NaviLire, Teaching French by Navigating in Texts

Lita Lundquist, Jean-Luc Minel, Javier Couto

NaviLire, Teaching French by Navigating in Texts

Lita Lundquist
Département de français
Institut F.I.R.S.T
Handelshøjskolen i København, CBS
ll.first@cbs.dk

Jean-Luc Minel
MoDyCO, UMR 7114 CNRS
200 Avenue de la République
92001 Nanterre
jean-luc.minel@u-paris10.fr

Javier Couto
Instituto de Computación
Herrera y Reissig 565
Montevideo
Uruguay
jcouto@fing.edu.uy

Abstract

In this paper, we describe *NaviLire*, a software based on text navigation to teach French as a foreign language. First, we discuss the design of reading knowledge. Second, we describe the text representation and its implementation framework. Finally, we present results from an experimentation, which is in progress.

Keywords: text navigation, teaching textual linguistics.

1 Introduction

In this paper, we describe *NaviLire*, a software used to teach French as a foreign language. The application is founded on *NaviTexte* [1], a workstation for visualizing and navigating in texts. The design of this workstation relies on two hypotheses. The first one makes a distinction between a text and the semiotic views of this text. The second hypothesis postulates that text reading is guided by knowledge which depends on the reader's competencies. Consequently, a declarative language has been defined in order to describe such knowledge [2].

With the *NaviLire* version of *NaviTexte*, we have conceived a tool that allows the student (reader) to perceive textual units that contribute to and maintain text coherence and to navigate between them in a text. We suppose that such a tool will be an outstanding didactic tool for teaching reading of foreign language texts, as well as producing written texts in the foreign language.

The remainder of this paper addresses the issues listed earlier. In the next section, we discuss the design of reading knowledge. The third section describes the text representation and the fourth section details its implementation framework. The final section describes the *Navilire* application and some first results.

2 Designing Reading Knowledge

The design of the *NaviTexte* workstation is built on the conclusions from the Regal Project [3, 4]. In that project, the purpose was to answer the following questions: What must be shown of a text, and how? Several works [5-7] emphasized that visualization should be indicative, in order to bring out parts of text which interest a reader, and informative to give him information about these parts of text. In addition, in keeping links between all these parts, the software designed for *Regal* [4] gives the possibility to move dynamically from one kind of information to another. This has been called "track" of textual navigation [8, 9]. Thus, for a text, potentially many textual navigation tracks exist and it is possible to compare them with ambulatory progression in hypertexts [10]. Nevertheless, our navigation principles differ from those in hypertexts in that they rely on semiotic and linguistic marks of text. As a consequence, the

navigation is not guided by the author, compared to hypertext navigation where s/he has to determine links; instead, the navigation is the result of an interpretation process relying on textual annotations.

Conceptually, like in a hypertext approach, navigation is an operation, which links a text unit (TU) called the source, to another one, called the target. But, in our approach, it is possible to specify for the *source* and the *target*, several *conditions* and a *span* of text (cf. Fig. 1).

Consequently, before processing a navigation operation, conditions on attributes of the source are checked. A simple condition expresses constraints on values of the annotations of TU (see section 3). The span limits the scope of the search of the target. Furthermore, each operation must specify a type of moving by using one of these pre-definite instructions: {*First, Last, Forward(i), Backward(i)*}. *First, Last,* indicates that the search of the target is absolute: the TU displayed will be the first (respectively the last) TU, in the specified span, which matched the conditions. *Forward(i), Backward(i),* indicate that the search is carried out relatively to the source (before or after) and indexed by the integer *i*. For example, {*Forward(3)}* is interpreted as the search of the third TU located after the source, provided that its attributes match the conditions. The value of the span must be chosen in the set declared in the *Head* of the text. Complex conditions must be expressed by using Boolean operators *{AND, OR, NOT}* or hierarchical ones *{IS_DESCENDANT, IS_ASCENDANT, IS_INCLUDED}* between simple conditions.

```
<Op_Nav Titre="Going to the next referent"
Type="Forward">
<Source>
<Condition Type="Simple">
<UT Type="Phrase">
<Attribut Nom="Referent">REF1</Attribut></UT>
</Condition>
</Source>
<Cible>
<Condition Type="Simple">
<UT Type="Phrase">
<Attribut Nom="Referent">REF1</Attribut></UT>
</UT></Condition></Cible>
</Op_Nav>
```

Figure 1: Example of a navigation operation

But there were some limitations in the first version of this language. For example, to express navigation operations between different lexical chains in a text (going from a referent to the next one), it is necessary to write one operation for each lexical chain although all these operations correspond to the same principle of navigation: going from a TU to another TU which has the same name of attribute (Referent) but different values (REF1, REF2, etc.).

Consequently, this first version has been upgraded [11] with the possibility to express conditions on the relation between the source and the target. A new tag <Relation_Source_Target> allows expressing only relations between the values of attributes. So far, operators equality(=), difference (!=), contains (~) have been defined. With these extensions, all the operations on lexical chains, described above, are replaced by one operation (cf. Fig. 2),

```
<Op_Nav Titre="Going to the next referent"
Type="Forward">
<Source>
<Condition Type="Simple">
<UT Type="Phrase">
<Attribut Nom="Referent" </Attribut></UT>
</Condition></Source>
<Cible>
<Condition Type="Simple">
<UT Type="Phrase">
<Attribut Nom="Referent</Attribut></UT>
</UT>
</Condition></Cible>
<Relation_Source_Target>
 Source.Referent = Cible.Referent
</Relation_Source_Target>
</Op_Nav>
```

Figure 2: Example of a navigation operation using upgraded language

3 Text Representation

The conceptual modeling is inspired by [12]: “A system for capturing documents structure should be flexible enough to accommodate the variations in structure that occur naturally” and improved by proposals from [13, 14]. As a consequence, our text model is based on typed units, which can be embedded. The description of a text is made up of two parts the *Head* and the *Body*. In the *Body*, typed units (TU) are marked up using XML format and it is possible to embed them (cf. fig. 3).

```
<UT Type="Proposition" Nro="1" >
<UT Type="SN" Nro="4" >
<Attribut Nom="Type Referent">RD</Attribut>
<Attribut Nom="Indice Referent"> 1 </Attribut>
<Chaine> le chef de l'Etat </Chaine> </UT>
<UT Type="SV" Nro="1" >
<Attribut Nom="Modalité"> Valorisée </Attribut>
<Attribut Nom=" Valeur Modalité"> Négative
</Attribut>
<Chaine> n'a pas su trouver un geste de réconfort ni un
mot de compassion pour </Chaine>
</UT>
……..
```

Figure 3: Example of text annotations

Each unit has one type and an unlimited number of attributes. Only innermost units have an attribute, named *Chaine*, which tags the string of character. Nevertheless, this kind of tagging has some well-known limitations: for example, impossibility to tag overlapping or discontinuous units.

Even so, X-Link or Xpointer offer to solve such problems, their utilization by non experts users is far too complex, consequently, it is possible to define new elements in the *Head*, composed with existing TU described in the *Body* part. To refer to an existing TU, the principle is to use a reference of the TU as it is defined in the *Body*. Four types of new elements can be created: *Set*, *Sequence*, *Reference* and *Graph*.

A *Set* is a set of TU for which there exists a relation of equivalence from the point of view of the annotator. For example, TU with different part of speech attributes can express a same topic (like verb and noun phrase), so a *Set* is the perfect structure to express that.

A *Sequence* is an ordered set of TU for which a relation of syntactic or semantic cohesion exists. For example, in the *Body* it is not possible to tag discontinuous verbal syntagm, like "sont stockées" in the expression "ne sont pas stockées". A *Sequence* allows to correctly tag such discontinuity.

A *Reference* defines an oriented relation between two TU and one navigation operation is associated with this object. This operation goes from the referred to the referent. The representation of the link between a pronominal anaphora and its referent is a typical example of the utilization of a *Reference*.

The last element, a *Graph* is used to build multiple relations between TU. This structure corresponds exactly at the mathematical notion of graph where nodes, which represent TU, are connected by oriented arcs, which represent relations between these TU. From our point of view, this complex structure could be very useful to represent complex discourse structures like coherence tracks (see § 5) or a thematic index like those put at the end of books[1].

4 General Software Organization

NaviTexte consists of several sub-systems [1]. The first one builds a text representation {Ta} from a text annotated by dedicated software [16, 17]. A second sub-system loads and compiles one or several cartridges, which describe the visualization and navigation knowledge (cf § 3) as well as the form of the display (linear, structured, graphical, etc.). The result of such compilation is a graph of potential tracks. This graph is projected on the text {Ta}. A third sub-system displays the text on the screen by applying the semiotic forms chosen by the reader. It must be pointed out that a reader has the possibility to load another cartridge of reading knowledge at any time and that its compilation is dynamically computed.

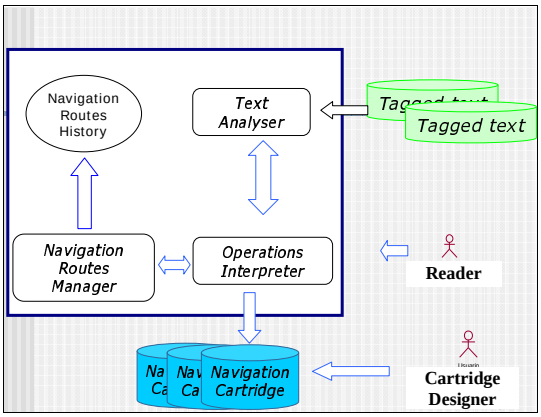


Figure 4: General software organization

5. NaviLire, Application in Text Linguistics

5.1 Teaching Text Linguistics

For the past thirty years, text linguistic researchers have worked on describing linguistic marks of textual coherence in order to bring out principles of text structuring [18]. A set of concepts and models of textual interpretation has been worked out, including for example,

[1] See [15] for more details.

anaphors, connectors, mental spaces, etc. In particular, these studies have shown that even for languages apparently close like French and Danish, texts are not organized in the same way [19]. Consequently, text linguistics has important implications in foreign language teaching, especially from a contrastive point of view, when language pairs are analyzed through texts used in authentic communication situations. It seems that the teaching of text linguistics contributes to sharpen the attention of students towards the building of well-formed texts and to stimulate their own production of texts. Consequently, a tool that allows the student (reader) to perceive textual units that contribute to and maintain text coherence and to navigate between them in a text, can be supposed to be an outstanding didactic tool for teaching reading of foreign language texts, as well as producing written texts in the foreign language.

In the reading process, a reader has to cope with two basics types of cognitive problems. First, identifying discursive referents in a text and choosing correct relations between noun phrases that refer to them. In other words, the reader has to decide between a co-reference relation, in which there is only one referent, and a referential disjunction, in which there are several referents. This cognitive competence is crucial for the building up of a coherent mental representation of the text and hence central in the learning process: "learning from text requires that the learner constructs a coherent mental representation of the text" [20:307]. Second, identifying the function and orientation intended by the sender. This orientation is generally marked from the beginning of the text and consequently acts as an "interpretation program" [21]. Identifying this orientation, which is essentially provided by the predications, is also crucial for a correct deciphering of the semantic and pragmatic coherence.

As a general rule, textual coherence can be divided into three types, viz. the referential, the predicative and the pragmatic coherence, [18, 22], which again are based on the three speech acts: reference, predication, and illocution [23]. In actual text navigation, however, the specific units to be focused on depend on the characteristics of the text in question.

Thus the title of the text example below *L'amnistie fiscale n'est pas immorale!* starts, by means of the negation and the expressive content and sentence form, an interpretation program of the argumentative type, in which participate two "voices", that of the protagonist and that of the antagonist, who would claim to the contrary, namely that *l'amnistie fiscale est immorale*. These two voices initiate two 'tracks' of coherence, which can be identified and linked in the navigation process of NaviLire, together with other argumentative markers. It is also possible to navigate along entities based on discourse referents, such as *la taxe*, which is maintained via anaphoric expressions such as *une telle mesure*, *la mesure envisagée par M. Raffarin*, which again lend themselves to ample linguistic observations upon differences with e.g. means of anaphorising in Danish. And last but not least, the frequent uses of *le conditionnel* could be identified and commented upon.

Example :
« L'amnistie fiscale n'est pas immorale!
Le gouvernement de Jean-Pierre Raffarin étudie l'opportunité d'instituer une taxe sur les fonds placés à l'étranger et rapatriés en France. Le produit de cette taxe financerait le plan que vient de dévoiler son ministre de la Cohésion sociale. De nombreux responsables politiques ont manifesté leur hostilité à une telle mesure. Elle serait inefficace et, surtout, immorale car elle blanchirait les «criminels en col blanc». […] Une telle somme ne contribuerait, il est vrai, que modestement au financement du plan de cohésion sociale dont le coût, en cinq ans, se monterait à 13 milliards d'euros. Mais faut-il pour autant déclarer la mesure envisagée par M. Raffarin inefficace ? Non, car d'autres avantages en résulteraient. Des capitaux importés, qu'ils soient prêtés ou investis, créent des emplois. » [Le Figaro, le 16 juillet 2004].[2]

In a subsequent task, students can be, and have been, asked to reproduce, in writing an essay, these features which are characteristic of an argumentative text in French.

5.2 Results

So far, *NaviLire* has been put into practice on a small scale only, viz. in the teaching of French

[2] This text forms part of the first text sample annotated and encoded in NaviLire.

texts and text linguistics to Danish language students at the 4th year of the studies of Language and Communication at the Copenhagen Business School. A pilot experiment was carried out in order to evaluate the effects of using the program. In this experiment, all students (14 subjects) were, in a first stage, pre-tested in a reading experiment, using the same "paper and pencil" method, in order to determine their general level of reading competence (which turned out to be quasi-similar for all subjects). In a second stage of the experiment, half of the students (7 subjects), performed a new reading experiment with pencil and paper (the "papiristes"), while the other half (the "navilistes") read the same text with *NaviLire*[3]; this last condition included guided 'navigations' (cf. fig. 5) along the 'tracks' of 1) the principal theme, 2) the argumentative operators, and 3) the personal pronoun *nous*, as well as a visualization of all these three features (cf. fig. 6)[4].

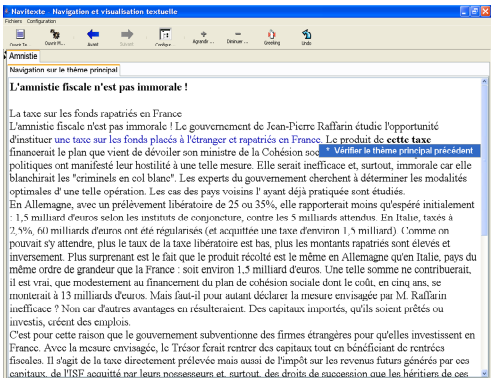


Figure 5: Example of a navigation operation in action

All subjects, encompassing both conditions, answered the same questions (paper and pencil), some of which addressed phenomenological perspectives ("How difficult is the text to read/remember, on a scale from 1-5"); others tested the number of terms remembered ("Quote as many terms you remember that refer to X", or "Which of the following terms appeared in the text: x, y …n"); and yet others aimed at evaluating the subjects' general understanding of the text ("What type of text is this?", "How do you argue for this characterization?").

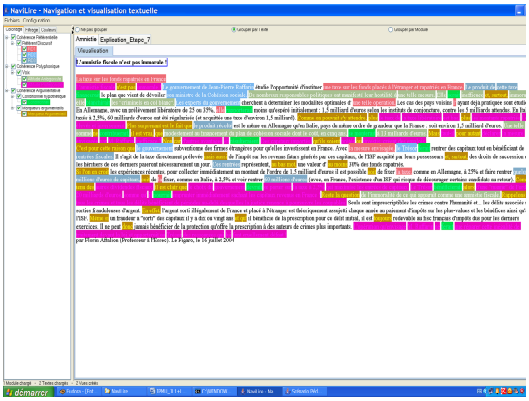


Figure 6: Example of final visualization

The first results (cf. table 1) are based on forty answers, of which 35 concern questions about the content of the text. These results show that the "navilistes" have a better comprehension performance than the "papiristes" for 14 questions, an identical performance for 16 other questions, and a poorer performance for 5 questions.

Table 1: Comparison of "navilistes" and "papiristes"

	NUMBER OF QUESTIONS	**PERCENTAGE**
"Navilistes" better than "Papiristes"	14	40
"Navilistes" the same as "Papiristes"	16	45,7
"Navilistes" worse than "Papiristes"	5	14,3
Total	35	100

The qualitative results in table 2, show that as far as the phenomenological questions is concerned, there is no difference between the two conditions: "Papiristes" and "navilistes" judge the text in approximately the same way, i.e., evaluate it to be equivalently easy to read (an average of 2,6 vs. 2,7) and remember (an average of 2,7 vs. 2,9); for the number of correct questions ("Which of the following terms appeared in the text: x, y …n"), the "navilistes" performed better than the "papiristes" (an average of 32,3 correct answers vs. 29,9), as

[3] The students had performed approx. 1 ½ hours training Navilire in beforehand.

[4] Larger vizualisations are shown at: http://infolang.uparis10.fr/modyco/membres/textes.cfm?IDchercher=350, in "References et Textes".

they did with the question “Reproduce as many argumentative operators as you remember” (an average of 4 vs. 2,6). The “papiristes” were, however, able to reproduce the highest number of terms referring to the theme (an average of 7 vs. 4,3).

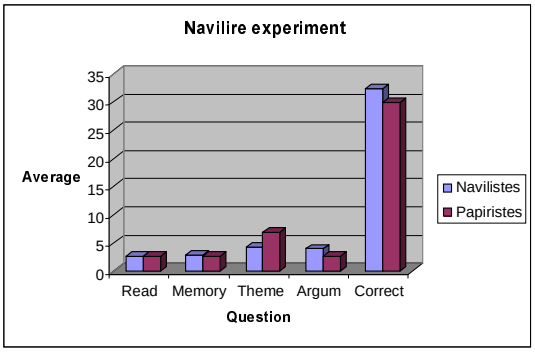
This last result is linked to a more remarkable difference, which was found in the qualitative analysis of the answers, showing that the papiristes seem to have perceived the general structure of the text better than the “navilistes”. Indeed, out of seven “papiristes”, four referred to the over-all structure of the text, which was argumentative, compared to none among the “navilistes”. Two examples of the “papiristes”’ answers to the questions: “What type of text is this?” and “How do you argue for this characterization” show this effect:

1. “Argumentative text type (because of) the structuration of the text : problem : What to do ? -> exposition of consequences -> conclusion with perspectives: “We go on as usual”.”
2. “Argumentative text type (because):
    - the first paragraph sets the agenda
    - the second paragraph functions as an argument/”an eye-opener”
    - and the last paragraph presents a warning and a moral.”

These differences found between “papiristes” and “navilistes” as far as the general understanding of the text and its structure is concerned, can, with some caution, be said to be in line with earlier psycholinguistic findings to the effect of e-learning programs [24-25].

The general conclusions to be drawn from the experiments is first that the effects of the *NaviLire* workstation needs a further and more refined experimentation in order to seize cognitive differences stemming from the two different teaching and reading conditions; and second, that it would be useful to reconsider the interface and general design of the *NaviLire* program in the light of the qualitative result mentioned above. In fact, it would be particularly constructive to make up for the apparent inconveniences as to missing the general lines of comprehending texts in e-learning programs, by building up some adaptative semiotic views.

Table 2: Partial qualitative results from the NaviLire experiment



References

[1] J. Couto, J.-L. Minel. (2004). Outils dynamiques de fouilles textuelles, *Actes de RIAO 2004,* pages 420-430, Avignon, France, 2004.

[2] J. Couto (2006). NaviTexte, plate-forme de visualisatiuon et de navigation textuelle, PhD, (in progress), Université Paris-Sorbonne, France, 2006.

[3] O. Ferret, B. Grau, J.-L Minel, S. Porhiel (2001). Repérage de structures thématiques dans des textes, *Actes de TALN 2001,* pages 163-172, Tours, 2001.

[4] J. Couto, O. Ferret, B. Grau, N. Hernandez, A. Jackiewicz, J.-L. Minel, S. Porhiel (2004). RÉGAL, un système pour la visualisation sélective de documents, *Revue d'Intelligence Artificielle*, vol 18, pages 481-514, Hermès, 2004.

[5] H. Saggion, G. Lapalme (2000). Concept Identification and Presentation in the Context of Technical Text Summarization, *RIAO 2000*, Paris, France, 2000.

[6] M.-Y. Kan, K. McKeown, J. Klavans (2001). Domain-specific informative and indicative summarization for information retrieval. *Proceedings of the first Document Understanding Conference*, pages 19-26, New Orleans, 2001.

[7] C. Boguraev, R. Kennedy, S. Bellamy, Y. Brawer, J. Wong, T. Swartz. (1998). Dynamic Presentation of Document content for Rapid On-Line Skimming. *AAAI Spring Symposium on Intelligent Text Summarisation,* Stanford, CA, 1998.

[8] G. Crispino, J. Couto (2004). Construction automatique de résumés. Une approche dynamique. *TAL*, Vol 45/1, pages 95-120, Hermès, Paris, 2004.

[9] J.-L. Minel (2005). Réflexions autour de l'identification, la modélisation et la visualisation de certaines organisations textuelles. I/n *L 'unité texte*, S. Porhiel et D. Klingler (eds), pages 231-250, Editions Perspectives, (2005).

[10] M. Gery. (2002). Un modèle d'hyperdocument en contexte pour la recherche d'information structurée sur le Web. *Revue des Sciences et Technologies de l'Information*, 7, pages 11-44, Hermès, Paris, 2002.

[11] J. Couto, J.-L. Minel (2005). Enrichissement du langage de navigation dans NaviTexte. *Rapport technique du MoDyCo,* France, 2005.

[12] L. Clarke, G.V. Cormack, F. J. Burkowski (1995). An Algebra for Structured Text Search and a Framework for its Implementation. *The Computer Journal*, Vol 38, n°1, 1995.

[13] G. Crispino (2003). Une plate-forme informatique de l'Exploration Contextuelle : modélisation, architecture et réalisation (ContextO). Application au filtrage sémantique de textes. PhD, Université Paris-Sorbonne, 2003.

[14] Text Encoding Initiative Website: http://www.tei-c.org.

[15] J. Couto, J.-L. Minel. (to appear in 2006). SEXTANT, un langage de modélisation des connaissances pour la navigation textuelle, *ISDD06,* 10 pages, Caen, France, 2006.

[16] J.-L Minel, E. Cartier, G. Crispino, J.P. Desclés, S. Ben Hazez, A. Jackiewicz (2001). Résumé automatique par filtrage sémantique d'informations dans des textes, Présentation de la plate-forme FilText. *Technique et Science Informatiques*, n° 3, pages 369-396, Hermès, Paris, 2001.

[17]. F. Bilhaut, M. Ho-Dac, A. Borillo, T. Charnois, P. Enjalbert, A. Le Draoulec, Y. Mathet, H. Miguet, M.P. Pery-Woodley, L. Sarda (2003) ; Indexation discursive pour la navigation. intradocumentaire : cadres temporels et spatiaux dans l'information géographique. *Actes de TALN*, pages 315-320, Batz-sur Mer, 2003.

[18] L. Lundquist (1980). *La cohérence textuelle, syntaxe, sémantique, pragmatique*, Copenhagen, Nordisk Forlag, 1980.

[19] L. Lundquist (2005). Noms, verbes et anaphores (in)fidèles. Pourquoi les Danois sont plus fidèles que les Français. *Langue française*. Vol 145. pages 73-92, 2005.

[20] W. Kintsch (2003). *Comprehension. A Paradigm for Cognition*. Cambridge, Cambridge University Press, 1998/2003.

[21] L. Lundquist (1993). La Cohérence textuelle argumentative, illocution, intention et engagement de consistance. *Revue québécoise de linguistique*, vol. 22-2, pages 109-138, 1993.

[22] L. Lundquist (2000). Knowledge, events and anaphors in texts for specific purposes. L. & R. Jarvella (eds.), *Language, text, and knowledge. Mental models of expert communication*. Mouton de Gruyter, Berlin, 2000.

[23] J. Searle (1969). *Speech Acts, An Essay in the Philosophy of Language*. Cambridge, Cambridge University Press, 1969.

[24] P. Coirier, D. Gaonac'h, J.-M. Passerault (1996). *Psycholinguistique textuelle. Approche cognitive de la compréhension et de la production des textes.* Armand Colin, Paris, 1996.

[25] T. Baccino (2004). *La lecture électronique.* Presses Universitaires de Grenoble, Grenoble, 2004.