Handling explanation in operational contexts

Patrick Saint-Dizier [1]

Abstract. In this document, we present some preliminary facets of a model for explanations and their realizations in language. We introduce the notion of explanation function realized by means of explanation schemes. We show how these schemes are realized, what purpose they serve and on what linguistic basis they can be recognized.

1 Introduction and Context

Explanation and its relations to Language is a vast area of investigation. Explanation analysis is quite well developed via writing recommendations in didactics [3]. In artificial intelligence, it is often centered around the notion of argumentation [4], [26], [32], but argumentation is just one (major) facet related to explanation. Let us also note studies on causality, and an emerging field around negotiation [4] and explanation in multi-agent systems. Two decades ago, explanations were used to produce natural language outputs for experts systems, often from predefined templates [29]. Quite interesting principles have emerged from this research. In ergonomy and cognitive science, the ability for humans to integrate explanations about a task (possibly via a guidance system) when they perform that task is measured as such and in relation with the document properties (typography, pictures) [18], [20], [7].

In linguistics, a lot of efforts have been devoted to the definition and the recognition of discourse frames [34], [23] [28] and rhetorical relations [21], [19], which are, for some of them, central to explanation. However, we now observe a proliferation of rhetorical relations, which, in general, turn out to be quite difficult to recognize from language marks since they involve some pragmatic interpretation. Finally, explanation is a field which is investigated in pragmatics (e.g. cooperative principles, dialogue principles) [25] and in philosophy (e.g. rationality and explanation, phenomenology of explanation, causality, etc.) [16],[36] and [10].

Explanations are in general structured with the aim of reaching a goal [1], [6]. 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 DIY, maintenance procedures, health, didactic texts) form an excellent source of corpus to observe how explanations are constructed, linguistically realized and what goals they target. Indeed, in procedures, style is often quite straightforward [12], procedures being essentially oriented towards action: there must be little space for inferences that may lead to misinterpretations, hence the need of explanations. Explanation occurs also in goal-driven but non procedural contexts, for example, as a means to justify a decision as in legal reasoning, or as a way to explain the reasons of an accident in accident reports. Explanation may also be associated with various pragmatic effects (irony, emphasis, dramatisation, etc.) for example in political discourse. In each case, explanation has a goal-oriented structure [9], [30].

Explanation analysis and production is essential in opinion analysis to make more explicit how a certain opinion is supported [17], it is also essential in question answering when the response which is produced is not the direct response: the user must then understand why the response provided is appropriate. Finally, it is central in a number of types of dialogues, clarification situations, persuation strategies, etc.

Our main goal is to identify a number of prototypical explanation schemes (See [32] for argumentation forms), their linguistic basis (e.g. prototypical language marks or constructs), and to categorize their communicative goals. We aim at identifying the language and pragmatic means which are used to support, motivate and convice the reader. We will basically concentrate here on explanations associated with procedures. We consider a large variety of procedures, from large public ones (cooking, DIY), to profesional ones (maintenance, health), with a large diversity of target readers and application domains. Our aim is to have a closer analysis of the forms and organizations explanations may take in operational contexts (installation, production, maintenance procedures) compared to general studies on discourse. Besides the theoretical dimension, the main aim is to have a better and concrete analysis of the difficulties, the gaps, the risks that one may encounter while realizing a procedure. This is a major issue in e.g. maintenance or production to prevent risks.

In this short paper, we first introduce the notion of **explanation function** that specifies communicative goal of explanations. Explanation functions are abstract constructs which are realized in language via **explanation schemes**. An important feature of explanation is that it is transcategorial: it includes syntactic and lexical semantics factors, as well as typographic and pragmatic factors. The boundaries of explanations are fuzzy: e.g. a whole procedure can indeed be viewed as an explanation, or just e.g. its illustrations, advice and warnings.

2 Corpus analysis and annotation

Our initial corpus is composed of 2000 procedural texts in French, from 24 very diverse domains. The backbone of procedures is based on the 'goal (titles) - instructions' structure. However, this structure is associated with a large diversity of explanations meant to guide, motivate, evaluate, warn users. The explanation structure is therefore crucial for the understanding of a procedure and its correct execution.

In order to identify explanation functions and general explanation principles and goals, we have first carried out an 'informal' annotation of the corpus. This was realized by 3 Master students in linguistics on 62 different texts (about 85 pages), with the same training and annotation instructions. This is a difficult task: identifying and categorizing rhetorical relation is almost never straightfor-

[1] IRIT-CNRS Toulouse France, email: sstdizier@irit.fr

ward. To help annotators, the task was realized on the outputs of our <TextCoop> system (see demo paper in this volume), that tags instructions, goals, various discourse markers, arguments, and within an instruction: some manners, instruments, conditional expressions and simple causal forms. The kappa test indicates an agreement level of 80types of elements (rhetorical is more difficult than arguments for example). We then had a discussion and constructed an annotated corpus which is the result of thsi discussion, where disagreement have been partly or totally resolved. This helped us refine our definitions for the different objects to tag.

The tags to use were first given a priori, from our experience of the domain (<TextCoop> started 5 years ago) and then gradually enriched when tagging the corpus. After an adaptation period, annotators could do an homogeneous task. The principles are to tag first the most crucial and less ambiguous structures (e.g. reformulation) which are most of the time indicated via clear linguistic marks. Once these settled, we moved on to more subtle ones, proceeding by contrast with the previously defined ones.

We basically have the following preliminary set of tags corresponding to frequently encountered structures (called hereafter Elementary Explanation Structures: EES):

- rhetorical: elaboration, illustration, reformulation, result, contrast, analogy, evidence, encouragement, hint, evaluation.
- 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 et al. 2008),
- conditions: involve at least 2 structures: condition and then - else, and the assumption structure, which is an hypothetical statement.
- cause: which involve a statement and a structure: cause or consequence. We limit annotations to trans-sentential causal expressions, i.e. those operating over instructions,
- concession: statement followed by a concession statement,
- goal expression: purpose,
- frame: circumstance, and some propositional attitudes such as: commitment, authority.

Annotations can be embedded, leading to complex structures. In our corpus, annotations are realized in XML with some attributes, such as e.g. the force (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. The / below is an 'or'.

An example, in readable form, from didactics, is the following (most marks produced by our <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...]]]

While the Internet should never be your only source of information, [$_{contrast}$ it would be ridiculous not to utilize its vast sources of information. [$_{advice}$ You should use the Internet to acquaint yourself with the topic more before you dig into more academic texts.]]]

Besides its illustrative purpose, this example shows the difficulty of the task and its subjectivity. Nevertheless, by going through a reasonable amount of pages of annotated texts, we can discover regularities related to (1) the communicative goals and (2) the ways these goals are realized. The analysis is here entirely manual. Annotated corpus is also used to test the implementation.

3 Explanation Functions

From our corpus investigations, we can propose a first, global classification of the functions realized by explanations. Our view on explanation is quite broad, with the introduction of quite a large number of functions that play a central role in explanation, even if their associated language realization may play several roles, besides explanation (e.g. guidance or expected result below). To carry out this task, our strategy was to identify the underlying communicative aims and to categorize them. Again, this is somewhat intuitive, but it was realized with the contribution of annotators, identifying, for each annotated fragment, what is its communicative goal(s). Within an operational context, explanation functions can schematically be subdivided into two fields: 'Why do action A?' and 'How-to do A?'. To avoid any confusion with existing terms, our explanation functions are prefixed with E-.

The first subset, **'why do A?'** functions, is composed of information providers: **E-structure** and **E-information**. The function 'structure' enhances the structure and the coherence of the text. It is in a large part composed of low-level goal and function expressions (*push to open the box*), indicating motivations and expected results. The aim of 'information', which operates at the ideational level, is to enhance or contradict the beliefs of the reader by providing precise information on some aspects of the action at stake. This very general function may be subdivided into more precise functions such as e.g. E-clarification or E-precision, 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, via some precautions or recommendations. This subset is formed of the various types of arguments, as usually found in argumentation classifications: **E-warnings** and **E-advice** when there is no implication on the author's part, and **E-threats** and **E-rewards** 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. Besides the recognition of arguments, evaluating their illocutionary or persuasion force is of much interest. This is realized in general via a series of marks, essentially adverbial.

The second subset, **'How-to-do A?'** functions, contains several families. The first one deals with functions related to controls on the user related to his actions and interpretations. In the first set, control on user actions, a first group of functions is associated with the notion of **E-guidance**. This function has quite fuzzy boundaries, it can just include temporal marks (Muller et al, 2004), but also possibly manners and instruments which offer a number of details and ways to realize and coordinate actions, up to instructions themselves. It is deeply related to the argument structure of the action verb of the instruction (mainly adjuncts and temporal marks). The next function is **E-framing**, which indicates via a statement the range of application

of an instruction or of a group of instructions (*for X25-01 pumps:....*); it often has the form of a low level title or a condition. **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 way. Finally, **E-elaboration** explains in more depth how to realize an action.

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 instructions. In this class fall relatively well know functions directly associated with rhetorical relations: **E-definition, E-reformulation, E-illustration, E-elaboration**. 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, E-evaluation**.

4 Explanation schemes

Explanation schemes allow us to introduce a grammar that quite globally describes the semantic and pragmatic structure of explanations. This grammar is based on EES and constrained statements, which form its terminal elements, explanation functions being the non-terminal symbols. We can associate with every explanation function a set of explanation schemes. There is no unique assignment: an explanation scheme may be shared by several explanation functions (Walton et ali. 2008).

As an illustration, we give below a few schemes associated with explanation functions (the * indicates multiple occurences):
E-warning: $[_{warning}]$,
$[[_{warning}]\ [_{illustration}]*\ [_{elaboration}]*]$, etc.
E-Definition: $[[_{definition}]\ [_{illustration}]*]$
E-Expected-Result: $[[_{circumstance}]\ [_{statement}$ expr(+modal,+probability) $]]$
this latter example requires a statement with a modal expression such as 'should' introducting a probability.
E-illustration: $[[_{circumstance}]\ [_{illustration}]*]$
Explanation functions can be complex compounds and may include other explanation functions; e.g.:
E-warning:
$[[_{warning}]$ E-illustration$*$ E-elaboration$*$.

The identification of each EES is based on patterns. These patterns are mainly based on marks and lexical constraints. The marks identified are basically relational: they bind two statements (basically, kernel and satellite). We developed subcategorization frames for these marks [28], [23]. The simplest example is probably *illustration* which as the following patterns in English:
S for example S/NP (NP can be plural) ; (this means any S followed by the string 'for example' followed either by another S or an NP)
S for instance S/NP ;
S such as NP, S e.g. NP ;
S like NP
Elaboration includes marks such as: *actually, that is (to say), namely, as a matter of fact, more precisely,* with similar subcats.

For the tag 'Illustration', tested over a set of 32 procedures has the following recognition rate, we get the following results where precision has been favored over recall:

recall	sprecision
85%	98%

The recall rate is now so high because in a number of contexts, the notion of illustration (not counting pictures) is not so clear, in particular it may overlap with elaboration, for which we have the following results:

recall	precision
79%	91%

However, these results are really good considering information retrieval systems for exemple. Elaboration is a complex relation, where there may be a large gap between the kernel and its satellite, amking the recognition by means of patterns quite challenging.

Some EES, such as warnings and advice are composite structures: they include a conclusion and a support for that conclusion. Warnings are basically organized around a unique structure composed of an 'avoid expression' combined with a proposition. The variations in the 'avoid expressions' capture the illocutionary force of the argument via several devices, ordered here by increasing force:
(1) 'prevention verbs like avoid' NP / to VP (*avoid hot water*)
(2) do not / never / ... VP(infinitive) ... (*never put this cloth in the sun*)
(3) it is essential, vital, ... to never VP(infinitive).

Supports are identified from various marks:
(1) via connectors such as: *otherwise, under the risk of,* or via verbs expressing consequence,
(2) via negative expressions of the form: *in order not to, in order to avoid, etc.*
(3) via specific verbs such as risk verbs introducing an event (*you risk to break*). In general the embedded verb has a negative polarity.
(4) via the presence of very negative terms, such as: nouns: *death, disease, etc.*, adjectives, and some verbs and adverbs. We defined a lexicon of about 200 negative terms found in our corpora.

Some supports may be empty, because they can easily be inferred by the reader. In that case, the argument is said to be truncated. Warning patterns are implemented on the <TextCoop> platform (see demo paper in this volume). We carried out an indicative evaluation (e.g. to get improvement directions) on a corpus of 36 procedural texts from professional domains, containing 262 warnings. We get the following recognition rates, compared to our manually annotated reference corpus:

conclusion recognition	support recognition	(3)	(4)
88%	91%	95%	95%

(3) conclusions well delimited (4) supports well delimited, w.r.t. warnings correctly identified.

5 Ongoing work

This work remains preliminary and exploratory. So far, we have fully implemented on the <TextCoop> platform (version V2, in Java) 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. 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 act verbs as a 'support' for explanation, considering e.g. the very detailed descriptions given in [35].

Out <TextCoop> platform is designed for text semantics. It offers a formalism to describe text semantics structures (discourse relations or simply semantic patterns) based on recognition rules.

<TextCoop> V.2 is implemented in Java, and it is in particular based on JFLEX and JCUP technology. Input-output structures tend to be close to UIMA recommendations to allow for a certain interoperability.

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

From an applied perspective, our aim is to evaluate the impact of explanations on users performances (how they understand them, use them, memorize them). In parallel, we are developing an analysis of risk expressions, since risk prevention is a major challenge in procedures, in particular in developing countries. From a theoretical perspective, we aim at integrating this work into Action theory to be able to formally define the complexity of a procedure.

REFERENCES

[1] Adam, J.M., *Types de Textes ou genres de Discours? Comment Classer les Textes qui Disent De et Comment Faire*, Langages, 141, pp. 10-27, 2001.

[2] Adam, J.M., *Types de Sequences Textuelles Elementaires*, Pratiques n56, Metz, 1987.

[3] Ashley, K.,D., Desai, R., Levine, J.M., Teaching case-based argumentation concepts using dialectic arguments vs. didactic explanations, ITS 2002, Lecture notes in computer science, 2002.

[4] Amgoud, L., Bonnefon, J.F., Prade, H., *An Argumentation-based Approach to Multiple Criteria Decision*, in 8th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'2005, Barcelona, 2005.

[5] Amgoud, L., Parsons, S., Maudet, N., *Arguments, Dialogue, and Negotiation*, in: 14th European Conference on Artificial Intelligence, Berlin, 2001.

[6] Aouladomar, F., Towards Answering Procedural Questions. Dans : KRAQ'05 - IJCAI workshop, Edinburgh, F. Benamara, P. Saint-Dizier (Eds.), p. 24-36, juillet 2005.

[7] Bieger, G.R., Glock, M.D., *The Information Content of Picture-text Instructions*, Journal of Experimental Education, 53, 68-76, 1984-85.

[8] Bouffier, A., From texts to structured documents: The case of Health Practice Guidelines, International Semantic Web Conference, Doctoral Consortium, Pusan, Novembre 2007.

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

[10] Davidson, D., *Essays on Actions and Events*, Oxford: Oxford University Press, 1980.

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

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

[13] Gardent, C., Discourse tree adjoining grammars, report nb. 89, Univ. Saarlandes, Saarbrucken, 1997.

[14] Greimas, A., *La Soupe au Pistou ou la Conservation d'un Objet de Valeur*, in Du sens II, Seuil, Paris, 1983.

[15] Grosz, B., Sidner, C., Attention, intention and the structure of discourse, Computational Linguistics 12(3), 1986.

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

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

[18] Lemarie, J., Lorch, R. F., Eyrolle, H., Virbel, J., SARA: A text-based and reader-based theory of text signaling. Educational Psychologist, 43, 27, 2008.

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

[20] Luc, C., Mojahid, M., Virbel, J., Garcia-Debanc, C., Pery-Woodley, M-P., *A Linguistic Approach to Some Parameters of Layout: A study of enumerations*, In R. Power and D. Scott (Eds.), Using Layout for the Generation, Understanding or Retrieval of Documents, AAAI 1999 Fall Symposium, pp. 20-29, 1999.

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

[22] Mann, W., Thompson, S.A. (eds), Discourse Description: diverses linguistic analyses of a fund raising text, John Benjamins, 1992.

[23] Miltasaki, E., Prasad, R., Joshi, A., Webber, B., Annotating Discourse Connectives and Their Arguments, new frontiers in NLP, 2004.

[24] Muller,P., Tannier, X., *Annotating and Measuring Temporal Relations in Texts*, In Proceedings of Coling 2004, volume I, pages 50-56, Geneve, 2004.

[25] Pollock, J.L., Knowledge and Justification, Princeton university Press, 1974.

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

[27] Rosner, D., Stede, M., *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, 1992.

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

[29] Schank, R., Explanation Patterns: Understanding Mechanically and Creatively, Laurence Erlbaum, 1986.

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

[31] Talmy, L., Towards a Cognitive Semantics, vol. 1 and 2, MIT Press, 2001. Vander Linden, K., *Speaking of Actions Choosing Rhetorical Status and Grammatical Form in Instructional Text Generation* Thesis, University of Colorado, 1993.

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

[33] Webber, B.L. and Di Eugenio, B. Free Adjuncts in Natural Language Instructions, Proceedings COLING-90, Helsinki, Finland, August 1990

[34] Webber, B., D-LTAG: extending lexicalized TAGs to Discourse, Cognitive Science 28, pp. 751-779, Elsevier, 2004.

[35] Wierzbicka, A., English Speech Act Verbs, Academic Press, 1987.

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