

A Linguistically-based Approach to Discourse Relations Recognition

Rodolfo Delmonte, Gabriel Nicolae, Sanda Harabagiu and Cristina Nicolae

Human Language Technology Research Institute
Department of Computer Science
The University of Texas at Dallas, Richardson, TX 75083

Abstract. We present an unsupervised linguistically-based approach to discourse relations recognition, which uses publicly available resources like manually annotated corpora (Discourse Graphbank, Penn Discourse Treebank, RST-DT), as well as empirically derived data from “causally” annotated lexica like LCS, to produce a rule-based algorithm. In our approach we use the subdivision of Discourse Relations into four subsets – CONTRAST, CAUSE, CONDITION, ELABORATION, proposed by [7] in their paper, where they report results obtained with a machine-learning approach from a similar experiment, against which we compare our results. Our approach is fully symbolic and is partially derived from the system called GETARUNS, for text understanding, adapted to a specific task: recognition of Causality Relations in free text. We show that in order to achieve better accuracy, both in the general task and in the specific one, semantic information needs to be used besides syntactic structural information. Our approach outperforms results reported in previous papers [9].

1 Introduction

In this paper we will address the task related to discourse processing, i.e. that of Discourse Relations Recognition (hence DRR), as well as the other related task, the one of building Discourse Structures. We take DRR to be of the outmost interest in all fields of text processing, from summarization to question-answering. We will then focus on the detection of a particular Discourse Relation, i.e. Causal Relations. The task referred to as the Detection of Causality Relations (hence DCR) is a precondition for the possibility that current Q/A systems may advance from factoid based performances to the answering of WHY questions.

In their paper, [9] report their experiment of Discourse Segmentation into Discourse Units in which they use the output of Charniak’s parser in one experiment and syntactic trees taken directly from the Penn Treebank in another. The recall disagreement is very small: 1.8%, as is the precision: 0.6%, something that speaks in favor of using syntactic output in a systematic way in the first place. We will return to the problem of segmentation in a section below. In our approach, Discourse Units will be segmented automatically, but no learning phase is forecast and the segmentation will be produced solely on the basis of syntactic structures. This is motivated by the fact that Discourse Units must be mapped onto semantically interpretable spans of text, and syntax provides the most natural and reliable interface to semantics available nowadays (but see [2]). The problem is addressed in more detail in the section on Evaluation below.

The learning algorithm presented both in [7] and [9] makes no use of semantic information, but only of lexical, grammatical and syntactic information. In particular, in the former the classifier makes use of what they call the “most representative words” of a sentence/discourse unit, by which they mean the nouns, verbs and cue phrases present in each sentence. This is just a crude erasing procedure of function words, including all adjectives and adverbs. On the contrary, in [9] the training section for the segmentation task consisted of triples that contained the syntactic tree and the discourse tree.

The discourse model the authors use for Discourse Parsing, on the contrary, contains information directly related to what they call “a lexicalized syntactic tree”, very much in line with Collins’ approach to parsing. In their model, each feature is a representation of the syntactic and lexical information that is found at the point where two EDUs are joined together. For each EDU they thus identify a word as the Head word, denoted H, which is the highest node in EDU. Thus, the verb will always be the Head of finite tensed EDUs and, in case a modal or an auxiliary precedes the main verb, it will be taken as Head, since it will be mapped onto a higher VP node than the main verb. This would be semantically unacceptable in our approach, where the syntax only plays a secondary role.

We carried out a thorough study of the three databases – Discourse Graphbank, hence DGB¹ [10], Penn Discourse Treebank, hence PDTB² [8], and RST- DT³ [1] - which contain manually annotated Discourse Relations, with the aim to propose a Model of Causal Relations (hence CRs) that encompasses different levels of linguistic representations and features. The Model will eventually be used by a question/answering system to enable it to deal with WHY questions. CRs have always constituted a very thorny issue in the linguistic and philosophical literature, and to cover all related issues we will address the whole gamut of linguistic information.

The paper is organized as follows: in section 2 we introduce the theoretical framework, the data analysis for the DCRs and the difference between explicit and implicit discourse relations; section 3 presents the Hypothesis and the Model; section 4 is devoted to the evaluation, and finally there is a short conclusion.

2 Causal Relations

The goal of this work was finding if and in what measure CRs are dependent on Knowledge of the World, and what kind of constraints surface syntactic structure and semantic interpretation play on the nature of CRs— eventually what kind of information needs to be stored in the lexicon in order to spot CRs.

As a general starting point, we noted that natural CRs – i.e. causality relations related to natural facts of the physical world in which we live - are usually left unexpressed unless the text is dealing with explanatory descriptions and experimental work in scientific literature and educational academic and school related literature. Whenever CRs are reported of physical events they must be exceptional or possibly of

¹ <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2005T08>

² <http://www.seas.upenn.edu/~pdtb>

³ <http://www ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC2002T07>

abnormal dimension, e.g., the tsunami is an abnormal ocean wave which is caused by landslide or earthquake.

In order to express CRs in text, two clauses/sentences must be used: one expressing Motivation or Cause, and another expressing Result or Effect. It is not always the case that Causes are expressed fully: in many cases they are left implicit as if they can be understood by the addressee. What is usually reported in texts is the effect of some previous event which might or might not be reported in the previous text, and may be adjacent or (rarely) non adjacent to the current sentence. We also noted that in DGB CRs expressed at a propositional level, thus intersententially, make use of a connective only in 30% of cases: the remaining examples are either untensed propositions – gerundives, infinitivals or adjuncts – and don't carry any specific or unambiguous discourse marker, or they are sentences without the presence of a connective, thus leaving the link implicit. The latter constitute certainly the most challenging type of CRs to detect automatically. We have also considered situations in which Causes are not expressed by clauses, but simply by Prepositional Phrases, as for instance in the example below that we take from DGB:

{The late Prime Minister Olof Palme was blacklisted from the White House}_{effect} {for his outspoken opposition to the Vietnam War.}_{cause}

Eventually, Cause may be simply part of the lexical meaning of the verb/noun, and may be extracted by some system of semantic decomposition in lexical primitives or templates. This is done using the LCS lexicon⁴, where Causatives and Inchoatives are clearly marked by the presence of a CAUSE template.

In order to isolate difficult to solve cases from the rest, we produced a first subdivision of possibly different types of CRs along the lines proposed by a number of previous works, and we came up with the following basic inventory that we adopted in our algorithm:

- Cause-Result
- Rationale-Outcome
- Purpose-Outcome
- Circumstance-Outcome
- Means-Outcome

We use the label Outcome to indicate an Event which is not in direct causal relation with another event but is simply explained (Rationale) by it, marked with the intention to carry it out (Purpose), or in a possible or conditional relation with it (Circumstance). We will treat other dubious cases as the ones indicated below as Circumstances:

Vitamin B enables the body to digest food.
Corn oil prevents butter from burning.

The reason for this choice is that cases like these contain Generic assertions and not Factual assertions, and we want to keep Factual assertions apart from Nonfactual ones. In this sense, we also consider conditionals and counterfactuals as belonging to a totally different class of Discourse Relations, namely the one definable as Hypothesis. Eventually, temporal discourse markers such as *after*, *when* will not be treated as

⁴ <http://www.umiacs.umd.edu/~bonnie/verbs-English.lcs>

possible causal cue words, but just as temporal ones or as Circumstances in case they imply some causal link, as for instance in the following two examples:

Liberman left the Republican Party after George W. Bush was nominated.
Mary stopped the car when she saw the elephant.

2.1 Data Analysis: The CRs Examples Annotated in DGB and PDTB

The data analysis has been concerned with the individuation of linguistic criteria to subdivide CRs as they have been selected and annotated in the two databanks available at HLT. DGB is the corpus of annotated Discourse Relations built by [10], which contains 8910 relations and clause pairs, 466 (thus constituting approximately the 5% of all relations) of which have been labeled as CE, i.e. Cause-Effect. The PDTB contains only 1515 discourse relations, 204 of which have been labeled as CAUSE (approximately 13%).

The two corpora are markedly different however: the PDTB only contains “implicit” or lexically unexpressed discourse markers. In other words, sentence pairs related by a given Discourse Relation are not connected by an explicit lexical connector. They are also usually made of two separate sentences; however, the structures selected may also respond to phonological criteria, as reported in the Technical Manual, and in some cases they may be constituted by PPs that are adjuncts contained in a parenthetical. In such cases, punctuation plays a role, together with intonation. The same applies to the other database we inspected.

As far as the DGB is concerned, discourse relations may come with both expressed and unexpressed connectors, both inter- and intrasententially organized. In particular, as far as CRs are concerned, we have the following distribution of data:

- 190 sentence pairs: 1) lexically unexpressed discourse marker;
- 136 clause pairs: 1) result/effect expressed as infinitival preceded by `to/in_order_to`;
- 89 sentence pairs: 1) lexically expressed discourse markers (68 because + 21 because_of + 4 cause, caused_by);
- 38 clause pairs: 1) result/effect expressed as gerundive preceded by `by/for`; 2) intrasententially;
- 35 sentence pairs: 1) lexically expressed discourse markers (`as, as_a_result, so, so_that, as_a_means_of, part_of, since, since_then, when, due_to`);
- 8 have not been classified because we were not in agreement with the annotators.

Overall, we ended up with 2/3 of all CRs with no specific lexical connectors and 1/3 having one specific discourse marker, 2/5 of all CRs constituted by intrasentential structures and 3/4 having an intersentential structure. Thus, by far, the majority of discourse segments annotated are not between two clauses with a finite verb. Other additional semantic and syntactic details we noted are:

- 70 contain negation markers (`not, no, none, nothing, never, neither, nor`);
- 39 sentences treated as DUs are Relative Clauses;
- 82 NPs intervene in the discourse relations either directly as one segment or indirectly;
- 13 PPs and APs are treated as DU segments.

As to PDTB, we analyzed all of the 204 CRs pairs and we discovered that they are highly paraphrastic, using copulative or light verbs together with extended nominalization to convey the meaning of events. Overall, we counted 218 cases of copulative verbs as main verbs (be, there_be, become, have), which makes at least one for each sentence pair. On the contrary, the same verbs in the DGB only amount to 126 cases, only about 27% of the sentence pairs. We also counted the number of clauses overall, and the mean length per discourse relation is much higher in PDTB (3.1), when compared to DGB (2.3).

In PDTB, Cause annotated sentence pairs show a consistent use of negation: there are 83 sentences containing negation, i.e. 40.7%; if we compare this with DGB, where we had 70 negation items – i.e. 15% of the all examples – we can see that it is much higher.

2.2 Explicit and Implicit CRs: Are they Interchangeable?

For computational purposes, we wanted to assume that the use of discourse markers is not optional. The strong hypothesis would be that in order to be able to omit/insert a causality discourse marker some structural and semantic conditions have to be obtained. In doing this work we discovered that inserting BECAUSE in implicit CR-DUs corresponds to testing the causality reading of the DR. Consider the following examples taken from the DGB:

DU1: to produce chemical weapons.

DU2: the plant was built

- The plant was built BECAUSE they/one wanted/intended to produce chemical weapons.

DU1: He died in Dec. 21 Pan Am plane crash in Scotland while on his way to New York

DU2: to observe the signing of an agreement for Namibian independence from South African rule.

- He died in Dec. 21 Pan Am plane crash in Scotland while on his way to New York BECAUSE he wanted/intended to observe the signing of an agreement for Namibian independence from South African rule.

However, to enforce our starting hypothesis of the non comparability of implicitly vs. explicitly marked CRs, no such reformulation is allowed when the relation is actually a CAUSE DR, i.e. there are two separate clauses with finite tense, and there is no discourse marker. Consider:

DU1: these early PCs triggered explosive product development in desktop models for the home and office.

DU2: PC shipments annually total some \$38.3 billion world - wide.

- These early PCs triggered explosive product development in desktop models for the home and office BECAUSE PC shipments annually total some \$38.3 billion world-wide.

So, eventually, we may safely affirm that untensed clauses cannot be classified as Causal Relations in case there is no introductory preposition – typical ones are: to, in_order_to, thanks_to, for, from, by. As to tensed clauses, the examples with no cue word are markedly different and contain special assertions, heavily paraphrastic and idiomatic. This justifies a model based on a bipartition of clauses, where finite ones will only convey Causal Relations in exceptional cases and only in presence of a semantic marker.

3 The Hypothesis

We assumed as our starting hypothesis that CAUSES must be expressed whenever the EFFECTS are unexpected, unwanted, unattested, in other words whenever the FACT constituting the effect clause is new, surprising, negates or contradicts the evidence. To achieve such a pragmatic result, EFFECT clauses must contain at least one of the major or more than one of the following linguistic components:

A. Minor components:

- NEGATION
- MODALITY OPERATOR (adverb, modal verb, adjective, etc.)
- QUANTITY OPERATOR (quantifiers, intensifiers, adjectives, etc.)
- DISCOURSE MARKERS indicating ADVERSE or CONTRARY meaning (but).

B. Major components:

- DISCOURSE MARKERS indicating cause (because_of, caused_by, due_to, as, etc.)
- DISCOURSE MARKERS indicating result (why, result_from, as_a_result, so_that, so)
- LEXICALIZED NEGATION (refuse, reject)
- ATTITUDINAL VERBS (think, believe, expect, criticize, attack, etc.)
- EMOTIONAL VERBS (concern, worry, doubt, fear, etc.)
- QUANTITY GRADING VERBS (rise, soar, slump, cut_down, etc)
- EVALUATIVE PREDICATES (wrong, right, improve, worse, etc.)
- NECESSITY PREDICATES (must, ought_to, need, have_to, etc.)
- EVENTIVE DEVERBAL NOUNS and ADJECTIVES indicating (natural) catastrophe, illegal or dangerous events (attack, crash, weapon, bribe, etc.)
- PREDICATES (adjectives, adverbials, verbs, nouns) indicating novelty, discovery (scientific and not)
- PREDICATES (adjectives, adverbials, verbs, nouns) indicating problematic and troublesome situations and events.

C. Structural components:

- RESULTATIVE INFINITIVALS
- ADJUNCT GERUNDIVES headed with for, by
- ADJUNCT PPs OR NPs whose nominal head is deverbal or indicates a state-result in aspectual terms.

D. Referential components:

- SUBJECT pronominals corefer
- Possessive pronominal corefers to a previous argument
- The second sentence in the discourse pair starts with a deictic pronoun (this/that).

3.1 Towards a General Model of CRs

Attempts at producing Automatic Discourse Relations Classification are very few in the literature: Ken Barter in 1995 produced an interactive algorithm and an evaluation, limited though to a small number of examples. As reported above, [7] carried out a study with the aim of detecting discourse relations automatically. They trained an automatic classifier to recognize the relations that were not signaled by a discourse marker. The classifier was trained on examples of actual relations versus examples of

non-relations (random pairs of units, sometimes each taken from different documents), using lexical patterns. It learnt to distinguish relations that were not signaled by a discourse marker, increasing accuracy over a discourse-marker-based method by as much as 77%. This result refers to an experiment carried out on automatically classified DRs. When tested on manually annotated data, the same algorithm has been reported to achieve an accuracy of 57%.

We now comment [9]'s paper, where the authors report on another experiment intended to improve on the results reported in [6]. The experiment is organized around two tasks: segmenting the text into Discourse Units, on the basis of DUs organized with lexical heads and syntactic constituency, and producing an automatic labeling of the test data. The final results fare better than the previous ones; F-measure is around 68%.

The model we are thinking of is illustrated in Figure 1 below, and is based on three main sources of information:

- structural information, to tell CAUSE apart from the remaining relations;
- lexical information, to detect complementation cases of complement nontensed clauses – infinitivals and gerundives – from ADJUNCT, which alone can contribute to NON-CAUSE relations;
- tense and reference as constraints on the definition of CAUSE relations.

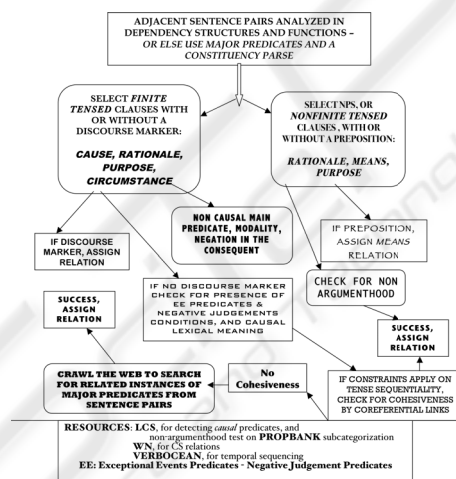


Fig. 1. A scheme of the algorithm to detect CRs with symbolic procedures.

In addition to these main factors, we will use knowledge of the world in the form of evaluative, subjective, Exceptional Event predicates, predicates of negative judgment, etc. as they have been defined in the Hypothesis section above.

The system uses the output of a parser – in our case GETARUNS' parser – made up of Functional Dependency relations, where we added a number a semantic items to complete the interpretation task. At first the clauses are subdivided into two structural types: tensed clauses are set apart from untensed clauses. Untensed clauses are checked for Non-Argumenthood before they may be considered as DUs. They contribute a Result interpretation that may trigger the Cause relation in the main clause. Else, they are interpreted as Rationale or Purpose and trigger an Outcome relation in

the main clause. Gerundives preceded by preposition *by* will be interpreted as Means; those preceded by *for* are interpreted as Rationale or Purpose according to additional conditions obtained in the main clause – see below.

As to tensed clauses, be they relative clauses, complement clauses, coordinate clauses, or simply main clauses, they may express all gamut of Causal Relations. The most difficult to be assigned, as we noted above, is Circumstance. It requires a certain number of constraints to apply before the interpretation is accepted by the algorithm. In particular, the SUBJECT argument must be non-animate non-human, or else the proposition must be under the scope of modal operators, or the main predicate must be a semantically opaque predicate. In the presence of subordinators, specific cue words or causality markers, the discourse relation is assigned straightforwardly. After that, the choice is still difficult to make: the algorithm will use the list of Exceptional Event predicates, Negative Judgment predicates, as well as the presence of negation or other indicators of some uncommon or unpredictable situation that requires the Cause to be explicitly expressed. Finally, if some of these are obtained, the algorithm looks for cohesive links by accessing the output of the anaphora resolution algorithm and checks for the presence of Resolved Nominals or Pronominals, i.e. whether the current clause contains nouns/pronouns which corefer to some previously mentioned entity present in the history list. In case this also fails and there is no cohesive link, the system will have to search WordNet, LCS and FrameNet for some indication that the Predicate-Argument Structures contained in the two clauses under examination can be taken to be explicitly in a Causality Relation. We usually start out by a lookup in LCS. LCS entries contain cross reference to Levin verb classes, to WordNet senses and to PropBank argument lists. These have been mapped to a more explicit label set of Semantic Roles, which can be regarded more linguistically motivated than the ones contained in FrameNet, which are more pragmatically motivated. However, for our purposes, LCS notation is more perspicuous because of the presence of the CAUSE operator, and is more general: from a total number of 9000 lexical entries, 5000 contain the CAUSE operator. On the contrary, in FrameNet, from a total number of 10000 lexical entries, only 333 are related to a CAUSE Frame; if we search for the word “cause” in the definitions contained in all the Frames, the number increases to 789, but is still too small compared to LCS.

3.2 The CRs Algorithm in Detail

We will now present in detail the contents of the algorithm from a technical and linguistic point of view. The algorithm performs these actions:

1. collects linguistic information from the parser output;
2. translates linguistic information into Semantic and Pragmatic Classes;
3. assigns Discourse Relations on a clause-by-clause level: some will be specific, some generic or default;
4. detects the presence of Causality Relations according to the algorithm traced above;
5. determines coreference relations between the Discourse Units thus classified.

The input to the algorithm is reported in Table 1: the list of linguistic items represents clause level information as derived from dependency structure – related to the example “he was not frozen in place by rigid ideology”:

Table 1. Linguistic information from dependency structure.

<i>Speech</i>	<i>Predicate</i>	<i>Tense</i>	<i>Voice</i>	<i>Modal</i>
direct	freeze	past	passive	nil
<i>Neg</i>	<i>Aux</i>	<i>Foc</i>	<i>Aspect</i>	<i>Mood</i>
not	nil	nil	Accomplishment	Indic

Lexical aspect is derived from LCS following the approach presented in [5]. The remaining linguistic items are quite straightforward to describe from the output of a dependency parser. Table 2 contains the semantic interpretation of the initial features, which is wrought by means of combining features together and the use of external knowledge to tell different classes of verbs apart, which we take from WordNet, LCS, FrameNet and PropBank. For instance, Auxiliary and Tense contribute to produce Comptense; Point of view depends on Speech and the semantic class of verb; Factuality depends on Mood and Modality; Relevance depends on negation and Viewpoint. A full set of rules is reported in [4] (to be published).

Table 2. Informational structure derived from linguistic information.

<i>ClauseType/SemRole</i>	<i>View-point</i>	<i>Comptense</i>	<i>Factuality</i>	<i>Relevance</i>
main/prop	external	past	no	Back-ground

The second step takes as input the forms shown above in Table 2 and the list of arguments and adjuncts, and produces the output of Table 3. The second step is represented below by three sets of rules:

1. the first rule looks for clauses governed by be/have, i.e. copulative constructions in which the verb has no semantic relevance;
2. the second rule looks for all cases of clauses, be it finite or nonfinite;
3. the third rule takes into account adjunct PPs governed by a preposition, which have already received their interpretation by the parser.

Table 3. Discourse structure derived from informational structure.

<i>DISCREL</i>	<i>DISCDOMAIN</i>	<i>DISCSTRUCTURE</i>
narration	objective	down (1-2)

Finally, as explained above and indicated in Figure 1, the rules elaborated for the detection of Causal Relations are fired after the other DRs have already been decided. The rules have the task of recovering the Cause/Circumstance whenever a Result/Effect has already been decided by the previous algorithm. To this aim, we try to take advantage of the information encoded in two resources: WordNet and FrameNet. WordNet contains a small list of Cause-Effect inferential links between verb predicates – some 200. FrameNet encodes the same information in a few Frame to Frame relations, 12 – 14 if we add Inchoatives. We also adapted the information available in VerbOcean⁵, which, however, only focuses on temporal restrictions. It has a list of HAPPENS-BEFORE relations amounting to 6500, but only a small percentage of those can actually be computed as entertaining some causal relation.

In order to spot Negative Judgment predicates we use the Laswell Value Dictionary and the Harvard IV dictionary⁶. As discussed above, DISCOURSEREL is recomputed after the first pass.

⁵ <http://semantics.isi.edu/ocean/>

⁶ <http://www.webuse.umd.edu:9090/tags/>

In case a RESULT or a RATIONAL clause has been previously computed, the system looks for the most adequate adjacent clause to become a CAUSE or an OUT-COME. This is done by checking Temporal and Coreferentiality constraints. The outcome of this pass is to relabel some previously labeled Discourse Relations that constitute default relations into the one required by the context.

4 Results and Evaluation

In order to carry out the evaluation, we assumed that the gap existing between Discourse Units (hence DUs) and Syntactic Units (hence SUs) should be minimal. To that end, we compared the three corpora at hand with the task to be realized. As it turned out, PDTB resulted unsuitable for both tasks, the DRR and the DSR. The reason for that is simply the arbitrariness of correspondences existing between Syntactic Units and Discourse Units. The annotators had syntactic structures available from PTB and did not impose themselves a strict criterion for the definition of what a DU might correspond to. The result is an average proportion of 3.2 clauses per DU, a fact that makes the corpus unsuitable for either tasks. In their presentation of annotation guidelines, downloadable from PDTB website, [8] summarize in one slide what counts as a legal Discourse Unit, which they refer to as Argument: a single clause (tensed/untensed); a single sentence; multiple sentences; NPs that refer to clauses; and some nominal forms expressing events or states. So, basically, any syntactic unit may actually become a discourse Argument, and the correspondence Clause-DU is completely lost. The problem is that, in order to take automatic decisions, the criteria to be used become hard to define, and refer to the pragmatic and semantic domain rather than simply and more safely to syntax. This situation should also affect machine learning methods, which would use general linguistic features to build their models and then would be at a loss when trying to scale them to different syntactic structures.

So, finally, we are left with two corpora, DGB and RST. The latter in particular is very strictly compliant with a syntax based unit correspondence. DGB, even though not intended to respond to such a criterion, has an average ratio of clauses per DU of 2.2. We decided to cope with this inconsistency by assigning a Discourse Relation to a corresponding larger Unit in case the smaller one was missing when computing the same sentence, on the basis of the presence of the same main predicate.

General data for the two corpora are tabulated below:

Table 4. General data and EDU/sentence ratio.

	<i>No. of sentences</i>	<i>No. of words</i>	<i>No. of EDUs</i>	<i>RatioEDU/Ss</i>
<i>RST-test</i>	799	19300	2346	2.9
<i>DGB</i>	3110	72520	8910	2.8

Table 5. Quantitative evaluation of DU segmentation.

	<i>No. of EDUs</i>	<i>My no. of EDUs</i>	<i>Recall</i>	<i>Precision</i>
<i>RST-test</i>	2346	2281	97.2	87.9
<i>DGB</i>	8910	8910	100	89.8

Results reported in [9] report an F-measure of 84.7% when using the PennTreebank structure and 83.1% when using Charniak's parser.

Table 6. F-measure values for the DGB experiment.

	<i>Causality relations</i>	<i>Condition</i>	<i>Contrast</i>	<i>Other relations</i>
DGB	38.13	48.25	44.34	59.13

The overall F-measure for DGB is 64.57%.

5 Conclusions

We presented an algorithm for the automatic detection of discourse relations, particularly focused on Causality Relations, that works on linguistically-based symbolic rules and needs no training. The paper was mainly concerned with the work carried out on characterizing Causality Relations from a syntactic, semantic and pragmatic point of view. The corpus we used to test our system, Discourse Graphbank, has not yet been used by other research teams, and there is no comparison we can make. However, we intend to use it and test it with other corpora. Our algorithm only contains general rules and no special procedure to cope with implicit Discourse Relations. We may safely assume that once these rules have been implemented and tested extensively on the corpora available, an increase in the accuracy is to be expected.

References

1. Carlson L., Marcu, D. and Okurowski M. 2002. *RST Discourse Treebank*. Philadelphia, PA:Linguistic Data Consortium.
2. Delmonte R. 2003. *Parsing Spontaneous Speech*. In proceedings of EUROSPEECH2003.
3. Delmonte R. 2005. *Deep & Shallow Linguistically Based Parsing*. In A.M. Di Sciullo (ed.), *UG and External Systems*, John Benjamins, Amsterdam/Philadelphia, pp. 335-374.
4. Delmonte R. 2007 (to be published). *Computational Linguistic Text Processing*. Nova Publishers, New York.
5. Dorr B. J. and Olsen M. B. 1997. *Deriving Verbal and Compositional Lexical Aspect for NLP Applications*. In proceedings of the 35th ACL, pp. 151-158.
6. Marcu, D. 2000. *The theory and practice of discourse parsing and summarization*. Cambridge, MA: MIT Press.
7. Marcu D. and Echiabi A., 2002. *An Unsupervised Approach to Recognizing Discourse Relations*. In proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, PA, pp. 368-375.
8. Miltsakaki E., Prasad R., Joshi A. and Webber B. 2004. *The Penn Discourse Treebank*. In proceedings of the 4th International Conference on Language Resources and Evaluation, Lisbon, Portugal.
9. Soricut R. and Marcu D. 2003. *Sentence level discourse parsing using syntactic and lexical information*. In proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1, pp. 149-156.
10. Wolf F., Gibson E., Fisher A., Knight M. 2005. *Discourse Graphbank*. Linguistic Data Consortium, Philadelphia.