

A unification-based approach to multiple VP Ellipsis resolution*

Claire Gardent

GRIL, Université de Clermont-Ferrand (France) and
Department of Computational Linguistics, Universiteit van Amsterdam
Spuistraat 134, 1012 VB Amsterdam (The Netherlands)
E-mail: claire@mars.let.uva.nl

Abstract

An assumption shared by many theories of discourse is that discourse structure constrains anaphora resolution (cf. [Grosz and Sidner 1986] for definite NPs, [Lascarides and Asher 1991], [Nakhimovsky 1988] for temporal anaphora, [Webber 1990] for deictic pronouns and [Gardent 1991], [Prüst and Scha 1990] for VP ellipsis). The aim of this paper is (i) to show that this assumption also applies to multiple VP ellipsis (VPE), (ii) to argue that other levels of linguistic information (such as syntax and semantics) interact with discourse structure in determining multiple VPE acceptability and (iii) to make these intuitions precise by providing a unification-based account of multiple VPE resolution.

1 Introduction

[Klein and Stainton-Ellis 1989] convincingly argue that VPE need not resolve to the nearest possible antecedent. The most intricate examples they give to support this claim involve what they dubbed **multiple VPE** and can be illustrated by the following discourses (square brackets surround antecedent VPs, \emptyset_i indicates VP ellipses and indices represent anaphoric dependencies)¹:

*The work reported here was partially carried out in the LRE Project 61-062, *Towards a declarative theory of discourse*.

¹Although this data often raises suspicion among linguistic audiences as to its credibility, the facts are that (1) it is real life data and (2) it can be understood and it is usually understood in an unambiguous fashion. Hence

- (1) I promised myself I [₁ wouldn't go to Manchester] unless I first [₂ opened a big stack of mail]. I didn't \emptyset_2 , so I didn't \emptyset_1 . (*Nesting*)
- (2) If you [₁ work hard, make the right choices and keep your nose clean], you [₂ get ahead]. If you don't \emptyset_1 , you don't \emptyset_2 . (*Crossing*)
- (3) I was [₁ really thin] then, and I tried some ski-pants that [₂ looked really good on me], and I [₃ should have bought them]. But I didn't \emptyset_3 , and now I 'm not \emptyset_1 and they wouldn't \emptyset_2 . (*Mixed*)

As these examples show, there is not one pattern relating multiple VPEs to their antecedents, but at least three: nesting, crossing and mixed. Nesting and crossing can be represented as follows (where VP_i and \emptyset_i represent antecedent and elliptical VPs respectively):

Nesting: $VP_1 \dots VP_n \ \emptyset_n \dots \emptyset_1$

Crossing: $VP_1 \dots VP_n \ \emptyset_1 \dots \emptyset_n$

while a **mixed** pattern simply is a configuration in which both crossing and nesting occur. According to this terminology, (1) illustrates a nesting pattern, (2) shows a crossing pattern and (3) a mixed pattern. Thus, it is clear that no unique dependency configuration constrains the resolution of multiple VPEs. On the contrary, it appears that all patterns are possible and thus that any configurational restriction on VPE resolution is doomed to failure. Interestingly however, despite the multiple ways in which each of the VPEs could be resolved, there is in actual fact no ambiguity as to how the global discourse should be understood. This suggests that some strong constraints come into play to help the hearer resolve

the question: What is it that constrains multiple VPE resolution in such a way that these "exotic" discourses are in fact intelligible?

adequately. In what follows, I argue that discourse structure (rather than surface ordering) is one of the main constraint regulating multiple VPE resolution.

2 Discourse grammar and VPE resolution

The discourse grammar used builds on [Polanyi and Scha 1984]. More specifically, I assume that discourse is a tree structured entity whose well formedness can be described by a unification based discourse grammar. Under such a grammar, a discourse constituent is either a discourse relation, a clause or a discourse relation together with one or more discourse constituent(s). The grammar associates with each constituent a complex category which for the purpose of this paper, I will assume to consist of the six main attributes PHO, CAT, SEM, IN, OUT and RESTR. PHO, CAT, SEM unsurprisingly denote the phonology, the category and the semantic representation of the constituent described by the complex category. IN and OUT are attributes which represent the flow of anaphoric information, that is, IN represents the in-going context (where a context is a sequence of potential antecedents i.e. a sequence of VP categories) and OUT, the out-going context. Finally RESTR is short for restriction and takes as value a constraint which must evaluate to true for the category to be well-formed.

Conventions: In what follows I will omit any information that is not relevant to the purpose of the discussion. In particular, I shall omit irrelevant attributes in categories and any anaphoric information not pertaining to VPE (i.e. anaphoric pronominal information is ignored). Furthermore, the values of IN and OUT attributes (which should be VP categories) will be abbreviated to the SEM values of these categories. Finally, I will use the term *a*-clause as an abbreviation for antecedent clause and *e*-clause, for elliptical clause.

A simple example will illustrate the workings of the discourse grammar with respect to VPE resolution. Consider the discourse in (4).

- (4) (a) Jon [₁ likes Mary] (b) and (c) Peter does \emptyset_1 too.

As indicated by the bracketed letters, this discourse includes three basic discourse constituents: the two clauses (a) and (c) and the discourse connective *and*. Consider first the category associated with (a). Ignoring irrelevant attributes, this category can be represented as follows²:

²For expository purposes, I assume here a sentential (rather than a discourse) semantics. In practice, however, the analysis is to be based on a discourse semantics and most importantly, the definitions of structural identity and of equivalence classes over relations (see below) are to apply to discourse semantics representations and to discourse relations respectively.

$$\boxed{a} \begin{bmatrix} \text{SEM} & \text{like:[j,m]} \\ \text{IN} & \text{ } \\ \text{OUT} & \text{[like:[m]]} \end{bmatrix}$$

With regard to VPE resolution, two points are relevant. First the IN value is a don't-care-value (symbolized here by the anonymous variable), thus signaling the fact that incoming anaphoric information is irrelevant in the case of non-elliptical clauses. Second, the OUT value contains the information associated with the sentence main VP thus signalling the fact that non elliptical clauses update the current outgoing context with new information. Note that anaphoric information concerning VPs is here assumed *not* to be cumulative, that is the OUT value of \boxed{a} is not "added" to the IN value - rather it constitutes the sole output of (a) independent of the preceding context. The intuition formalised here is that the discourse entity providing the interpretation of an elided VP is not as persistent as an individual discourse entity and thus should remain local to the discourse constituent that introduced it (although in some particular cases such as e.g. parallelism, anaphoric information pertaining to VPs can be percolated by the discourse grammar rules). For more details on this point, the reader is referred to [Gardent 1991], pages 141-142.

Now consider the category assigned by the discourse grammar to the elliptical clause (c). Again ignoring irrelevant attributes, this category can be represented as:

$$\boxed{c} \begin{bmatrix} \text{SEM} & \text{R:[p | As]} \\ \text{IN} & \text{[R:As]} \\ \text{OUT} & \text{[R:As]} \end{bmatrix}$$

where R and As are unification variables over relations and arguments respectively. The important point to note here is that the variables R and As are shared by the IN value on the one hand, and by the SEM value on the other. This in effect implements VPE resolution. To see this, suppose that we have a discourse rule of the following form (AND abbreviates the category for *and*):

$$\begin{bmatrix} \text{SEM} & \text{Sem}_1 \\ \text{IN} & \text{In} \\ \text{OUT} & \text{Out}_1 \end{bmatrix}, \text{ AND }, \begin{bmatrix} \text{SEM} & \text{Sem}_2 \\ \text{IN} & \text{Out}_1 \\ \text{OUT} & \text{Out}_2 \end{bmatrix} \\ \Rightarrow \begin{bmatrix} \text{SEM} & \text{and:[Sem}_1, \text{Sem}_2] \\ \text{IN} & \text{In} \\ \text{OUT} & \text{[Out}_1, \text{Out}_2] \end{bmatrix}$$

Application of this rule to the categories of (a), (b) and (c) above will trigger the unification of Out₁ with [like:[m]] on the one hand, and [R:As] on the other. Thus [R:As] is unified with [like:[m]] and the semantic representation R:[p | As] of (c) will become like:[p,m], just as required.

3 Discourse structure and Multiple VPE resolution

3.1 Some data

The claim this paper makes about multiple VPE resolution is that the same discourse relation must hold between the multiple VP ellipses on the one hand and the multiple antecedents on the other. The present section has for object to substantiate this claim. As a first case in point, consider the following example.

- (5) I never go swimming because I don't look good in a swimming suit. (*causal*)
- a. I might if I did. (*causal*)
 - b. If I did, I probably would. (*causal*)
 - c. Sarah does and so she does. (*causal*)
 - d. ? I might after I did. (*temporal*)
 - e. ? I might but I did. (*contrast*)

Example (5) gives a case of a-clauses which are related by a causal relation. Several possible continuations are then given, some of them are acceptable, some of them are not. The relevant observation is that in those cases where the relation holding between e-clauses also is a causal one, the continuation is acceptable; however, in those cases where the relation holding between e-clauses is of a different nature, the continuation is unacceptable.

As a second case in point, consider example (6):

- (6) I was thin then and the trousers looked good on me and I should have bought them,
(*THIN* → *LG*) ∧ *BT*
- a. but I didn't and now I am not and they wouldn't.
 $\neg BT \wedge (\neg THIN \rightarrow \neg LG)$
 - b. but now I am not and they wouldn't and anyway I didn't.
 $(\neg THIN \rightarrow \neg LG) \wedge \neg BT$
 - c. ? but now I am not and I didn't and they wouldn't.
 $\neg THIN \wedge \neg BT \wedge \neg LG$

Here the antecedent discourse unit consists of three clauses, the first two can be said to be related by a causal relation (*because I was thin, the trousers looked good on me*) whereas the third clause is conjoined to the first two. Again, several possible continuations are given, some of them are acceptable, some of them are not. This time, the observation is that in the case where no causal relation can be established between the appropriate e-clauses (i.e. when those clauses corresponding to the cause and to the result of the cause are not adjacent), the continuation is unacceptable³. That is, in the case where an identical relational pattern cannot be established

³This observation was originally made in [Stainton-Ellis 1988], page 75.

for e- and a-clauses, multiple VPE becomes hard to understand, if not unacceptable.

In what follows, I take these examples to suggest that the same discourse relation must hold between a- and e-clauses respectively. I characterise this observation in terms of parallelism, make this notion precise and show how it interacts with other grammar components (e.g. syntax and semantic) to determine multiple VPE resolution. It should be stressed however that the approach can only be as precise as the definition of discourse relations and unfortunately, this notion is notoriously elusive. Nonetheless the hope is that this paper captures an important intuition about multiple VPE resolution namely, the intuition that parallelism constitutes one of the (many) factors affecting multiple VPE acceptability and interpretation.

3.2 Formal analysis

Assuming a discourse grammar of the type described in section (2), the claims this paper makes about multiple VPE resolution are (i) that whenever a discourse contains multiple VPEs, the clauses containing the VPEs and those containing the antecedent VPs form two complex discourse constituents which are related together by the relation of parallelism and (ii) that parallelism constrains VPE resolution in that each VPE will resolve to the "parallel VP" in the complex discourse constituents formed by the a-clauses. We now make these claims precise. First, we define the semantic representation language \mathcal{L} used by the grammar described in section (2). \mathcal{L} consists of the wffs described by the following syntax:

wff	→	{ term, formula, polarity:rel:[wff ₁ ... wff _n]
term	→	{ variable, constant }
formula	→	polarity:predicate:[arg ₁ ... arg _n]
arg	→	{ term, formula }
rel	→	constant
predicate	→	constant
polarity	→	{ 1, 0 }

The intuition is that \mathcal{L} is a quantifier free language where variables are unification variables and polarity (i.e. absence or presence of negation) is always explicit (that is, non-negated wffs are described as positive i.e. marked with 1). Thus for instance, the expression 0:and:[0:p, 1:q] is a wff of \mathcal{L} , which one can think of as the more traditional propositional logic formula $\neg(\neg p \wedge q)$. We call *PROP* the set of wffs of the form polarity:predicate:[arg₁ ... arg_n]⁴. Given this language \mathcal{L} , the discourse relation of *parallelism* is said to hold between two propositions represented by the \mathcal{L} wffs Φ and Ψ (written, *parallelism*(Φ, Ψ)) iff Φ is structurally identical with Ψ . Structural identity is defined as follows:

⁴Note that contrary to tradition *negated* propositions are assumed to be atomic wffs.

Definition 1 (Structural identity between \mathcal{L} formulae)

If $\Phi, \Psi \in \mathcal{L}$, then Φ is structurally identical (or *s-identical*) with Ψ (written $\Phi =_s \Psi$) if:

- (i) $\Phi, \Psi \in PROP$
- or (ii) $\Phi = [\Phi_1 \dots \Phi_n], \Psi = [\Psi_1 \dots \Psi_n]$
 $\Phi_1 =_s \Psi_1$ and $\dots \Phi_n =_s \Psi_n$
- or (iii) $\Phi = \rho_1 : \Phi_1, \Psi = \rho_2 : \Psi_1,$
 $\rho_1 = \rho_2$ and $\Phi_1 =_s \Psi_2$

That is, structural identity is identity up to propositional level (where negation is taken to be part of propositional information). To give two simple examples:

$$1 : p =_s 0 : q$$

and

$$1 : \text{implies}[1 : p, 0 : q] =_s 1 : \text{implies}[1 : r, 0 : s]$$

To state the constraint regulating multiple VPE resolution, we first define the notion of a *yield*.

Definition 2 (Yield)

If $\Phi \in \mathcal{L}$, then the yield of this semantic representation Φ , written $\mathcal{Y}(\Phi)$, is:

- If $\Phi \in PROP$, $\mathcal{Y}(\Phi) = \langle \Phi \rangle$
- If $\Phi = [\Phi_1, \dots, \Phi_n]$, $\mathcal{Y}(\Phi) = \mathcal{Y}(\Phi_1) \cdot \dots \cdot \mathcal{Y}(\Phi_n)$
 where \cdot denotes
 sequence concatenation
- If $\Phi = \rho : \Phi_1$ $\mathcal{Y}(\Phi) = \mathcal{Y}(\Phi_1)$

Thus the yield of an \mathcal{L} wff Φ consists of the sequence of atomic propositions contained in Φ . Finally, we state the constraint as follows:

Definition 3 (Constraint on multiple VPE resolution)

Let Φ be the semantic representation associated with the discourse segment formed by the *a*-clauses and Ψ be that associated with the *e*-clauses. Then, if

$$\begin{aligned} \Phi &= \Psi, \\ \mathcal{Y}(\Phi) &= \langle Pol_1 : \mathcal{P}_1 : [s_1 | ss_1], \dots \\ &\quad , Pol_2 : \mathcal{P}_n : [s_n | ss_n] \rangle \text{ and} \\ \mathcal{Y}(\Psi) &= \langle Pol_3 : \emptyset_1 : [t_1 | tt_1], \dots \\ &\quad , Pol_4 : \emptyset_n : [t_n | tt_n] \rangle, \\ \text{then} &\quad \text{for } 1 \leq i \leq n, \emptyset_i = \mathcal{P}_i \text{ and } ss_i = tt_i. \end{aligned}$$

That is, each elided predicate \emptyset_i and argument list tt_i ⁵ in $\mathcal{Y}(\Psi)$ resolves to the parallel predicate \mathcal{P}_i and argument list ss_i in $\mathcal{Y}(\Phi)$. To see how this constraint works, consider example (1). Suppose that the discourse grammar assigns to the *a*- and the *e*- part of this discourse the following (simplified) semantic representations:

$$\begin{aligned} \text{A-clauses:} & \quad 0:\text{and}:[0:\text{OM}:[i], 1:\text{GtM}:[i]] \\ \text{E-clauses:} & \quad 1:\text{and}:[0:\text{R}_1:[i], 0:\text{R}_2:[i]] \end{aligned}$$

⁵The first argument in the list corresponds to the subject NP and is thus ignored.

Then definition 3 adequately predicts that $R_1 = \text{OM}$ and $R_2 = \text{GtM}$. That is, the constraint embodied in definition 3 implements the fact that multiple VPE resolution is sensitive to the semantic- rather than to the surface-ordering of the antecedents.

3.3 Implementation

The above analysis can be implemented in the discourse grammar described in section (2) as follows. The parallelism rule will be:

$$\left[\begin{array}{cc} \text{SEM} & \text{SEM}_1 \\ \text{IN} & \text{IN} \\ \text{OUT} & \text{OUT}_1 \\ \text{RESTR} & - \end{array} \right] \left[\begin{array}{cc} \text{SEM} & \text{SEM}_2 \\ \text{IN} & \text{OUT}_1 \\ \text{OUT} & \text{OUT}_2 \\ \text{RESTR} & - \end{array} \right]$$

$$\Rightarrow \left[\begin{array}{cc} \text{SEM} & 1:\text{parallelism}:[\text{SEM}_1, \text{SEM}_2] \\ \text{IN} & \text{IN} \\ \text{OUT} & [\text{OUT}_1, \text{OUT}_2] \\ \text{RESTR} & \text{SEM}_1 =_s \text{SEM}_2 \end{array} \right]$$

This rule has two effects. First, it requires that the semantic representations of the constituting discourse constituents be *s*-identical – this implements the restriction stated in defining parallelism. Second, it unifies the *OUT* value of the first discourse constituent with the *IN* value of the second – this ensures that the antecedents provided by the first (possibly complex) discourse constituent are accessible to any VPEs occurring in the second constituent. Now consider the rule for the connective *unless* (where *UNLESS* abbreviates the category associated with *unless*):

$$\left[\begin{array}{cc} \text{SEM} & \text{SEM}_1 \\ \text{IN} & \text{IN}_1 \\ \text{OUT} & \text{OUT}_1 \end{array} \right], \text{UNLESS}, \left[\begin{array}{cc} \text{SEM} & \text{SEM}_2 \\ \text{IN} & \text{IN}_2 \\ \text{OUT} & \text{OUT}_2 \end{array} \right]$$

$$\Rightarrow \left[\begin{array}{cc} \text{SEM} & 0:\text{and}:[0:\text{SEM}_2, 0:\text{SEM}_1] \\ \text{IN} & [\text{IN}_1, \text{IN}_2] \\ \text{OUT} & [\text{OUT}_2, \text{OUT}_1] \end{array} \right]$$

Note that the order of the resulting *OUT* value is $[\text{Out}_2, \text{Out}_1]$ (and not $[\text{Out}_1, \text{Out}_2]$ as suggested by the surface ordering). This reflects the fact that multiple VPE resolution is sensitive to the logical- rather than the surface-ordering of its antecedents. Application of the *UNLESS* rule to the *a*-clauses⁶ (*I wouldn't go to Manchester unless I open my mail*) in example (1) will yield the category (recall that irrelevant attributes and attribute values are omitted):

$$\left[\begin{array}{cc} \text{SEM} & \boxed{1} \quad 0:\text{and}:[0:\text{open}:[i, \text{mail}], 0:\text{go}:[i, \text{toM}]] \\ \text{IN} & \boxed{2} \\ \text{OUT} & \boxed{3} \quad [[\text{open}:[\text{mail}]], [\text{go}:[\text{toM}]]] \end{array} \right]$$

⁶Here, I do not consider the problem raised by the embedding clause *I promised myself that*.

5.1 Syntax

Consider again example (8) where the discourse formed by the e-clauses is of the form *If P, Q* and the associated semantic representation may be either $p \rightarrow q$ or $\neg q \rightarrow \neg p$. Now look at the syntax of antecedent and elliptical VPs. The first elliptical VP is the perfective auxiliary *has* and thus subcategorises for a past participle whereas the second ellipsis consists of copula *be* and thus selects a predicative phrase. Correspondingly, the antecedent VPs are (1) a predicative phrase (*lucky*) and (2) a past participle (*ordered his software from a house that can help*). If we assume that VPE acceptability is sensitive to the syntactic information associated with the antecedent, then the above observations explain why the discourse relation holding between a- and e-clauses must be $\neg q \rightarrow \neg p$ rather than $p \rightarrow q$. For in the first case *hasn't* indeed resolves to a past participle (namely *ordered his software from a house that can help*) and *isn't* to a predicative phrase (i.e. *lucky*); whereas in the second case, the subcategorisation requirements of the auxiliaries are systematically violated. Thus if we assume that the (or at least some) syntactic properties of the antecedent VPs are relevant in determining VPE acceptability, then we can account for the fact that despite of the ambiguity introduced by semantic equivalences between discourse relations, there is only one reading for (8) i.e. the reading which is compatible both with the discourse requirement of parallelism between a- and e-clauses and with the syntactic constraints between antecedent and elliptical VP. As already mentioned (cf. section 2), the present discourse grammar makes precisely this assumption since it takes anaphoric information to be sequences of VP categories i.e. feature structures containing *inter alia* syntactic information about admissible antecedent VPs.

5.2 Semantics

[Sag 1980] argues that VPE is subject to a constraint on semantic representations, which is dubbed the *alphabetical variant constraint*. The analysis is convincing in that it accounts for a wide range of facts about VPE and its interaction with other linguistic phenomena such as quantification, extraction, pseudo-clefts, *ready* constructions and equisentences. For instance, the alphabetic variant constraint will account for the inacceptability of (9)⁸:

- (9) If every boy thinks that Mary is in love with him, the party will be a success. * If they don't, it won't.

Note that in this case, discourse parallelism does hold between a- and e-clauses. So if discourse parallelism (as defined in this paper) was taken to be the only constraint regulating VPE acceptability, this

⁸To be compared with the well formed: *If every boy brings a bottle, the party will be a success. If they don't, it won't.*

(ill formed) discourse could not be rejected by the grammar. However, if Sag's constraint is assumed then the ill-formedness of (9) can be accounted for as follows. Sag's constraint states that VPE is acceptable iff the semantic representation of the antecedent VP (which he assumes to be a lambda abstraction over individuals) is identical up to renaming of bound variables with the semantic representation of the ellipsis and furthermore, all occurrences of a free variable occurring both in the representation of the antecedent and of the ellipsis are bound by the same operator. Given this, the ill-formedness of (9) is explained by the fact that the pronoun *him* is represented by a variable (say, *y*) which is free in the semantic representation associated with the antecedent VP (i.e. $\lambda x.think(x,love(m,y))$) and cannot be bound by the same operator (i.e. the universal quantifier introduced by the subject NP *every boy*) when occurring in the semantic representation of the elliptical VP (because it occurs outside the scope of *every*).

Here again, the assumption that the antecedent of a VPE is represented by a monostratal category means that Sag's alphabetic variant constraint can easily be integrated in the present account. This can be done in two ways. The first possibility consists in adopting Sag's view and adding a constraint in the category associated with VP ellipsis auxiliaries to the effect that the semantic representation of the antecedent VP and that of the ellided VP must be alphabetic variants of each other. This has the inconvenience of requiring a global check over the semantic representation of the whole discourse segment containing a- and e-clauses, a check which is essentially non compositional in nature⁹. A second possibility is to adopt a dynamic semantics (i.e. a semantics where meaning is taken to be a relation between contexts and where a context contains information about pronoun denotations). Under such an assumption, it can be shown that the inacceptability of any discourse violating the alphabetic variant constraint comes out as a failure to interpret this discourse (model theoretic interpretation simply fails) so that the semantic representation of a- and e-clauses need not be checked upon. Such an approach is described in [Gardent 1990] and could easily be integrated in the present framework: it suffices to replace the static semantics whose syntax is described in 3, by the dynamic semantics given in [Gardent 1990].

5.3 Pragmatics

Just as syntax and semantics, pragmatics can interact with discourse constraints to determine multiple VPE acceptability. A particularly clear illustration of this interaction comes from the pragmatics of discourse connectives i.e. words such as *but*, *unless*, etc. Consider for instance the discourse in (10).

⁹For more details concerning this point, see [Gardent 1990].

- (10) I gave her some questions to ask you if you rang her.
- a. I did but she didn't.
 - b. * I did but she did.

Although both continuations can be viewed as parallel to the a-clauses (cf. section 6), only continuation (a) is acceptable. Continuation (b) is unacceptable because the pragmatics of *but* (which requires some contrastive relation to hold between the propositions it relates) is violated.

The discourse grammar sketched here does not integrate pragmatic information and thus cannot account for the difference in acceptability between (a) and (b). Whether it can be extended to do so remains an open question although recent work in pragmatics (such as [Elhadad and McKeown 1990]) suggests that the monostratal, unification based approach to discourse grammar is fully compatible with a comprehensive treatment of the semantics and pragmatics of discourse connectives.

6 Taking stock

While section (3) argues that multiple VPE resolution is subject to the discourse constraint of parallelism, section (5) shows that it is also sensitive to other linguistic components such as syntax and semantics. The present section (i) discusses how the resulting overall analysis accounts for the examples given so far, (ii) introduces some additional data and (iii) summarises how the various linguistic modules interact in determining VPE acceptability for the set of cases presented throughout the paper.

We start by examining the examples given so far. Examples (2) and (5a) are simple cases of discourse parallelism where a- and e-clauses translate to the same canonical LF and no extraneous factor blocks resolution so that each VPE resolves to the parallel element in the antecedent discourse constituent. Example (3) is more intricate and can actually be explained in two different ways. A first possibility is to assume that *I should have bought them* and *but I didn't* form a discourse constituent and, *I was really thin and ... the ski-pants looked really good on me* and *now I'm not and they wouldn't* another (the intuition here would be that discourse constituents reflect the temporal structure of discourse, that is, temporally related events must be part of the same discourse constituent). Under this first hypothesis, we have on the one hand a case of (single) VP ellipsis where *but I didn't* resolves to *I didn't buy them* and on the other hand a simple case of parallelism between complex discourse constituents¹⁰. The second possibility is to consider that the three a-clauses form a discourse constituent which is parallel with the discourse constituent formed by the three e-clauses. In

¹⁰Thanks to an anonymous referee for pointing out this possible interpretation.

this case, the semantic representations of a- and e-clauses can be symbolised as:

$$\begin{aligned} \text{A-clauses: } & (T \rightarrow LG) \wedge BT \\ \text{E-clauses: } & \emptyset_1 \wedge (\emptyset_2 \rightarrow \emptyset_3) \end{aligned}$$

This clearly does not obey parallelism. In this case, syntax imposes the choice of an equivalent LF (i.e. $(\emptyset_2 \rightarrow \emptyset_3) \wedge \emptyset_1$). As in (8), this syntactic constraint stems from the subcategorisation requirement of a VPE auxiliary, namely '*m not*' which requires a predicative phrase as antecedent.

For completeness, consider now the following additional examples.

- (11) I gave her some questions to [₁ ask you] if you [₂ rang her]. I did \emptyset_2 but she didn't \emptyset_1 .
- (12) It was preposterous. It [₁ couldn't possibly work]. There [₂ must have been some other precautions]. But there weren't \emptyset_2 and it did \emptyset_1 .
- (13) *Xenophobia pestis*, like the hard native perennial it is, bourgeons as lordly young Mediterranean male cyclists sail into oncoming traffic with such signorial arrogance that even as we swear and skid, we look round wildly for street signs to see if he [₁ 's right], and we [₂ are wrong] and the one-way system [₃ 's undergone one of its periodic reversals]. (He isn't \emptyset_1 . We aren't \emptyset_2 . It hasn't \emptyset_3 .)

(11) illustrates a case where parallelism constrains the choice of an alternative semantic representation with the result that the a-clauses semantics is represented by a wff of the form $\neg(p \wedge \neg q)$ rather than the canonical semantic translation for discourses of the form *If P, Q* i.e. $p \rightarrow q$. Example (12) provides one more illustration of the interaction of syntax with discourse in determining multiple VPE resolution whereas example (13) illustrates a simple case of discourse parallelism.

The following table summarises these observations. The first column (Ex.) indicates the number of the example being referred to together with a mention of the linguistic module, if any, which forced the choice of an equivalent semantic representation: *D* stands for Discourse and *S* for syntax. The second column (Canonical LF) indicates the "canonical" semantic representations (or Logical Forms) of a- and e-clauses: a-clauses are represented by capital letter abbreviations which are mnemonic for their propositional content, whereas the semantics of elliptical clauses is represented by \emptyset_i where *i* reflects surface ordering. Finally, the third column indicates an equivalent semantic representation for both e- or a-clauses (or none when this is superfluous). The intuition is that this column also indicates anaphoric dependencies whereby it indicates for each ellipsis which is the parallel element in the final semantic representation of the a-clauses. To take an example, consider the discourse in (1). For this discourse

the table indicates that discourse forces the choice of a non-canonical semantic representation for the a-clauses. That is, the choice of the non-canonical semantic representation is determined in this case by the discourse requirement that a- and e-clauses stand in a parallelism relation. As a result, each ellipsis will resolve to its parallel element in the equivalent LF (rather than the canonical one) i.e. θ_1 resolves to *OM* (i.e. *open a big stack of mail*) and θ_2 to *GtM* (i.e. *go to Manchester*).

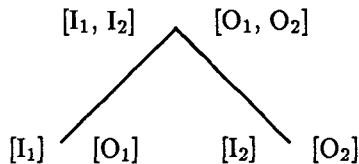
Ex.	Canonical LF	Equivalent LF
1 _D	$\neg OM \rightarrow \neg GtM$	$\neg(\neg OM \wedge GtM)$
	$\theta_1 \wedge \theta_2$	$\theta_1 \wedge \theta_2$
2	$WH \rightarrow GA$	
	$\theta_1 \rightarrow \theta_2$	
3 _S	$(T \rightarrow LG) \wedge BT$	$(T \rightarrow LG) \wedge BT$
	$\theta_1 \wedge (\theta_2 \rightarrow \theta_3)$	$(\theta_2 \rightarrow \theta_3) \wedge \theta_1$
5a	$LG \rightarrow GS$	
	$\theta_2 \rightarrow \theta_1$	
6 _S	$L \rightarrow OS$	$L \rightarrow OS$
	$\theta_1 \rightarrow \theta_2$	$\theta_2 \rightarrow \theta_1$
11 _D	$RH \rightarrow AY$	$\neg(RH \wedge \neg AY)$
	$\theta_1 \wedge \neg \theta_2$	$\theta_1 \wedge \neg \theta_2$
12 _S	$W \wedge P$	$W \wedge P$
	$\theta_1 \wedge \theta_2$	$\theta_2 \wedge \theta_1$
13	$R \wedge W \wedge UPR$	
	$\theta_1 \wedge \theta_2 \wedge \theta_3$	

7 Problems and further research

A first problem concerns the propagation of anaphoric information throughout the discourse tree. To see what the problem is, consider the discourse in (14).

(14) Jon won't dance unless Mary does.

In the absence of any additional context, the antecedent of the VPE in the second clause is the VP of the first clause i.e. *dance*. Now let us examine again the discourse rule for *unless* sketched in section 3. For this rule, the distribution of anaphoric information can be pictured as follows:



Note that anaphoric information is only shared between mother and daughters, not between sisters. This means that the rule sketched in section 3.3 will fail to resolve the VPE in example (14) because in

this case, resolution can only obtain if $O1 = I2$ i.e. if anaphoric information is shared between sisters. An obvious fix would be to modify the *unless* rule so that I_2 unifies not only with the IN value of the rightmost daughter but also with the OUT value of the leftmost daughter. The modified rule would then be:

$$\left[\begin{array}{cc} \text{SEM} & \text{SEM}_1 \\ \text{IN} & I_1 \\ \text{OUT} & \boxed{1} \end{array} \right], \text{UNLESS}, \left[\begin{array}{cc} \text{SEM} & \text{SEM}_2 \\ \text{IN} & \boxed{1} \\ \text{OUT} & O_2 \end{array} \right]$$

$$\Rightarrow \left[\begin{array}{cc} \text{SEM} & 0:\text{and}:[0:\text{SEM}_2,0:\text{SEM}_1] \\ \text{IN} & [I_1, \boxed{1} I_2] \\ \text{OUT} & [O_2, O_1] \end{array} \right]$$

However, although this would solve the problem raised by example (14), it would still fail to account for cases such as (15).

(15) (a) Jon won't [₁ dance] unless (b) Mary does θ_1 and (c) Bob won't [₂ come] unless (d) Sarah does θ_2 .

Here the problem is that the new *unless* rule requires the IN value of (d) to unify both with the OUT value of (b) i.e. *dance* and with the OUT value of (c) i.e. *come*. Clearly unification fails and thus example (15), although perfectly well-formed, is rejected by the grammar.

In more general terms, the problem is that anaphoric information can come to be instantiated both in a top-down and in a bottom-up fashion (i.e. through sharing of information between mother and daughter or through sharing of information between sisters)¹¹. When the two types of information conflict, unification fails and a perfectly well formed discourse may be rejected by the grammar. In other words, the grammar will undergenerate.

There are several possible solutions to this problem. A first one would be to privilege one source of information over the other, say by means of priority union. In this way, one anaphoric flow would overwrite the other. But apart from the computational problems involved in using such rewrite operations at run time, it is also unclear which information should be privileged. Thus although in (15), bottom-up (or local) information seems to prevail, example (16) shows that in some cases, top-down information may be strongest:

(16) (a) Jon won't go to Manchester unless (b) he opens his mail and (c) Bob won't go to Paris unless (d) he does.

¹¹The first type of anaphoric flow is top-down in that anaphoric information on the mother may be required to unify with the anaphoric information of some other node *higher up* in the discourse tree, whereas the second type is bottom-up because the anaphoric information specified on the sisters may in turn be required to unify with the anaphoric information carried by some other node *lower down* in the discourse tree.

Here, there is at least one reading where the ellipsis in (d) resolves to the parallel element (b) (i.e. *opens his mail*) rather than to the immediately preceding VP (i.e. *go to Paris*). Furthermore it is easy to find cases where the overall discourse is ambiguous between a "top-down reading" and a "bottom-up" one. Thus perhaps a better solution would be to always allow both possibilities and to let the various modules of the grammar decide which reading is actually available. The details and the adequacy of such an approach, I leave here as an open research question.

A second problem concerning the present paper concerns the definition of discourse relations and of equivalence classes over discourse relations. Here it is perhaps worth stressing that although logical connectives have been used throughout the paper to represent discourse relations, these are definitely not a sufficient means of characterization. As a simple case in point, consider a natural language discourse of the form $P \text{ so } Q$. In section 3.2, such a discourse is translated as $p \wedge q$ (where p and q represent the propositional content of the natural language discourses P and Q respectively). Clearly this translation does not exhaust the meaning of the discourse connective *so*: for instance, the causal link between p and q is not accounted for. More generally, it is clear that much work remains to be done on the semantics of discourse relations before the present analysis of multiple VPE resolution can be adequately tested.

Finally, a third question involves the interaction of discourse grammar with anaphora resolution in general. As already mentioned, the resolution of most types of anaphora can be argued to be influenced by discourse structure. It would be interesting to investigate in how far the various mechanisms developed to express this constraint are compatible. More specifically, it would be interesting to see whether the discourse grammar sketched in section 2 could be made to account for the complex interaction of VPE with other anaphoric phenomena such as strict/sloppy identity, pronominal and temporal anaphora.

8 Conclusion

A model has been proposed of how discourse structure influences multiple VPE resolution. However, the suggestion is that the analysis generalises to all cases of VPE, that is, that discourse structure is one of the main factors determining VPE resolution in general. In this sense, the analysis proposed here fits well with one of the mainstream ideas in discourse theory, which is that discourse structure constrains anaphora resolution. It should also be pointed out that this analysis includes a treatment of parallelism similar to that developed in [Asher forthcoming] and is as such likely to be compatible with the treatment of sloppy/strict ambiguity proposed there.

The model proposed is characterised by two main

properties: reversibility and monostratality. It is **reversible** because it is characterised in a purely declarative manner. Note in particular that the definition of structural identity is entirely independent of any notion of processing and is as such strictly declarative. In practical terms, this means that this model can be used both for analysis and for generation. **Monostratality** (i.e. the fact that different levels of linguistic information can be stated within a category) is another important aspect of the model in that it allows for different knowledge sources to interact in determining VPE acceptability and resolution. A typical example of this interaction is involved in the treatment of cases of multiple VPE involving semantically equivalent wffs: in such cases, syntax often interacts with discourse information to determine the correct resolution. More generally, it can be argued that VPE is a phenomenon which simultaneously involves phonology, syntax, semantics and discourse (cf. [Lappin and McCord 1990], [Gardent 1991]). The present model allows for such a simultaneous interaction and thus improves on serial models of VPE resolution (i.e. models where the various levels of linguistic information interact in a serial rather than a simultaneous fashion) such as [Webber 1978].

The model described in this paper has been implemented in SICSTUS PROLOG and runs on a SUN 4 computer. It has been tested in analysis as well as in generation mode.

Acknowledgements: I would like to thank Martin van den Berg, Patrick Blackburn, Remko Scha and Henk Zeevat for many helpful comments and suggestions.

References

- [Asher forthcoming] Asher, N.: forthcoming, Reference to abstract objects in English: a philosophical semantics for Natural Language metaphysics. Book ms.
- [Elhadad and McKeown 1990] Elhadad, N. and McKeown, K.R.: 1990, Generating connectives. *Proceedings of COLING-90*, Helsinki.
- [Gardent 1990] Gardent, C.: 1990, Dynamic Semantics and VP Ellipsis. In *Proceedings of the European Workshop on Logics for Artificial Intelligence*, J. van Eijck (ed.), Amsterdam.
- [Gardent 1991] Gardent, C.: 1991, *Gapping and VP Ellipsis in a Unification-Based Grammar*. PhD thesis, University of Edinburgh.
- [Grosz and Sidner 1986] Grosz, B. and Sidner, C.: 1986, Attention, Intention and the Structure of Discourse. *Computational Linguistics*, 12(3), July-September 1986, 175-204.
- [Klein and Stainton-Ellis 1989] Klein, E. and Stainton-Ellis, K.: 1989, A note on multiple VP

ellipsis. Centre for Cognitive Science, University of Edinburgh, Research Paper EUCCS/RP-30.

- [Lascarides and Asher 1991] Lascarides, A. and Asher, N.: 1991, Discourse relations and defeasible knowledge. *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, 55-63.
- [Nakhimovsky 1988] Nakhimovsky, A.: 1988, Aspect, aspectual class and the temporal structure of narrative. *Computational Linguistics*, 14(2), 29-43.
- [Polanyi and Scha 1984] Polanyi, L. and Scha, R.: 1984, A syntactic approach to discourse semantics. *Proceedings of the 10th International Conference on Computational Linguistics and the 22nd Annual Meeting of the Association for Computational Linguistics*, Stanford University, 413-419.
- [Prüst and Scha 1990] Prüst, H. and Scha, R.: 1990, A discourse approach to Verb Phrase Anaphora. *Proceedings of ECAI*.
- [Sag 1980] Sag, I.A.:1980, *Deletion and Logical Form*. New York and London: Garland Publishing.
- [Lappin and McCord 1990] Lappin, S. and McCord, M.: 1990, Anaphora Resolution in Slot Grammar, *Computational Linguistics*, vol. 16, no 4.
- [Stainton-Ellis 1988] Stainton-Ellis, C.S.:1988, *A processing perspective on Verb Phrase Ellipsis*, MPhil dissertation, University of Edinburgh.
- [Webber 1978] Webber, B.: 1978, A formal approach to discourse anaphora. PhD Thesis, Harvard University.
- [Webber 1990] Webber, B.: 1990, Structure and ostension in the interpretation of discourse deixis. To appear in *Language and Cognitive Processes*, 1991. Research report MS-CIS-90-58, University of Pennsylvania, Philadelphia.