CTML: a context transport mark-up language for formalization and verification of legal, medical, bureaucratic and safety procedures and protocols

Graziella Tonfoni [1,2]

University of Bologna,
Bologna, Italy
and
SEAS - The George Washington University,
Washington D.C., U.S.

Abstract

CPP-TRS [2] stands for Communicative Positioning Program - Text Representation Systems and is a visual language based upon a consistent set of dynamic visuals, which may be used in combination to represent qualitative reasoning about the nature of information packaged in textual format. A subset of CPP-TRS visuals, indicating those categories which create communicative context, precisely textual signs and textual symbols, have been recently isolated and converted into functions and commands of a mark-up language for documentation annotation. The newly derived special purpose language is called CTML, which means Context Transport Markup Language, [4] and the main objective to be achieved is accurate incorporation of contextual features into documentation. Communicative function carried by signs, together with communicative intention carried by symbols and communicative turn-taking carried by turn-taking symbols create context, which may be easily incorporated within each single instruction, that has been produced in textual format. Instructions displayed according to a set of categories selected as the most relevant ones will provide users with those interpretive clues which are needed to support accurate and effective understanding of each instructional sequence. Strategic information, such as legal, medical, bureaucratic text and safety protocols are, if designed and packaged according to the CTML standard [5] may be easily and effectively stored, retrieved, accessed, distributed and updated; therefore significantly enhancing power and speed in the search process as well.

1 Introduction

CPP-TRS [2] is an integrated environment, showing different levels of complexity, meant to represent textual operations.

In technical documentation and instructional text management, recognition and visualization of communicative patterns, both at a global level which means by paragraph, and a local level which means by sentence, is very useful. CTML is a derivative language meant to make textual operations possible at the highest possible degree of reliability.

In order to be able to operate text consistently, a common standard for both instructional text production and instructional text understanding is extremely important.

Text production has been defined as text actualization and text understanding as text sensing and text perceiving [2].

Before turning into an instructional text made of a sequence of organized paragraphs and sentences, the originating information flow is made of unprocessed sequences of information. The information flow will undergo a process of texture configuration based upon visualization of information distribution, made possible by the activation of information quality and quantity detectors.

The text actualization process suggests that a text-operators' crew, made of information tailoring experts [1], who have been monitoring the configuration process, will be able to convert the information flow into an actual instructional text.

Once actualized, the text shows the active sensors and perceptors [2], which act as macro-instructions (sensors) to extract global communicative patterns (GCP) and as micro-instructions (perceptors) to extract the local communicative patterns (LCP) from the text. As a result of a consensually shared standard, text sensors will be able to produce text sensing and text perceptors will be able to produce text perception.

Text sensing means that the technical writers' team will be able to indicate the global communicative position (GCP) within the text, indicated by signs.

Visually processed text units are meant to incorporate text and both sensors and perceptors. This means that each text unit is transparent to any text operator because of the common sharing of a same system of representation.

The actual text may undergo further processes of reformulation, which means that each text unit may be further expanded or even radically transformed.

Teams of textual operators, like technical writers, are may operate on text-sensors by modifying their communicative function also called trajectory in order to design material for different kinds of users/readers. They may also

---

[1] Graziella Tonfoni teaches Computational Linguistics at the University of Bologna Italy. She is also a Visiting Professor at the School of Engineering and Applied Science at the George Washington University, Washington D.C., U.S.
[2] Email: `tonfoni@alma.unibo.it`

operate on text perceptors by modifying their communicative intention and turn-taking also called energy. The result of such a set of operations is a more or less in depth reformulated and reconfigured text.

The text will show both active effectors and manipulators which act as macro-instructions, (effectors) to modify the GCP global communicative positioning of the actual text, and as micro-instructions (manipulators) to modify the LCP local communicative positioning of the actual text. As a result of the consensually shared system, text effectors and text manipulators will modify text sensing and text perception in fully transparent ways.

To summarize, we can say that text reconfiguration means precisely the following: a technical writer, after having sensed and perceived an actualized text, will decide to operate on the patterned and perceived text as to extend it into a communicative environment which is different than that in which it exists and leave it transparent. Text virtualization means that the technical writer may create different communication spaces for an actualized text, which becomes a virtual text with a set of visible operations. It is very important to distinguish between information flows which have not yet been actualized, and a virtual text, which is an extension and expansion of an actualized text within a virtual space of communication.

We all know that the punctuation system allows readers to distinguish between an interrogative sentence and a declarative one. In a similar way, the technical writer, who is the information tayloring expert , will be able to identify and use an appropriate visual representation provided by the annotation system in order to make GCP and LCP explicit. In some ways, CTML can be defined as a communicative traffic control system.

As with any system, by knowing the basic rules, an infinite number of operations may be performed and understood in a quick, unambiguous and easy way.

CTML, a derivative language, allows any text operator to actualize any kind of text by following very simple instructions leading toward more complex explanations. By conveying GCP and LCP visually, interactions occurring in different languages at the same time may be facilitated.

The same system may be applied to verbal interactions also converted into a written format.

CTML is based on the concept that you may enhance and mutually control communication by visually processing text. The starting point is a visible global organization of text, and the ending point is the actual language of the text. This approach is soundly grounded in cognitive research that is explicitly aimed at explaining the complexities of how our minds apprehend, process and communicate knowledge.

2 CTML: a powerful mark-up language derived from CPP-TRS

CTML is a mark up language for consistent and context sensitive packaging of information based upon a visual system for text visualization. The originating communicative context in which an instruction was first generated is made accessible and allows for full visibility of different communicative operations which were performed or need to be performed both globally and locally, synchronously and asynchronously.

Operations that each part of a text is likely to undergo are visually represented in order to be consistently interpreted.

Legal and medical texts, bureaucratic and safety procedures and protocols, just like any document, come from "somewhere in time and space and leads toward somewhere else" [3].

They may therefore be defined as a piece of information that has been derived from a dynamically evolving information flow before it is converted into a more stable form.

Enhanced encoding procedures for supporting accurate interpretation are aimed toward conveying effectively those relevant clues, which represent context throughout a combination of specific icons, which are of three kinds :

- **Instructional texts annotation signs**: to indicate the **communicative function** of a technical text;

- **Instructional texts annotation symbols**: to indicate the **communicative style** of a technical text;

- **Instructional texts annotation turn taking symbols**: to indicate **roles and interplay** between the technical text producer and the user.

Instructional texts annotation signs, which represent the various communicative functions, each technical text may convey, paragraph by paragraph, are the following ones:

Square: for an informative paragraph which carries information about a specific fact, to be linked up with another paragraph made available, in order to extend topical continuity and context consistency.

Square within the Square: for a summary of a certain paragraph which has been produced to reinforce contextual consistency between an original instruction and its own abstract.

Frame: for a paragraph which is found to be analogous in content to other paragraphs and previously stored cases, it is meant to reinforce contextual consistency between and among different cases.

Triangle: for a reminder of a previous paragraph found to be of relevance to the current one, meant to establish topical continuity with background information, which has not been previously introduced, because not yet available.

Circle: for a main concept conveyed by a certain paragraph, which has been abstracted as to be linked to other paragraphs, showing topical continuity. It is meant to reinforce topical keywords identification and to effectively link together paragraphs which show the same keyword, in graphically explicit ways.

Grouped Semicircles: for main concepts, which are abstracted out of an originating paragraph meant to establish both topical continuity and context consistency between the originating paragraph and a set of topical keywords.

Semicircle: for a locally identified concept, abstracted out of a paragraph and meant to reinforce context consistency by establishing further links to other paragraphs, which may be triggered by the same keyword, in graphically explicit ways.

Inscribed Arcs: for indicating the need for an upgrade and update of a certain paragraph; it indicates that a revision process is likely to occur, though it does not declare if such revision will be a major or a minor one.
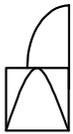
Opened Text Space: for indicating that an upgrade and update has indeed occurred within a certain paragraph; it indicates that the paragraph has now reached a new revision state. It does not declare if the revision has been a major or a minor one.

Right Triangle: for a comment made to a certain paragraph coming in, when more contextual information is needed, which has to be derived from other external sources, not previously available, based on topical continuity.

Instructional texts annotation symbols are meant to indicate communicative intentions and styles, locally within a certain paragraph, sentence

by sentence.

They represent different modes of information packaging, which if activated at different times or at the same time, may be combined and used dynamically for repackaging purposes. They effectively indicate transitional states within the same paragraph, by declaring explicitly the nature of those changes which have occurred or are likely to occur next.

They are the following ones:

Describe: from Latin *describo*: write around. It means complementing the original sentence with as much information as may be found interesting to add, without any specific constraints. It is represented by a spiral, which starts from a central point of the spiral indicating the original sentence and proceeds toward expanding the sentence at various degrees, linking it with other sentences coming in from different sources and found to be relevant to facilitate interpretation.

Define: from Latin *definio*: put limits. It means complementing the sentence with limited information about a very defined topic, which has been previously selected and identified as the most relevant one, represented by the middle point of the square. It indicates that there is a specific need to incorporate specific information, which is made available and implies accurate and most selective focusing on a very limited package of highly specific information.
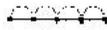
Narrate: from Latin *narro*: tell the story. It means complementing the sentence with various facts and events by following a logical and chronological order. It indicates a set of major points or facts representing different diachronic stages, which are strictly linked up together in a sequence.

Point out: take a single point out of a story chain. It means isolating a specific event or fact among those reported within a sentence, focusing on just that one, and adding more detailed information, by expanding it significantly.
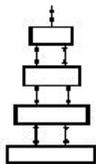
Explain: from Latin *explano*: unwrap, open up. It means that facts and reasons are given to support interpretation of a certain event, within a certain sentence. We may start by indicating the originating cause and proceed toward showing the effects or start with effects and go back to the cause, according to what is found to be more significant.

Regress: from Latin *regredior*: go back. It means that more information about a certain topic, presented within the sentence, is absolutely needed to gain a deeper understanding. It represents a specific topic focusing process and an in depth information expansion, which is activated only for that precise topic . The reader may want to consider if further information is needed and ask for availability of further resources.

Inform: from Latin *informo*: put into shape, shape up. It means that any sentence is the result of some information packaging and that the specific sentence indicated is packaged in the most unconstrained way; it may therefore be subject to many and various kinds of repackaging. It leads toward two different kinds of further specification, which are respectively conveyed by the "inform synthetically and the "inform analytically" indication.
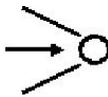
"**inform synthetically**" means departing from a larger sentence and proceed toward a summary related to a specific topic, identified as being the most relevant one emerging from the originating sentence.
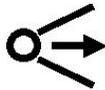
"**inform analytically**" means departing from a given sentence as to expand or add more information, which needs to be previously converted into the form of a sentence which is not available yet.

Reformulate: from Latin *reformo/reformulo*: change shape and shape again. It means changing the kind of information packaging, which was adopted before, and substituting a certain information request with a different one, still related to the same sentence. It may turn into a more or less radical transformation of the originating sentence, according to a precisely defined request or set of requests.

Express: from Latin *exprimo*: push out and press out. It means adding personal opinions and individual feelings related to facts and events within a certain sentence. It indicates the most subjective mode of information packaging, which is openly seen as bound to very personal evaluations, judgements and emotional states.

Instructional texts annotation turn taking symbols are meant to

define the mode of accessing and reading a paragraph or a sentence requested at each given time; they are suggested by the producer to be followed by the user.

They are the following ones:

Major Scale: it shows that literal interpretation is needed and that those pieces of text indicated and marked off, should be extracted and quoted literally, the way they were first intended to be.

Minor Scale: it shows that accurate interpretation may need a further process of abstraction and that pieces of texts indicated and marked off, may undergo significant reconfiguration processes, up to abstraction.

Open or Unsaturated Rhythm: it shows that accessing the text may lead the user toward incomplete interpretation of those facts and events, which are presented. It is meant to suggest that the user access more and various kinds of sources, which are made available.

Tight or Saturated Rhythm: it shows that accessing the text will lead the user toward complete interpretation of those facts and events, which are presented. It is meant to suggest that the user sticks to the interpretation provided, though access to other sources is still available, to support evidence.

Instructional texts annotation amplifier symbols come last and may be added only after the previously illustrated ones have been used; they apply to larger instructional documentation territories and indicate specific operations, which are to be performed to connect sets of instructions, which have been previously encoded and accurately stored.

They are the following ones:
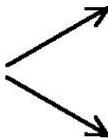
Choose: it is meant to represent the dynamic process of first identifying and then deciding between optional contexts for interpretation, which are mutually exclusive, given a certain set of instructions.
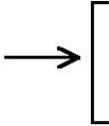
Identify: it is meant to represent a definition of a more specific context within a broader context, for interpretation of a set of instructions; it naturally occurs before "search" and "select".
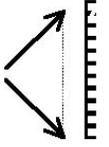
Search: it is meant to represent the dynamic process of choosing among different contexts for interpretation of a set of instructions, which are many and compatible as to find the most appropriate one.
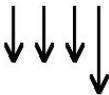
Select: it is meant to represent multiple contexts, which may evolve either synchronously or asynchronously and may be modified, once a certain decision making process has been performed, and stored and kept as an example.
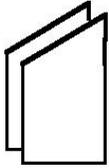
Copy/Replicate: it is meant to represent the dynamic process of duplication and repetition of a certain context, which, if lost, would affect understanding and accurate interpretation of a set of events and facts, described and explained by a set of instructions.

Ahead: it is meant to represent the progression of a certain set of instructions, which are linked together by context consistency or harmoniously shifting contexts.

Back: it is meant to represent the need to go back to delete and replace the originating context, which has been radically shifted, in the course of various transition states, such that, if not eliminated, would indeed affect consistent interpretation of a whole set of instructions.
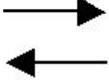
Conflict: it is meant to represent an emerging inconsistency and incompatibility between various context attributions to a set of instructions, which needs to be cleared as to proceed toward any further interpretation.

Instructional texts annotation amplifier symbols are meant to enhance significantly navigation throughout large corpora of texts as well.

3   Conclusions

Just like geographic maps only show those features, which become relevant according to the nature and purpose assigned to the map itself, the same "way of thinking" may be extended to technical texts according to various

packaging and repackaging priorities, in continuously shifting contexts of use and users' profiles and needs. Time dedicated to quite an expensive process, such as enhanced encoding is, may become time and cost effective, because each encoded instruction will provide an enormous amount of examples meant to effectively support any knowledge acquisition process.

In order to better clarify concepts and methodologies illustrated in the present paper the author will provide practical examples during her verbal presentation.

References

[1] Rimer, B. K., M. Conaway, P. Lyna, B. Glassman, K. S. H. Yarnall, I. Lipkus and L. T. Barber, *The impact of tailored interventions on a community health center population*, Patient Education and Counseling **37** (1999), pp. 125–140.

[2] Tonfoni, G., "Communication Patterns and Textual Forms," Intellect, Exeter, U.K., 1996.

[3] Tonfoni, G., "Information Design: The Knowledge Architect's Toolkit," Scarecrow Press, Lanham, Maryland, U.S., 1998.

[4] Tonfoni, G., *Intelligent control and monitoring of strategic documentation: a complex system for knowledge miniaturization and text iconization*, in: *Proceedings of the ISIC/CIRA/ISAS 98 Conference* (1998), pp. 869–874.

[5] Tonfoni, G., *On augmenting documentation reliability through communicative context transport*, in: *Proceedings of the 1999 Symposium on Document Image Understanding Technology SDIUT99*, Annapolis, Maryland, U.S., 1999, pp. 283–286.

Acknowledgments