# Robust VP Ellipsis Resolution in DR Theory

*Johan Bos*

## 1. Introduction

The phenomenon of Verb Phrase Ellipsis (henceforth VPE[1]) that manifests itself in the English language has been a popular topic of research in formal semantics and computational linguistics. In fact, it has been studied in great detail regarding various issues: whether the level of resolution should take place on the syntactic or semantic level; how VPE interacts with quantifier scope; and, especially, how to account for ambiguities resulting from the so called sloppy and strict interpretations that occur when VP ellipsis interacts with anaphoric pronouns.

Empirical approaches to VPE, that is, studying or automatically processing VPE on the basis of occurrences in corpora, have almost been completely ignored, with two notable exceptions: Hardt 1997, and Nielsen 2005. Hardt (1997) studied several hundred examples of VPE automatically found in the Penn Treebank, and implemented and evaluated a system for finding antecedents. Similar work was carried out by Nielsen (2005), but in addition he also discussed various ways to identify VPE in open-domain texts, and manually annotated occurrences of VPE in a corpus. Both Hardt and Nielsen provide algorithms for resolving VPE on the surface level, i.e., by expanding the elided verb phrases into its full form.

The goal of this article is to investigate how an existing implementation of a wide coverage NLP system for the semantic analysis of open-domain natural language texts can be extended to deal with the detection and resolution of VPE. In contrast to Hardt 1997 and Nielsen 2005, ellipsis resolution will proceed on the semantic level. As Nielsen (2005) does, we will also deal with the problem of VPE detection in open-domain texts.

The framework proposed in this article for processing VPE is robust, in the sense that it achieves high coverage on open-domain texts. Yet, it is based on formal linguistic theory. The syntactic analysis is carried out in the context of Combinatorial Categorial Grammar (Steedman 2001). The semantic analysis follows Discourse Representation Theory (Kamp 1981). Resolution of anaphoric expressions is inspired by the binding and accommodation theory of presupposition (Van der Sandt 1992).

The article is organised as follows. We will first introduce VPE for those not familiar with it, sketch the basic approach, present the computational formalism that we employ, and describe the corpus that we use. Then we focus on the task of VPE detection and the task of VPE antecedent location, report our results on both tasks, and list various problematic cases encountered. This is followed by a description of a new algorithm for VPE resolution in Discourse Representation Theory.

## 2. Preliminaries

### 2.1. VP Ellipsis

For those readers unfamiliar with the phenomenon of VPE and the terminology used in the literature we will provide a brief introduction on the topic.

VPE manifests itself in English when verb phrases are abbreviated to an auxiliary verb (do, have, be, will) or deleted in an infinite clause, where the interpretation of the elided VP depends on an earlier introduced verb phrase in the discourse, usually the previous one. Consider the following examples of VPE, where the auxiliary escorting the elided VP is typeset in boldface, and the intended antecedent is underlined:

(1) *Carlo lives in San Lorenzo and so **does** Marco.*

(2) *Carlo hates his boss and Marco **does** too.*

The first sentence is interpreted as meaning that both Marco and Carlo live in San Lorenzo. Using the terminology used in (Dalrymple, Shieber, and Pereira 1991), we distinguish between the **source clause** (*Carlo lives in San Lorenzo*) and the **target clause** (*so does Marco*), with *Carlo* and *Marco* being **parallel elements**.

Example (2) introduces a source of ambiguity. It has a strict interpretation, where Marco and Carlo both hate Carlo's boss; and a sloppy interpretation, where Carlo hates Carlo's boss and Marco hates Marco's boss. Although this is an interesting problem, and many theoretical approaches to VPE have been devoted to it, the sloppy-strict ambiguity triggered by VPE seems to occur only sporadically in real data (see Section 5.3).

### 2.2. Basic Approach

VPE resolution is a complex task and involves various aspects of processing. In order to quantifiy the performance of an VPE resolution algorithm with respect to the various stages of processing involved, I will follow Nielsen 2005 and divide the problem of VPE resolution into three tasks:

1. VPE detection;
2. VPE antecedent location;
3. VPE resolution.

The first problem is the task of determining whether a sentence of English contains an elliptical verb phrase. Given an elliptical verb phrase, the second problem is concerned with finding the correct antecedent in the text. The third problem constitutes the resolution of the elliptical verb phrase: whether the right material (not too much, not too little) of the source clause is abstracted and applied correctly to the target.

For practical reasons, within the scope of this article, we will only consider cases of VPE generated using forms of the auxiliary verb *do*. We have no reason to believe that the analysis put forward in this article does not extend to other forms of VPE, but demonstrating so will be left for future work. Following Nielsen 2005, we will exclude cases of *do-it* and *do-so* anaphora from our analysis, arguing that these are principally different from VPE.

### 2.3. Syntactic and Semantic Formalism

As computational framework we use the C&C wide-coverage parser (Clark and Curran 2004) and Boxer (Bos 2005) to produce semantic representations for open-domain English texts. The C&C parser implements Combinatory Categorial Grammar (CCG, following Steedman 2001) using a statistical model trained on CCGbank, an annotated treebank for derivations of CCG (Hockenmaier and Steedman 2002). As CCG is not central to this article, we assume some familiarity with type-logical grammars and won't go into details of CCG theory. Instead, in Figure 1 we illustrate CCG with a simple example derivation as output by the C&C parser.

```
bapp('S[dcl]',
   lex('N','NP',
      leaf('John', 'John', 'NNP', 'I-PERSON', 'N')),
   fapp('S[dcl]\NP',
      leaf('has', 'have', 'VBZ', 'O', '(S[dcl]\NP)/NP'),
      fapp('NP[nb]',
         leaf('a', 'a', 'DT', 'O', 'NP[nb]/N'),
         leaf('car', 'car', 'NN', 'O', 'N')))).
```

*Figure 1.* CCG derivation for *John has a car* as output by the C&C parser.

The example derivation in Figure 1 shows the lexical categories for each word (for instance, `NP/N` for the determiner *a*, `(S[dcl]\NP)/NP` for the transitive verb *has*). It further demonstrates how combinatory rules combine the categories (for instance, forward application, `fapp`, combines a determiner with a noun, resulting in a category `NP` for the phrase *a car*, and backward application, `bapp`, combines the noun phrase with a verb phrase). The complete analysis receives the category `S[dcl]`, a declarative sentence.

The output of the parser is used to construct Discourse Representation Structures (DRSs), as proposed in Discourse Representation Theory (Kamp and Reyle 1993). In fact, we follow DRT closely, but nevertheless deviate on four points:

1. We restrict ourselves to a first-order fragment of the DRS language;

2. We use a Neo-Davidsonian analysis of events and thematic roles;[2]

3. We extend standard DRT with Van der Sandt's theory of presupposition projection (Van der Sandt 1992);

4. We employ an explicit sentence merge-operator "+" to combine smaller DRSs into larger ones.[3]

Boxer implements DRT on top of the CCG derivations output by the C&C parser. A DRS generated by Boxer is the one below in Figure 2, when given as input the sample derivation shown in Figure 1. This DRS exemplifies the use of discourse referents of type event, and the use of two-place relations to represent thematic roles.

```
 ____________________
| x3 x2 x1           |
|____________________|
| named(x3,john)     |
| car(x2)            |
| sell(x1)           |
| event(x1)          |
| agent(x1,x3)       |
| patient(x1,x2)     |
|____________________|
```

*Figure 2.* DRS for *John sold a car* as output by Boxer.

### 2.4. Analysis of VP Ellipsis in CCGbank

In CCGbank, the relation of a VPE sentence to its antecededent is assumed to be anaphoric in nature, and is analysed by giving the auxiliary of an elided VP clause the category `S\NP` (Hockenmaier 2003, p. 61). Put differently, sentences infected with VPE are considered complete sentences, rather than sentences with a VP missing, in which one would expect a category of the type `(S\NP)/(S\NP)` for the auxilary. In cases of elliptical inversion, Hockenmaier assigns the category `S[inv]/NP` to the auxiliary to analyse phrases such as *and so does John*. In either case, the DRS generated for a sentence with an elided VP, without applying any form of ellipsis resolution, is basically of the form illustrated in Figure 3.

```
  ____________________   ____________________
 | x3 x2 x1           | | x5 x4              |
 |____________________| |____________________|
(| named(x3,john)     |+| named(x5,bill)     |)
 | car(x2)            | | do(x4)             |
 | like(x1)           | | too(x4)            |
 | event(x1)          | | event(x4)          |
 | agent(x1,x3)       | | agent(x4,x5)       |
 | patient(x1,x2)     | |____________________|
 |____________________|
```

*Figure 3.* DRS for *John likes a car. Bill does, too*, as output by Boxer, without performing VPE resolution.

### 2.5. Data

The data that we considered for our corpus study and system development comprises most of the Wall Street Journal sections found in the Penn Treebank. All occurrences of VPE constructed with the auxiliary verb *do* were manually annotated.[4] The data was arbitrarily divided into development data[5] and test data[6]. The development data (in total 89 cases of VPE with the auxiliary "to do") was used to develop and tune the system. The test data (50 examples of VPE) wasn't uncovered until the system was in a stable state, and then used for evaluation of VPE detection and VPE antecedent location.

## 3. VPE Detection

### 3.1. Method

As the DRS in Figure 3 suggests, a simple baseline for VPE detection on the level of logical form (here: DRS) is traversing information in DRSs for events symbolised by `do` that only have one argument role: that of agent. This baseline should have a rather good coverage, but it will also wrongly identify cases as VPE, namely those where the parser identifies forms of *to do* as an intransitive verb when it is part of fixed expressions such as *having to do something*, *doing good/well/better/best* and variations thereof, such as *doing little/much*. Our improved system is a simple extension of the baseline where such cases are filtered out. In addition, it also checks whether the auxiliary "do" is part of a wh-question, in which case it is not considered to be an instance of VPE.[7]

### 3.2. Evaluation and Results

Evaluation of VPE detection is conducted by measuring precision (P, the number of correctly identified VPE divided by the number of identified VPE) and recall (R, the number of correctly identified VPE divided by the total number of VPE in the data set). As usual, in empirical approaches to natural language processing, we also compute the F-score, the harmonic mean of precision and recall. For completeness we show the results for both the development data and the (unseen) test data (Table 1).

*Table 1.* Results (precision, recall and F-score) of VPE detection on Development and Test Data comparing the baseline and improved system.

| **Development** | **P** | **R** | **F** |
|---|---|---|---|
| Baseline | 0.62 | 0.82 | 0.71 |
| Improved | 0.77 | 0.81 | 0.79 |

| **Test** | **P** | **R** | **F** |
|---|---|---|---|
| Baseline | 0.54 | 0.87 | 0.67 |
| Improved | 0.74 | 0.87 | 0.80 |

As Table 1 shows, recall of both the baseline and improved system is similar, but the improved system excells in precision. Note that the performances of the systems on the development and test data do not deviate strongly—this is an indication that we did not overtune the improved system on the development data.

Comparing these results to other work on VPE detection gives us the following figures. Hardt (1997) reports a recall of 44% and precision of 53% for VPE detection (achieving an F-score of 48%), including all types of VPE, for a small part of the Penn Treebank dataset (in total 48 cases of VPE). Nielsen (2005) outperforms these numbers using a method based on part-of-speech tags and other syntactic features: his system obtains an F-score of 82% on (a part of the gold standard) the Penn Treebank dataset, and 71% on re-parsed data.

A straight comparison of our results with that of Nielsen is not straightforward. Even though the overall corpus is the same (Penn Treebank), the subsets of studied data and types of VPE differ: this study take more data into account, but Nielsen covers a wider variety of VPE cases. However, we many tentavily conclude that our system, with its performance on re-parsed data, shows promising results (F-score of 80%) compared to that of Nielsen's (F-score of 71%) in the task of VPE detection.

### 3.3. Problematic Cases

Even though our system can be said to perform well in the task of VPE detection, it still fails to recognise instances of VPE in certain cases. It also wrongly detects VPE, roughly every fourth case it considers one. A basic error analysis could reveal the causes of these mistakes and perhaps suggests a way to improve the system.

Let's consider first the false positives. These are mainly due to parsing mistakes, where the parser tries to deal with difficult cases such as long-distance dependencies or relatively complex modifier verb interactions. For instance, in (3) the main verb in the embedded question is mistaken for an intransitive rather than a transitive verb. Similarly, in (4) and (5) the auxiliary verb is incorrectly analysed as an intransitive verb.

(3) *Index traders [...] don't even know what the companies they own actually **do**, complains Andrew Sigler, chairman of Champion International Corp.* [WSJ section 01]

(4) *But Mr. Smith said [...] the oral agreement **did** in fact exist, and that even [...] at another studio.* [WSJ section 04]

(5) *Makoto Utsumi [...] said the ministry **did**n't in any way suggest to Japanese banks that they stay out of the UAL Corp. leveraged buyout.* [WSJ section 14]

In many of the 'missed' cases it is the other way around — in those the auxiliary was analysed by the parser as a transitive, instead of an intransitive verb. Because the parser that we use is a stochastic parser, the obvious way to improve the analyses is to provide more training data by extending CCGbank.

## 4. VPE Antecedent Location

### 4.1. Method

The baseline algorithm for finding proper antecedents we will start to work with is straightforward: consider the closest VP that occurs before the elided VP as antecedent. The distance between source and target is measured using the token position of the VPE and that of the head of the verb phrases considered as antecedent. Token positions are represented by natural numbers.

Technically, this is implemented as follows. Once we detect a case of VPE, we record the token position $P_E$ of the auxiliary verb that forms the elided VP. Then we search through the DRS representing the entire context provided, and attempt to locate an event with at least an agent role, with corresponding token position $P_A$, such that $P_A < P_E$ and there is no $P_i$ corresponding to an event such that $P_A < P_i \wedge P_i < P_E$.

Even though source and target clauses of VPE are normally in close proximity to each other, for simplicity, the search is carried out at the entire previous context, rather than, say, within a two-sentence window. In our implementation, the context is defined by the complete newspaper article as defined in the PennTreebank corpus. For each newspaper article, one DRS spanning all its sentences is constructed. Obviously, this is not an efficient method, but once we know more about promixity of antecedent VPs we can restrict the search space in future versions of our algorithm.

### 4.2. Evaluation and Results

We evaluated VPE Antecedent location by checking whether the main verb of the selected event overlaps with that of the annotated antecedent VP. This is the method that Hardt (1997) defines as "Head Match: the system choice and coder choice have the same head." Using this measure, the system achieves an accuracy of 73% on the development data, and 72% on the test data. These figures are in the same ballpark as those reported by Hardt, who reports a success rate of 62% for a baseline system based on recency only, and an accurracy of 84% for an improved system taking recency, clausal relations, parallelism, and quotation into account.

### 4.3. Problematic Cases

Choosing the closest VP is a strategy that will work in a lot of cases, but obviously not in all. Here we list some examples that our system got wrong, with the wrongly selected antecedents in square brackets. To start, examples illustrating that selecting the most recent antecendent event as antecedent for an elided VP is not always a good strategy, are (6)–(8).

(6) *You either believe Seymour can [do it again] or you **do**n't.* [WSJ section 00]

(7) *They say insurance companies use policies aimed at [excluding bad risks] because their competitors **do**.* [WSJ section 05]

(8) *Wells Rich declined to [comment on the status of the account], as **did** the other agencies.* [WSJ section 05]

It seems that, at least for these cases, selecting a VP as antecedent which is itself part of a larger VP, is not always a good idea, and choosing the outermost VP would have yielded the correct antecedent. In DRS terms, this rule could be implemented by disregarding events which are themselves are arguments of thematic roles of other events.

As Asher (1993) argues, for choosing a correct antecedent VP one often needs to take discourse structure into account. A case in point is (9), which our system got wrong by selecting the closest antecedent.

(9) *The arbs may recoup some of their paper losses if the UAL deal gets patched up again, as they **did** in 1982 when Occidental Petroleum Co. rescued them with a $4 billion takeover of Cities Service.* [WSJ section 14]

Although in most cases the antecedent of an elided VP is found in the preceeding text, it sometimes is found in the text following the ellipsis. Parallel to the usage of the term VP-anaphora for VPE, one could call such cases instances of VP-cataphora. We found only two cases of VP-cataphora in the corpus:

(10) *As they **did** when the Philippines was a colony of the U.S., teachers for the most part teach in English, even though it is a foreign language for most Phillipine children.* [WSJ section 08]

(11) *As she has **done** in the past, she stated her support for Mr. Lawson but insisted on keeping on an advisor who opposed and disparaged his policies.* [WSJ section 08]

Even though cases of VP-cataphora are rare, closer inspection might reveal whether there are certain syntactic constructions that license VP-cataphora: it is striking that both examples (10) and (11) start with the adverbial "as". Perhaps there are clear syntactic clues signalling VP-cataphora which are easy to integrate into the VPE antecedent location algorithm.

Finally, we consider a linguistically interesting example of VPE found in the corpus, repeated here as (12).

(12) *If Brazil devises an economic strategy allowing it to resume growth and service debt, this could lead it to open up and deregulate its sheltered economy, analysts say, just as Argentinian President Carlos Saul Menem has been **doing** even though he was elected on a populist platform.* [WSJ section 04]

It is perhaps not cristal clear what the correct antecedent is, but it certainly not the one found by the system. It seems to me that the antecedent is *allowing it to resume growth and service debt*. If this is so, this would be also an instance of a sloppy reading, but a special one, as the sloppy interpretation of the pronoun refers to Argentina, a concept only mentioned indirectly in the text by the noun phrase *Argentinian President*.

## 5. VP Ellipsis Resolution

### 5.1. Resolving Strategies

With the computational framework we have at our disposal, we can choose to resolve VPE either on the syntactic level (i.e., the level of CCG derivation) or on the semantic level (i.e., the level of DRS). While a comparison of these two possibilities would be an interesting topic of study, in the scope of this article, we choose to restrict ourselves to the latter option—that is, resolving elliptical VPs by reconstruction on the level of discourse representation structure.

This choice immediately raises another question: will resolution take place before or after resolution of anaphoric expressions, such as proper names, pronouns, and definite descriptions. We will argue that it is at least technically simpler to perform VPE resolution after resolving anaphoric expressions.

Various ways of reconstruction have been proposed in the literature: by predication or abstraction over the subject in the antecedent verb phrase (Sag 1976; Klein 1987), by employing higher-order resolution (Dalrymple, Shieber, and Pereira 1991), or by semantic abstraction over constituents (Asher 1993). We implement a novel copying and renaming approach, which is inspired by a method first proposed by Bäuerle (1988) within the framework of DRT.

### 5.2. A Copying and Renaming Algorithm

In order to explain our resolution method, we will need to introduce some new terminology. We will call the DRS associated with the source clause the **source DRS**, and the DRS in which the VPE is detected the **target DRS**. (For simplicity we assume that each clause corresponds to exactly one DRS.) The discourse referent of type event corresponding to the antecedent VP in the source DRS is called **source event**, and the discourse referent of the elided

VP **target event**. Finally, the semantic material to be copied in the process of VPE resolution is called the **antecedent DRS**.

Basically, the way our resolution algorithm works is by constructing an antecedent DRS based on the source DRS, abstract over the source event, and apply the result to the target event. What enters the antecedent DRS depends on the parallel relations between source and target DRS. The parallel relations denote a non-empty set of two-place relations, the **P-set**. The elements of the P-set are determined by the thematic roles in the target DRS, as well as by relations introduced by additional modifiers of the elided VP. For example, in Figure 3, the P-set is the singleton $\{agent\}$. But for an elided phrase with a temporal modifier such as *Bill did yesterday* the P-set would be $\{agent, temploc\}$.

The semantic material that enters the antecedent DRS are those relations which have the source event as internal argument but are not members of the P-set. In addition, discourse referents and properties belonging to external arguments could be copied to the antecedent DRS. We distinguish three cases for each external argument:

1. If it is declared in the domain of the source DRS: copy the discourse referent and associated conditions;

2. If it is declared in the global DRS: don't copy the discourse referent;

3. If it is part of a conditional DRS: copy the antecedent of the conditional.

Case 1 is triggered by anaphoric expression in the source clause such as proper names and definite descriptions that are accommodated or bound to the global DRS (see Figure 4). Case 2 applies to indefinite noun phrases in the source clause (Figure 5), and the third case to universally quantified noun phrases (Figure 6).

Note that renaming of variables is done using standard tools of $\alpha$-conversion and $\beta$-reduction, that are also used during semantic construction. This means that there is no need to rename any discourse referents when they are copied from the source DRS to the antecedent DRS; instead, renaming will be carried out as a side-effect of $\beta$-reduction after applying the antecedent DRS to the target DRS.

```
  _________________    _________________   _________________
 | x2 x1           |  | x4 x3           | | x6 x5           |
 |_________________|  |_________________| |_________________|
(| named(x2,john)  |+(| car(x4)         |+| car(x5)         |))
 | named(x1,bill)  |  | sell(x3)        | | too(x6)         |
 |_________________|  | event(x3)       | | event(x6)       |
                      | agent(x3,x2)    | | agent(x6,x1)    |
                      | patient(x3,x4)  | | sell(x6)        |
                      |_________________| | patient(x6,x5)  |
                                          |_________________|
```

*Figure 4.* DRS for *John sold a car. Bill did, too* as output by Boxer, performing VPE resolution. Note that both proper names are accommodated in the global DRS, and that the semantic material for *a car* has been copied to the target DRS.

```
  _________________    _________________   _________________
 | x3 x2 x1        |  | x4              | | x5              |
 |_________________|  |_________________| |_________________|
(| named(x3,john)  |+(| see(x4)         |+| see(x5)         |))
 | car(x2)         |  | event(x4)       | | too(x5)         |
 | named(x1,bill)  |  | agent(x4,x3)    | | event(x5)       |
 |_________________|  | patient(x4,x2)  | | agent(x5,x1)    |
                      |_________________| | patient(x5,x2)  |
                                          |_________________|
```

*Figure 5.* DRS for *John saw the car. Bill did, too* as output by Boxer, performing VPE resolution. Note that both proper names are accommodated in the global DRS. Note that the semantic material for *the car* has not been copied to the target DRS.

```
  _________________    ____________________________________   ____________________________________
 | x2 x1           |  |                                    | |                                    |
 |_________________|  |____________________________________| |____________________________________|
(| named(x2,john)  |+(|  __________      _________________ |+|  __________      _________________ |))
 | named(x1,bill)  |  | | x3       |    | x4              || | | x5       |    | x6              | |
 |_________________|  | |__________|    |_________________|| | |__________|    |_________________| |
                      | | car(x3)  | ==> | drive(x4)       || | | car(x5)  | ==> | patient(x6,x5)  | |
                      | |__________|    | event(x4)       || | |__________|    | too(x6)         | |
                      |                 | agent(x4,x2)    || |                 | event(x6)       | |
                      |                 | patient(x4,x3)  || |                 | agent(x6,x1)    | |
                      |                 |_________________|| |                 | drive(x6)       | |
                      |____________________________________| |                 |_________________| |
                                                             |____________________________________|
```

*Figure 6.* DRS for *John drove every car. Bill did, too* as output by Boxer, performing VPE resolution.

### 5.3. Sloppy Interpretations

Our algorithm always produces a strict reading for a pronoun appearing in the antecedent verb phrase. As sloppiness in VPE interpretation appears to be a true rarity in real data, this is hardly a concern for a system aiming for wide-coverage rather than theoretically accurate analyses. Indeed, in the corpus (financial newspaper articles) used for this study, we only found two examples in the development data (89 cases of VPE) that allow for sloppy interpretation:

(13) *IBM, though long a leader in the Japanese mainframe business, didn't introduce its first PC in Japan until five years after NEC **did**, and that wasn't compatible even with the U.S. IBM standard.* [WSJ section 04]

(14) *Neil Kinnock, Labor Paryy leader, dubbed the 46-year-old Mr. Major a lap dog unlikely to veer from his boss's strongly held views, as Mr. Lawson sometimes **did**.* [WSJ section 08]

## 6. Conclusion

We presented a wide-coverage NLP system for a deep semantic analysis of text and in particular its coverage and performance on analysing VP ellipsis. We conducted an evaluation of the system on examples of VPE annotated in the Wall Street Journal part of the Penn Treebank. The system detects VP ellipsis and resolves them on the level of logical form using a copying-and-renaming approach. It achieves high precision and recall on detecting elided VPs, and a reasonable accuracy on locating correct antecedents.

Nonetheless, clearly more research is required for improving the algorithm for VPE antecedent location, especially in dealing with VP-cataphora, and antecedents that are part of nested verb phrases. In addition, taking discourse structure into account to constrain possible antecedents for VPE, as suggested by Asher (1993), could improve the performance of the algorithm further.

The success of the resolution algorithm for VPE proposed in this paper depends mainly on the ability to determine parallel elements between source and target clause. We have claimed that this can be done on the basis of thematic roles. As detecting parallelism is a non-trivial job, a quantified assessment of our resolution algorithm is left as a topic for future work.

## Notes

1. Sometimes referred to as VP-deletion or VP-anaphora, especially in the earlier literature on VP ellipsis.
2. We limit ourselves to a small inventory of thematic roles, namely *agent*, *patient* and *theme*.
3. More formally, if $B_1$ and $B_2$ are DRS, then so is $(B_1+B_2)$. The merge is dynamic, in the way that discourse referents in $B_1$ bind free occurrences in $B_2$.
4. A large portion of this annotated data was kindly provided to me by Leif Nielsen, who used it in his thesis (Nielsen 2005).
5. Sections 00, 01, 02, 03, 04, 05, 06, 07, 08, 10, 12, and 14 were used as development data.
6. Sections 09, 11, 13 and 15 were used as test data.
7. The standard model used by the C&C parser is not trained on questions, and therefore often gets an unwanted analysis in wh-questions with do-support. A way to solve this problem is to select a different model, indeed a model trained on questions, when a question is given to the parser, but this idea will be left for future work.

## References

Asher, Nicholas
1993 *Reference to Abstract Objects in Discourse*. Dordrecht: Kluwer Academic Publishers.

Bäuerle, Rainer
1988 Ereignisse und Repräsentationen. Technical Report LILOG Report 43, IBM Deutschland GmbH.

Bos, Johan
2005 Towards wide-coverage semantic interpretation. *Proceedings of Sixth International Workshop on Computational Semantics IWCS-6*. 42–53.

Clark, S., and J.R. Curran
2004 Parsing the WSJ using CCG and Log-Linear Models. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics (ACL '04)*. Barcelona, Spain.

Dalrymple, Mary, Stuart M. Shieber, and Fernando C.N. Pereira
1991 Ellipsis and Higher-Order Unification. *Linguistics and Philosophy* 14: 399–452.

Hardt, Daniel
1997 An Empirical Approach to VP Ellipsis. *Computational Linguistics* 23 (4): 525–541.

Hockenmaier, J., and M. Steedman
2002 Generative Models for Statistical Parsing with Combinatory Categorial Grammar. *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics*. Philadelphia, PA.

Hockenmaier, Julia
2003 Data and Models for Statistical Parsing with Combinatory Categorial Grammar. Ph.D. diss., University of Edinburgh.

Kamp, H., and U. Reyle
1993 *From Discourse to Logic; An Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and DRT*. Dordrecht: Kluwer.

Kamp, Hans
1981 A Theory of Truth and Semantic Representation. In *Formal Methods in the Study of Language*, Jeroen Groenendijk, Theo M.V. Janssen, and Martin Stokhof (eds.), 277–322. Mathematical Centre, Amsterdam.

Klein, Ewan
1987 VP Ellipsis in DR Theory. *Studies in Discourse Representation Theory and the Theory of Generalised Quantifiers*.

Nielsen, Leif Arda
2005 A corpus-based study of Verb Phrase Ellipsis Identification and Resolution. Ph.D. diss., King's College London.

Sag, Ivan
1976 Deletion and logical form. Ph.D. diss., MIT.

Steedman, M.
2001 *The Syntactic Process*. The MIT Press.

Van der Sandt, R.A.
1992 Presupposition Projection as Anaphora Resolution. *Journal of Semantics* 9: 333–377.