ISKO. CAPÍTULO ESPAÑOL. CONGRESO (7º: 2005: BARCELONA) 136-147

The example within the Technical Document
Which unit for which statute?

Ahmed CHAIBI, zchaibi@yahoo.fr

Université Stendhal-Grenoble 3. Équipe CRISTAL-GRESEC (Grenoble, France).

Abstract
In this paper we suggest taking the example unit in the Technical Document as a topic of study, whose objective is to determine surface indexes (linguistic, typographic, structural, lexical and punctuation), making it possible to identify and delimit the unit example within a Technical Document corpus in order to conceive an adequate and rational method that makes it possible to identify the example automatically. The result of the experimentation conducted among judges (experts, non-experts) confirmed the relevance of surface indexes defined according to our approach. We conclude with the example unit statute in a process of information research in a Technical Document. Indeed, we wonder if the example unit could be an answer to a request for a targeted and convenient information that technicians need.

Keywords
examples in documents, technical documents, information retrieval

1 Introduction

Our study belongs to the domain of information research in voluminous text documents such as technical documents. These documents are distinguished by their heterogeneity, often including disparate objects (text objects, non-text objects), among which example units are to be found. S. CARO (1995) considers these units as background elements of the technical text. Their function is to provide explanations and illustrations of information transmitted by foreground text. According to the study by C. PAGANELLI (2000), information transmitted by examples is highly searched by users, because it contributes directly to the execution of a task. Thus, the importance given by users to the example unit in the technical document led us to address this research topic.

In this article we intend to look for surface indexes (linguistic, typographic, structural, lexical and punctuation) in order to identify and delimit the example unit within a technical corpus. Subsequently, we attempt to define this unit, as an answer to possible user request, in a

process of information research in a technical document. Within this scope, and to afford a rigorous methodological foundation to our study, we will attempt to find answers to the following research questions:

— What shapes can example units take?
— Do surface indexes allow us to identify and delimit the example unit?

2 Functions of the example

In old rhetoric, the example served the rhetorical induction, since the example consists in proceeding «d'un particulier à un autre particulier par le chaînon implicite du général: d'un objet, on infère la classe, puis de cette classe on défère un nouvel objet» according to Barthes, quoted by Thyrion (1997).

Nowadays, the use of the example has functions other than the induction. In his study of the function of examples in thesis texts, D. Coltier distinguished two major categories of examples, where each assumes a fully-fledged function: «dans les textes à thèse, le recours à l'exemple permet au scripteur de régler la bonne marche de la communication à deux niveaux différents, il lui sert d'une part, à légitimer les arguments de la thèse qu'il présente favorisant ainsi l'adhésion, recherchée, du destinataire à la thèse proposée, il permet par ailleurs au scripteur d'éliminer des obstacles à la compréhension du destinataire. Ces deux fonctions (argumentative et explicative)» (Coltier 1988).

In the case of argumentative examples, the function of the example consists in transforming a subjective assertion into an objective affirmation. According to D. Coltier, the argumentative example allows possible to anchor, to legitimise, what has been said. Whereas the explanatory example is intended to clarify a note, a paradoxical phenomenon or not, its main function is to eliminate obstacles, promoting fluent communication, without acting on the subjects to whom it is connected.

It has been acknowledged that the two major functions of examples, explanation and argumentativeness, are closely associated, and examples may be found where the two functions closely overlap. However, it is not easy to distinguish between them because of the fragility of criteria.

The example constitutes one of the basic components of a technical corpus, as it is often used in order to clarify further technical discourse. It is then judicious to pay more attention to the characteristics of technical documentation, in order to understand the main role assigned by the writer to the example unit in a technical corpus. In this context, we are going to present the definition as well as the essential properties of technical documentation.

3 The example in a technical document

The technical document is defined here as all work, be it material or digital, containing an ensemble of technical information concerning the conception or the working and scheduling of a product or service and its maintenance, etc.

The technical document is a heterogeneous document where information varies according to the textual unit nature, something that brought S. Caro (1995) to make the distinction in a technical text between foreground and background textual unit; the function of background units is to provide details, comments, descriptions, less important information compared to the main text: they help the reader and have no effect on the chronological progress and sequences of facts in the text. Weinrich was one of the first to use this kind of textual distinction, he used the notion of backdrop instead of background: «L'arrière-plan s'est ce qui à lui seul n'éveillerait pas l'intérêt, mais qui aide l'auditeur à s'orienter dans le texte et lui en rend la lecture plus aisée.» Weinrich, quoted by S. Caro (1995). The theme attributed to background textual units may vary; it may be about the text, the way of reading or an example from another domain that is going to facilitate understanding, etc.

The example frequently appears in the background of the text; its role is to clarify, to illustrate, to specify information carried by the foreground text (Caro 1995). It can also be part of the textual units of descriptive category; these units have an argumentative purpose (Paganelli 1997).

In his study of procedural texts, L. Heurley (1994) was interested in examples, while considering them as being useful information unit for the user. He qualifies examples as powerful organizers calling on the user's pre-existing knowledge. These examples make it possible to improve the understanding of instructions while using the principle of analogy. Hence, one of the procedural text particularities compared with other types of texts, is that it often relies on concrete examples.

4 The location of examples in a technical corpus

4.1 Corpus object of the study

The corpus that was the object of our study is: *«AIX version 3.20; guide de gestion de stations de travail sans disque»* (*AIX version 3.20; guide of diskless workstation management*), which is a user's manual in two volumes of the system *AIX 3.20* for scientists using this system, whose objective is to allow users to use *AIX 3.20,* and have a more or less in-depth knowledge of the system.

After locating the different example cases in such a manual, we noted that examples come in various shapes: textual or non-textual examples (figures, tables, graphics), indeed we counted 3 distinct shapes of example units:

— the example unit is made of text,
— the example unit contains language of command,
— the example unit is a figure.

4.2 Location of examples

In technical documentation the example can take many different forms: textual and non-textual (figures, tables, graphics). Therefore, to identify and delimit these example units is necessary to take the properties of every case of example into consideration.

The indications present in the surface of the technical text make it possible to identify and delimit every example unit. These indications are linguistic, typographic, structural, and lexical or rely on punctuation. It is important to underline that these surface indications are presented under different levels of the text and give rise to recognition difficulties more or less important.

In our approach to example location, we begin with cases of examples associated with the explicit surface indications and then with those with non-explicit indications.

4.2.1 Cases of examples with explicit surface indications

It is acknowledged that examples are generally introduced explicitly into a technical text by expressions that are self-repeating in most cases. Some expressions are often associated to the particular forms of examples. We could find three expression categories of the example announcement:

1. Simple expressions associated in most cases to punctuation marks (period, colon, semicolon) such as:
   «Exemple:», «Un exemple», «Par exemple,», «Par exemple:», «L'exemple n°.»,
   «L'exemple ci-dessous…», «L'exemple ci-dessus…», «L'exemple ci-joint…», «L'exemple ci-après…»,
   «L'exemple suivant…», etc.
2. Expressions combined with other terms such as:
   «Comme l'illustre l'exemple ci-dessous», «La figure x fournit un exemple de…»,
   «On trouve au tableau x des exemples de…», «La courbe / l'algorithme / la figure / le schéma /le tableau ci-dessous /… est un exemple de…», etc.
3. Example units are sometimes announced by titles that are generally introduced at the end of the example unit such as:
   «Exemple de système de fichier», «Exemple de cadre de sélection», etc.

According to our corpus of study, we counted four different categories of examples units introduced by explicit surface indications.

— *The example is referenced in the logical structure of the document*
This is the case where the example is announced by a title that is part of the hierarchical structure of the document summary. The typographic organization in this condition should facilitate the location of the example since the title is generally presented under certain typographic processes (bold, underlined, italic). In the extract below, the end of the example is indicated by a line break to the left margin.

Exemple

Pour afficher les informations sur le client sans disque syzygy en format deux-points, entrez:

lsdclient -cL syzygy

Les informations suivantes s'affichent:

```
#name:hostname:ip:bhost:bdir:rhost:root:hhost:home:
shost: \
spot:sname:shhost:share:mhost:micro:swhost:swap:ssize:
dhost:\
dump:sclient:
syzygy:syzygy.paris.ibm.com:129.35.23.56:\
dlong.paris.ibm.com:/tftpboot:dlong.paris.ibm.com:\
/export/root/syzygy:dlong.paris.ibm.com:/export/home/
syzygy:\
dlong.paris.ibm.com:/usr:SPOT2:dlong.paris.ibm.com:/
usr/share: \
dlong.paris.ibm.com:/usr/lib/microcode:dlong.paris.
ibm.com:\
/export/swap/syzygy/swapfs0:32768:dlong.paris.ibm.com:\
/export/dump:no:
```

[Retour au texte…]

EXTRACT NO. 1

— *The example is separated from the text*
This is the case where the example is explicitly announced by the syntagm «exemple» or «par exemple», and the body of the actual example is quite detached from the body of the text, by simple returns to the line, indentations, and changes in the typographic processes in comparison with the surrounding text. In the box below, the example is announced by the syntagm «exemple» and the body of the example is quite separated by a line break and an indentation with a change of font size in comparison with the remainder of the text. A line break separates the end of the example from the remainder of the text.

Pour accéder à votre répertoire personnel à partir d'un répertoire quelconque, entrez:

```
cd
```

Exemple:

```
$ pwd
/u/marion/travail/93
$ cd
$pwd
/u/marion
$
```

[Retour au texte…]

EXTRACT NO. 2

— *The example is part of the text body*

The typographic disposition of the example impedes to highlight the example unit in comparison with the text, as it is attached to the text body. In this typical case, the example may include some sentences, a single sentence or only one word. In such conditions, the identification of the example becomes a delicate, or even difficult, task. We assert that the research of other surface indications in the technical text is necessary, in order to easily locate the example unit. Paralinguistic organizers are regarded as indications that help to categorize information. Thus, they can contribute to identifying and delimiting the example unit «Le fonctionnement des organisateurs est double. D'une part, ils permettent de distinguer les informations en fonction de leur importance. Ensuite, ils permettent d'établir une différence entre les informations selon leur nature» (CARO 1995).

In the case below, the example is delimited entirely by the two parentheses, which constitute, in this situation, one of the surface indications that permit the location of examples, «*Le fait d'être défini grâce à la notion d'importance fait que la parenthèse est jugée comme un outil privilégié dans la mise en relief*» (CARO 1995). The body of the example is presented by a font size different from that of the remainder of the text.

1. Exécutez la commande *mkdataless* sur le client. L'utilisateur doit indiquer le nom d'unité logique du disque à utiliser (par exemple, hdisk1). Un espace local de cliché et de pagination sera créé sur le disque spécifié.

EXTRACT NO. 3

An example can also be a sentence (delimited by a capital and a period) that fits normally in the continuity of the text, with the same typographic disposition as illustrated by the following example.

> Si une erreur se produit au cours d'une étape, les opérations précédentes sont annulées. Par exemple, si la copie de l'image d'amorçage n'aboutit pas, la commande *mkspot* supprime le répertoire créé pour le SPOT et le serveur reprend son état initial.

EXTRACT NO. 4

— *The example is referenced indirectly*

The indirect access to the example somehow complicates its location, it being referenced by an intermediary (generally by a title). In our corpus, we found an example unit introduced via the «*Remarque*», here the actual example is announced below using «*l'exemple ci-dessous*».

> Remarque: Pour éviter tout risque d'erreur, ces modifications doivent être apportées à chaque serveur, avant de lancer des commandes DWM. L'exemple ci-dessous illustre les définitions de variables.
> – Les machines suivantes fournissent des ressources:
> syzygy avalon dlong titres avalon.
> – Le répertoire */etc/aixdwm* est spécifié comme emplacement central de la base de données, dans le fichier *dwm_platform* de chaque serveur.
> – Les variables suivantes sont définies:
> DWM_SERVERS= «syzygy,avalon,dlong,titres:»
> DWM_HOST=avalon
> REMOTE_DWM_DIR=/etc/aixdwm
> Ces modifications permettront aux serveurs DWM de dialoguer les uns avec les autres.
>
> [Retour au texte…]

EXTRACT NO. 5

4.2.2 Cases of examples without explicit surface indications

Examples introduced by non-explicit expressions are not easy to find, because they are not associated to syntagms as «*exemple*» or «*par exemple*». These expressions refer to real examples, «Cependant, plusieurs cas de figures peuvent se présenter, rendant l'identification de l'exemple plus difficile, notamment parce que l'exemple peut être introduit par d'autres expressions (illustration, représentation, remarque, etc.)» (PAGANELLI 2000). This kind of expression often announces example units, without necessarily implying the effective existence of an example in any way.

The following example is introduced via a paragraph title, whereas the example itself is not announced explicitly in the text, although it is an

example. The typographic disposition of the body of the example in relation to the text (change of font size in comparison with the remainder of the text) makes it possible to spot it, the end of the example is indicated by a line break.

Format de fichier

Les noms d'unité figurant dans le fichier indiqué par la variable *Fich* doivent être séparés par des espaces:

hdisk0 hdisk1 cd1

ou bien chaque ligne ne doit contenir qu'un nom d'unité:

hdisk0
hdisk1
cd1

[Retour au texte…]

EXTRACT NO. 6

4.3 The syntagm «example» does not refer to a real example

We would like to emphasise that expressions of example announcement can appear in a technical corpus without referring to a real example. It is a situation often encountered in technical documents, and this is confirmed in the user's manual: occurrences of the term «*exemple*» or «*par exemple*» exist in the technical corpus, but do not refer to real examples in any way.

5 VALIDATION OF EXAMPLE LOCATION BY THE METHOD OF JUDGES

The main objective of our experimentation consists of confirming the cognitive validity of our surface indications. We aim to validate these indications by the method of judges. Consequently, judges (experts, non-experts) are going to confirm or invalidate the relevance of our surface indications, making it possible to identify and delimit the example unit inside a technical text. Every subject is asked to mark and delimit the example unit presented in each one of the assigned extracts, while indicating the surface indications (explicit, implicit) that helped them to identify the example.

5.1 The subjects

Ten adult judges participated in our experience. In order to choose the subjects participating in the experience, we took two categories of judges:

— five experts in data processing who frequently use some computing systems but are not experts on the *AIX system version 3.20*. Three judges are university professors and the other two are data processing engineers.
— five non-expert judges who are beginners in data processing. They are all students (2nd and 3rd cycle).

Besides, we decided that the number of judges in the two categories (experts, non-experts) would be exactly the same, so we could compare the two samples.

5.2 The experimental material used

The material includes 15 extracts containing examples, which judges have to assess. These selected extracts represent different types of examples appearing in the two manuals that are the subject of our study (*AIX version 3.20; guide of diskless workstation management without disk*). Example units can be textual or non-textual (figures). Each example extract is presented in its text environment to allow the judges to have an idea of the context and the framework surrounding the example. This would be helpful in their assessing work. The selected extracts are placed in an arbitrary order without a logical link between the different extracts, so as not to impact the judges' assessment. The choice of example extracts is made according to the shape that each example can take, as well as the presence or absence of surface indexes (explicit, implicit) that make it possible to identify the example unit.

5.3 Discussion of results

The results of the experimentation are very significant, and should be interpreted on several levels (type of example, shape of the example, surface indexes and type of judge). We shall now comment these results, talking only about the influence that surface indexes may have on the judges' decision.

As we have defined different example types and shapes out of the corpus, the subject of our study, our analysis approach is going to take into account the kind of example unit, despite the fact that every type can be presented under two different shapes (text, graphic), so it is necessary to add the explicit and implicit surface indexes, which define each example unit.

In fact, surface indexes usually make it possible to spot and delimit the example unit, the latter acts on different levels in the text, thus genera-

ting more or less important recognition difficulties. The experimentation has made possible to observe that there are examples that are automatically identifiable; such as when the example unit is announced by an explicit title referenced in the summary. However, example units cannot be obviously identified if they are not associated to syntagms such as «exemple» or «par exemple». In this case, it seems necessary to search for other surface indexes allowing to identify and delimit the example unit. Other cases may be considered; the example unit is easily identified but its demarcation presents some difficulties, especially when it is composed of textual and non-textual objects at the same time (pictures, figures, etc.).

The surface indexes in the technical text are one of the tools that make possible the identification and delimitation of the example unit. However, there are cases where it is difficult to identify the example unit in spite of the presence of explicit surface indexes. This means resorting to other indications such as the meaning of the technical speech surrounding the example unit. All of this makes the demarcation of the reply unit a more delicate matter than might have been thought.

Indeed, results of the experimentation reveal that surface indexes may be pertinent and essential indicators of example unit location in technical documents. Nevertheless, there are cases where only one explicit surface index is not sufficient to permit the location of the example unit. The results of the surface index assessment by the method of judges confirm the relevance of our choices.

The knowledge acquired by the experienced judges contributed to facilitating the example unit location operation, as they showed a very notable facility in identifying examples compared to non-expert judges, particularly in extracts presenting ambiguous cases (the case of the example unit without explicit surface indexes).

6 Is the example unit a relevant information unit?

Technical documents are characterized by their complexity due to their very strong logical, structural and hierarchical organization. They are generally voluminous. All of this makes it rather difficult for users (experts, beginners) to conduct a more specific and varied information research, operative by nature.

Examples are among the most sought-after informational units by users in technical documents, because they immediately provide useful information (Paganelli 2000). Generally, users in working environments do not have time to read several pages to reach the required information, which is crucial to decision making or solving a technical problem. By way of example, a technician who wants to identify and repair a radar has to look in 41 different places on 165 pages across 8 different documents.

The example unit is regarded as a privileged source of information by the user as it is easy to understand and contributes directly to the execu-

tion of a task. In this sense, an experienced user in an operational framework needs fast and direct information, and the example unit is the best suited to give this kind of information.

The example unit can be used as an answer to a possible user request in isolated fashion as an autonomous and applicable information unit. However, in some cases the text framework of the example unit is indispensable to understand it. This often complicates the demarcation of the unit answer. Example units can function as answers to isolated information queries; some are announced in the hierarchical structure of the summary, while other example units are indirectly referenced, and should be presented to the user with their textual context.

7 Conclusion

Surface indexes present in the technical text are one of the tools that allow to identify and define example units. However, there are cases where it is difficult to identify the example unit in spite of the presence of explicit surface indexes. In this situation, other indications besides the surface indexes have to be taken into account, such as the meaning of the technical speech that surrounds the example unit.

The example in the technical document holds a major position and facilitates the understanding of technical discourse. This unit is characterized by lexical and typographic criteria, permitting its automatic location. However, the mixing of certain examples with other types of information in the technical corpus gives rise to the problem of their location and their detachment from the text framework.

For certain authors, the example unit can be regarded as a relevant information unit so it may serve as an answer to a possible user request. Before confirming this idea, it would be interesting to study the needs and the profile of the user, since these are two fundamental factors in order to determine the answer to give after a process of information research. Therefore, further consideration should be given to the topic; in other words, how could the example unit, as an answer to an information query, be presented?, and does user profile have an impact on this query?

References

(Caro 1995) Caro, Stéphane. «Rôle des organisateurs para-linguistiques dans la consultation des documents électroniques». Thèse de doctorat soutenue en décembre 1995, (Université Stendhal – Grenoble 3).

(Chaibi 2001) Chaibi, Ahmed. «Le repérage des exemples dans les documents techniques». DEA soutenue en juin 2001, (Université Stendhal – Grenoble 3, Institut de Communication et des Médias).

(Coltier 1988) Coltier, Danièle. «Introduction et gestion des exemples dans les textes à thèse». *Pratiques*. n. 58 (juin 1988), p. 23-41.

(Froissart 1999) Froissart, Christel; Lallich-Boidin. «Le document technique: unicité et pluralité». In: Colloque du chapitre français de l'isko (2e: 1999: Lyon). *L'indexation à l'ère d'Internet, deuxième colloque du chapitre français de l'ISKO, Lyon 21 et 22 octobre 1999*, texte 18, p. 1-8.

(Heurley 1994) Heurley, Laurent. «Traitement de textes procéduraux: étude de psycholinguistique cognitive des processus de production et de compréhension». Thèse de doctorat, (Université de Bourgogne, mention psychologie, janvier 1994).

(Morizio 2002) Morizio, Claude. *La recherche d'information*. Paris: Nathan: ADBS, 2002, 126 p.

(Mounier 2003) Mounier, Evelyne; Paganelli, Céline. «The extraction and representation of knowledge contained in a technical document». In: Communication reliée extraite du IV International Colloquium on library and information science, Université de Salamanque, mai 2003, 13 p.

(Ouerfelli 2001) Ouerfelli, Tarek. «La segmentation des documents techniques composites dans une perspective d'indexation: vers la définition d'un modèle dans une optique d'automatisation». Thèse de doctorat soutenue en juillet 2001, (Université Stendhal – Grenoble 3, Institut de Communication et des Médias).

(Ouerfelli 2002) Ouerfelli, Tarek. «La segmentation des documents techniques on amont de l'indexation: définition d'un modèle». In: Colloque international sur le document electronique, CFD (5e: 2002: Hammamet, Tunis), p. 37-50.

(Paganelli 1997) Paganelli, Céline. «La recherche d'information dans des bases de documents techniques en texte intégral. Etude de l'activité des utilisateurs». Thèse de doctorat soutenue le 10 juillet 1997, (Université Stendhal – Grenoble 3).

(Paganelli 2000) Paganelli, Céline; Mounier, Evelyne. «Le repérage des exemples dans les documents techniques». *Informatique documentaire.* n. 77 (2000), p. 23-40.

(Thyrion 1997) Thyrion, Francine. *L'écrit argumenté*: *questions d'apprentissage.* Louvain-la-neuve: Peeters, 1997. (Série pédagogique de l'Institut de Linguistique de Louvain).