A controlled language at Airbus

L. Spaggiari, F. Beaujard, E. Cannesson

laurent.spaggiari@airbus.com florence.beaujard@airbus.com
emmanuelle.cannesson@airbus.com

ABSTRACT

In 1998, Airbus started a project dedicated to the creation of a controlled language for industrial use; in this case enhancing warning texts quality in the cockpit of Airbus aircraft. Another objective was to provide designers a means of facilitating their job while respecting the stringent safety criteria. This project was divided into three parts conducted in the frame of R&D activities, the first one dealing with the terminological aspects, the second devoted to syntax and the third one focused on the acronyms for computers naming.

Throughout this innovative linguistic project, Airbus proves to be aware of the constant need to enhance safety.

Keywords: Controlled language – Word order structuring – Terminological standardization – Abbreviations – Procedural texts – Acronyms.

Introduction : what is a CL ?

Controlled languages (CLs) are of vital interest (for safety and economic reasons, etc.) for industry. Indeed, they have been created in order to resolve problems of readability (reducing the complexity of syntactic structures of a text increases its readability), of comprehensibility (a lexical disambiguation increases the comprehensibility of a text) and of translatability (a syntactic and semantic control facilitates the shift between two languages) but not of grammaticality (a grammatical text written in a given CL will not necessarily be considered as grammatical in the corresponding natural language).

As Goyvaerts (1996) wrote, "Industry does not need Shakespeare or Chaucer, industry needs clear, concise communicative writing - in one word Controlled Language".

"A restricted or controlled language refers to a system that limits language to a set number of core vocabulary words, and usually, a set of writing guidelines for grammar, mechanics, and style. […] A controlled language attempts to reduce ambiguities, colloquialisms, and synonyms" (AECMA (1995)).

English is a very productive natural language for CLs' creation as it is the current international language used for trade and science. Nevertheless, other natural languages such as German, Chinese, Swedish, French, etc. have generated CLs. A CL is not "simple" or "baby" English, German, French, etc. but simplified English, German, French, etc.

The Airbus project

In aeronautics, pilots must daily use procedures in any type of situations (normal or abnormal). It was observed that some incidents aboard commercial planes were due to the non compliance with procedures. Whether it be oral or written, a message will be considered successful and effective when it is in keeping with the mental process implemented to reconstruct and interpret the information contained in this message. But, because one does not expect any individual to master and speak a number of languages with the same level of competence as the one he has reached in his own mother tongue, industry prefers a precise and concise language to the use of a natural language which would allow non-parallel grammatical constructions, possessing inherent ambiguities of various types, etc.

In 1998, Airbus started a project dedicated to the creation of a controlled language for industrial use; in this case enhancing warning texts quality in the cockpit of Airbus aircraft. Another objective was to provide designers a means of facilitating their job while respecting the stringent safety criteria. This project was divided into three parts conducted in the frame of R&D activities, the first one dealing with the terminological aspects, the second devoted to syntax and the third one focused on the acronyms for computers naming.

For industries willing to create a CL need to be aware of what has already been done, we first built up an overview which could give instant access to information. To achieve it, we had a close look at what has been written in the field of CLs and tried to get in touch with the persons involved in different projects (K. Barthe, E. Johnson, K. Godden, B. Arendse, E. Adolphson, etc.) We encountered different domains such as aircraft, meteorology, emergency services (police, fire, maritime, ambulance, etc.), etc.

Our overview is meant to be a help for work. It is open-ended and can be added to. A user with queries about a particular CL can easily and quickly (in a click) get concise and succinct answers such as the rules applied, the company involved in the project, etc. by consulting the related ID card. These ID cards (about 40 that required half a year to be compiled) consist of the following headings: complete name of the CL, date (of creation, or duration of the study), organization (a company, university, etc. which owns the product), designer (an individual or a pool, a company, etc. who is in charge of the project), objectives / application, elaboration / content, and bibliographical references. Because our aim was not to provide extensive information, we added an appropriate bibliography on each ID card so that the user can find helpful references for more precise information.

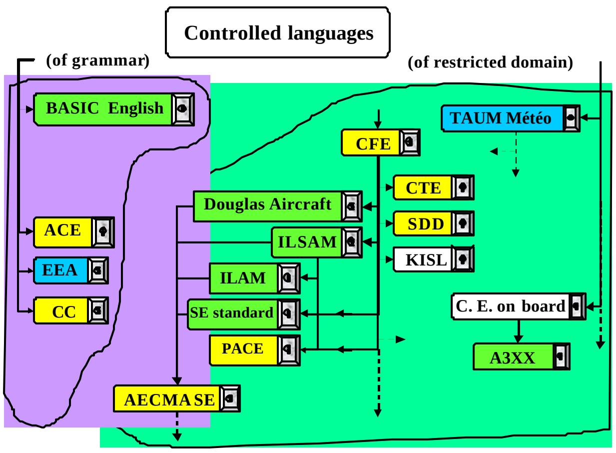


Figure 0: CLs' overview

Name	DOUGLAS AIRCRAFT
Date	1979
Society	McDonnell Douglas Corp.
Designer	*McDonnell Aircraft Company*
Purpose Application	Standardization of readability and translatability of technical and A/C maintenance manuals.
Elaboration Content	2000 words taken from the list of the preferred verbs used in the Navy, in the Air Force, and in McDonnell 50's technical manuals. This technical vocabulary was one of the sources studied for the creation of the AECMA SE lexicon.
Biblio References	**Gingras B. (87)** **Huijsen W.O. (98a&b)** **Stewart K.M. (98)**

Figure 1: An ID card

Main characteristics of our CL

It will not be a translation tool. This CL will be presented in the form of a writing guide (including recommended structures and vocabulary). These constructed sentences are not intended to evolve. Nevertheless, improvements by modification of messages or addition of new ones will be possible, if necessary. Moreover, it is not meant to be computerized for the moment.

It will improve comprehensibility by improving the abbreviations and the acronyms, deleting synonyms, reducing and standardizing syntactic structures, making some ellipses explicit.

It will take into account the interference between languages because even though texts are written in English (official language for exchange in the cockpit), pilots are from various linguistic origins. Thus, different crew configurations exist: both pilots are native English speakers, or only one of them, or both are non-native English speakers (possibly not sharing the same mother tongue).

Theoretical and practical (both semantic and syntactic) choices will be clearly justified at each step of the establishment of this language. Its validation will consist in the checking of objectives and of assessments done by different persons (domain acquainted or not). Theoretical choices will comply with bibliographical references, existing theories and new ones, etc. Practical choices will comply with pilots' assessments, workgroups, former studies, etc. These choices will be justified and written in order to ensure a good traceability of design rationale.

It will respect the Airbus family concept (commonality between aircraft). It is to say the existing use of terms that participates to the upholding of knowledge in a community of speakers and also eases the cross crew qualification.

It will deal with technical constraints such as for instance the restricted space dedicated to the display of warning texts on screen (between 20 to 36 characters according to the type of text concerned).

It will be crew oriented. To make sure that it achieves its objectives, all the persons involved (from the designers to the pilots) will meet for workgroup sessions. These sessions will help to collect the comments of potential users and to guarantee homogeneity between cockpit crew-machine interfaces, operational documentation and maintenance. This feedback will be very useful, because “the sooner user requirements are integrated into the design, the quicker it will be possible to iron out snags with the end-user’s help” (Patri (1998)).

Figure 2: The working group

<u>How we did it?</u>

1. terminological methodology

Always concerned by the respect of objectives and constraints, we developed an innovative method intended to determine terms and the form under which they should appear in future warning texts. The terminological study has been conducted along two major thrusts: the terminological standardization principle and the morphological reduction process. It is to be noticed that at each stage of these two axes users are taken into account (via interviews and assessments) in order to formulate recommendations that correspond to their operational need and experience.

1.1. The terminological standardization principle

This principle aimed at normalizing the existing terminology and was based on two postulates:

- existing terms are reusable in the future terminology but on the condition they fulfill the different fixed criteria;
- existing terms are not implied in a synonymy phenomenon.

With the help of the analysis of the corpus (about 3000 sentences and 700 words), a decisional tree (figure 3) has been created with different criteria. These are of different kinds:

a) derivational and flexional through:

- the keeping of final morphemes such as –ed and –ing as explicit visual marks of in course or accomplished processes;
- or the keeping of the final –s as the mark of the plural form;
- or the suppression of the negative prefix in favor of the explicit negative expression through the use of not.

b) homophonic and homographic through the location of homophones and homographs in the corpus, but also homographs with various languages.

c) geographical and genealogical through the preference of terms with an American English tendency and/or terms presenting a Latin root.

d) documentary through the checking of the use of these terms in different aeronautical references such as aeronautical regulations, maintenance references, operational documentation and air traffic control references.

These criteria permit to tag each term the same way in order to create the decisional matrix. When a term respects all the criteria, it is immediately reused in the new terminology. If it does not respect at least one criterion, a candidate (respectful of that criterion) is proposed. Both term and its candidate undergo the decisional tree. Then, it must be decided which one of the term or candidate must be kept in the new terminology; this is done through the observation of the decisional matrix and the comparison of all criteria.

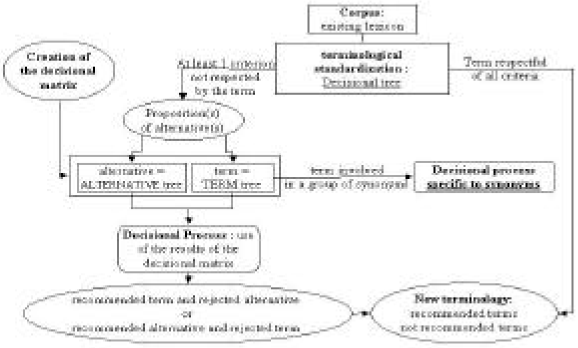


Figure 3: Existing terms are reusable

When a term is involved in a group of synonyms, a specific treatment has been established (figure 4). With the help of an expert of the domain, three categories have been isolated: invalid, valid synonyms and groups of synonyms to be confirmed.

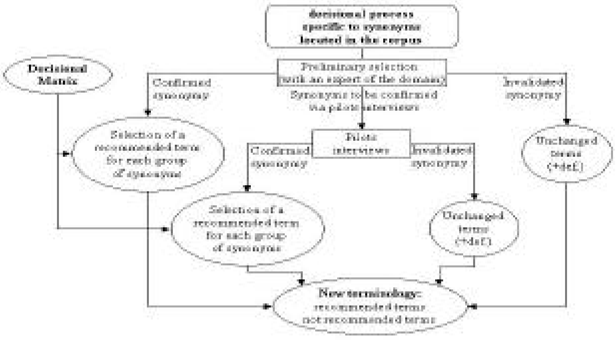


Figure 4: Treatment of synonyms

a) invalid synonymy: When a group of synonyms is invalidated, the expert gives a precise definition for each term and they are introduced in the new terminology.
b) valid synonymy: Using the contexts of the terms, the expert is able to validate the synonymy and to recommend the use of only one of these terms. Therefore, with the help of the decisional matrix it is possible to select one term representative of the group which would be introduced in the new terminology. The other terms appear in the new terminology as not recommended terms.
c) synonymy to be confirmed: Because the expert was not able to confirm some synonymies, it was decided to consult more experts of the domain in order to take a decision whether these groups were synonyms or not. This was realized by interviewing eight pilots (airline, flight test and instructors) in the same spirit as it was done with the first expert.

After the establishment of the new terminology it was necessary to determine under which form (short or full) the terms should appear in the future warning texts.

1.2. The process of morphological reduction

The process of morphological reduction (figure 5) permits to respect the uniqueness criterion: “one word – one meaning – one short form”.

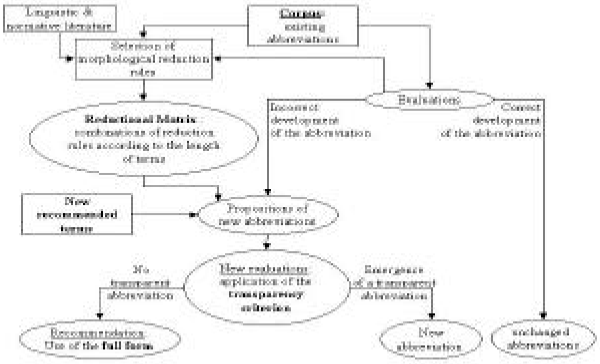


Figure 5: Morphological reduction process

With the help of the corpus analysis and different linguistic references we created a reductional matrix which is the compilation of abbreviating rules applicable depending on the length of terms. Abbreviated forms are generated by the matrix and submitted to assessment with pilots selected as explained in the syntactic part. Speed and correctness criteria were measured in order to characterize the transparency criterion defined as follows: “an abbreviation is transparent when correctly developed in a minimum of time without any context”. It is important to note that previous assessments have been conducted on existing abbreviations in order to respect the commonality principle; therefore every existing short form which has been judged as transparent was kept in the new terminology. New short forms generated by the matrix were proposed only for those whose existing short forms failed at the transparency criterion. In the end, for each recommended entry of the new terminology it is recommended to use either a short form, or the full form when we were unable to generate a transparent abbreviation or for specific reasons (safety/rarity/commonality, etc).

2. syntactic process

As for terminology, syntax was concerned by the respect of objectives and constraints. Due to the lack of room, one of the main characteristics of the corpus is the quasi systematic lack of grammatical words such as in, of, by, etc. In this context, the parameter "word order" is crucial and must be taken into account. As Slobin (1985) notices, "It is likely that elements such as case inflections, verb inflections, pre- or postpositions, and conjoining and subordinating particles provide major orienting points for the perception of structure". In figure 6, the shortest[1] English sequence "young horse breaker" has two meanings whereas the incorrect sequence "horse young breaker" would have only one. Indeed, depending on what the adjective "young" defines "the horse breaker" or "the horses", the grammatical nominal phrase "a young horse breaker" can be understood differently, respectively "a young breaker of horses" or "a breaker of young horses". On the contrary, when "young" is close to "breaker", the ambiguity disappears and only the first meaning "the young breaker of horses" is now possible.

[1] Without any room constraint, a possibility would be to use the « of » construction : the configuration of... or the last configuration of..., thus avoiding the ambiguity.

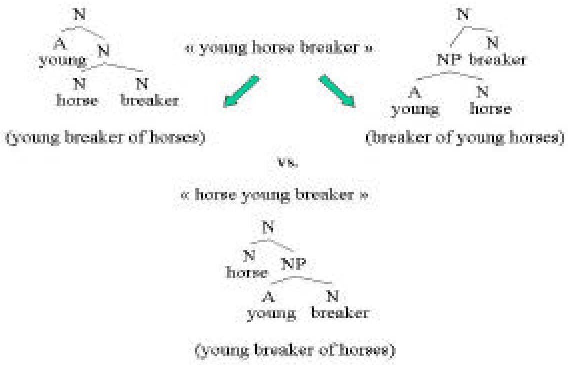


Figure 6: Word ordering and comprehension

Thus, an efficient controlled language must be based on "a consistent usage of conventions and restrictions related to grammars, vocabularies, and styles […] The consistency of a given text depends on the absence of illegal constructions (proscriptive and prescriptive rules) and on imposed regularities in the presentation of textual information" (Declerck in CLAW 2000). To achieve it, we analyzed the corpus within the scope of potential ambiguity, adopting a syntactic standardization principle because consistency is one of the most basic usability principles; Therefore the same information should be formatted in the same way to facilitate recognition.

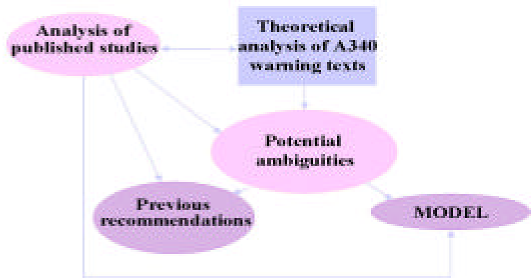


Figure 7: Syntactic process (theoretical part)

We dealt with different phenomena such as the expression of the negative, the coordination, the place of a condition in a procedure, etc. This corpus analysis plus a close look at the published studies enabled us to cope with the potential ambiguities by providing a set of rules (about 20) "generating" a unique and homogeneous information format and content (figures 7&8).

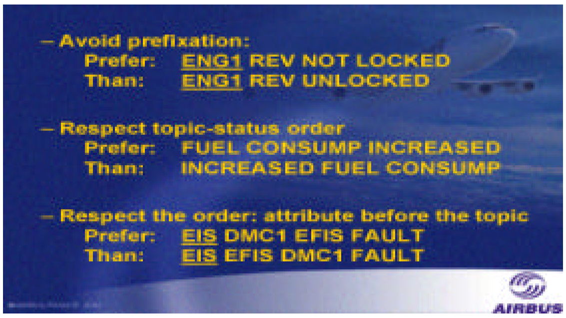


Figure 8: Example of syntactic rules

But, because this language had to be applied, we conducted assessments with different populations (airline, flight test and flight instructor pilots and people outside the aeronautic domain) with different mother tongues to compare the different solutions in terms of information usefulness for the pilot and also of understanding when word ordering is concerned (figure 9).
To cover this last point, we built up a multiple sentence (similar to that one) reflecting all the plausible orders:

young horse breaker tired
horse young breaker tired
tired young horse breaker
tired horse young breaker

and submitted it to everyone's understanding.

We first selected the languages spoken by Airbus customers. Then we applied a genetic criterion for the selection of the languages which were not covered by the first filter. And finally, inside the different groups of languages, we chose those that presented significant structural differences. As a result, 64 persons have been interviewed which represented some 16 languages.

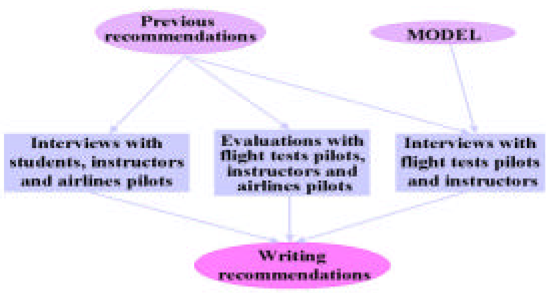


Figure 9: Syntactic process (practical part)

The statistical analysis of the collected data allowed us to observe a general tendency for people to follow the rule : What can be meant to be associated must be syntactically close in a sentence.

To sum it up, the assessments on rewritten texts we conducted with pilots from different mother tongues confirmed the syntactic rules we proposed.

3. computers naming

From the database we created, we analyzed the structure of the current acronyms and also conducted assessments with airlines pilots, instructors and flight test pilots in order to determine how acronyms were learnt and used in the cockpit. Assessments were also performed to detect the types of confusions likely to appear. For that, we first built up one questionnaire in order to master the acronyms training, the learning support, what the pilot does when an unknown acronym is read in the cockpit. Then, the assessment included questions on some acronyms, asking for:

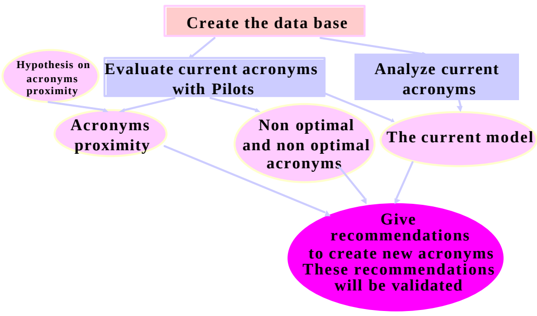
- giving the full form
- explaining what the function of the corresponding equipment is
- giving the system concerned

This part was useful to know what the pilot intuitive expectations are concerning acronyms model, i.e. the kind of information expected concerning the first letters, for letters in the middle of the acronym and at the end. Data analysis was done thanks to a confrontation of the current acronyms full forms with what has been given by pilots.

After these theoretical (structure analysis) and pragmatic (assessments) approaches, we were able to master what an optimal acronym should be. The overall method described above is summarized in the hereafter figure (figure 10).

In parallel, we hypothesized that there are different levels of letter proximity from a graphical and phonetic point of view. So we used different published studies to develop grids of interference between letters. When two acronyms have only one letter of difference, these grids allow us to know whether these acronyms proximity is high or low.

Figure 10: Computers naming process



At the end of this process we were able to give recommendations in order to help designers to create appropriate acronyms. This shall be part of one “Airbus Language Reference”. The hereafter slide shows examples of recommendations. Based on these recommendations, designers propose acronyms. The final choice is done after comparative assessments of these proposals .

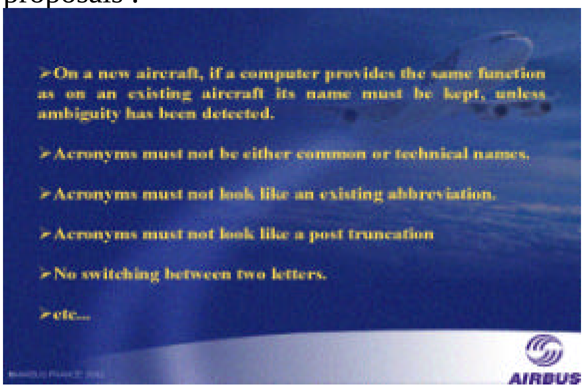


Figure 11: Sample of rules

Conclusions and perspectives

The terminological methodology permitted the creation of the terminology of a controlled language based on the existing lexicon, applicable for the moment to warning texts. This precise and concise language is guaranteed by an homogeneous structuring of the message at different levels (term selection and abbreviation, word ordering, etc.) A validation protocol in operational contexts is in progress in order to validate the new terminology, and the syntax developed. As far as objectives are concerned, the first results indicate that this language is valuable. Indeed, the rewritten warning texts appear to be quicker and better understood than the current ones. The final rules stem from a construction which is not only justified at every stage but also assessed by the various actors (pilots and designers) involved in the project, i.e. texts users and writers.

Throughout this innovative linguistic project, Airbus proves to be aware of the constant need to enhance safety. It has been decided to extend this linguistic activity to other cockpit functions: ATC and FM interfaces, cockpit commands and to extend it to the operational documentation as well.

Bibliographical elements

AECMA (1995) : Simplified English: a guide for the preparation of aircraft maintenance documentation in the international aerospace maintenance language. Bruxelles : Gulledelle.

Cabre M. T. (1998) : La terminologie : théorie, méthode, et applications. Ottawa : Ottawa University Press.

CLAW (1996, 1998, 2000) : Proceedings of the First (Second and Third) International Workshop on Controlled Language Applications.

Doppagne A. (1998) : Majuscules, abréviations, symboles et sigles, Duculot : Paris, Bruxelles.

Gavieiro-Villatte E. & Spaggiari L. (1999) : Open-ended overview of controlled languages. In BULAG 24 (1999), pp. 89-100.

Gavieiro-Villatte E. (2001) : Elaboration de la terminologie d'une langue contrôlée visant à améliorer la compréhension des textes d'alarmes en aéronautique:

application aux futurs postes de pilotage. PhD, University of Franche-Comté.

Gingras B. (1987) : Simplified English in maintenance manuals. In Journal of the society for Technical Communication N°34(1), pp. 24-28.

Goyvaerts P. (1996) : Controlled English, curse or blessing ? A user's perspective. In CLAW 1996.

Grishman R. & Kittredge R. (1986) : Analyzing language in restricted domains : sublanguage description and processing. New Jersey : Lawrence Erlbaum.

ISO 704 E (1987) : Principles and methods of terminology (first ed.).

Huijsen W.O. (1998a) : Completeness of compositional translation. Utrecht : Drukkerij Elinkwijk B.V.

Huijsen W.O. (1998b) : Controlled language : an introduction. In CLAW 98, pp. 1-15.

Jonassen D. H. (1982) : The technology of text : principals for structuring, designing and displaying text (tomes 1&2). Englewood Cliffs : Educational Technology Publications.

Kittredge R. & Lehrberger J. (1982) : Sublanguage : studies of language in restricted semantic domains. New York : Walter de Gruyter.

Monnier P. (1993) : Les sigles: Approches linguistiques, applications au domaine du spatial. DEA université de Toulouse Miraïl.

Patri G. (1998) : A340-500/600 : un développement sur maquette numérique à Toulouse, Revue Aérospatiale 153.

Pinker S. (1999) : Words and rules: the ingredients of language. New York : Basic Books.

PRESCOT (1997) : Le texte procédural : langage, action et cognition. Toulouse : May.

Schriver K. A. (1997) : Dynamics in document design : creating texts for readers. New-York: J. Wiley & Sons Inc.

Slobin D. I. (1985) : The crosslinguistic study of language acquisition (volumes 1 et 2). Hillsdale : Lawrence Erlbaum Associates.

Spaggiari L. (2002) : Formalisation des rapports syntaxiques d'une langue contrôlée: application aux messages d'alerte dans les postes de pilotage avion. PhD, University of Poitiers.

Stewart K. M. (1998) : Effect of AECMA Simplified English on the comprehension of aircraft maintenance procedures by non-native English speakers. PhD, University of British Columbia.

Wieringa D. & al. (1993) : Procedure writing : principles and practices. Colombus : Battelle Press.