Computing Parallelism in Discourse

Claire Gardent
Computational Linguistics
Universität des Saarlandes
Saarbrücken, Germany
`claire@coli.uni-sb.de`

Michael Kohlhase
Computer Science
Universität des Saarlandes
Saarbrücken, Germany
`kohlhase@cs.uni-sb.de`

Abstract

Although much has been said about parallelism in discourse, a formal, computational theory of parallelism structure is still outstanding. In this paper, we present a theory which given two parallel utterances predicts which are the parallel elements. The theory consists of a sorted, higher-order abductive calculus and we show that it reconciles the insights of discourse theories of parallelism with those of Higher-Order Unification approaches to discourse semantics, thereby providing a natural framework in which to capture the effect of parallelism on discourse semantics.

1 Introduction

Both Higher-Order Unification (HOU) approaches to discourse semantics [Dalrymple *et al.*, 1991; Shieber *et al.*, 1996; Gardent *et al.*, 1996; Gardent and Kohlhase, 1996] and discourse theories of parallelism [Hobbs, 1990; Kehler, 1995] assume parallelism structuration: given a pair of parallel utterances, the parallel elements are taken as given.

This assumption clearly undermines the predictive power of a theory. If parallel elements are stipulated rather than predicted, conclusions based on parallelism remain controversial: what would happen if the parallel elements were others? And more crucially, what constraints can we put on parallelism determination (when can we rule out a pair as not parallel?)

In this paper, we present a theory of parallelism which goes one step towards answering this objection. Given two utterances, the theory predicts which of the elements occurring in these utterances are parallel to each other. The proposed theory has one additional important advantage: it incorporates HOU as a main component of parallelism theory thereby permitting an integration of the HOU approach to discourse semantics with discourse theories of parallelism. The resulting framework permits a natural modelling of the often observed effect of parallelism on discourse semantics [Lang, 1977; Asher, 1993]. We show in particular that it correctly captures the interaction of VP-ellipsis and gapping with parallelism. More generally however, the hope is that it also provides an adequate basis for capturing the interaction of parallelism with such discourse phenomena as deaccenting, anaphora and quantification.

Our approach departs from [Grover *et al.*, 1994] in that it genuinely can predict parallel elements. Although both approaches rely on a sorts/types hierarchy and on some kind of overwriting to carry out this task, our approach readily extends to gapping cases whereas as we shall see (cf. section 3.2) [Grover *et al.*, 1994] doesn't.

We proceed as follows. First we present a sort-based abductive calculus for parallelism and show that it predicts parallel elements. We then show how this abductive calculus can be combined with HOU thus yielding an integrated treatment of parallelism and discourse semantics. We then conclude with pointers to further research and related work.

2 Defining discourse parallelism

In linguistic theories on discourse coherence [Kehler, 1995], ellipsis [Dalrymple *et al.*, 1991] (henceforth DSP) and corrections [Gardent *et al.*, 1996], the notion of parallelism plays a central role. In particular, the HOU-based approaches presuppose a theory of parallelism which precomputes the parallel elements of a pair of utterances. For instance, given the utterance pair *Jon likes golf. Peter does too*, DSP's analysis of ellipsis presupposes that *Jon* and *Peter* have been recognised as being parallel to each other.

Similarly, discourse theories of parallelism also assume parallelism structuration. According to [Hobbs, 1990; Kehler, 1995] for instance, there is a class of discourse relations (the *resemblance relations*) which involve the inferring of structurally parallel propositions and where arguments and predicates stand in one of the following configurations:

Relation	S-Ent	T-Ent	Reqts
Parallel	$p(\vec{a})$	$p(\vec{b})$	a_i, b_i similar
Contrast	$p(\vec{a})$	$\neg p(\vec{b})$	a_i, b_i similar
	$p(\vec{a})$	$p(\vec{b})$	a_i, b_i contrastive
Exemplification	$p(\vec{a})$	$p(\vec{b})$	$a_i \in b_1$ or $a_i \subseteq b_i$
Generalisation	$p(\vec{a})$	$p(\vec{b})$	$b_i \in a_1$ or $b_i \subseteq a_i$

where $\vec{a}, \vec{b}$ represent argument sequences; a_i, b_i are any elements of these sequences; and S- and T-Ent are the propositions entailed by the two (source and target) parallel utterances. Furthermore, entities are taken to be *similar* if they share some reasonably specific property and *contrastive* if they have both a shared and a complementary property.

Again, the parallel elements (a_i and b_i) are taken as given that is, the way in which they are recognised is not specified. In what follows, we present a computational theory of parallelism which predicts these parallel elements. The model is a simple abductive calculus which captures Hobbs and Kehler's notions of parallelism and constrast as they are given above. We make the simplifying assumptions that contrast and parallelism are one and the same notion (we speak of contrastive or *c-parallelism*) and that the properties p used in determining them are restricted to sorts from a given, domain-specific sort hierarchy. Thus we can use sorted type theory [Kohlhase, 1994] to model similarity and contrastive parallelism.

2.1 Sorted Logic

Sorts correspond to the basic cognitive concepts. Logically they can either be seen as unary predicates or as refinements of the types. The intuition behind this is that the universe of objects of a type α is subdivided in subsets which are represented by sorts $\mathbb{A}, \mathbb{B}, \ldots$. Since these can in turn be subdivided into subsets, the sorts are ordered by a partial ordering relation $\leq$ in a so-called **sort hierarchy**[1]

Just as in the case of types, every formula has a sort, that can be computed from the sorts of the constants and variables occurring in it. In fact, formulae can have multiple sorts, corresponding to the fact that the intersection of the sets represented by their sorts can be non-empty.

For this paper we assume a fixed finite set of sorts for each type. For the base type e, we will use the following sort hierarchy in our examples.

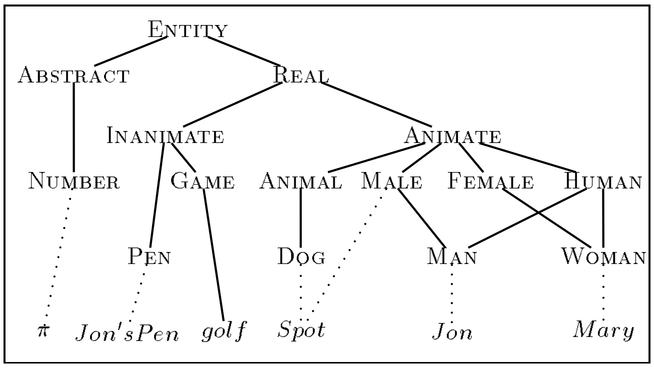

Figure 1: Sort hierarchy of type e

Note that the intersection of the sorts Male and Dog is non-empty, since the constant *Spot* has both sorts. If we want to make this explicit, we can give *Spot* the **intersection sort** Male&Dog. Even though we assume the **simple sorts** (i.e. the non-intersection ones) to be non-empty, the intersection sorts can in general be. For instance, the sorts Animate and Inanimate are disjoint, since they are complementary. The existence of complementary sorts allows us to model the requirements for parallel elements quite naturally. Two formulae **A** and **B** (of any type) are **similar** if they have a common sort; they are **contrastive**, if they have a **distinguishing sort** $\mathbb{D}$, i.e. if **A** has sort $\mathbb{D}$ but **B** has sort $\neg\mathbb{D}$ or vice versa and finally they are **c-parallel**, iff they are both.

For instance *Jon* and *Mary* are parallel, since both are of sort Human, but *Jon* has sort Man, whereas *Mary* has sort Woman = ¬Man&Human&Female $\leq$ ¬Man and therefore *Mary* also has the distinguishing sort ¬Man. This supports DSP's analysis of

> *Jon likes golf, and Mary likes golf.*

For the higher-type, the sort hierarchies of lower type induce further sorts: For any sorts $\mathbb{A}$ and $\mathbb{B}$ of types α and β, $\mathbb{A} \to \mathbb{B}$ is a **functional** sort of type $\alpha \to \beta$. We call sorts that do not contain an arrow **basic sorts**. Similarly, the sort hierarchy of lower type induces subsort relations: $\mathbb{B} \to \mathbb{C} \leq \mathbb{A} \to \mathbb{D}$, is entailed by $\mathbb{A} \leq \mathbb{B}$ and $\mathbb{C} \leq \mathbb{D}$. Furthermore, the resulting sorts can be further subdivided by **functional base sorts**, i.e. sorts that do not contain an arrow, but are of functional type.

For our examples we will use the following sort hierarchy of type $e \to e \to t$

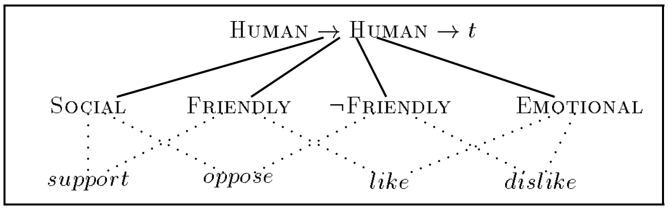

Figure 2: Sort hierarchy of type $e \to e \to t$

2.2 Computation of parallelism

Given the above analysis, the relations *support* and *oppose* are c-parallel, since they have both a common

---

[1] For the purposes of this paper, we assume the sort hierarchy to be given. For applications, hierarchies could be generated from domain representations in KL-ONE like formalisms commonly used in NL systems.

sort (Social) and a distinguishing sort (Friendly). Further, in

Jon supported Clinton, but Mary opposed him.

parallelism theory should predict that the full first utterance *Jon supported Clinton* is c-parallel to the second namely, *Mary opposed him.* However, the sort t does not have subsorts that license this. Rather than dividing t into cognitively unplausible sorts, we propose an abductive equality calculus that generates all possible explanations, why a pair of formulae could be c-parallel, based on the respective sort hierarchies. The calculus manipulates two equalities $=^s$ for similarity and $=^p$ for c-parallelism. The inference rules given in figure 3, give the derivation from figure 4 that explains the parallelism in terms of assumed contrastivity and similarity of the components. We have put the justifications of the abducibles in boxes. Note that this calculus gives us the explanation that *Jon supported Clinton* is c-parallel to *Mary opposed him*, since *Jon* is c-parallel to *Mary* and *support* is c-parallel to *oppose* and finally, we can make *Clinton* and *him* similar by binding *him* to *Clinton.*

$$\frac{\mathbf{AB} =^p \mathbf{CD}}{(\mathbf{A} =^p \mathbf{C} \wedge \mathbf{B} =^s \mathbf{D}) \vee (\mathbf{A} =^s \mathbf{C} \wedge \mathbf{B} =^p \mathbf{D}) \vee (\mathbf{A} =^p \mathbf{C} \wedge \mathbf{B} =^p \mathbf{D})}$$

$$\frac{\lambda X.\mathbf{A} =^p \lambda X.\mathbf{B}}{\mathbf{A} =^p \mathbf{B}} \qquad \frac{\lambda X.\mathbf{A} =^s \lambda X.\mathbf{B}}{\mathbf{A} =^s \mathbf{B}}$$

$$\frac{\mathbf{A} =^p \neg\mathbf{B}}{\mathcal{A} =^s \mathbf{B}} \qquad \frac{\mathbf{A} =^s \neg\mathbf{B}}{\mathcal{A} =^p \mathbf{B}}$$

$$\frac{\mathbf{A}::\mathbb{A} \quad \mathbf{B}::\mathbb{A}}{\mathbf{A} =^s \mathbf{B}}\, just-s \qquad \frac{\mathbf{A}::\mathbb{A} \quad \mathbf{B}::\neg\mathbb{A} \quad \mathbf{A} =^s \mathbf{B}}{\mathbf{A} =^p \mathbf{B}}\, just-c$$

$$\frac{\mathbf{A}::\mathbb{B} \to \mathbb{C} \quad \mathbf{B}::\mathbb{B}}{\mathbf{AB}::\mathbb{C}} \qquad \frac{\mathbf{A}::\mathbb{A}}{\lambda X_{\mathbb{B}}\mathbf{A}::\mathbb{B}\mathbb{A}}$$

Figure 3: The abductive Calculus for Parallelism

Of course, there is a similar derivation that makes *Mary* and *Jon* similar and finally one that makes *support* and *oppose* similar but *Mary* and *Jon* c-parallel. Thus we have the problem to decide which of the different sets of abducibles is the most plausible.

For this it is necessary to give a measure function for sets of abducibles. For instance the three pairs

$$Jon =^p Peter \qquad Jon =^p Spot \qquad Jon =^p \pi$$

are obviously ordered by increasing plausibility. We observe that this plausibility coincides with the distance (the length of the connecting path) from the least sorts of the objects to common sort. Therefore, our approach is to derive plausibility values for abducibles from the justifications of abducibles by calculating distances in the sort hierarchies.

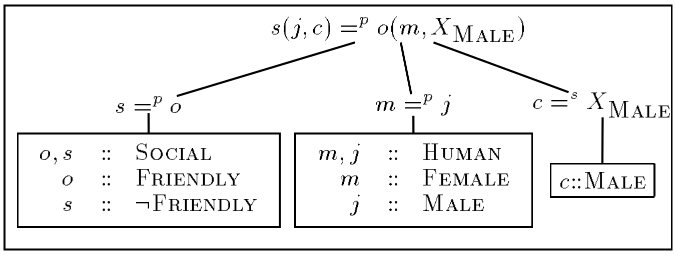

Figure 4: Jon supported Clinton, but Mary opposed him.

3 HOU with Parallelism

In recent approaches to ellipsis [Dalrymple *et al.*, 1991] and deaccenting [Gardent *et al.*, 1996; Gardent, 1997], both parallelism and higher-order unification are central to the semantic analysis. For instance DSP analyze a VP-ellipsis such as

Jon likes golf, and Mary does too.

as being represented by $l(j,g) \wedge R(m)$ where R represents the ellipsis *does*, whose semantic value is then determined by solving the equation $R(j) = l(j,g)$. The motivation for having j occurr in the left-hand side of the equation is that j represents a c-parallel element. This is where parallelism and the assumption of parallelism structuration come in. On the other hand, Higher-Order Unification is also essential in that it is used to solve the equation and furthermore, it is shown to be a crucial ingredient in attaining wide empirical coverage (in particular, it is shown to successfully account for the interaction of ellipsis with quantification, anaphora and parallelism).

However, it is also known that a pure form of HOU is too powerful for natural language and that a more restricted version of it namely, Higher-Order Coloured Unification (HOCU) is more adequate in that it helps prevent over-generation i.e. the prediction of linguistically invalid readings [Gardent and Kohlhase, 1996]. To see this, consider again the example just discussed. Given the stipulated equation, HOU yields two values for R namely, $\lambda X.l(X,g)$ and $\lambda X.l(j,g)$, of which only the first value is linguistically valid. To remedy this, DSP postulate a **Primary Occurrence Restriction** (POR): the term occurrence representing the element which is parallel to the subject of the elliptical utterance, is a primary occurrence and any solution containing a primary occurrence is discarded as linguistically invalid. For instance, j is a primary occurrence in the equation $l(j,g) = R(j)$, so the solution $R = \lambda X.l(j,g)$ is invalid. [Gardent and Kohlhase, 1996] show that DSP's POR can be expressed within HOCU because it uses a variant of the simply typed λ-calculus where symbol occurrences can be annotated with so-called *colours* and substitutions must obey the following constraint:

> For any colour constant c and any c–coloured variable V_{c}, a well–formed coloured substitu-

tion must assign to V_{c} a c–monochrome term i.e., a term whose symbols are c–coloured.

In this setting the POR can be expressed by coloring the primary occurrence j with a colour pe but R with a colour $\neg\mathsf{pe}$. Due to the constraint above, this in effect, enforces the POR.

More generally, [Gardent and Kohlhase, 1996] argue that HOCU rather than HOU, should be used for semantic construction as it allows a natural modelling of the interface between semantic construction and other linguistic modules. In what follows, we therefore assume HOCU as the basic formalism and show how it can be combined with the abductive calculus for parallelism, thereby providing an integrated framework in which to handle parallelism, ellipsis and their interaction.

3.1 Abductive Reconstruction of Parallelism

As just mentioned, we need a basic inference procedure that is a mixture of higher-order colored unification and the sorted parallelism calculus introduced above. The problem at hand is to make colored sorted formulae similar or c-parallel. For an algorithm ARP we build up on a sorted version of HOCU (which can be obtained by a straightforward combination of color techniques from [Hutter and Kohlhase, 1995] with sorted methods from [Kohlhase, 1994]) but instead of simply having equations for sorted $\beta\eta$-equality, we also add the equations for c-parallelism and similarity to the unification problem as special equations $=^p$ and $=^s$.

The ARP calculates sufficient conditions for a given set of input equations by transforming systems of equations to a normal form from which these can be read off.

Decomposition rules consist in those from figure 3 with the difference that the rule for abstractions transforms equations of the form $\lambda x.A =^t \lambda y.B$ to $[c/x]A =^t [c/y]B$, and $\lambda x.A =^t B$ to $[c/x]A =^t Bc$ where c is a new constant, which may not appear in any solution. Furthermore, there is a rule for colored constants that decomposes an equation $c_{\mathsf{a}} =^s c_{\mathsf{b}}$ into the color equation $\mathsf{a} = \mathsf{b}$.

The variable elimination process for colour variables is very simple, it allows to transform a set $\mathcal{E} \cup \{\mathsf{A} = \mathsf{d}\}$ of equations to $[\mathsf{d}/\mathsf{A}]\mathcal{E} \cup \{\mathsf{A} = \mathsf{d}\}$, making the equation $\{\mathsf{A} = \mathsf{d}\}$ solved in the result. For the formula case, elimination is not that simple, since we have to ensure that $|\sigma(x_{\mathsf{A}})| = |\sigma(x_{\mathsf{B}})|$ to obtain a $\mathcal{C}$-substitution σ. Thus we cannot simply transform a set $\mathcal{E} \cup \{x_{\mathsf{d}} = M\}$ into $[M/x_{\mathsf{d}}]\mathcal{E} \cup \{x_{\mathsf{d}} = M\}$, since this would (incorrectly) solve the equations $\{x_{\mathsf{c}} = f_{\mathsf{c}}, x_{\mathsf{d}} = g_{\mathsf{d}}\}$. The correct variable elimination rule transforms $\mathcal{E} \cup \{x_{\mathsf{d}} = M\}$ into $\sigma(\mathcal{E}) \cup \{x_{\mathsf{d}} = M, x_{\mathsf{c}_1} = M^1, \ldots, x_{\mathsf{c}_n} = M^n\}$, where c_i are all colours of the variable x occurring in M and $\mathcal{E}$, the M^i are appropriately coloured variants (same colour erasure) of M, and σ is the $\mathcal{C}$-substitution that eliminates all occurrences of x from $\mathcal{E}$.

Due to the presence of function variables, systematic application of these rules can terminate with equations of the form $x_{\mathsf{c}}(s^1, \ldots, s^n) = h_{\mathsf{d}}(t^1, \ldots, t^m)$. Such equations can neither be further decomposed, since this would loose unifiers (if G and F are variables, then $Ga = Fb$ as a solution $\lambda x.c$ for F and G, but $\{F = G, a = b\}$ is unsolvable), nor can the right hand side be substituted for x as in a variable elimination rule, since the sorts would clash. The sorted, colored variant of Huet's classical solution to this problem is to instantiate x_{c} with a c-monochrome formula that has the right sort $\overline{\mathbb{B}_n} \rightarrow \mathbb{A}$ (that of x_{c}) and the right head h_{d} (which we assume to have sort $\overline{\gamma_m} \rightarrow \mathbb{A}$). These so-called **general bindings** have the following form:

$$\mathcal{G}^h_{\mathsf{d}} = \lambda z^{\mathbb{A}_1} \ldots z^{\mathbb{A}_n}.h_{\mathsf{d}}(H^1_{\mathsf{e}_1}(\overline{z}), \ldots, H^m_{\mathsf{e}_m}(\overline{z}))$$

where the H^i are new variables of sort $\overline{\mathbb{B}_n} \rightarrow \gamma_i$ and the e_i are either distinct colour variables (if $\mathsf{c} \in \mathcal{CV}$) or $\mathsf{e}_i = \mathsf{d} = \mathsf{c}$ (if $\mathsf{c} \in \mathcal{C}$). If h is one of the bound variables $z^{\mathbb{A}_i}$, then $\mathcal{G}^h_{\mathsf{d}}$ is called an **imitation binding**, and else, (h is a constant or a free variable), a **projection binding**.

The general rule for flex/rigid equations transforms $\{x_{\mathsf{c}}(s^1, \ldots, s^n) = h_{\mathsf{d}}(t^1, \ldots, t^m)\}$ into $\{x_{\mathsf{c}}(s^1, \ldots, s^n) = h_{\mathsf{d}}(t^1, \ldots, t^m), x_{\mathsf{c}} = \mathcal{G}^h_{\mathsf{c}}\}$, which in essence only fixes a particular binding for the head variable x_{c}. It turns out (for details and proofs see [Hutter and Kohlhase, 1995]) that these general bindings suffice to solve all flex/rigid situations, possibly at the cost of creating new flex/rigid situations after elimination of the variable x_{c} and decomposition of the changed equations (the elimination of x changes $x_{\mathsf{c}}(s^1, \ldots, s^n)$ to $\mathcal{G}^h_{\mathsf{c}}(s^1, \ldots, s^n)$ which has head h). This solution for pure equations has to be adapted to the more general similarity and contrastivity relations $=^s$ and $=^p$, where we have to provide further imitation rules. In particular, for an equation $X^{\mathbb{A}}_{\mathsf{a}}\overline{\mathbf{U}} =^p h\overline{\mathbf{V}}$ we have to allow imitation bindings $\mathcal{G}^k_{\mathbb{A},\mathsf{a}}$ for $X^{\mathbb{A}}_{\mathsf{a}}$ for any constant k that is contrastive to h and analogously for $=^s$.

3.2 Gapping and ARP

We now illustrate the workings of ARP by the following example

Jon likes golf, and Mary too.

where the second clause is a gapping clause in that both the verb and a complement are missing. This example clearly illustrates the interaction of parallelism with semantic interpretation: if the parallel elements are *Jon* and *Mary*, the interpretation of the gapping clause is *Mary likes golf*, but if conversely *Mary* is parallel to *golf*, then the resulting interpretation is *Jon likes Mary*. Although the first reading is clearly the default, the second can also be obtained – in a joke context for instance. In what follows, we show that ARP predicts both the ambiguity and the difference in acceptability between the two possible readings. Additionally, we show that DSP's a-priori labelling of occurrences as primary or not primary can now be reduced to a more plausible constraint namely, the constraint that *Mary* is a parallel element which has exactly one parallel counterpart in the source (or antecedent) clause.

The analysis is as follows. First, we follow DSP and assign the above example the representation

$$l(j,g) \wedge R(m)$$

where R stands for the missing semantics. However, in contrast to DSP, we do not presuppose any knowledge about parallelism in the source utterance and determine the meaning of R from the equation

$$l(j_{\mathsf{A}}, g_{\neg\mathsf{A}}) =^p R_{\neg\mathsf{pe}}^{\text{WOMAN}\to t}(m_{\mathsf{pe}})$$

which only says that the propositions expressed by *Jon likes Mary* and *golf too* stand in a c-parallel relation[2]. The rationale for the colors in this equation is that *Mary* must be a parallel element in the target utterance. For *Jon* and *golf* in the source utterance, we do not know yet which of them will be a parallel element, but it can be at most one of them, which we code by giving them unspecified but contradictory colors[3]. Finally, R gets the color $\neg\mathsf{pe}$, since it may not be instantiated with formulae that contain primary material (POR).

Since the elided material in gapping constructions and VPE may only copy material from the source utterance (and may not introduce new material) we add the constraint to ARP that $=^p$ and $=^s$ imitations may only be applied to equations, where the head is pe-colored. We call this the **copying constraint** for gapping and VPE. It ensures that whenever two elements are similar but not identical, then they must be primary, since they are parallel.

Let us now go through the ARP computation to see that our analysis obtains exactly the desired readings and to gain an insight of the mechanisms employed therein.

The initial equation is a flex/rigid pair, where only the strict imitation[4] rule is applicable (there is no projection binding of sort WOMAN → t). So, we obtain the binding $\lambda Z.l(H_{\neg\mathsf{pe}}Z)(K_{\neg\mathsf{pe}}Z)$, where H and K are new variables of sort WOMAN → HUMAN. Eliminating this equation yields the equation

$$l(j_{\mathsf{A}}, g_{\neg\mathsf{A}}) =^p l(H_{\neg\mathsf{pe}}m_{\mathsf{pe}})(K_{\neg\mathsf{pe}}m_{\mathsf{pe}})$$

which can be decomposed to the equations

$$H_{\neg\mathsf{pe}}m_{\mathsf{pe}} =^p j_{\mathsf{A}} \qquad\qquad K_{\neg\mathsf{pe}}m_{\mathsf{pe}} =^p g_{\neg\mathsf{A}}$$

For the variable $H_{\neg\mathsf{pe}}$ in the first equation both the imitation binding $\lambda Z.j_{\neg\mathsf{pe}}$ and the projection binding $\lambda Z.Z$ are possible.

In the first case, we have the equation $j_{\mathsf{A}} =^p j_{\neg\mathsf{pe}}$, which entails that $\mathsf{A} = \neg\mathsf{pe}$ leaving us with the second equation (we can eliminate double negations on colors)

$$K_{\neg\mathsf{pe}}m_{\mathsf{pe}} =^p g_{\mathsf{pe}}$$

Again we have the possibility of imitate or project. Since the imitation binding $\lambda Z.g_{\mathsf{pe}}$ for K_{pe} leads to a color clash in $g_{\mathsf{pe}} =^p g_{\neg\mathsf{pe}}$, only the projection binding $\lambda Z.Z$ yields a solution, since the resulting equation $m_{\mathsf{pe}} =^p g_{\mathsf{pe}}$ is valid, since *golf* and *mary* share the sort REAL.

If, on the other hand, we choose the projection binding for H_{pe}, then variable elimination yields the equation $j_{\mathsf{A}} =^p m_{\mathsf{pe}}$, which is valid, since *Jon* and *Mary* share the sort HUMAN and which entails that $\mathsf{A} = \mathsf{pe}$ leaving us with the second equation

$$K_{\neg\mathsf{pe}}m_{\mathsf{pe}} =^p g_{\neg\mathsf{pe}}$$

Again we have the possibility of imitate or project. This time, the imitation binding $\lambda Z.g_{\neg\mathsf{pe}}$ for $K_{\neg\mathsf{pe}}$ leads to the trivialy valid equation $g_{\neg\mathsf{pe}} =^p g_{\neg\mathsf{pe}}$, while the projection binding $\lambda Z.Z$ yields the equation $m_{\mathsf{pe}} =^p g_{\neg\mathsf{pe}}$, which must be unsolvable, since the colors clash.

If we collect the bindings, we arrive at the two solutions $\lambda Z.l(Z,g)$ and $\lambda Z.l(j,Z)$, which correspond to the readings *Mary likes golf* and *Jon likes Mary*. Note that since the similarity of *Jon* and *Mary* is stronger than that between *Mary* and *golf*, the first reading is preferred, while the second reading may only be obtained in the context of a joke. Note also that the use of colours (i.e. the constraint that *Mary* has exactly one parallel counterpart in the source) correctly rules out the mathematically valid solution $\lambda Z.l(Z,Z)$ where *Mary* would be analysed as contrasting with both *Jon* and *golf*.

We now briefly compare our approach with [Grover *et al.*, 1994] where an account of VP-ellipsis is given which uses first-order default unification on feature-structure semantic representations and claims to predict the parallel elements. The underlying idea is that parallel elements have identical thematic roles. Although the approach works well for VP-ellipsis, it is only because in the VPE case, the parallel element of the elliptical clause is known to fill the agentive role. However in gapping cases, part of the resolution problem is to determine the thematic role of the constituents appearing in the elliptical clause. For instance, in *Jon likes golf, and Mary too*, the thematic role of *Mary* is either agent or patient. Unfortunately, the only obvious way to incorporate this into a default unification approach would be to augment it with variables over features – an essentially higher-order construct.

---

[2] By contrast, an extension of DSP's analysis to gapping would posit the equations $l(j,g) = R(j)$ and $l(j,g) = R(g)$ thereby postulating both the parallel elements, and the ambiguity of the gapping clause.

[3] Clearly, this coding is not general enough for the general case, where there are more than one parallel elements in the target utterance, we leave a general treatment to further work.

[4] Note the copying constraint is at work here.

3.3 Controlling ARP

Clearly, a naive implementation of the ARP calculus as sketched above will be intractable, since the set of abducibles is much too large. However, most abducibles are very implausible and should not be considered at all. As in all implementations of abductive processes, the search for abducibles has to controlled, which in turn calls for a quality measure of abduced equations. A standard (but not very imaginative[5]) measure would be the conceptual distance of the sorts justifying the equation, (i.e. the number of subsorts crossed to reach the common and discerning sorts). In our example, the rating of $m =^p g$ is 6, while that of $m =^p j$ is 2, justifying the claim that the reading *Mary likes golf.* is more plausible than *Jon likes Mary.* Since all other readings are either ruled out by the colors or are even more implausible, e.g. an A^* implementation of ARP will only derive these, iff given an appropriate threshold. Since the aim of this paper is to establish the principles of parallelism reconstruction, we will not pursue this here.

4 Conclusion

We have given a sketch of how to develop a computational framework for calculating parallelism in discourse. This approach is based on the HOCU variant of DSP's HOU account of ellipsis, but unlike that approach does not presuppose knowledge about the parallel elements. Instead, it computes them in the analysis.

Parallelism can be seen as affecting the interpretation of the second of two parallel utterances in mainly two ways: it can either constrain an anaphor to resolve to its source parallel counterpart (this is the case for instance, in the gapping example discussed above); or it can add to its truth conditional content. For instance, in

> *Jon campaigned hard for Clinton in 1992. Young aspiring politicians often support their party's presidential candidate*

parallelism enforces a reading such that *Jon* is understood to be a *young aspiring politician* and *Clinton* is understood to be Jon's *party's presidential candidate*.

In future work, we plan to investigate these two aspects in more details. As for the interaction of parallelism with binding, one important question is whether our proposal preserves DSP's insights on the interaction of parallelism with ellipsis, anaphora and quantification. On the other hand, to account for the incrementing effect of parallelism on semantic interpretation, the proposal will have to cover the discourse relations of exemplification and generalisation. Note however that the proposed interleaving between HOU and abductive calculus gives us a handle on that problem: mismatches between semantic structures can be handled by having the calculus be extended to abstract away irrelevant structural differences (this would account for instance for the fact that in our example, a temporal modifier occurs in the source but not in the target) whereas sorted HOU can be used to infer information from the most specific common sort (in this case, the sort of *young aspiring politicians*).

References

[Asher, 1993] Asher, Nick 1993. *Reference to abstract objects in discourse.* Kluwer, Dordrecht.

[Dalrymple *et al.*, 1991] Dalrymple, Mary; Shieber, Stuart; and Pereira, Fernando 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy* 14:399–452.

[Gardent and Kohlhase, 1996] Gardent, Claire and Kohlhase, Michael 1996. Higher–order coloured unification and natural language semantics. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics.* ACL, Santa Cruz.

[Gardent *et al.*, 1996] Gardent, Claire; Kohlhase, Michael; and van Leusen, Noor 1996. Corrections and Higher-Order Unification. In *Proceedings of KONVENS96.* De Gruyter, Bielefeld, Germany. 268–279.

[Gardent, 1997] Gardent, Claire 1997. Parallelism, HOU and deaccenting. CLAUS Report 85, University of Saarbrücken.

[Grover *et al.*, 1994] Grover, Claire; Brew, Chris; Manandhar, Suresh; and Moens, Marc 1994. Priority union and generalisation in discourse grammar. In *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics.*

[Hobbs, 1990] Hobbs, Jerry 1990. *Literature and Cognition.* Number 21 in CSLI Lecture Notes. CSLI.

[Hutter and Kohlhase, 1995] Hutter, Dieter and Kohlhase, Michael 1995. A coloured version of the λ-calculus. SEKI-Report SR-95-05, Universität des Saarlandes.

[Kehler, 1995] Kehler, Andrew 1995. *Interpreting Cohesive Forms in the Context of Discourse Inference.* Ph.D. Dissertation, Harvard University.

[Kohlhase, 1994] Kohlhase, Michael 1994. *A Mechanization of Sorted Higher-Order Logic Based on the Resolution Principle.* Ph.D. Dissertation, Universität des Saarlandes.

[Lang, 1977] Lang, Ewald 1977. Semantik der koordinativen verknüpfung. *Studia Grammatika XIV.*

[Shieber *et al.*, 1996] Shieber, Stuart; Pereira, Fernando; and Dalrymple, Mary 1996. Interaction of scope and ellipsis. *Linguistics and Philosophy* 19:527–552.

---

[5] Clearly, a more sophisticated measure would include concepts like the specificity of the solution.