4 Re-Evaluating the Net Hypothesis

We are currently working on correcting the semantics components of the eleven faulty rules by hand. If all problematic rules are corrected in a way that only nets are derived, we expect that of the well-formed MRS structures 94.5 % (Rondane) and 91.5 % (Redwoods) of the syntactic structures as annotated in the treebanks derive nets. A first experiment shows that with our partly corrected version of the ERG, almost 92 % of the derivations annoted with well-formed MRS structures (89.5% of all sentences) in the Rondane treebank produce nets (Fig. 7).[1] It is

[1]For technical reasons, the treebank for the partly corrected grammar contains slightly fewer sentences. Note that if we remove the missing sentences from the classification for the original treebank, we obtain results similar to the ones shown in

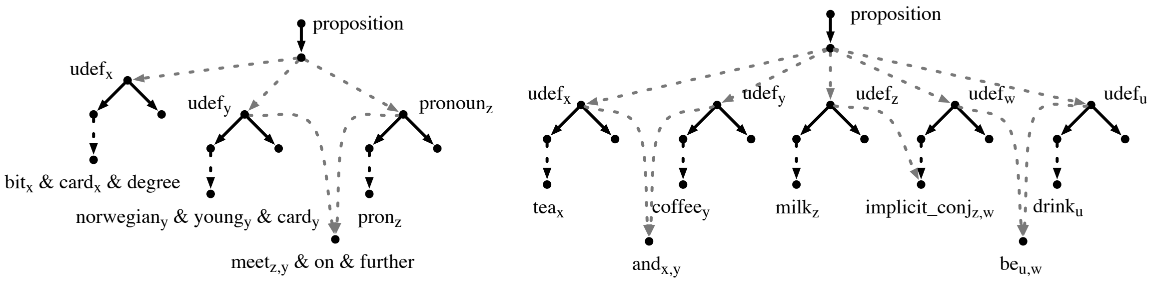


Figure 5: MRS expressions for the annotated derivation for "a bit further on we meet two young Norwegians" (left) and "Drink is tea, milk and coffee" (right) in the Rondane treebank

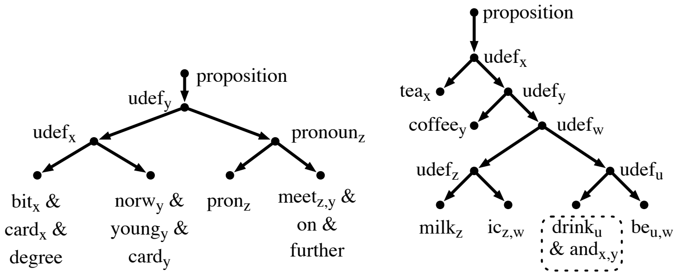


Figure 6: Example solutions for the MRS structures in Fig. 5

important to note that in particular measure noun phrases with degree modifications, which are relatively often used, are not yet fully corrected. Note also that the number of ill-formed MRS structures has been considerably reduced.

It is important to note that at least some applications of each of the eleven rules above passed the well-formedness checks in the LKB, which shows that nets can allow us to identify semantically problematic rules which shallower checks can't find. In addition, non-nets make up a larger portion of the MRS structures in the original grammar than the ill-formed structures; so they are likely to capture classes of errors that are at least as prevalent as those that the existing checks do.

5 Conclusion

We have shown that nets can be a useful tool for debugging the semantics component of a large-scale grammar. All eleven rules in the ERG that computed only non-nets turned out to be semantically problematic, typically in that they were missing a variable name coindexation, or some fragments (EPs) were unconnected; also, none of these rules would have been easily found by the existing well-formedness tests in the LKB. A partly corrected version of the ERG derived 89.5 % nets on the Rondane corpus.

In order to make the net criterion practically useful, we have developed an efficient algorithm that checks whether a given MRS is a net in linear time. A portable open source implementation of this algorithm is publically available from `http://utool.sourceforge.net`.

There are various ways in which the work we report here could be extended. On the one hand, it would be interesting to see whether a similar debugging methodology would yield problem rules based on the LKB's well-formedness tests, and it would be natural to look not just for problematic *rules*, but also for problematic *lexicon entries* this way. On the other hand, we suspect that some semantically problematic MRS structures are derived not by a single rule, but by a combination of rules. One way of finding such rule combinations would be to analyse the MRSs for a corpus with a decision tree learner, which would try to derive rules that capture such combinations.

Fig. 4.

References

Emily M. Bender, Dan Flickinger, and Stephan Oepen. The grammar matrix: An open-source starter-kit for the rapid development of cross-linguistically consistent broad-coverage precision grammars. In John Carroll, Nelleke Oostdijk, and Richard Sutcliffe, editors, *Procedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics.*, Taipei, Taiwan, 2002.

M. Bodirsky, D. Duchier, J. Niehren, and S. Miele. A new algorithm for normal dominance constraints. In *Proc. SODA*, 2004.

J. Bos. Predicate logic unplugged. In *Proc. 10th Amsterdam Colloquium*, pages 133–143, 1996.

A. Copestake and D. Flickinger. An open-source grammar development environment and broad-coverage english grammar using HPSG. In *Proc. LREC*, 2000.

A. Copestake, D. Flickinger, C. Pollard, and I. Sag. Minimal recursion semantics: An introduction. *Journal of Language and Computation*, 2004. To appear.

A. Copestake, A. Lascarides, and D. Flickinger. An algebra for semantic construction in constraint-based grammars. In *Proc. 39th ACL*, Toulouse, 2001.

M. Egg, A. Koller, and J. Niehren. The constraint language for lambda structures. *Journal of Logic, Language, and Information*, 10:457–485, 2001.

R. Fuchss, A. Koller, J. Niehren, and S. Thater. Minimal Recursion Semantics as dominance constraints: Translation, evaluation, and analysis. In *Proc. 42nd ACL*, Barcelona, 2004.

J. Niehren and S. Thater. Bridging the gap between underspecification formalisms: Minimal recursion semantics as dominance constraints. In *Proc. 41st ACL*, Sapporo, 2003.

S. Oepen, K. Toutanova, S. Shieber, C. Manning, D. Flickinger, and T. Brants. The LinGO Redwoods treebank: Motivation and preliminary applications. In *Proc. 19th COLING*, 2002.