The language of explanation dedicated to technical documents[1]

Sarah Bourse - IRIT-CNRS Toulouse, France
sarah.bourse@gmail.com
Patrick Saint-Dizier - IRIT-CNRS Toulouse, France
stdizier@irit.fr

Resum. El llenguatge de l'explicació dedicat als documents tècnics. Aquest article presenta un estudi de les explicacions tal com apareixen en documents tècnics, especialment en procediments on aquestes explicacions resulten crucials. El nostre estudi es basa en l'anàlisi d'un corpus ampli. S'hi introdueix la noció de funció explicativa amb objectius comunicatius i es presenten els esquemes d'explicació que són les realitzacions lingüístiques de les funcions. Així, es proporcionen alguns detalls sobre els patrons lingüístics que permeten la identificació d'aquests esquemes.

Paraules clau: semàntica textual, explicació, procediments, estructures lingüístiques.

Abstract. The language of explanation dedicated to technical documents. In this paper, we introduce a study of explanations as they appear in technical documents, in particular procedures where they are crucial. Our study is based on an extensive corpus analysis. We introduce the notion of explanation function, serving communication goals, and then explanation schemes which are the language realizations of the functions. We give some details about linguistic patterns that allow for the identification of these various schemes.

Keywords: text semantics, explanation, procedures, linguistic patterns.

---

1. Introduction and context

Explanation and its relations to language and linguistics is a relatively new but vast area of investigation. It involves the taking into account of a large number of linguistic aspects, from syntax to pragmatics, but also typography. At the moment, explanation is essentially developed in various sectors where didactics is involved, e.g. writing recommendations for producing essays or technical documents or, 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. Let us also note studies on causality, and an emerging field around negotiation and explanation in multi-agent systems associated with abstract notions of belief (Amgoud *et al.* 2001). Two decades ago, explanations were used to produce natural language outputs for experts systems, often from predefined templates. The goal was to justify why a certain proof was carried out and why a certain solution was proposed as a result of a query. In the same range of ideas, and more generally, natural language generation planning made some use of explanation structures (e.g. McKeown 1985).

Interesting principles have emerged from these works which have motivated the emergence of more foundational research. For example, in ergonomy and cognitive science, the ability for humans to integrate explanations about a task described in a document or on an electronic device (possibly via a guidance system) when they perform that task is investigated and measured in relation with the document properties (typography, pictures) (Lemarié *et al.* 2008, Bieger and Glock 1984). In linguistics, a lot of efforts have been devoted to the definition and the recognition of discourse frames (Webber *et al.* 1990, 2004; Miltasaki *et al.* 2004; Saito *et al.* 2006) and the linguistic characterization of rhetorical relations (Mann and Thompson 1988, 1992; Longacre 1982), which are, for some of them, central to explanation (Rösner and Stede 1992, Van der Linden 1993). However, we now observe a proliferation of rhetorical relations with various subtleties, which, for some of them, turn out to be quite difficult to recognize from language marks since they involve quite a lot of pragmatic considerations and domain knowledge. Finally, explanation is a field which is investigated in pragmatics (e.g. cooperative principles, dialogue principles) (Pollock 1974) and in philosophy (e.g. rationality and explanation, phenomenology of explanation, causality, etc.) (Keil and Wilson 2000, Wright 2004, Davidson 1980).

Explanations are in general structured with the aim of reaching a goal. This goal may be practical (e.g. how to reach a certain location) or more interpersonal or epistemic (e.g. convince someone to do something in a certain way, negotiate with someone while providing explanations about one's point of view). Explanations are in fact often associated with a kind of instructional style, explicit or implicit, 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, didactic texts) form an excellent source of corpus to observe how explanations are constructed, linguistically realized and what aims they target. Procedures are composed of a skeleton, a goal (possibly subgoals) and sequences of instructions designed to reach this goal. This skeleton is paired with a very large number of forms of explanations meant to guide, support, help, etc. the user. This is very useful to build a corpus of explanations since the language is often simple and direct. Indeed, in procedures, style is often straightforward (Di Eugenio and Webber 1996), procedures being essentially oriented towards action: there must be little space for inferences. This kind of corpus is particularly well-adapted for our investigation as it covers quite a large proportion of situations, reproducible, e.g. in authoring tools, helpdesks or in natural language generation systems.

Explanation also occurs in goal-driven but non-procedural contexts, for example, as a means to justify a decision in legal reasoning or in political discourse, or as a way to explain the reasons of an accident in insurance accident reports. Explanation may also be associated with various pragmatic effects (irony, emphasis, dramatisation, etc.) for example in political discourse. In each of these cases, explanation does keep a goal-oriented structure (Carberry 1990, Takechi *et al.* 2003).

Explanation analysis and production is essential in opinion analysis to make more explicit how a certain opinion is supported (Kosseim and Lapalme 2000), it is also essential in question answering systems when the response which is produced is not the direct response: the user must then understand via appropriate explanations why the response provided is appropriate. Finally, it is central in a number of types of dialogues, clarification situations, persuasion strategies, etc.

Our main goal is to identify a number of prototypical, widely used, explanation schemes, their linguistic basis (e.g. prototypical language marks or constructs, planning issues), and to categorize their communicative goals. We aim at identifying the language and pragmatic means, given a certain aim, which are used to help, support, motivate or convince a reader. This is obviously a very large task with a number of semantic as well as pragmatic issues which are very difficult to capture and to model. We will basically concentrate in this paper on explanations found in procedural texts because the language which is used is simple and direct, allowing for a clear and relatively extensive study of the phenomenon. This greatly narrows down the difficulties, while allowing us to generalize later to other types of documents. Our aim is to have a closer analysis of the forms and organizations explanations may take in operational contexts and to integrate them into general studies on discourse and on action theory. In a subsequent stage, we will explore how our results extend to other contexts where explanation is central, including issues such as planning in language production.

For that purpose, we consider a large variety of procedures, from large public ones (cooking, DIY), to professional ones (maintenance, health), with a large diversity of authors, target readers and application domains. We have a corpus of about 8000 pro-

cedures from 24 different domains out of which we have randomly extracted a development corpus of 1700 procedures dedicated to this study. The work was carried out for French and English, with, at the moment, about the same volume of texts; for the sake of readability, all examples are given here in English.

In this paper, we first present our corpus analysis strategy and methodology, and then introduce the notion of *explanation function* that specifies communicative goals of explanations. Explanation functions are abstract constructs which are realized in language via what we call *explanation schemes*. These notions applied to explanations and the way they are formalized are new —to the best of our knowledge— and constitute our contribution, although they owe much to general considerations of rhetorical relation theory. An important feature of explanation is that schemes are transcategorial: they include syntactic and lexical semantics factors, as well as typographic (e.g. icons) and pragmatic factors. However, images, diagrams, and other multimedia devices are not considered here.

2. Corpus analysis and annotation

In order to identify explanation functions and then in order to propose general explanation construction principles, we have first carried out an annotation of a portion of our corpus. This was realized by 3 Master students in linguistics on 74 different texts from our development corpus (about 88 pages), with the same training and annotation instructions.

Constructing a corpus in general is a very challenging task: it is first necessary to identify the parameters we want to measure, while keeping the others constant. Then, the scope of each parameter needs to be defined (which values they can take) and the distribution we want to have (the percentage of texts for each parameter value). The next step is the corpus construction, and its validation with regard to these parameters. Our parameters include the taking into account of the diversity of domains, authors, type of style, target audience and the difficulty of realization of the procedure (i.e. complex temporal sequences of actions, use of complex equipments, often associated with a large number of warnings). The 74 selected texts allow us to make an indicative analysis that gives useful research directions and helps to confirm our working method. This small corpus of texts is composed as follows, with the following parameters and distribution rates among their values:

- Domain: cooking: 7, gardening: 9, DIY: 12, health care: 8, didactics: 8, video game solutions: 5, social relations management: 9, beauty: 4, technology: 12.
- Authors: teenagers: 15, professionals: 42, others: 17.
- Target: large public: 47, students: 17, professionals: 10.
- Style: poor: 25, standard: 39, high level: 10.
- Difficulty of procedure: basic: 8, simple: 12, average: 24, difficult: 30.

The second step is the annotation. This is a difficult task: identifying and categorizing rhetorical relations is almost never straightforward. To help annotators, the task was realized from the outputs of the <TextCoop> system, that tags, with quite a good accuracy, instructions, goals, various discourse markers, arguments (previously studied and implemented), some manners, instruments, conditional expressions and simple causal forms (Delpech and Saint-Dizier 2008).

The tags have received a definition from previous works in discourse analysis, and have been made more precise and as unambiguous as possible considering the type of document we are dealing with, which are objective and explicit and which may follow authoring instructions. Annotators have been trained, and after an adaptation period, needed for a comprehensive assimilation of the notions used in discourse analysis, annotators could do a homogeneous task with a relatively good consensus of 86%, measured by a Kappa test. Training may be time consuming because notions in discourse analysis are less clear-cut than in e.g. predicate-argument analysis. Training ends when there is little hesitation between these notions when annotating texts.

Over the 74 texts, 1127 structures related to explanation have been annotated, a ratio of 15 annotations per text. Distribution over texts ranges from 3 to 28 which often depend on the length of the text but also on the domain and the target audience. We kept in our analysis those annotations that occur at least 10 times over the whole set of texts, which corresponds to a frequency of about 1%, which sounds minimal to us to be really of interest. However, some non-frequent structures may be of interest because they introduce interesting views on explanation structures.

We basically consider the following tags, called hereafter Elementary Explanation Structures (EES), definitions are given here informally, and they follow from commonly admitted definitions that we have somewhat simplified considering the type of document used:

- *Elaboration and evaluation*: elaboration (adds new information), illustration (provides related examples), reformulation (simply restates, does not add any information), result (specifies the outcome of an action), contrast (introduces a comparison by means of a major difference between two methods, objects or situations), analogy (form of comparison by image to help understanding), encouragement, hint, evaluation (explains a user what kind of result he should get or evaluates his performances).
- *Arguments:* warning, advice (very few threats or rewards, where the author is involved). These often have the form of a causal expression, formally: do X 'because/ otherwise' Z, with quite a large variety of causal marks (Fontan and Saint-Dizier 2008) following Amgoud *et al.* (2001, 2005).

- *Condition:* It involves at least 2 structures: condition and then - else, and the assumption structure, which is an hypothetical statement. Condition is in general introduced by 'if' or 'when'.
- *Cause:* It involves a statement and an ensuing situation as defined in Talmy (2001). We limit annotations to trans-sentential causal expressions, i.e. those operating over instructions.
- *Concession:* statement associated with a concession statement (a general rule is given followed by an exception that could be admitted).
- *Goal expression*: purpose, following Talmy (2001).
- *Frame:* circumstance (of the realization of an action), and some propositional attitudes such as: commitment, authority.

Annotations can be embedded, leading to complex structures. In our corpus, annotations are in XML with some attributes, such as the strength or weight of arguments, and meta-annotations such as the certainty of the annotator. A given structure may receive several annotations in case of ambiguity, overlap or multiple functions, this is expressed by the symbol /.

An example, in readable form, borrowed from didactics, is the following (most marks produced by the <TextCoop> system are omitted, only EES are given, instructions appear on new lines):

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

From these pages of annotated texts, we can induce more abstract regularities related to (1) the communicative goals and (2) the ways these goals are realized. For example, behind 'advice' or 'purpose' there is a communicative goal, designed respectively to help the user via a suggestion and to explain to him the reasons of an action.

3. Explanation functions

From our corpus investigations, and from our experience of how procedures are written and used by technicians or casual users, we can propose a first, global classification of quite a large number of 'conceptual' functions that realize the communicative goals mentioned above. These are called *explanation functions*. To carry out this task, our strategy was to identify and then categorize and structure the underlying communicative aims associated with our annotations or groups of annotations, producing a second level of annotations. Again, this is somewhat intuitive, but this was realized as a collective task with the aim of reaching a consensus. We present here the main categories and their organisation, which seems to be quite stable. To go deeper into more specialized explanation functions, we obviously need to carry out further investigations, testing and refinements. Our investigations must also be enlarged in the future to other types of texts where explanation is present, besides procedural texts.

Within an operational context in a broad sense, explanation functions can schematically be subdivided into two fields: the motivations of doing something (Why do action A?) and the way to make something (How-to do A?). This view sets a more global analysis of action with the dichotomy intentions or motivations on the one hand and realization and its facets on the other, as can be found in Action Theory under a more philosophical view.

To avoid any confusion with EES names or with any existing terms, our explanation functions, which are abstract constructs, possibly complex, are prefixed with E-. Let us now introduce these functions. These remain so far abstract: their language realizations are accounted by schemes, given in the next section. Since we worked on French and English texts in parallel, for the sake of readability, examples in English are given here. Classification is stable over these two languages (and probably over a large number of languages) and linguistic realizations are often very similar.

The (why do A?) category of EES, the motivations functions, is basically composed of two subsets. The first subset, is composed of information providers: E-structure, E-information and E-arguments.

The function E-structure enhances or makes more explicit the structure and the coherence of the text. It is in a large part composed of low-level goal and function expressions (*push the white button [to open the box]*), indicating motivations and expected results. Goal expressions have a low scope and appear in general at the end of the statement, as confirmed by ergonomic studies. This function also includes several forms of low level structures, among which temporal structures realized by means of temporal connectors, punctuation or typography.

The aim of the E-information function, which operates at the ideational level, is to enhance, reinforce, weaken or contradict the beliefs of the reader, as anticipated by the

author of the text, by providing more specific information on some aspects of the action at stake (*Adding salt to your sauce is unnecessary because fish sauce is already salted*). This very general function may be subdivided into more accurate functions such as E-clarification (*Poke the wire into the bottom of the flower [(where the stem was)] as far as you can without it coming out the other side*) or E-precision (*Hang the hanger in a dark area and wait for the flowers to dry. [A full drying process will take anywhere from 2 to 3 weeks.]*), still to be investigated.

The second subset of this group operates at the inter-personal level, and aims at motivating the user to realize the action at stake as required and as accurately as possible, via some precautions or recommendations. This subset is composed of the various types of arguments, as usually found in argumentation classifications: E-warnings (*Carefully plug in the mother card vertically otherwise you risk to damage its connectors*) and E-advice (*We recommend professional products for your leathers, they will offer a stronger protection while repairing some minor damages*), when there is no implication on the author's part, and E-threats (*You must confirm your connection within 30 minutes otherwise we will cancel it*) and E-rewards (*We suggest you to pay when you book your flight since a discount coupon will be offered to you for your next purchase*) otherwise. These are designed to justify the importance of an action and the necessity of doing it as required (warnings) or to indicate the optional character of an action (advice) and the benefits of doing it. Several authors, among which Walton *et al.* (2008), have shown that arguments are composed of two parts: the *conclusion*, which is the statement, and its *support*, which is the reason of the following, e.g. the advice or the warning. Besides the recognition of arguments, evaluating their illocutionary or persuasion strength is of much interest. This is realized in general via a series of marks, essentially adverbial, denoting various levels of intensity.

The second category develops the way to realize an action, the How-to-do A? functions. It contains several EES families. The first family deals with functions related to controls on the user related to his actions, while the second family is related to the control of his interpretations. The third family is composed of help functions.

In the first family, control on user actions, we first have the notion of E-guidance. This function has quite fuzzy boundaries, it only includes temporal marks (Muller *et al.* 2004) guiding the organization of functions, but also manners, durations and instruments which offer a number of details and ways to realize and coordinate actions (*Open the box: [use a 2.5 inch key and a screw-driver]*). These are deeply related to the argument structure of the action verb of the instruction (but they are mainly adjuncts or connectors). The next function, E-framing, indicates via a statement (often starting a sentence or a paragraph) the context of application of an instruction or of a group of instructions (*[for X25-01 pumps]: disable first...*); it often has the form of a low level title or a frame. Next, E-expected-result describes the target result, it is a means for the user to evaluate his performance and to make sure he is on the right track (*at this stage, the sauce must be dark brown*). Finally, E-elaboration explains in more depth

how to realize an action (*Hold the seedling by the stem with your palm facing the roots of the plant, and turn the soda bottle upside down, [lightly shaking the soil out and the plant with it]*).

The second family is related to the control of the interpretations made by the user. The goal is to make sure he correctly understands the text. In this class fall relatively well-known functions directly associated with rhetorical relations: E-definition (gives a definition of a certain concept, e.g. *The transmission in your car is a gearbox that transmits powers from the engine to the live axle*), E-reformulation (says the same thing with different words or constructions, there is no new informational contents, e.g. *Before starting, make sure you have the right experience and skills, [in other words that you can do the job]*), E-illustration (gives one or more examples). Their goal is to ensure that the user correctly understands the terms of the procedure.

The third family is composed of two functions which provide basic help to the user: E-encouragements (supports in some way the user action, e.g. *at this stage, the most difficult operations have been realized*), E-evaluation (provides a precise evaluation of what should be observed at this point so that the user can check if he did well or not, e.g. *If the paste is really crunchy, then you are an excellent cook, you can move on to the next step)*.

These functions can be organized into groups leading to forms of explanation plans, e.g. make sure the task is correctly understood (E-definition, E-reformulation) and then warn about risks (E-warning). Planning explanations is the ultimate goal of an author while taking into account the prototypical profile of the readers.

4. Explanation schemes

Explanation schemes introduce the level of language realizations for explanation functions. These schemes are defined from the corpus by coupling EES annotations with the identification of explanation functions in these texts. Generalizing over schemes associated with a given explanation function then allows us to introduce a grammar for each explanation function. This grammar is based on EES and constrained statements, which form its pre-terminal elements, explanation functions can appear as non-terminal symbols since they may enter into schemes. There is no unique assignment: an explanation scheme may occur in several explanation functions, as can be seen in (Walton *et al.* 2008) for argumentation schemes, with a different perspective.

As an illustration, we give below a few schemes associated with explanation functions (the * indicates multiple occurrences). These are given here in standard bracketed notation and each bracketed term corresponds to an EES:

E-warning ➔ $[_{\text{warning}}$],
$[[_{\text{warning}}$] $[_{\text{illustration}}$]*],
$[[_{\text{warning}}$] $[_{\text{illustration}}$]* $[_{\text{elaboration}}$]*], etc.

These schemes indicate that an E-warning is composed of the following EES: either a warning alone, or a warning followed by one or more illustrations, or a warning followed by one or more illustrations and elaborations.

E-definition   ➔ $[[_{\text{definition}}$] $[_{\text{illustration}}$]*]
E-illustration   ➔ $[[_{\text{circumstance}}$] $[_{\text{illustration}}$]*]
E-expected-result ➔ $[[_{\text{circumstance}}$] $[_{\text{statement}}$ expr(+modal, +probability)]]

This latter example requires a statement with a modal expression such as 'should' which introduces a probability.

Explanation functions can be complex compounds and may include other explanation functions:

E-warning ➔ $[[_{\text{warning}}$]   E-illustration*   E-elaboration*]

This refers to an EES followed by two explanation functions.

The identification of each EES is based on patterns or rules, induced by corpus analysis. Pattern definition and testing is an active research area (see Auger and Barrière 2008, Sierra *et al.* 2008), in particular for knowledge extraction, relation extraction between terms, web search via keywords and for the analysis of relatively small size constructions (e.g. a verb and its arguments). Termino (2008) shows a large number of deep and very insightful approaches for the extraction of various types of semantic structures, including rhetorical relations such as definition. Furthermore, it has been shown that the number of discourse markers in a typical text is approximately one marker for every two clauses (Redeker 1990). This is sufficiently large to enable the derivation of rich rhetorical structures. Discourse markers are also used in a way which is consistent with the syntax and pragmatics of the discourse segments they relate to (Marcu 1997).

Our investigations for elaborating patterns were realized for each EES separately, on a large corpus of text, besides the 74 texts advocated above. For each EES, the task was to identify and isolate common features: typical words or classes of words (e.g. negative polarity nouns), morphological aspects, typographic marks, etc. From a global point of view, we developed patterns or rules that account for the structure of each EES (Saito *et al.* 2006, Miltasaki *et al.* 2004). Let us consider the case of the EES E-illustration which has 20 patterns for English. Table 1 presents a short extract.

A 'gap' denotes a finite string of words of no present interest, punctuation marks are relevant:

Pattern 1	, FE , gap . *, for example, gap.*
Pattern 2a	(FE : gap) *(for example : gap)*
Pattern 2b	(gap , FE) *(gap , for example)*
Pattern 8	let's take gap BWE . *let's take gap by way* *of illustration .*
Pattern 10	HERE is gap EXA gap . *here is another example gap .*
Pattern 13	let's take gap EO gap . *let's take gap illustrations of gap .*
Pattern 15	EG gap . *e.g. / such as / including gap .*

FE : "for example", "for instance"
BWE : "by way of example", "by way of illustration"
EXA : "an example", "example(s)"
EO : "an example / illustration of", "illustration(s) of", "example(s) of"
EG : "e.g.", "such as", "including"

The EES related to warnings and advice are more complex structures: they include a conclusion and a support for that conclusion; these are described and evaluated in Fontan and Saint-Dizier (2008). Additional patterns are given in the appendix.

These patterns allow for an automatic tagging of EES in texts. For that purpose, we use our <TextCoop> platform which is dedicated to discourse analysis. The formalism of a pattern in this environment somewhat extends the formalism of regular expressions. It has the following characteristics:

- Left-hand part: structure identifier.
- Right-hand part: a finite sequence of the following symbols:

(1) Terminal strings of words, punctuation or elements such as html or XML tags, since typography is taken into account;

(2) Non-terminal symbols, these symbols refer to various linguistic or *ad hoc* structures as observed in corpora; they may be associated with typed feature structures;

(3) Gaps, which are, roughly, symbols that stand for a finite sequence of words of no present interest;

(4) Insertion points, which indicate where to insert tags related to the recognition of the discourse structure at stake. These may be inserted anywhere in the recognized structure.

- Resulting structure: specifies the annotated structure. In general the string of words (represented by the symbols and the gaps) found is reproduced with tags inserted wherever appropriate.

In the patterns, symbols may be specified as being optional, or as having multiple occurrences, as in any regular expression.

For example, roughly, the first example above would be coded as:

illustration → [,], fe, gap, [.].
fe   → [for, example].
fe   → [for, instance].

The <TextCoop> platform is based on a meta-interpreter in Prolog, which, roughly, scans a text, word by word, and tags all structures it recognizes with regard to the patterns it has been given. <TextCoop> extends logic-based grammars to discourse analysis.

5. Ongoing Work

This work remains preliminary and exploratory. So far, we have implemented on the <TextCoop> platform the recognition forms within procedures for advice, warnings, purpose, concession, condition, result, elaboration and illustration. Cause is ongoing. EES such as circumstance, analogy and contrast are more difficult to characterize, indeed there is almost no unambiguous linguistic mark to characterize them. The result is an enriched semi-structured document with explicit explanation marks, together with procedural marks. This allows us to have a more accurate analysis of explanation, yet to be improved. In particular it is of much interest to consider speech acts as a support for explanation (Wierzbicka 1987).

Explanation schemes need further work to characterize in more detail their structure, and conceptual, explicative functions. We will also investigate structural regularities that would reveal forms of know-how in explanation organisation.

References

Amgoud, L., J. F. Bonnefon and H. Prade (2005). "An Argumentation-Based Approach to Multiple Criteria Decision". In L. Godo (ed.), *Proceedings of the 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU 2005*, 3571, 269-280. Berlin: Springer.

Amgoud, L., S. Parsons and N. Maudet (2000). "Arguments, Dialogue, and Negotiation". In W. Horn (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence*, 338-342. Amsterdam: IOS Press.

Auger, A. and C. Barrière (2008). "Pattern-based approaches to Semantic Relation Extraction", Special issue of *Terminology*, 14: 1.

Bieger, G. R. and M. D. Glock (1984-85). "The Information Content of Picture-text Instructions", *Journal of Experimental Education*, 53: 2, 68-76.

Carberry, S. (1990). *Plan Recognition in Natural Language Dialogue.* Cambridge, MA: MIT Press.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.

Delin, J., A. Hartley, C. Paris, D. Scott and K. Vander Linden (1994). "Expressing Procedural Relationships in Multilingual Instructions". In *Proceedings of the Seventh International Workshop on Natural Language Generation*, 61-70. Stroudsburg: The Association for Computational Linguistics.

Delpech, E. and P. Saint-Dizier (2008). "Investigating the structure of procedural texts for answering How-to Questions". In *Proceedings of Language Resources and Evaluation Conference (LREC 2008)*, 46-51. Paris: European Language Resources Association (ELRA).

Di Eugenio, B. and B. Webber (1996). "Pragmatic Overloading in Natural Language Instructions", *International Journal of Expert Systems*, 9: 1, 53-84.

Fontan, L. and P. Saint-Dizier (2008). "Analyzing the Explanation Structure of Procedural Texts: Dealing with Advices and Warnings". In J. Bos and R. Delmonte (eds.)*, Proceedings of Semantics in Text Processing (STEP 2008)*, 1, 115-127. London: College Publications.

Huddleston, R. D. and G. K. Pullum (2002). *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Keil, F. C. and R. A. Wilson (2002). *Explanation and Cognition*. Cambridge, MA: MIT Press.

Kosseim, L. and G. Lapalme (2000). "Choosing Rhetorical Structures to Plan Instructional Texts", *Computational Intelligence*, 16: 3, 408-445.

Lemarié, J., R. F. Lorch, H. Eyrolle and J. Virbel. (2008). "SARA: A Text-Based and Reader-Based Theory of Text Signaling", *Educational Psychologist*, 43: 1, 1-23.

Longacre, R. (1982). "Discourse Typology in Relation to Language Typology". In Sture Allen (ed.), *Proceedings of the Nobel Symposium 51*, 457-486. Stockholm: Almquist and Wiksell.

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

Mann, W. and S. Thompson (ed.) (1992). *Discourse Description: Diverse Linguistic Analyses of a Fund Raising Text*. Amsterdam: John Benjamins.

Marcu, D. (1997). "The Rhetorical Parsing of Natural Language Texts". In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL 97)*, 96-103. Stroudsburg: The Association for Computational Linguistics.

McKeown, K. (1985). *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge: Cambridge University Press.

Miltasaki, E., R. Prasad, A. Joshi and B. Webber (2004). "Annotating Discourse Connectives and Their Arguments". In A. Meyers (ed.), *Proceedings of the NAACL/HLT Workshop on Frontiers in Corpus Annotation*. Stroudsburg: The Association for Computational Linguistics. CD version.

Muller, P. and X. Tannier (2004). "Annotating and Measuring Temporal Relations in Texts". In *Proceedings of the 20th International Conference on Computational Linguistics (Coling 2004)*, 1, 50-56. Stroudsburg: The Association for Computational Linguistics.

Pollock, J. L. (1974). *Knowledge and Justification*. Princeton: Princeton University Press.

Reed, C. (1998). *Generating Arguments in Natural Language*. Doctoral thesis. University College London.

Redeker, G. (1990). "Ideational and Pragmatic Markers of Discourse Structure", *Journal of Pragmatics*, 14: 3, 367-381.

Rösner, D. and M. Stede (1992). "Customizing RST for the Automatic Production of Technical Manuals". In R. Dale, E. Hovy, D. Rösner and O. Stock (eds.), *Aspects of Automated Natural Language Generation, Lecture Notes in Artificial Intelligence*, 199-214. Berlin: Springer-Verlag.

Saito, M., K. Yamamoto and S. Sekine (2006). "Using Phrasal Patterns to Identify Discourse Relations", *Proceedings of the Human Language Technology Conference of the NAACL,* 133-136. Stroudsburg: The Association for Computational Linguistics.

Sierra, G., R. Alarcon, C. Aguilar and C. Bach (2008). "Definitional verbal patterns for semantic relation extraction", *Terminology*, 14: 1, 74-98.

Takechi, M., T. Tokunaga, Y. Matsumoto and H. Tanaka (2003). "Feature Selection in Categorizing Procedural Expressions". In *Proceedings of the Sixth International Workshop on Information Retrieval with Asian Languages (IRAL2003)*, 49-56. Stroudsburg: The Association for Computational Linguistics.

Talmy, L. (2001). *Towards a Cognitive Semantics*. Cambridge, MA: MIT Press.

Van der Linden, K. (1993). *Speaking of Actions Choosing Rhetorical Status and Grammatical Form in Instructional Text Generation*. Doctoral thesis. University of Colorado.

Walton, D., C. Reed and F. Macagno (2008). *Argumentation Schemes*. Cambridge: Cambridge University Press.

Webber, B. and B. Di Eugenio (1990). "Free Adjuncts in Natural Language Instruction". In H. Karlgren (ed.), *Proceedings of the 13th International Conference on Computational Linguistics (Coling 1990)*, 2, 395-400. Stroudsburg: The Association for Computational Linguistics.

Webber, B. (2004). "D-LTAG: Extending Lexicalized TAGs to Discourse", *Cognitive Science,* 28: 5, 751-779.

Wierzbicka, A. (1987). *English Speech Act Verbs*. Sydney: Academic Press.

Wright, von G.H. (2004). *Explanation and Understanding*. Ithaca: Cornell University Press.

Appendix

Here is a sample of marks related to a number of EES defined above, these are defined here as subcategorization frames, although these structures are self-explanatory, we add examples[2] for the sake of readability:

Note: C stands for "clause".

WARNING	**avoid** [VP (gerundive)] / [NP]	***Avoid*** *using common colloquial words and expressions.* *Try to **avoid** the first and second person.*
	{C} **do not** [VP (imperative)] **never** [VP (imperative)] **pay attention to** [NP] or [VP (infinitive)] **be careful to** [VP (infinitive)]	***Do not*** *open the box.* ***Never*** *open the box.* *You should **be careful to** leave the door open.*
	take care (not) to [VP (infinitive)]	***Take care not to*** *burn your fingers!* ***Take care to*** *follow pesticide use regulations.*
	take care that [NP] (**do not**) [VP (infinitive)] **make sure / ensure (that)** [C] **be sure to** [VP (infinitive)] **you should not** [VP (infinitive)]	***Take care that*** *you **don't** catch cold.* ***Make sure*** *all of your supplies are within arms reach.*
PURPOSE	[C] **to** [VP (infinitive)] [C] **in order to** [VP (infinitive)]	*Leave it open **to** allow air to come in.*
	[C] **so** [C (purpose)]	*He spoke softly **so** she wouldn't overhear the conversation.*
	[C] **so that** [C (purpose)] [C] **so as to** [VP (infinitive)]	*I'll close the windows **so that** the rain doesn't come in.*
	[C] **for** [NP]/[VP (gerundive)] [C] **for the sake of** [NP]/[VP (gerundive)] [C] **for the purpose of** [NP]/ [VP (gerundive)]	*I did it **for** you.* *Use a bold font **for the sake of** readability.*
	[C] **for fear of** [NP]/[VP (gerundive)]	*He delayed his departure **for fear of** an accident./ He delayed his departure **for fear of** missing something.*
	[C] **with the intention of** [VP (gerundive)] [C] **with a view to** [VP (gerundive)]	*I came to this school **with the intention of/with a view to** getting a degree.*
	[C] **with an eye to** [VP (gerundive)]	*The mayor took office **with an eye to** improving the town.*
	[C] **with the intent of** [VP (gerundive)]	*He left his country **with the intent of** changing his life.*

---

2. Most of the examples come from Huddleston and Pullum (2002).

CONCESSION	**although** [C (concession)]{,}[C] or [C] **although** [C (concession)] **though** [C (concession)]{,}[C] [C] **though** [C (concession)] **even though** [C (concession)][,][C] [C] **even though** [C (concession)] **despite** [NP]/[VP (gerundive)][,][C] **in spite of** [NP]/[VP (gerundive)][,][C] **notwithstanding** [NP][,][C] **notwithstanding** [that][C] [C][.]/[;] **nevertheless** [C (concession)] [C][.]/[;] **nonetheless** [C (concession)] [C][.]/[,] **still** [C] [C][.]/[,] **yet** [C] [C][.]/[,] **however** [C] [C][.] **for** [all NP (concession)] [C][,] **albeit** [verbless clause (concession)]	*Sonia doesn't speak French **although** she grew up in Paris.* ***Though** living in Holland he works in Germany.* *He knew they were there **even though** he couldn't see them.* ***In spite of** / **Despite** the recession, travel agents seem to be doing well.* ***Notwithstanding** Ed's reservations, the agreement is the best I could hope for.* *Sonia doesn't speak French; **nevertheless**/**nonetheless**, she grew up in Paris.* *I am very happy here, **still**/**and yet** I can't help feeling a little homesick at times.* *It is strange, **yet** true (= but it is true).* *It was pouring with rain.* ***However** they decided to go.* ***For** all our good intentions, the meeting soon broke up in acrimony.* *The book covers the whole field, **albeit** somewhat superficially.*
CONDITION	**if** [C (antecedent)][,][C (consequent)] [C (consequent)] or [C (consequent)] [C(consequent)] **unless** [C (antecedent)] **providing** [C (antecedent)][,][C (consequent] or [C (consequent)] **providing** [C (antecedent)] **provided (that)** [C (antecedent)] [C (consequent)] **on condition (that)** [C (antecedent)] **in case** [C (antecedent)][,][C (consequent)] or [C (consequent)] **in case** [C (antecedent)] [C (consequent)]{[,]} **as long as** [C (antecedent)] [C (consequent)]{[,]} **so long as** [C (antecedent)] **in the event (that)** [C (antecedent)][,] [C (consequent)] or [C (consequent)] **in the event (that)** [C (antecedent)] **supposing (that)** [C (antecedent)][,][C (consequent)] **assuming (that)** [C (antecedent)][,][C (consequent)]	*John will go with you **if** you ask him to.* *I won't speak to him **unless** he apologizes.* ***Provided that**/ **Providing** you water this plant every year, it will live for years.* *You may borrow the book **on condition (that)** you return it tonight.* *I will make a cake **in case** he is coming.* *I'll lend you my car **as long as** you bring it back by midnight.* *I'll lend you my car **so long as** you need it.* ***In the event (that)** the attacks continue, we will retaliate.* ***Assuming (that)** / **Supposing (that)** everybody agrees, the project will get under way next month.*

ADVICE	[Pronoun] **recommend/suggest** [VP (gerundive)] / **that** [NP][VP (subjunctive)]	*We would always **recommend** that you instruct solicitors as soon as possible.*
	[Pronoun] **advise** [NP] **to** [VP (infinitive)] **you may wish to** [VP (infinitive)] **you may just want to** [VP (infinitive)]	*We always **advise** you **to** lock the door before you leave.* ***You may wish to** inquire as to the identities of local attorneys who accept the plan before making the decision to purchase any given plan.*
	you'd better [VP (infinitive)] **try to** [VP (infinitive)] **try** [VP (gerundive)]	***You'd better** come back before noon.* ***Try to** be here before noon.* ***Try** eating less meat.*
CONSEQUENCE or RESULT	[C] **so** [C (consequence)] [C] **so that** [C (consequence)] [C] **with the result that** [C (consequence)] [C][.]/[,] **that is why** [C] [C][.]/[,] **because of this** [C] [C][.]/[,] **for this/that reason** [C] [C][.]/[,] **as a result** [C] [C][.]/[,] **as a consequence** [C] [C][.]/[,] **in consequence** [C] [C][.]/[,] **consequently** [C]	*The airline had overbooked, **so that** two of us couldn't get on the plane.* *They had gambled away all their money, **so** / **with the result that** they didn't have the fare to get home.* *The airline had overbooked; **for this reason** / **as a result** / **because of this** / **consequently** two of us couldn't get on the plane.*
ELABORATION	[C][.]/[,] **actually** [C] **namely** [NP] [C][.]/[,] **more precisely** [C] **viz.** [NP]	*Red fruits, **namely** strawberries, blueberries and raspberries…* *Logic-based programming languages, **more precisely** Prolog and Oz…*
ILLUSTRATION	[C][.]/[,] **for example** [C]/[.] [C][.]/[,] **for instance** [C]/[.] **such as** [NP] **like** [NP] **including** [NP] [C]{,} **e.g.** [C]	*The benefits are obvious: efficiency and ease of participation, **for example**.* *However, nuts **such as** chestnuts, root vegetables **like** potatoes, seeds **such as** buckwheat, and pulses **such as** chickpeas can also be made into flour.*
REFORMULATION	[C][.]/[,] **in other words** [,][C] [C][.]/[,] **that is (to say)** [C] [C][,] **that means** [C] [C][,] **i.e.** [C] [C][.] **to put it another way** [,][C]	*Decide exactly what you want to keep in the medicine cabinet and don't deviate from the plan. **In other words/that is to say/ to put it another way**, don't give in to the temptation to cram stray items into the cabinet when you're in a rush to clean up.*