A Repository of Rules and Lexical Resources for Discourse Structure Analysis: the Case of Explanation Structures

Sarah Bourse, Patrick Saint-Dizier

IRIT-CNRS, 118 route de Narbonne
31062 Toulouse cedex France, sarah.bourse@univ-tlse2.fr, stdizier@irit.fr

Abstract

In this paper, we present an analysis method, a set of rules, lexical resources dedicated to discourse relation identification, in particular for explanation analysis. The following relations are described with prototypical rules: instructions, advice, warnings, illustration, restatement, purpose, condition, circumstance, concession, contrast and some forms of causes. Rules are developed for French and English. The approach used to describe the analysis of such relations is basically generative and also provides a conceptual view of explanation. The implementation is realized in Dislog, using the <TextCoop> logic-based platform, that also allows for the integration of knowledge and reasoning into rules describing the structure of explanation.

Keywords: discourse analysis, logic programming, discourse markers.

1. Challenges of Discourse Analysis

Discourse structure analysis is a very challenging task because of the large diversity of discourse structures, the various forms they take in language and the impact of knowledge and pragmatics in their identification (Longacre, 1982; Keil, 2000). Recognizing discourse structures cannot in general only be based on purely lexical or morphosyntactic considerations: subtle kinds of knowledge associated with reasoning schemas are often necessary. These latter capture the various facets of the influence of pragmatic factors in our understanding of texts (Kintsch, 1988; Di Eugenio et al., 1996). The importance of structural and pragmatic factors does depend on the type of relation investigated, on the textual genre and on the author and targeted audience. In our context, didactic or technical texts such as procedures or requirements are obviously much easier to process than free-style texts.

Rhetorical structure theory (RST) (Mann et al., 1988, 1992) is a major attempt to organize investigations in discourse analysis, with the definition of 22 basic structures. Since then, almost 200 relations have been introduced which are more or less clearly defined. Background information about RST, annotation tools and corpora are accessible at `http://www.sfu.ca/rst/`. A recent overview is developed in (Taboada et al., 2006). Very briefly, RST posits that coherent texts consist of minimal units, which are all linked to each other, recursively, through rhetorical relations. No unit is left pending: all units are connected to others. Some text spans appear to be more central to the text purpose, these are called nuclei (or nuclei), whereas others are somewhat more secondary, these are called satellites. Satellites must be associated with nuclei: they get their meaning from the nucleus they are associatred with. Relations between nuclei and satellites are often one-to-one or one-to-many.

The literature on discourse analysis is particularly abundant from a linguistic point of view. Several approaches, based on corpus analysis with a strong linguistic basis are of much interest for our purpose. Relations are investigated together with their linguistic markers in works such as (Delin et al. 1994), (Marcu 1997, 2002), (Kosseim and Lapalme, 2000) with their usage in language generation in (Rösner and Stede, 1992), and in (Saito et al. 2006) with an extensive study on how markers can be quite systematically acquired. A deeper approach is concerned with the cognitive meaning associated with these relations, how they can be interpreted in discourse and how they can trigger inferential patterns (Wright, 2004; Moschler, 2007; Fiedler, 2001), just to cite a few works).

Within Computational Linguistic circles, RST has been mainly developed in natural language generation for content planning purposes, e.g. (Kosseim et al. 2000), (Reed et al. 1998). Besides this area, (Marcu 1997, 2000) developed a general framework and efficient strategies to recognize a number of major rhetorical structures in various kinds of texts. The main challenges are the recognition of textual units and the identification of relations that hold between them. The parsing algorithm he introduced relies on a first-order formalization of valid text structures which obey a number of structural assumptions. These, however, seem to be somewhat too restrictive w.r.t. our observations. In particular our observations show that the following assumptions are too restrictive: relations occur between non-overlapping text spans, relations are either vertical or horizontal (they can involve non parent nodes), text structure is a binary-branching tree in most cases (we have many situations with more than two nodes). Marcu's work is based on a number of psycholinguistic investigations (Grosz et al., 1986) that show that discourse markers are used by human subjects both as cohesive links between adjacent clauses and as connectors between larger textual units. An important result is that discourse markers are used consistently with the semantics and pragmatics of the textual units they connect and they are relatively frequent and unambiguous.

In this paper, we first explore and define the facets of explanation: what it is, which purposes it serves, and what are the discourse relations which are involved. We then give a definition of these relevant relations, within the perspective of explanation, and illustrate them by means of a few prototypical rules. We also focus on the lexical re-

sources which are needed to develop them. These rules are given in a readable format, these are implemented in Dislog, running on the <TextCoop> platform (Saint-Dizier, 2012 a,b). This repository of rules will be shortly available in French and English under a free licence, together with the <TextCoop> platform.

2. A generative Approach to Discourse Analysis

Discourse analysis, and the analysis of the structure of explanation in particular, is quite a difficult task. Our approach is somewhat generative:
- We first define the structure of nuclei and satellites for the different relations we investigate, focussing on their fundamental structure, something comparable to a base form.
- We then develop binding principles and offer the possibility to express various types of constraints.
Each system is relatively simple, and captures interesting linguistic observations and generalizations. The complexity arises from the interactions between these components. Then, as illustrated below, it is clear that discourse analysis is a complex task, that includes, among others, the identification of ambiguities as well as the processing of complex constructions.
To be able to recognize complex organizations of discourse structures, we have developed the <TextCoop> platform (Saint-Dizier 2012 a,b) that allows, besides rules recognizing base forms for discourse relations (encoded in the DISLOG language), to introduce various well-formed constraints expressed as principles. Next, <TextCoop> offers a relatively complex processing strategy that can detect complex language structures, in particular:

- several structures may be embedded,
- others may be chained (when a satellite is a nucleus for another relation),
- nuclei and related satellites may be non-adjacent,
- nuclei may be linked to several satellites of different types,
- some satellites may be embedded into their nucleus.

As a result, discourse relations receive a relatively simple description, with well-identified lexical resources; the strategy implemented in <TextCoop> manages the hard recognition tasks, under well-formedness constraints. Ressources used are essentially lexical, e.g.: connectors, verbs, semantic features. A few grammatical and morphological considerations are also used in rules.
The second aspect offered in the Dislog language is the possibility to include reasoning and knowledge constraints into the rules. Dislog being interpreted in Prolog, this is quite natural and easy, provided that links can be established with e.g. standard ontologies or terminologies or knowledge bases. An example is developed below.
The approach developed here for discourse analysis makes the distinction between two phases: recognizing, via rules or patterns, basic structures (satellites or nucleui), and then the binding of a nucleus with one or more satellites by means of binding rules. These rules introduce a very powerful mechanism. They have the same syntax in Dislog as basic rules. There are many advantages to this approach: first it is more modular, since the various components of discourse analysis are developed separately, next it allows to bind a nucleus with several satellites, possibly not adjacent. It also allows binding rules to develop more complex configurations, as can be found in domain dedicated texts. No additional complexity is then added into the linguistic description. The interactions and priorities between binding rules and principles can also be investigated and specified independently from the remainder of the linguistic descriptions.
Another observation is that it is much easier to recognize satellites than nuclei, which, in the case of explanation, are quite neutral from the language point of view. In a number of situations, it is necessary to develop inferences to be able to accurately identify a nucleus. A prototypical example is developed below for the case of illustration. While the satellite is clearly marked, the nucleus can only be identified on the basis of knowledge. Another advantage of having binding rules specified apart is that dedicated reasoning aspects can be dealt with at the appropriate level of the analysis. A few binding rules and examples where knowledge is used are also developed in this paper.
From a foundational point of view, our analysis of discourse, and explanation in particular, aims at defining a kind of conceptual or cognitive analysis of discourse: while text spans involved in discourse relation convey a certain meaning, it is also of much interest to precisely identify the semantics conveyed by the relations themselves, taking into account syntactic considerations such as the position of the various text spans.
To conclude this section devoted to methodological considerations, consider the following example that illustrates some typical discourse structures and their interactions:

[$procedure$ [$purpose$ Writing a paper: [$advice$ Read light sources, then thorough]]
[$assumption/circumstance$ Assuming you've been given a topic,]
[$circumstance$ When you conduct research], [[$advice\_concl$ move from light to thorough resources [$advice\_support$ to make sure you're moving in the right direction]].
[$instruction$ Begin by doing searches on the Internet about your topic [$purpose$ to familiarize yourself with the basic issues;]]
[$temporal-sequence$ then] [$instruction$ move to more thorough research on the Academic Databases];
[$temporal-sequence$ finally], [$instruction$ probe the depths of the issue by burying yourself in the library.]
[$warning$ [$warning\_concl$ Make sure that despite beginning on the Internet, you don't simply end there.
[$warning\_support$ A research paper using only Internet sources is a weak paper, [$consequence$ which puts you at a disadvantage...]]]]
[$advice$ [$advice\_support$ While the Internet should never be your only source of information], [$contrast$ it would be ridiculous not to utilize its vast sources of information. [$advice\_concl$ You should use the Internet to acquaint

yourself with the topic more before you dig into more academic texts.]]]]

3. Explanation in action

3.1. Elements of a definition

Explanation and its relations to language and linguistics is a relatively new but vast area of investigation. It requires to take into consideration a large number of linguistic aspects, from syntax to pragmatics, but also typography. At the moment, explanation is developed in a number of sectors as diverse as didactics, procedures, health and safety, requirement engineering and, in interactive environments, systems such as helpdesks. In artificial intelligence, explanation is often organized around the notion of argumentation (Amgoud et al., 2005; Reed, 1998; Walton et al., 2008), but argumentation is just one facet of explanation. Two decades ago, explanation was used to produce natural language outputs for expert systems from predefined templates.

Explanation is a concept which is difficult to define. Briefly, it is a way, given a concept, an event or a goal, to provide the reader or listener with more information about it via e.g. elaboration, illustration, argumentation, etc. behind explanation, there is always one or more communicative goal(s) (Bourse and Saint-Dizier, 2011).

Explanation is composed of a sequence of informational elements organized via discourse structures, they are in general structured with the intent of reaching a goal. This goal may be practical or more interpersonal or epistemic (e.g. convince someone to do something in a certain way, negotiate with someone while providing explanations about ones point of view). Explanations are often associated with a kind of instructional style which ranges from injunctive to advice-like forms. Procedures of various kinds (social recommendations, as well as do-it-yourself (DIY), maintenance procedures, health care advice) and didactic texts form an excellent source of corpus to observe how explanations are constructed and linguistically realized.

Explanation occurs also in goal-driven but non-procedural contexts, e.g. as a means to justify a decision in legal reasoning or in political discourse, in opinion expression, in cooperative question-answering. Explanation may also be associated with various pragmatic effects (irony, emphasis, dramatization, etc.) for example in political discourse. In each of these cases, explanation does keep a goal-oriented structure (Carberry, 1990; Takechi et al., 2003). Finally, it is central to a number of types of dialogues, negociation, clarification situations, persuasion strategies, etc.

Our main objective is to identify a number of prototypical, widely used, explanation schemes as well as their linguistic structure (e.g. prototypical language markers or constructs, planning issues), and to categorize their communicative goals.

3.2. Introducing reasoning aspects into discourse analysis

Discourse relations identification often requires some forms of knowledge and reasoning. This is the case to resolve ambiguities in a relation identification when (1) there are several candidates or (2) to clearly identify the text span at stake. While some situations are extremely difficult to resolve, others can be processed e.g. via lexical inference or reasoning over ontological knowledge. Dislog allows the introduction of reasoning, and the <TextCoop> platform allows the integration of knowledge and functions to access it and reason about it.

This problem is very vast and largely open, with exploratory studies e.g. reported in (Van Dijk, 1980), (Kintsch, 1988), and more recently some debates reported in (`http://www.discourses.org/UnpublishedArticles/SpecDis&Know.htm`) .

Let us give a simple motivational example. The utterance (found in our corpus):

... red fruit tart (strawberries, raspberries) are made ...

contains a structure: (*strawberries, raspberries*) which is ambiguous in terms of discourse functions: it can be an elaboration or an illustration, furthermore the identification of its nucleus is ambiguous:

red fruit tart, red fruit ?

A straightforward access to an ontology of fruits tells us that those berries are red fruits, therefore:

- the unit *strawberries, raspberries* is interpreted as an illustration, since no new information is given (otherwise it would have been an elaboration)
- its nucleus is the '*red fruit*' unit only,
- and it should be noted that these two constituents, which must be bound, are not adjacent.

Similarly, the relation between an argument conclusion and its support (the reasons) may not necessarily be straightforward to identify and may involve various types of domain and common-sense knowledge:

Do not park your car at night near this bar: it may cost you fortunes.

Women's living standards have progressed in Nepal: we now see long lines of young girls early morning along the roads with their school bags. (Nepali Times).

In this latter example, *school bag* means going to school, then *school* means education, which, in turn, means better living conditions.

3.3. Processing complex constructions: the case of Dislocation

Similarly to syntax, we identified in relatively 'free style' texts (i.e. not as controlled as technical procedures) phenomena similar to quasi-scrambling situations, free-structure ordering or cleft constructions. This is in particular the case for arguments which are semantically complex constructs, subject to syntactic variations due to pragmatic considerations such as focus or foregrounding. These issues are 'deep' syntactic discourse constructions that need to be explained and modeled from a language point of view.

As an illustration, let us consider a relatively frequent situation that we call **dislocation**, which is close in the surface to cleft constructions in syntax (Lasnik et al., 1988), which occurs when, in a two segment construction, one segment is embedded into the other, as in:

Strawberries and raspberries are red fruits, for example.

'red fruits' is the nucleus of the relation while the illustration is split into two parts: 'strawberries and raspberries' and 'for example'. Here the nucleus is included into the

satellite.
In the following example:
Products X and Y, because of their toxicity, are not allowed in this building.
the support (its motivation or goal) of the argument is embedded into the conclusion (the main utterance), probably to add some stress on the toxicity of the products.
Finally, we observed in our corpora quasi-scrambling situations, a simple case being the illustration relation. Consider again the example above, which can also be written as follows:
Strawberries are red fruits similarly to raspberries, for example.
where the enumeration itself is subject to dislocation.

4. Corpus analysis, rule authoring and lexical resources

Our rule system is based on the analysis and tagging of quite a large set of texts: 255 procedures from 14 domains (do-it-yourself, gardening, health care, social behaviour, and professional procedures in aeronautics, energy, transportation, finance, communications, etc.) often with requirement specifications and 75 didactic texts. About half of the texts are in French and the other half in English. We have the equivalent of 400 pages (or 140 000 words) of procedures and 110 pages (or 45 000 words) of didactic texts. About 3400 occurences of discourse structures related to explanation (nucleus and satellite) have been collected for both French and English and will be used to develop rules and resources.
Our approach is based on the following method:

1. Collecting and tagging the corpus. Three trained annotators, fluent in French and English, tagged the same texts in order to limit misinterpretations (Kappa test shows an agreement of 78%, higher for procedures which are easier to interpret than didactic texts). Manual analysis is necessary because of the complexity of the observed data, as advocated above.

2. Tags are defined for a number of relevant discourse relations: illustration, elaboration, definition, circumstance, purpose, etc. Kernels and satellites are analyzed apart. For relations such as illustration, the satellite is the most prominent, its nucleus (what is exactly illustrated) must often be inferred from the satellite.

3. For a given discourse structure, corpus realizations are investigated and rules of an appropriate level of generalization are manually produced while keeping the linguistic features as explicit as possible (e.g. no ad hoc linguistic category is created), related lexical resources are created and generalized (to have as many realizations as possible). Although automatic methods could be used, we believe that the complexity of these rules requires a manual analysis to get a better linguistic accuracy.

4. Discourse structure rules are then written in the Dislog formalism (Saint-Dizier, 2012b). Each discourse structure may give rise to several rules, which are grouped into a cluster of rules.

5. Then, selective binding rules may be recursively defined to bind discourse structures, in particular to bind a nucleus with its satellite(s). Selective binding rules allow to express the complex situations presented above.

6. Dislog runs on the <TextCoop> logic-based platform. This platform offers the possibility to express constraints and priorities. In particular, it is possible to specify which rule cluster must be executed before others, via the description of a cascade of rules in the execution schema. It is also possible to indicate precedence and dominance constraints.

5. A repository of rules and lexical resources

The following elements have been designed for French and English for the rules we have designed and with the intent of authoring new rules:

- lists of connectors, which are organized by general types: time, cause, concession, etc.,
- list of terms which are specific to certain discourse functions,
- lists of verbs organized by semantic classes, close to those found in WordNet, that we have adapted or refined for discourse analysis, e.g. propositional attitude verbs, report verbs (Wierzbicka, 1987), etc.,
- list of terms with positive or negative polarity,
- some already defined clusters of discourse rules to recognize general purpose discourse functions (these are given and evaluated below),
- some predefined functions and predicates to access knowledge and control features (e.g. subsumption).

The result is the initial text, tagged by the different relations which have been identified. This is automatically realized in Dislog. Dislog also offers the possibility to develop partial dependency representations instead of tags; representations which turn out to be more appropriate in a number of situations, in particular when long-distance dependencies are involved.
The following subsections provide examples, among **the most common and the most simple** of the rules and lexical resources we have developed. For the sake of readability, some minor simplifications have been introduced. The rule base will be shortly available on demand.
For each discourse relation, a definition is given in addition to the examples of rules and resources. Then some linguistic realizations of discourse relations coming from our corpus of English didactic texts or procedures are provided. In the rules, “eos” stands for “end of sentence”. The curly brackets show that an element is optional. Resources given here are in general samples.
The details of the syntax of the rules is given in (Saint-Dizier 2012a).

5.1. Instruction

Definition: An instructive is a statement, often in an imperative form, that expresses the need to realize an action. This action can possibly be associated with various elements such as instruments, equipments, manners, etc. The main verb of an instruction is often in the imperative or infinitive form in French.
Number of specific rules : 12.
A few structures:
`Advice →`
`gap(not(neg), verb(action, infinitive), gap, eos. /`
`gap(not(neg), verb(_, faire), gap, verb(action, infinitive), gap, eos.`
'infinitive' denotes a veb in the infinitive form (without 'to'), 'faire' is a light verb in French, 'action' denotes an action verb, which is in general domain dependent.
Resources:
Besides modals and a few terms like pronouns, the main resource is a list of action verbs. However, in most cases, there is a need for only a limited set of verbs, about 100.
Example:
Write titles in bold font.

5.2. Title

Definition: Titles introduce a document or a part of it. In our context they express a high level goal, where the instructions that follow and the other explanation elements describe a way to reach this goal. Subtitles introduce a goal sub-goal hierarchy.
The recognition of titles is a specific problem. It can be done via the typography of the document if there are dedicated markers. Title identification in procedures is developed in (Delpech and Saint-Dizier 2008).

5.3. Advice

Definition: Relation between a conclusion and a support, the conclusion inviting the reader to perform an optional action to obtain better results, and the support giving a motivation for realizing this action.
Number of specific rules : 6 conclusions, 3 supports.
Structures:
`Advice →`
`verb(pref,infinitive), gap(G), eos. /`
`[it,is], adv_prob, gap(G1), exp(advice1), gap(G2), eos. /`
`exp(advice2), gap(G), eos.`
Resources:
verb(pref): *choose, prefer*
exp(advice1): *a good idea, better, recommended, preferable*
exp(advice2): *a X tip, a X advice, best option, alternative*
adv_prob: probably, possibly, etc.
Examples:
Choose aspects or quotations that you can analyse successfully for the methods used, effects created and purpose intended.
Following your thesis statement, it is a good idea to add a little more detail that acts to preview each of the major points that you will cover in the body of the essay.
A useful tip is to open each paragraph with a topic sentence.

5.4. Warning

Definition: Relation between a conclusion and a support, the conclusion drawing the attention of the reader to an action which is compulsory to perform, and the support giving a motivation for realizing this action or the risks which may arise.
Number of specific rules : 9 conclusions, 9 supports.
Structures:
`Warning-conclusion →`
`exp(ensure), gap(G), eos. /`
`[it,is], adv(int), adj(imp), gap(G), verb(action,infinitive), gap(G), eos.`
Resources:
exp(ensure): *ensure, make sure, be sure*
adv(int): *very, absolutely, really*
adj(imp): *essential, vital, crucial, fundamental*
Examples:
Make sure your facts are relevant rather than related.
It is essential that you follow the guidelines for each proposal as set by the instructor.

5.5. Binding rules for warnings

Let us give here a simple example of a binding rule. Warnings are composed of a conclusion and a support (not developed above). These two structures are recognized separetely by dedicated rules. Then, it is necessary to bind these two structures to get a warning. Let us assume that both supports and conslusions are explicitly tagged, then, a simple binding rule is:
`Warning →`
`<warning-concl>,gap(G1),< /warning-concl>, gap(G2), <warning-supp>, gap(G1), < /warning-supp>, gap(G3), eos.`
Then, the whole structure is tagged e.g. <warning>. Similar rules are defined to bind nucleus with their related satellites.

5.6. Cause

Definition: Relation where segment B (traditionally called the antecedent) provokes the realization of an event (the consequent).
Only a small number of cases have been investigated, number of specific rules : 6.
Structures:
`Cause →`
`conn(cause), gap(G), ponct(comma). /`
`conn(cause), gap(G), eos.`
Resources:
conn(cause): *because, because of, on account of*
ponct(comma): *, ; :*
Examples:
Because books are so thorough and long, you have to learn to skim.
Long lists result in shallow essays because you don't have space to fully explore an idea.
Many poorly crafted essays have been produced on account of a lack of preparation and confidence.

5.7. Condition

Definition: Relation where the segment B refers to a situation which is necessary for A to be realized.
Number of specific rules : 8.
Structures:
```
Condition →
conn(cond), gap(G), ponct(comma).  /
conn(cond), gap(G), eos.
```
Resources:
conn(cond): *if*
Examples:
If all of the sources seem to be written by the same person or group of people, you must again seriously consider the validity of the topic.
If you put too many different themes into one body paragraph, then the essay becomes confusing.
For essay conclusions, don't be afraid to be short and sweet if you feel that the argument's been well-made.

5.8. Concession

Definition: Relation where the segment B contradicts part of the segment A, or contradicts the implicit conclusion which can be drawn from segment A.
Number of specific rules: 9.
Structures:
```
Concession →
conn(opposition_alth), gap(G1),
ponct(comma), gap(G2), eos.  /
conn(opposition_alth), gap(G), eos.  /
conn(opposition_how), gap(G), eos.
```
Resources:
conn(opposition_alth): *although, though, even though, even if, notwithstanding, despite, in spite of*
conn(opposition_how): *however*
Examples:
An essay can be immaculately written, organized, and researched; however, without a conclusion, the reader is left dumbfounded, frustrated, confused.
Though the word essay has come to be understood as a type of writing in Modern English, its origins provide us with some useful insights.
Your paper should expose some new idea or insight about the topic, not just be a collage of other scholars' thoughts and research – although you will definitely rely upon these scholars as you move toward your point.

5.9. Contrast

Definition: Symmetrical relation where one segment is opposed to another segment.
Number of specific rules: 5.
```
Contrast →
conn(opposition_whe), gap(G), ponct(comma).
/
conn(opposition_whe), gap(G), eos.  /
conn(opposition_how), gap(G), eos.
```
Resources:
conn(opposition_whe): *whereas, but whereas, but while*
Examples:
The periodic sentence is one in which the main clause is considerably delayed, whereas the cumulative sentence opens quickly with the main clause.

5.10. Circumstance

Definition: Relation where the segment B refers to a frame in which A is to be realized by the reader of the procedure.
Number of specific rules: 12.
```
Circumstance →
conn(circ), gap(G), ponct(comma).  /
conn(circ), gap(G), eos.
```
Resources:
conn(circ): *when, once, as soon as, after, before*
Examples:
Before you put your outline together, you need to identify your argument and analyze it.
Once you use a piece of evidence, be sure and write at least one or two sentences explaining why you use it.

5.11. Purpose

Definition: Relation where segment B provides the aim targeted by the realization of the action expressed in segment A.
Number of specific rules: 14.
```
Purpose →
conn(purpose), verb(action, infinitive),
gap(G), ponct(comma).  /
conn(purpose), verb(action, infinitive),
gap(G), eos.
```
Resources:
conn(purpose): *to, in order to, so as to*
Examples:
To write a good essay on English literature, you need to do five things [...].
In order to make the best of a writing assignment, there are a few rules that can always be followed [...].

5.12. Illustration

Definition: Relation where segment B instantiates a member of segment A, used a representative sample for the class represented by segment A.
Number of specific rules: 20.
```
Illustration →
exp(illus_eg), gap(G), eos.  /
[here], auxiliary(be), gap(G1),
exp(illus_exa), gap(G2), eos.  /
[let,us,take], gap(G), exp(illus_bwe), eos.
```
Resources:
exp(illus_eg): *e.g., including, such as*
exp(illus_exa): *example, an example, examples*
exp(illus_bwe): *by way of example, by way of illustration*
Examples:
This is a crucial point for other types of writing such as fiction or personal essay writing.
Here are some examples of how they can be used well, so long as they are relevant to the essay: [...].

5.13. Restatement

Definition: Relation where segment B rephrases segment A without adding further information.

Number of specific rules: 9.
`Restatement →`
`ponct(opening_parenthesis), exp(restate), gap(G), ponct(closing_parenthesis). / exp(restate), gap(G), eos.`
Resources:
exp(restate): *in other words, to put it another way, that is to say, i.e., put differently*
Examples:
If you must say something in a complicated way spanning several sentences, try adding a sentence to summarize the idea. In other words, make every effort possible to be clear about each point in the essay. When you revise your essay, you'll need to ask yourself, is this argument well made; are there are any gaps in my argument; am I making the case as precisely as I can; are there are any premises or points that I make which aren't integrated into the whole paper. In other words, you'll continue to analyze your essay from the organizational and precision perspectives we've already discussed.

6. Results and performances

Rules and lexical resources are stored in a logic-based database in Dislog. The format can be adapted to various parser environments.
Results are reported here for English, and due to space limitations, only some relations are reported here. As a summary, discourse structures rules, elaborated from corpora over various domains, require the following resources:
In the following table, (1) stands for discourse markers, (2) connectors, (3) modal and non-modal auxiliaries, (4) negation operators, (5) pronouns, (6) prepositions, (7) punctuation, typography.

structure	1	2	3	4	5	6	7
instruction			X				X
advice concl.			X		X		
advice support		X	X		X		
warning concl.			X	X	X	X	
warning support		X	X	X	X	X	
illustration	X		X				X
restatement	X		X				X
purpose		X					X
condition		X					X
circumstance		X	X				X

Table 1: Resources, closed categories

Dislog offers the possibility to integrate knowledge to help resolve ambiguities. These concern mainly: resolving relation identification or ambiguities between various relations, and identifying the exact text span involved in a discourse unit. Modals, auxiliaries and open categories which are involved are given below. Then, knowledge refers e.g. to ontological data to resolve scope ambiguities. In the following table, (1) stands for action verbs and other verb classes, (2) adverbs, (3) expressions with negative (-) or positive (+) polarity, (4) ad hoc expresssions, (5) knowledge.

structure	1	2	3	4	5
instruction	action verbs	X			
advice concl.	communication	X		X	
advice support	change verbs		+	X	
warning concl.	communication	X		X	
warning support	change verbs		-	X	
illustration				X	X
restatement	epistemic			X	X
purpose				X	X
condition					X
circumstance				X	

Table 2: Resources, open categories

As can be noted, rules mainly require re-usable data. Action verbs as well as well ad hoc terms can be specialized.
From a test corpus (31 500 words), with the same distribution as above, we have the following coverage and accuracy rates, expressed in terms of recall and precision. Our strategy was to favour precision over recall since some discourse structures may be somewhat ambiguous or close to each other.
The following figures are based on a comparison of the system performances w.r.t. manual annotations. A structure is correct if it is correctly identified and well-delimited.

structure	number manually annotated	precision (%)	recall (%)
instruction	554	98	96
advice concl.	49	87	76
advice support	42	91	82
warning concl.	112	91	88
warning support	88	93	90
illustration	38	92	87
restatement	47	86	79
purpose	101	89	86
condition	168	93	82
circumstance	121	95	92

Table 3: Textcoop performances for the identification of discourse relations

Selective binding rules allow to bind nuclei and satellites or conclusions and supports for arguments. Since these rules are based (1) on previously assigned discourse tags and (2) some forms of reasoning over knowledge, these are relatively efficient. We get for arguments, advice and warnings together, a precision of 94 % for a recall of 89 %.

7. Conclusion and Perspectives

This paper is a first step towards defining a conceptual and linguistic analysis of explanation, as it can be found in a number of types of texts: procedures, requirements, didactic textes, etc. We presented here an analysis method, a set of rules and lexical resources dedicated to discourse relation identification, in particular for explanation analysis. The following relations are described with prototypical

rules: instructions, advice, warnings, illustration, restatement, purpose, condition, circumstance, concession, contrast and some forms of causes.

Rules are developed for French and English. The appraoch used to describe the analysis of such relations is basically generative and also provides a conceptual view of explanation. The implementation is realized in Dislog, using the <TextCoop> logic-based platform, that also allows for the integration of knowledge and reasoning into rules describing the structure of explanation.

Explanation analysis requires the taking into account of communication goals and intentions, this is a vast area of investigation in pragmatics, for which the tools and resources we have developed here constitute a first step. An important issue is the central role played by argumentation, and the role the other discourse relations presented here are used to deepen in various ways these arguments.

Resources (rules and lexical data) will be made available shortly via a repository, where updates can be uploaded by users. TextCoop and Dislog will also be shortly provided with a free licence.

8. Acknowledgements

This project is supported by the French ANR project LELIE and partly by an IFCPAR Indo-French project. We are also very grateful to a number of colleagues for discussions about this work, including Lionel Fontan and Estelle Delpech and three anonymous reviewers.

9. References

Bourse, S., Saint-Dizier, P., 2011, *The language of explanation dedicated to technical documents*, Sintagma, vol. 23.

Carberry, S., 1990, Plan Recognition in natural language dialogue, Cambridge university Press, MIT Press.

Delin, J., Hartley, A., Paris, C., Scott, D., Vander Linden, K., 1994, *Expressing Procedural Relationships in Multilingual Instructions*, Proceedings of the Seventh International Workshop on Natural Language Generation, pp. 61-70, Maine, USA.

Delpech, E., Saint-Dizier, P., 2008, Investigating the structure of procedural texts for answering How-to Questions, Language Resources and Evaluation Conference (LREC 2008), Marrakech, European Language Resources Association (ELRA).

Di Eugenio, B. and Webber, B.L., 1996, Pragmatic Overloading in Natural Language Instructions, International Journal of Expert Systems.

Fiedler, A., Horacek, H., 2001, Argumentation in Explanations to Logical Problems, in Proceedings of ICCS 2001, Springer LNCS 2073, pp. 969978.

Keil, F.C., Wilson, R.A., 2000, Explanation and Cognition, Bradford Book.

Kintsch, W., 1998, *The Role of Knowledge in Discourse Comprehension: A Construction-Integration Model*, Psychological Review, vol 95-2.

Kosseim, L., Lapalme, G., 2000, *Choosing Rhetorical Structures to Plan Instructional Texts*, Computational Intelligence, Blackwell, Boston.

Lasnik, H., Uriagereka, J., 1998, *A Course in GB syntax*, MIT Press.

Longacre, R., 1982, *Discourse Typology in Relation to Language Typology*, Sture Allen éd., Text Processing, Proceeding of Nobel Symposium 51, Stockholm, Almquist and Wiksell, 457-486.

Mann, W., Thompson, S., 1988, *Rhetorical Structure Theory: Towards a Functional Theory of Text Organisation*, TEXT 8 (3) pp. 243-281.

Marcu, D., 1997, The Rhetorical Parsing of Natural Language Texts, ACL 1997.

Marcu, D., 2002, An unsupervised approach to recognizing Discourse relations, ACL 2002.

Moeschler, J., 2007, The role of explicature in communication and in intercultural communication, in Kecskes I. et al. (eds), Exporations in Pragmatics, Linguistic, Cognitive and Intercultural Aspects, Berlin, Mouton de Gruyter.

Reed, C., 1998, Generating Arguments in Natural Language, PhD dissertation, University College, London.

Rösner, D., Stede, M., 1992, *Customizing RST for the Automatic Production of Technical Manuals*, in R. Dale, E. Hovy, D. Rosner and O. Stock eds., Aspects of Automated Natural Language Generation, Lecture Notes in Artificial Intelligence, pp. 199-214, Springler-Verlag.

Saito, M., Yamamoto, K., Sekine, S., 2006, Using Phrasal Patterns to Identify Discourse Relations, ACL, 2006.

Takechi, M., Tokunaga, T., Matsumoto, Y., Tanaka, H., 2003, *Feature Selection in Categorizing Procedural Expressions*, The Sixth International Workshop on Information Retrieval with Asian Languages (IRAL2003), pp.49-56.

Saint-Dizier, P., 2012(a), Processing Natural Language Arguments with the <TextCoop> Platform, Journal of Argumentation and Computation.

Saint-Dizier, P., 2012(b), DISLOG: A logic-based language for processing discourse structures, LREC 2012.

Taboada, M., Mann, W.C., 2006, Rhetorical Structure Theory: Looking back and moving ahead. Discourse Studies, 8(3), 423-459.

Van Dijk, T.A., 1980, *Macrostructures*, Hillsdale, NJ: Lawrence Erlbaum Associates.

Walton, D., Reed, C., Macagno, F., 2008, Argumentation Schemes, Cambridge University Press.

Wierzbicka, A., 1987, English Speech Act Verbs, Academic Press.

Wright, von G.H., 2004, Explanation and understanding, Cornell university Press.