Underspecified Modelling of Complex Discourse Constraints

Markus Egg
egg@let.rug.nl
University of Groningen

Michaela Regneri
regneri@coli.uni-sb.de
Saarland University

Abstract

We introduce a new type of discourse constraints for the interaction of discourse relations with the configuration of discourse segments. We examine corpus-extracted examples as soft constraints. We show how to use Regular Tree Gramamrs to process such constraints, and how the representation of some constraints depends on the expressive power of this formalism.

1 Introduction

Discourse structures cannot always be described completely, either because they are ambiguous (Stede, 2004), or because a discourse parser fails to analyse them completely. In either case, *underspecification formalisms* (UFs) can be used to represent partial information on discourse structure. UFs are used in semantics to model structural ambiguity without disjunctive enumeration of the readings (van Deemter and Peters, 1996).

Underspecified descriptions of discourse must handle two kinds of incomplete information, on the *configuration* of discourse segments (how they combine into larger units), and on the discourse *relations* that bring about this configuration: Our corpus studies on the RST Discourse Treebank (Carlson et al., 2002) showed *interdependencies* between relations and configuration, a phenomenon first noted by (Corston-Oliver, 1998). These interdependencies can be formulated as constraints that contribute to the disambiguation of underspecified descriptions of discourse structure.

E.g., in discourse segments constituted by the relation *Condition*, the premiss tends to be a discourse atom (or at least, maximally short).[1] Similarly, there is evidence for an interdependency constraint for the relation *Purpose(1)* [2]. In most cases, *Purpose(1)* has a discourse atom as its nucleus.

The corpus evaluation furthermore shows that those patterns never occur exclusively but only as tendencies. Realised as *soft* constraints, such tendencies can help to sort the set of readings according to the established preferences, which allows to focus on the best reading or the n-best readings. This is of high value for an UF-based approach to discourse structure, which must cope with extremely high numbers of readings. To model interdependency constraints, we will use Regular Tree Grammars (RTGs) (Comon et al., 2007). RTGs can straightforwardly be extended to *weighted* Regular Tree Grammars (wRTGs), which can represent both soft and hard constraints.

Apart from our corpus-extracted examples, we also consider a hard interdependency constraint similar to the Right Frontier Constraint. We show that we can integrate this attachment constraint with our formalism, and how its representation depends on the expressiveness of RTGs.

2 Underspecified Discourse Structure

We describe (partial) information on discourse structure by expressions of a suitable UF, here, dominance graphs (Althaus et al., 2003). Consider e.g. Fig. 1(a), the dominance graph for (1):

(1) [C_1 I try to read a novel] [C_2 if I feel bored]
[C_3 because the TV programs disappoint me]
[C_4 but I can't concentrate on anything.]

[1] Following Rhetorical Structure Theory (Mann and Thompson, 1988), most discourse relations have a central nucleus argument, and a peripheral satellite argument. For *Condition*, the premiss is the satellite, the nucleus, the conclusion.

[2] '(n)' as part of a relation name indicates that the nucleus is its n-th argument; relations with names without such an affix are multinuclear, i.e., link two segments of equal prominence. We sometimes omit the numbers where the position of the nucleus is clear from the context.

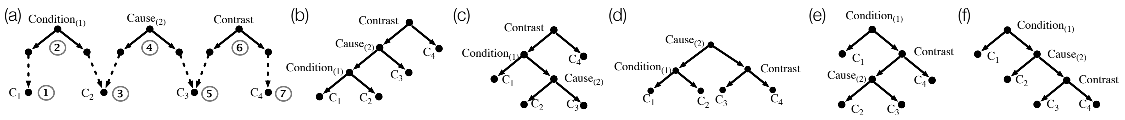

Figure 1: An underspecified discourse structure and its five configurations

$\{1\text{-}7\} \rightarrow$ *Condition*$(\{1\}, \{3\text{-}7\})$ [1]	$\{1\text{-}7\} \rightarrow$ *Cause*$(\{1\text{-}3\}, \{5\text{-}7\})$ [1]	$\{3\text{-}7\} \rightarrow$ *Contrast*$(\{3\text{-}5\}, \{7\})$ [1]
$\{3\text{-}5\} \rightarrow$ *Cause*$(\{3\}, \{5\})$ [1]	$\{1\text{-}7\} \rightarrow$ *Contrast*$(\{1\text{-}5\}, \{7\})$ [1]	$\{1\text{-}5\} \rightarrow$ *Cause*$(\{1\text{-}3\}, \{5\})$ [1]
$\{5\text{-}7\} \rightarrow$ *Contrast*$(\{5\}, \{7\})$ [1]	$\{1\text{-}5\} \rightarrow$ *Condition*$(\{1\}, \{3\text{-}5\})$ [3]	$\{1\text{-}3\} \rightarrow$ *Condition*$(\{1\}, \{3\})$ [9]
$\{3\text{-}7\} \rightarrow$ *Cause*$(\{3\}, \{5\text{-}7\})$ [1]	$\{1\} \rightarrow C_1$ [1] $\{3\} \rightarrow C_2$ [1]	$\{5\} \rightarrow C_3$ [1] $\{7\} \rightarrow C_4$ [1]

Figure 2: A wRTG modelling the interdependency constraint for Fig. 1

Such constraints describe a set of discourse structures (formalised as binary tree structures). Their key ingredient are (reflexive, transitive and antisymmetric) *dominance relations*, which are indicated by dotted lines. Dominance of X_1 over X_2 means that X_2 is part of the structure below (and including) X_1, but there might be additional material intervening between X_1 and X_2.

Fig. 1(a) states that C_1 is linked to a part of the following discourse (including at least C_2) by *Condition*, *Cause(2)* connects two discourse segments (comprising at least C_2 and C_3, respectively), and *Contrast* links a discourse segment to its left (including at least C_3) to C_4.

This constraint describes (is compatible with) exactly the five tree structures in Fig. 1(b-f), if described tree structures may only comprise material that is already introduced in the constraint. They model the potential discourse structures for (1) (see Webber (2004)). Dominance graphs like Fig. 1a. are *pure chains*. Pure chains describe all binary trees with the same leaf language, here the discourse segments, in their textual order. Pure chains define a left-to-right order, in that not only the leaves always form the same sequence, but also the inner nodes: If a labelled node X is further to the left in the chain than another node Y, in every described tree, X will either be Y's left child, or Y will be X's right child, or there will be a fragment F of which X is a successor on the left and Y is a right successor. Henceforth we will refer to fragments with their index in the chain (indicated by encircled numbers in Fig. 1a).

3 Representing Soft Interdependencies

The interdependency constraint for *Condition(1)* is that its satellite tends to be *maximally short*, i.e., mostly consists of only one discourse atom, and in most remaining cases, of two atoms. Thus, (b) and (d) are preferred among the configurations in Fig. 1, (c) is less preferred, and (e) and (f) are the least preferred. *Regular Tree Grammars* (RTGs) as UF (Koller et al., 2008) can express such complex constraints straightforwardly, and provide a convenient framework to process them. They allow to extract a best configuration with standard algorithms very efficiently.

Koller et al. (2008) show how to generate an RTG describing the same set of trees as a dominance graph. Similar to a context free grammar, an RTG uses production rules with terminal symbols and nonterminal symbols (NTs), whereby the left-hand side (LHS) is always a nonterminal and the right-hand side (RHS) contains at least one terminal symbol. One NT is the start symbol. A tree is accepted by the grammar if the grammar contains a derivation for it. An example for an RTG is given in Fig. 2, which describes the same trees as the dominance graph in Fig. 1a. The start symbol is $\{1\text{-}7\}$. To derive e.g. the tree in Fig. 1d, we first select the rule $\{1\text{-}7\} \rightarrow$ *Cause*$(\{1\text{-}3\}, \{5\text{-}7\})$ that determines *Condition* as root for the whole tree. The left child of *Condition* is then derived from $\{1\text{-}7\}$, and the right child from $\{5\text{-}7\}$ respectively. To emphasize the association with the dominance graph, we mark nonterminals as the subgraphs they represent, e.g., $\{1\text{-}7\}$ denotes the whole graph. The terminal in the RHS of a grammar rule determines the root of the LHS subgraph.

Koller et al. (2008) also use *weighted* RTGs (wRTGs, an extension of RTG with weights) to express soft dominance constraints (which, unlike hard constraints, do not restrict but rather rank the set of configurations). We use wRTGs to model the soft interdependency constraints. The grammar in Fig. 2 is also a wRTG that assigns a weight to each derived tree: Its weight is the product over all weights of all rules used for the derivation. Weights appear in squared brackets after the rules.

The (merely expository) weights in our example encode the preference of *Condition* for a maximally short right child: There are three grammar rules that establish *Condition* as the root of a subgraph (shaded in Fig. 2), which are distinguished by the size of the right child of the root (one ($\{3\}$), three ($\{$3-5$\}$) or five ($\{$3-7$\}$) nodes). The shorter the right child, the higher the weight associated with the rule. (1 is a neutral weight by definition.) The grammar thus assigns different weights to the trees in Fig. 1; (b) and (d) get the maximum weight of 9, (b), a medium weight of 3, and (e) and (f), the lowest weight of 1.

4 Expressive Power of RTGs

As Koller et al. (2008) show, the expressive power of RTGs is superior to other common underspecification formalism. We show an important application of the increased expressiveness with Ex. 2, where a. can be continued by b. but not by c:

(2) a. [C_1 Max and Mary are falling apart.] [C_2 They no longer meet for lunch.] [C_3 And, last night, Max went to the pub] [C_4 but Mary visited her parents.]

b. [C_{5a} She complained bitterly about his behaviour.]

c. [C_{5b} He left after his fifth pint of lager.]

Segment C_{5a} continues the preceding clause about Mary's visit with additional information about the visit, it thus attaches directly to C_4. To find a coherent integration of C_{5b}, we would have to connect it to C_3, as it provides more details about Max' night at the pub. However, in the given constellation of C_3 and C_4, that form a *Contrast* together, C_3 is not available any longer for attachment of further discourse units. (This constraint is reminiscent of the *Right Frontier Constraint*, as it is used by Asher and Lascarides (2003). However, it is unclear how the Right Frontier Constraint in its exact definition can carry over to binary trees.)

The given attachment constraint is not expressible with dominance graphs: it excludes the configurations of its dominance graph (Fig. 3) in which *Contrast* shows up as a direct left child, e.g., (3b/e/f) as opposed to (3c/d). For instance, the excluded structure emerges in (3e/f) by choosing *Cause* as root of the the subgraph 5-9 (i.e., including the *Contrast*- and *Sequence*-fragments). For convenience, we will talk about this constraint as the "left child constraint" (LCC).

$$\begin{array}{lllllll} S & \to & Contrast(S,S) & L & \to & Evid(S,S) \\ S & \to & Sequ(L,S) & L & \to & List(S,S) \end{array}$$
$$S \to L$$
$$L \to C_1 \quad L \to C_2 \quad L \to C_3 \quad L \to C_4 \quad L \to C_5$$

Figure 5: A filter RTG corresponding to Ex. 2

This additional constraint, however, can be expressed by an RTG like Fig. 4. We explicitly distinguish between subgraphs (referred to with numbers) and their associated NTs here. Crucially, some subgraphs can be processed in different derivations here, e.g., $\{$5-9$\}$ (as right child of *List*, irrespective of the relative scope of *Evidence* and *List*), or $\{$3-7$\}$ (in the expansions of both $\{EvLiCo\}$ and $\{LiCoSe\}$, like in (3c) as opposed to (3d)). Sometimes this derivation history is irrelevant, like in the case of $\{$5-9$\}$ (here, only *Contrast* may be chosen as root anyway), but there are cases where it matters: If $\{$3-7$\}$ is the left child of *Sequence*, as in (3b/d), the choice of *Contrast* as its root is excluded, since this would make *Contrast* the left child of *Sequence*, as in (3b). In contrast, $\{$3-7$\}$ as the right child of *Evidence*, like in (3c), allows both *Contrast* and *List* as root, because *Contrast* emerges as a right child in either case. Thus, the two occurrences of $\{$3-7$\}$ are distinguished in terms of different NTs in the grammar, and only in the NT for the latter occurrence is there more than one further expansion rule.

Regular tree languages are closed under intersection. Thus, one can derive a grammar like Fig. 4 by intersecting a completely underspecified RTG (here, the one derived from Fig. 3a) with a suitable *filter grammar*, e.g., Fig. 4. The filter grammar produces an infinite language, containing the fragments of Fig. 3a and excluding any derivation in which *Sequence* is the direct parent of *Contrast*. This is guaranteed by introducing the nonterminal L (the left child NT for *Sequence*), for which there is no derivation with *Contrast* as its root.

For an arbitrary pure chain with n fragments, the filter grammar generating the LCC is constructed as follows: S is the start symbol. For every fragment i s.t. $0 < i < n$, there is a derivation rule with S as its LHS and i in its RHS, thus either $S \to i$, for singleton fragments, or $S \to i(A, S)$, for binary fragments. If i is binary, we must determine A: If there is at least one fragment $f < i$ s.t. the LCC is assumed for f, we create a new NT L_i; every derivation rule with i on its RHS follows the pattern $X \to i(L_i, S)$ (thus $A = L_i$ in particular). If there is no LCC fragment to the left

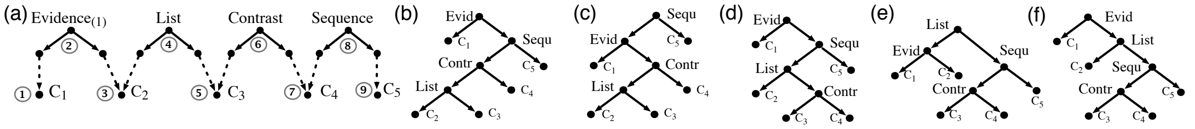

Figure 3: An underspecified discourse structure for Ex. 2 and five of its configurations

$$\begin{array}{rcl|rcl|rcl}
\{EvLiCoSe\} & \rightarrow & Evid(\{C_1\},\{LiCoSe\}) & \{EvLiCo\} & \rightarrow & List(\{Ev\},\{Co\}) & \{Ev\} & \rightarrow & Evid(\{C_1\},\{C_2\}) \\
\{EvLiCoSe\} & \rightarrow & List(\{Ev\},\{CoSe\}) & \{CoSe\} & \rightarrow & Cont(\{C_3\},\{Se\}) & \{Li\} & \rightarrow & List(\{C_2\},\{C_3\}) \\
\{EvLiCoSe\} & \rightarrow & Cont(\{EvLi\},\{Se\}) & \{EvLi\} & \rightarrow & Evid(\{C_1\},\{Li\}) & \{Co\} & \rightarrow & Cont(\{C_3\},\{C_4\}) \\
\{EvLiCoSe\} & \rightarrow & Sequ(\{EvLiCo\},\{C_5\}) & \{EvLi\} & \rightarrow & List(\{Ev\},\{C_3\}) & \{Se\} & \rightarrow & Sequ(\{C_4\},\{C_5\}) \\
\{LiCoSe\} & \rightarrow & Sequ(\{LiCo\}_L,\{C_5\}) & \{LiCo\}_L & \rightarrow & List(\{C_2\},\{Co\}) & & & \\
\{LiCoSe\} & \rightarrow & List(\{C_2\},\{CoSe\}) & \{LiCo\}_S & \rightarrow & Cont(\{Li\},\{C_4\}) & \multicolumn{3}{c}{\{C_1\} \rightarrow C_1 \quad \{C_2\} \rightarrow C_2} \\
\{LiCoSe\} & \rightarrow & Cont(\{Li\},\{Se\}) & \{LiCo\}_S & \rightarrow & Li(\{Li\},\{C_4\}) & \multicolumn{3}{c}{\{C_3\} \rightarrow C_3} \\
\{EvLiCo\} & \rightarrow & Evid(\{C_1\},\{LiCo\}_S) & & & & \multicolumn{3}{c}{\{C_4\} \rightarrow C_4 \quad \{C_5\} \rightarrow C_5}
\end{array}$$

Figure 4: A RTG integrating the attachment constraint for *Contrast* from Ex. 2 into Fig. 3

of i, $A = S$. If a new NT L_i was created, we need to create its RHSs: For every fragment h s.t. $0 < h < i$ and there is no LCC for h, there is a rewrite rule directly deriving h from L_i. If h is a singleton fragment, the rule is $L_i \rightarrow h$. Otherwise the rule is $L_i \rightarrow h(A', S)$, whereby $A' = S$, if there is no L_h, or $A' = L_h$ if there is some LCC fragment on the left of h.[3]

The grammar in Fig. 4 can be generated with that scheme; it has been reduced afterwards in that a general rule $S \rightarrow L$ substitutes for all rules of the form $S \rightarrow NT$ for which there is a corresponding rule $L \rightarrow NT$ (e.g., $S \rightarrow Evid(S, S)$).

5 Conclusion

Interdependency constraints that arise from the interaction of discourse relations and their surrounding structures are introduced as a new technique for disambiguating discourse structure. We integrate those constraints in underspecified discourse structures by exploiting the expressive power of Regular Tree Grammars as UF. As the corpus analysis yields in many cases only soft interdependency constraints, we use the weighted extension of RTGs, which allows to sort the readings of an underspecified representation and to identify preferred discourse structures. We then showed that the representation of some discourse constraints depend on the expressive power of RTGs. For notes on implementation and tractability of our approach, see Regneri et al. (2008).

References

Althaus, Ernst, Denys Duchier, Alexander Koller, Kurt Mehlhorn, Joachim Niehren, and Sven Thiel. 2003. An efficient graph algorithm for dominance constraints. *Journal of Algorithms*, 48:194–219.

Asher, Nicholas and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge UP, Cambridge.

Carlson, Lynn, Daniel Marcu, and Mary Ellen Okurowski. 2002. RST Discourse Treebank. LDC.

Comon, H., M. Dauchet, R. Gilleron, C. Löding, F. Jacquemard, D. Lugiez, S. Tison, and M. Tommasi. 2007. Tree Automata Techniques and Applications. Available on: `http://www.grappa.univ-lille3.fr/tata`. Release 12-10-2007.

Corston-Oliver, Simon H. 1998. *Computing Representations of Discourse Structure*. Ph.D. thesis, Dept. of Linguistics, University of California, Santa Barbara.

van Deemter, Kees and Stanley Peters, editors. 1996. *Semantic ambiguity and underspecification*. CSLI, Stanford.

Koller, Alexander, Michaela Regneri, and Stefan Thater. 2008. Regular tree grammars as a formalism for scope underspecification. In *Proceedings of the ACL 08*.

Mann, William C. and Sandra A. Thompson. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8:243–281.

Regneri, Michaela, Markus Egg, and Alexander Koller. 2008. Efficient Processing of Underspecified Discourse Representations. In *Proceedings of the ACL 08 (Short Papers)*.

Stede, Manfred. 2004. The Potsdam Commentary Corpus. In Webber, Bonnie and Donna K. Byron, editors, *ACL 2004 Workshop on Discourse Annotation*.

Webber, Bonnie. 2004. D-LTAG: extending lexicalized TAG to discourse. *Cognitive Science*, 28:751–779.

[3] To model this as a preference rather than as a hard constraint, no rules for the L-NTs are omitted, but rather weighted low. An intersection with a preference-neutral wRTG would rank the configurations violating the constraint low, and all others with neutral weights.