



Utilisation du standard SGML pour la représentation de bandes dessinées

Valérie Mapelli

► To cite this version:

Valérie Mapelli. Utilisation du standard SGML pour la représentation de bandes dessinées. [Travaux universitaires] 1994, pp.76. inria-00538332

HAL Id: inria-00538332

<https://inria.hal.science/inria-00538332>

Submitted on 22 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

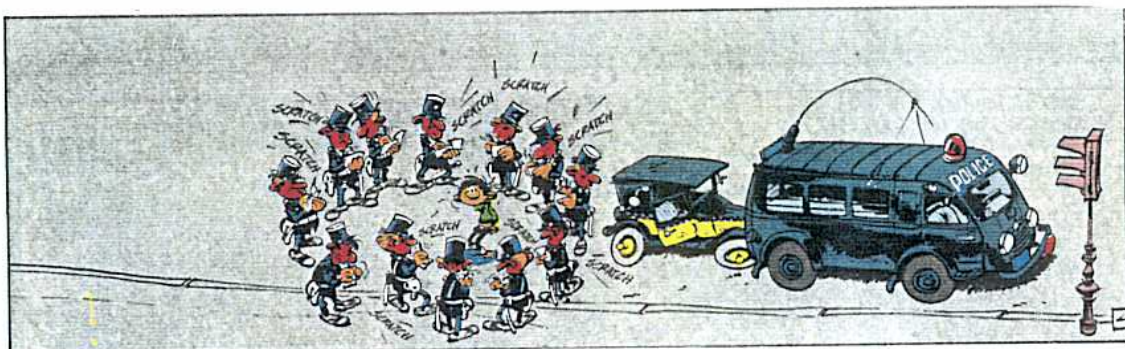
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DESS "Industries de la Langue"
Faculté de Lettres et Sciences Humaines
Université de Metz

PROJET - DESS "Industries de la Langue"

Utilisation du Standard SGML
(Standard Generalized Markup Language)
pour la Représentation des
Bandes Dessinées

Auteur: Valérie MAPELLI
Responsable du projet: Laurent ROMARY



Années 1994-1995

AVANT-PROPOS

Ce projet, intitulé *Utilisation du Standard SGML pour la Représentation des Bandes Dessinées*, fait partie intégrante du diplôme "DESS Industries de la Langue" que j'ai effectué à la Faculté de Lettres et Sciences Humaines de l'Université de Metz. Il aura contribué à l'obtention de ce diplôme.

Ce document a essentiellement un but initiatique quant à la compréhension du format SGML (Standard Generalized Markup Language) et de l'usage que l'on peut en faire. L'étude qui s'ensuit a pour objet le codage de la bande dessinée au moyen de l'outil SGML.

Je remercie Monsieur Laurent Romary qui a contribué à l'élaboration de ce projet et qui m'a donné les conseils utiles et nécessaires au bon déroulement de celui-ci.

Valérie MAPELLI.

SOMMAIRE

1. INTRODUCTION	5
2. SGML	7
2.1. En bref	7
2.2. Spécificité de SGML	7
2.2.1. Le balisage descriptif	7
2.2.2. Le type de document	8
2.2.3. L'indépendance des données	8
2.3. Les structures SGML	8
2.3.1. Les éléments	8
2.3.2. Les modèles de contenu	9
2.3.3. Les structures de document: les DTD	10
2.3.3.1. Les règles de minimisation	11
2.3.3.2. Le modèle de contenu	11
2.3.3.3. Les indicateurs d'occurrences	12
2.3.3.4. Les opérateurs de connexion	12
2.3.3.5. Les groupes de modèle	12
2.3.3.6. Informations supplémentaires	13
2.3.3.6.1. Exceptions au modèle de contenu	13
2.3.3.6.2. Les structures concurrentes	13
2.3.4. Les attributs	15
2.3.5. Les entités	17
2.3.6. Le balisage des sections	17
2.4. Ce qu'il faut savoir du schéma de codage de la TEI	18
3. LES PROBLEMES POSES PAR LA BANDE DESSINEE EN VUE D'UN CODAGE SGML	21
3.1. La BD en tant que document illustré	21
3.1.1. Définition	21
3.1.2. Historique	21
3.2. Les composants d'une BD	23
3.2.1. Les données graphiques	24
3.2.1.1. Les cadres	24
3.2.1.2. Les phylactères	26
3.2.2. Les données textuelles	28
3.2.2.1. Les indications temporelles et scéniques	28
3.2.2.2. Les onomatopées	29
3.2.2.3. Les symboles	29
3.2.2.4. Le contenu des dialogues dans les bulles	30
3.2.2.5. Autres données textuelles	31
3.2.2.6. Les notes infrapaginales	32

4. PROPOSITION DE CODAGE D'UNE BD SOUS FORMAT SGML	33
4.1. Codage des différentes parties d'une BD	33
4.2. Codage des caractères spéciaux	36
4.3. Les moyens nécessaires à l'implémentation de la BD sous format électronique	38
5. CONCLUSION	41
SOURCES DOCUMENTAIRES	43
TABLE DES ANNEXES	47

1. INTRODUCTION

Depuis que les systèmes informatiques offrent une forme nouvelle de compilation d'informations, ceux-ci se veulent être un support de taille pour toutes données textuelles, sonores, imagées, etc. Les versions électroniques de textes imprimés se sont révélées extrêmement utiles de par les avantages qu'elles proposent à leurs utilisateurs, entre autres:

- recherche par mot, phrase,
- recherche par oeuvre, genre, date, auteur,
- disponibilité des données directement depuis le lieu de travail ou le lieu de logement.

Les textes électroniques n'ont pas fonction de remplacer les textes imprimés, mais plutôt d'offrir un autre type d'accès et une autre méthode d'exploration des documents. Cependant, certaines difficultés doivent être gérées avant de parvenir à échanger ces précieuses archives à un niveau autre que local.

La diversité des réseaux informatiques, liée aux différences de conception de ceux-ci en sont un parfait exemple. Notamment, le choix d'une représentation des données unique forme l'un des obstacles à surmonter. Une telle représentation rendrait plus aisé le transfert des versions électroniques d'un programme ou d'une application informatique à une autre.

Confrontés à un nombre croissant de systèmes de codage appliqués à divers documents (théâtre, roman, poésie, etc.), un groupement d'associations linguistiques et scientifiques s'est accordé pour définir un ensemble de directives sous le label de TEI (*Text Encoding Initiative*), visant à développer un format standard facilitant ce transfert. Ces directives (*guidelines*), doivent fournir des outils permettant de représenter explicitement les parties d'un texte nécessitant une identification à des vues de traitement informatique.

Dans ce désir d'uniformiser, le choix d'un métalangage unique s'est porté sur le format SGML (Standard Generalized Markup Language). La norme SGML est en fait un schéma de codage permettant de structurer un texte électronique à l'aide de balises pouvant contenir des informations métatextuelles diverses (découpage en paragraphes ou phrases, caractéristiques bibliographiques du texte, commentaires, etc.).

La TEI, conjuguée à l'utilisation du format SGML semble relativement bien s'appliquer à des formes "courantes" de littérature, mais elle reste encore vague et incomplète en ce qui concerne des formes annexes telles que la bande dessinée. L'objectif de ce projet est donc de spécifier, sur la base des indications déjà fournies par la TEI, un ensemble de pratiques pour le codage des bandes dessinées en SGML.

Le chapitre 2 est une initiation à la norme SGML et aux TEI Guidelines.

Le chapitre 3 traite des particularités présentées par la bande dessinée, en comparaison avec d'autres documents textuels.

Le chapitre 4 présentera les résultats de l'étude quant au codage des bandes dessinées sous format SGML.

2. SGML

Avant d'aborder les difficultés que peut poser la représentation des bandes dessinées sous format électronique, il est nécessaire de présenter le standard de balisage qui sera utilisé à cet effet, en l'occurrence, la norme SGML. Nous verrons dans un premier temps l'évolution de son usage, puis les différentes possibilités qu'elle nous propose. Par ailleurs, une partie sera réservée à l'introduction à divers schémas de codage normalisés dans la TEI.

2.1. En bref

SGML est l'abréviation de *Standard Generalized Markup Language* (pour langage normalisé de balisage généralisé). Ce langage, ou plutôt métalangage, est un standard international, défini dans *ISO 8879: Information processing --- Text and office systems --- Standard Generalized Markup Language (SGML)*, (Genève, 1986). Il a pour but de décrire un texte électronique au moyen de règles de balisage.

Le balisage, aussi appelé étiquetage, décrivait à l'origine les annotations et autres marques de distinction apparaissant dans un texte afin d'informer les personnes intéressées sur la manière de présenter ou d'imprimer le texte. Par exemple, pour indiquer qu'une phrase devait être représentée en caractère gras, on la soulignait. Ce système a été conservé et amélioré lors de l'informatisation des textes, en créant toutes sortes de balises insérées dans les versions électroniques afin de permettre formatage, impression ou autre type de traitement [réf.1].

Un langage de balisage n'est autre qu'un ensemble de balises conventionnées, utilisées en combinaison pour le codage des textes. Il doit préciser quelles balises sont autorisées, lesquelles sont nécessaires, la manière dont elles se distinguent du texte et leur signification. Un document tel que les TEI Guidelines [réf.2] se propose de s'occuper de la signification des balises. SGML offre les moyens de réaliser le reste.

2.2. Spécificité de SGML

2.2.1. Le balisage descriptif

Contrairement à d'autres langages de balisage fonctionnant de manière procédurale, SGML est un système descriptif. Cela signifie que les codes utilisés sont représentés par des noms identifiant les différentes parties d'un document, comme la balise **<para>** signifiant *"ce qui suit est un paragraphe"*.

Un balisage procédural définit le type de traitement qui doit être appliqué à un point particulier du texte: *"nommer une procédure PARA avec les paramètres 1, b et x"* ou *"déplacer la marge gauche de 2 cases vers la gauche"*, etc.

2.2.2. Le type de document

Tout document est qualifié d'un type. Un type de document est défini par les parties qui le constituent, ainsi que leur structure.

Par exemple, un rapport sera composé d'un titre et d'un auteur, suivis d'un extrait et d'une séquence d'un ou plusieurs paragraphes. D'après cette définition, tout document faisant abstraction de l'un ou l'autre de ces éléments ne pourra être considéré comme un rapport.

Lorsque les documents sont jugés conformes (c'est-à-dire de type connu), alors un programme spécial appelé *parseur* permet de les traiter directement, après avoir vérifié que tous les éléments requis d'un type particulier sont effectivement présents.

2.2.3. L'indépendance des données

L'un des buts principaux de SGML est de faciliter le transfert des documents codés d'un environnement informatique à un autre sans risque de perdre des informations et conserver ainsi une certaine homogénéité de la nomenclature. A ces fins, SGML fournit un mécanisme de substitution indépendant du support machine, permettant de remplacer un ensemble de caractères par un autre. Ces ensembles sont connus sous le nom d'entités.

2.3. Les structures SGML

Un texte sous format SGML est composé d'éléments et de modèles de contenu.

2.3.1. Les éléments

Comme vu précédemment, un texte n'est pas seulement une séquence de mots, mais il est aussi structuré sous forme de titres, de paragraphes, de dates, etc., qui délimitent chaque séquence de caractères et de mots.

Des éléments sont ces constituants structurels auxquels on associe un nom de type **<nom>**. A partir de ces éléments, on peut alors définir une grammaire textuelle, sorte de schéma décrivant au moins trois aspects du morceau de texte considéré: son "étendue", c'est-à-dire les points bornant le début et la fin de la séquence, son "type", correspondant à la catégorie à laquelle appartient la séquence et enfin son "contexte", indiquant la catégorie à laquelle les instances d'objet sont associés.

Par exemple, un élément entre guillemets sera représenté par le type **quote** inséré au début et à la fin de la citation:

... Rosalind's remarks **<quote>** This is the silliest stuff that ere I heard of **</quote>** clearly indicate...

Afin de ne pas confondre ces éléments de démarcation avec le reste du document, on utilise une syntaxe dite "concrète". Elle est formée de caractères particuliers, distincts des éléments de structure (syntaxe "abstraite"). Si l'on reprend l'exemple précédent, la balise marquant le début de la séquence est exprimée par la **<quote>** où les équerres "<" et ">" indiquent respectivement le début et la fin de la balise et **quote** identifie l'élément délimité. La balise **</quote>** spécifie la fin de l'élément avec le caractère "/" rendant possible la différenciation entre la balise ouvrante et la balise fermante.

2.3.2. Les modèles de contenu

Un élément de texte est codé à l'aide d'étiquettes de début et de fin qui spécifient la catégorie ou le type d'élément dont il s'agit au moyen d'un "nom" particulier. Le contenu d'un élément d'un type particulier (e.g. **para**) peut être vide. De manière générale, les éléments d'un certain type contiendront ou seront eux-mêmes contenus dans des éléments d'un type différent. Nous aborderons dans une autre section l'un des atouts de SGML qui est sa capacité à former des règles indiquant comment ces éléments s'imbriquent les uns dans les autres.

Prenons en guise d'exemple la représentation SGML d'une anthologie qui est constituée de l'élément **anthology** (pour anthologie), auquel peuvent correspondre les éléments **title** (pour titre), **poem** (pour poème ou poésie, comprenant également un élément **title**), **stanza** (pour strophe) et **line** (pour ligne). Ainsi, *L'Albatros* de Charles Baudelaire [Baudelaire 89], codé sous SGML se présentera sous une forme comme:

```

<anthology><title>Les Fleurs du Mal</title>
  <poem><title>L'Albatros</title>
    <stanza>
      <line>Souvent, pour s'amuser, les hommes d'équipage</line>
      <line>Prennent des albatros, vastes oiseaux des mers,</line>
      <line>Qui suivent, indolents compagnons de voyage,</line>
      <line>Le navire glissant sur les gouffres amers.</line>
    </stanza>
    <stanza>
      <line>A peine les ont-ils déposés sur les planches,</line>
      <line>Que ces rois de l'azur, maladroits et honteux,</line>
      <line>Laissent piteusement leurs grandes ailes blanches</line>
      <line>Comme des avirons traîner à côté d'eux.</line>
    </stanza>
    <stanza>
      <line>Ce voyageur ailé, comme il est gauche et veule!</line>
      <line>Lui, naguère si beau, qu'il est comique et laid!</line>
      <line>L'un agace son bec avec un brûle-gueule,</line>
      <line>L'autre mime, en boitant, l'infirme qui volait!</line>
    </stanza>
    <stanza>
      <line>Le Poète est semblable au prince des nuées</line>
      <line>Qui hante la tempête et se rit de l'archer;</line>
      <line>Exilé sur le sol au milieu des huées,</line>
      <line>Ses ailes de géant l'empêchent de marcher.</line>
    </stanza>
  </poem>
  <!-- more poems go here -->
</anthology>

```

On remarque donc qu'une anthologie est constituée d'un titre et d'une série de poèmes eux-mêmes comprenant un titre et une série de strophes découpées en un certain nombre de lignes. Précisons que cet exemple n'est pas conforme aux documents de la TEI. De plus, la ligne **<!-- more poems go here -->** est un commentaire SGML et ne peut être traité en tant qu'élément du texte. D'autre part des règles en rapport avec les différents types de document doivent être explicitement précisées. C'est l'une des propriétés des TEI Guidelines.

2.3.3. Les structures de document: les DTD

Ceci est une introduction à l'expression de la structure de document sous SGML, connue sous le sigle DTD, abréviation de *Document Type Definition* (pour définition de type de document).

Une DTD est comparable à une grammaire: elle a pour objectif de réguler le langage de balisage en établissant un certain nombre de définitions propres aux éléments d'un type de document donné. Elles se présentent sous la forme d'une syntaxe simple recueillant un ensemble de déclarations en SGML.

ex:

<!ELEMENT anthology	- - (poem+)>
<!ELEMENT poem	- - (title?, stanza+)>
<!ELEMENT title	- O (#PCDATA)>
<!ELEMENT stanza	- O (line+)>
<!ELEMENT line	O O (#PCDATA)>

Les déclarations, comme les éléments, sont délimitées par des équerres "<" et ">"; le premier caractère qui suit l'équerre ouvrante est obligatoirement un point d'exclamation "!", immédiatement suivi de l'un des ensembles de mots-clés définis dans SGML, spécifiant le type d'objet à déclarer. Les cinq déclarations de notre exemple sont toutes du même type:

- elles commencent toutes par le même mot-clé ELEMENT qui introduit un élément,
- elles sont toutes constituées de trois parties: un nom ou groupe de noms, deux caractères précisant les règles dites de "minimisation" (cf. section 2.3.3.1.) et un modèle de contenu.

Les différents composants de la déclaration sont séparés par un espace (un ou plusieurs blancs, tabulations ou passages à la ligne).

2.3.3.1. Les règles de minimisation

Ces règles, formées de deux caractères séparés par un blanc, précisent la présence ou le caractère optionnel d'une balise en début ou en fin d'un élément donné. Le premier caractère réfère à la balise initiale et le second à la balise finale. Dans les deux cas, ils sont représentés soit par un tiret "-" soit par la lettre "O" (pour "optionnel"), correspondant respectivement à la présence et l'omission de la balise.

2.3.3.2. Le modèle de contenu

Les modèles de contenu sont indiqués entre parenthèses et ont pour objet de spécifier les occurrences de l'élément contenu. Ce sont soit d'autres éléments, soit des mots dits "réservés". Le mot réservé le plus usité est #PCDATA, abréviation de *Parsed Character Data* (données textuelles analysées par le parseur). Il signifie que l'élément défini contient obligatoirement des données textuelles. Dans le cas où un élément correspond au "père" d'un arbre de parenté, tel que <anthology>, alors celui-ci est presque toujours suivi des branches "fils" (e.g. <poem>) renvoyant inévitablement en bout de l'arbre à des fils ne contenant que des #PCDATA (e.g. l'élément <stanza> renvoie au modèle de contenu (line+) qui est repris dans la déclaration de l'élément <line> et qui elle-même

renvoie à (#PCDATA). Dans l'exemple précédent, les éléments **<title>** et **<line>**, marqués par #PCDATA ne peuvent donc que contenir d'autres éléments.

2.3.3.3. Les indicateurs d'occurrences

Dans le modèle de contenu, on utilise également des signes indiquant les occurrences de chaque élément (e.g. **line+** ou **title?**). Il existe trois indicateurs d'occurrences:

- Le signe plus "+" signifie qu'un élément peut apparaître une ou plusieurs fois (mais il ne peut être omis).
- Le point d'interrogation "?" signifie que l'élément aura au maximum une occurrence.
- L'astérisque ou l'étoile "*" signifie que l'élément aura de zéro à un nombre infini d'occurrences (il peut donc être omis).

2.3.3.4. Les opérateurs de connexion

Un modèle de contenu peut contenir plusieurs éléments. Lorsque des éléments sont séparés par une virgule ",", ceux-ci doivent tous être présents dans le texte. Si c'est un et commercial "&", les éléments doivent être présents dans l'ordre où ils sont indiqués dans le modèle. Enfin, si une barre verticale "|" les sépare, alors un et un seul d'entre eux doit être présent dans le texte.

2.3.3.5. Les groupes de modèle

Les modèles de contenu peuvent comprendre, outre des éléments simples et des #PCDATA, des listes d'éléments en combinaison avec les connecteurs précédemment cités. Ce sont alors des groupes de modèle.

Soit la déclaration suivante d'un couplet (strophe composée de deux lignes):

`<!ELEMENT couplet O O (line1, line2) >`

Les éléments **line1** et **line2** ont exactement le même modèle de contenu que la déclaration de l'élément **line**, i.e. #PCDATA, ce qui nous permet d'établir la déclaration ci-dessous:

`<!ELEMENT (line, line1, line2) O O (#PCDATA) >`

Ainsi, on introduit la notion de groupe de noms, qui est la combinaison de plusieurs éléments séparés par des connecteurs.

De cette manière, la nouvelle déclaration pour un poème pourra être définie par un modèle de contenu composé de zéro ou un titre et d'une ou plusieurs strophes, d'un ou plusieurs couplets ou d'une ou plusieurs lignes:

<!ELEMENT poem - O (title?, (stanza+ | couplet+ | line+)) >

Noter la différence avec la déclaration:

```
<!ELEMENT poem - O (title?, (stanza | couplet | line)+ ) > ,
```

où l'indicateur d'occurrence porte non pas sur un élément particulier mais sur tous les éléments inclus dans les parenthèses.

2.3.3.6. Informations supplémentaires

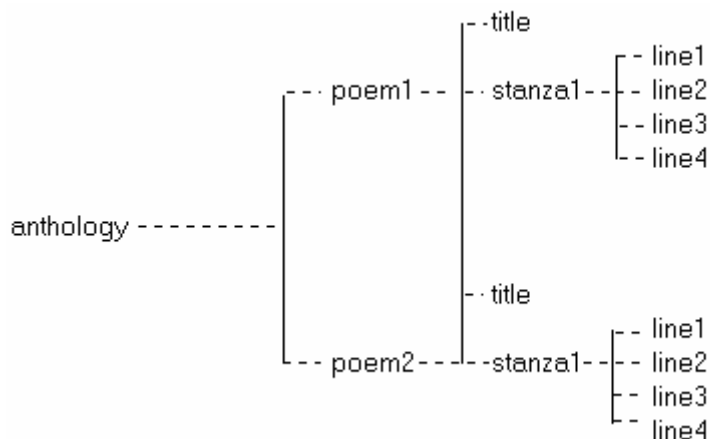
2.3.3.6.1. Exceptions au modèle de contenu

Certains éléments, tels que les annotations peuvent apparaître à n'importe quel endroit du document. Pour ce cas précis, SGML permet de modifier ultérieurement le modèle de contenu au moyen d'une liste d'exceptions. Il existe deux types d'exceptions:

- l'inclusion permet à des éléments supplémentaires d'être "inclus" à n'importe quel point du texte; elle est représentée par un signe "+" placé devant le modèle à inclure,
- l'exclusion évite que des éléments soient inclus dans le modèle; elle est représentée par un signe moins "-" placé devant le modèle à exclure.

2.3.3.6.2. Les structures concurrentes

Toutes les structures précédemment citées se présentent, comme dans l'exemple ci-dessous, sous une liste hiérarchisée.



Cependant, certains éléments propres à un type de document ne peuvent apparaître dans une telle liste. Un élément **<page>**, déclaré comme:

```
<!ELEMENT page - - ( (title?, line+)+ ) > ,
```


parvient difficilement à avoir sa place entre deux éléments spécifiques tels que **anthology** et **poem** ou **poem1** et **poem2**. Il faut donc instaurer une seconde structure, intervenant en concurrence (i.e. en même temps) avec la précédente liste.

Pour ce, on a dû créer deux types de documents: un document définissant une anthologie et un autre définissant la pagination d'une anthologie:

(1)

```
<!DOCTYPE anthology [
  <!ELEMENT anthology          - - (poem+)>
  <!ELEMENT poem               - - (title?, stanza+)>
  <!ELEMENT stanza             - O (line+)>
  <!ELEMENT (title | line)     - O (#PCDATA)>
]>
```

(2)

```
<!DOCTYPE p.anth [
  <!ELEMENT p.anth             - - (page+)>
  <!ELEMENT page               - - ((title?, line+)+)>
  <!ELEMENT (title | line)     - O (#PCDATA)>
]>
```

Un même texte peut ainsi être considéré selon plusieurs points de vue.

Afin de permettre le balisage des deux aperçus, il est nécessaire de préciser à quelle hiérarchie appartient tel ou tel élément. On inclut donc le nom du type de document entre parenthèses juste avant l'identifiant considéré (à la fois en début et en fin de balisage). Par conséquent, les pages (considérées dans le type de document **<p.anth>**) seront étiquetées par **<(p.anth)page>** au début et **</(p.anth)page>** à la fin de la partie concernée. Les éléments de la structure **<anthology>** seront marqués de la même façon par le type de document **<anthology>** (e.g. **<(anthology)poem>**).

Imaginons que le poème de Baudelaire est marqué par un saut de page apparaissant au milieu de la première strophe:

```
<(anthology)anthology>
<(p.anth)p.anth>
<(p.anth)page>

<!-- other titles and lines on this page here -->

<(anthology)poem<title>L'Albatros
<(anthology)stanza>
  <line>Souvent, pour s'amuser, les hommes d'équipage
  <line>Prendent des albatros, vastes oiseaux des mers,
</(p.anth)page>
<(p.anth)page>
  <line>Qui suivent, indolents compagnons de voyage,
```

```

        <line>Le navire glissant sur les gouffres amers.
    <(anthology)stanza>
        <line>A peine les ont-ils déposés sur les planches,
        <line>Que ces rois de l'azur, maladroits et honteux,
        <line>Laissent piteusement leurs grandes ailes blanches
        <line>Comme des avirons traîner à coté d'eux.

    <!-- other stanzas and lines go here -->

    </(anthology)poem>

    <!-- rest of material on this page here -->

</(p.anth)page>
</(p.anth)p.anth>
</(anthology)anthology>

```

Remarque: tous les systèmes informatiques ne traitent pas les structures concurrentes. Le trait CONCUR de SGML est facultatif et la TEI propose généralement des méthodes alternatives.

2.3.4. Les attributs

Les attributs ont pour but (1) de spécifier par un nom ou un nombre toute occurrence d'un élément particulier dans un texte et (2) de permettre l'ajout d'informations sans altérer les données du texte. Ce ne sont pas des parties du contenu d'un élément du texte, mais plutôt un moyen de le décrire qualitativement. L'attribut sert par exemple à indiquer le degré de fiabilité du texte ou à référer à une occurrence particulière d'un élément. Une balise finale peut ne pas contenir de paire d'attribut-valeur.

ex: <poem id=P1 status="draft">.....</poem>

Les attributs se présentent sous la forme de déclarations utilisant une syntaxe identique à celle utilisée pour déclarer les éléments.

ex: <!ATTLIST poem
 id ID #IMPLIED
 status (draft|revised|published) draft >

Le symbole ATTLIST introduit les particularités d'une liste d'attributs. La première partie spécifie l'élément ou les éléments considérés. A la suite des éléments, trois autres parties précisent le nom de l'attribut, le type de valeur qu'il peut prendre et une valeur par défaut.

Un certain nombre de mots-clés ont été créés dans SGML. Ce sont ceux qui apparaissent en lettres majuscules dans la deuxième partie de la déclaration. Les plus courants sont:

- ID: utilisé pour indiquer que l'attribut ID doit fournir une seule valeur d'identification pour toute instance de l'élément donné.
- CDATA: la valeur de l'attribut contient des données textuelles; des balises peuvent être incluses dans la valeur mais elles ne seront pas reconnues par le parseur.
- IDREF: contient une référence à un autre élément.
- NMTOKEN: contient un ensemble de caractères alphanumériques.
- NUMBER: est composé uniquement de données numériques.

La dernière partie de la déclaration contient une indication sur la manière dont un parseur doit interpréter l'absence d'un attribut donné. Elle peut se présenter sous la forme de mots-clés:

- #REQUIRED: indique la nécessité de la présence d'une valeur.
- #IMPLIED: indique l'optionnalité de la spécification d'une valeur.
- #CURRENT: précise que si aucune valeur n'est fournie, alors c'est la dernière qui a été utilisée qui doit être prise en compte.

On peut également insérer une valeur spécifique.

Dans les deux cas, cela permet d'attribuer une valeur par défaut.

Outre l'association de l'occurrence d'un élément à un nom ou nombre spécifique, les attributs permettent le croisement de références (en anglais, "cross-reference"). Afin d'illustrer cette notion, nous allons coder un extrait de registre familial:

Michel Dupond, boulanger.
Pauline Dupond, secrétaire, femme du susmentionné.

Soient l'élément considéré pour le codage de deux entrées <entry>, contenant un nom <name> et une profession <trade>. la deuxième entrée contient un élément complémentaire précisant sa relation avec la première entrée.

```
<register>
  <entry><name>Michel Dupond<trade>boulangier
  <entry><name>Pauline Dupond<trade>secrétaire
  <relation>femme du susmentionné
```

Le croisement de référence se présente sous la forme d'un attribut et d'une valeur arbitraire permettant de renvoyer la phrase de relation à l'entrée à laquelle elle se réfère:

```
<register>
  <entry id=E1234><name>Michel Dupond<trade>boulangier
  <entry><name>Pauline Dupond<trade>secrétaire
```

<relation target=E1234>femme du susmentionné

Ainsi, nous avons associé un nom arbitraire "E1234" à l'entrée du boulanger. En donnant le même identifieur à la valeur de l'attribut <**target**> (cible), nous confirmons à la fois l'existence de la relation et sa cible.

2.3.5. Les entités

Une entité a pour but de coder et nommer des parties arbitraires du contenu d'un texte. Une entité peut être un ensemble de caractères ou un fichier entier. Pour l'insérer dans le document codé, on utilise une référence d'entité (en anglais, "entity reference"). Par exemple, la déclaration ci-dessous définit une entité dont le nom est "tei" et dont la valeur est "Text Encoding Initiative":

```
<!ENTITY tei      "Text Encoding Initiative">
```

Lorsque l'entité rapporte un fichier complet, cela signifie que c'est le contenu du fichier qui doit être inséré dans le texte. Quand une entité est déclarée, on peut insérer son nom directement dans le texte en le précédant d'un et commercial "&" et le terminant par un point-virgule ";". On peut omettre le point-virgule si la référence de l'entité est suivie d'un espace ou se présente en fin de texte.

Quand le parseur rencontre une telle entité, il la substitue par sa référence. Un passage comme *"The work of the &tei has only just begun"* sera interprété par *"The work of the Text Encoding Initiative has only just begun"*. Cette méthode permet également la substitution de caractères ne pouvant être représentés dans certains systèmes informatiques.

Il existe aussi un autre type d'entité: les entités de paramètre. Ces entités sont utilisées dans les déclarations de balise et sont délimitées par un symbole de pourcentage "%" (précédé et suivi d'un espace) et un point-virgule ";". Le "%" est inséré entre le mot-clé ENTITY et le nom de l'entité.

2.3.6. Le balisage des sections

Il peut être utile de baliser une partie d'un texte en vue d'un traitement spécial par le parseur SGML. Par exemple, un morceau de phrase tel que "Liability is limited to \$50,000.", assertion légale dans le Delaware, USA, devra être changé en une nouvelle assertion qui, elle, serait légale pour l'état du Maryland.

A ces section étiquetées en SGML, on associe différents types correspondant aux mots-clés suivants:

- INCLUDE: permet d'inclure cette partie dans le texte et de le traiter normalement.
- IGNORE: exclut du texte la section marquée.
- CDATA: la section marquée peut contenir un ensemble de caractères ressemblant à des balises SGML mais ne devant pas être traités comme tels.
- RCDATA: similaire à CDATA, mais les références d'entité doivent être reconnues.

- TEMP: la section marquée est une partie incluse temporairement dans le texte, mais qui devra être enlevée ou changée ultérieurement.

2.4. Ce qu'il faut savoir du schéma de codage de la TEI

Un texte conforme à la TEI contient:

- (a) un TEI header (dont la balise est l'élément **<teiHeader>**);
- (b) la transcription du texte (marqué par l'élément **<text>**).

Le TEI header contient des informations similaires à celles contenues dans la page de titre d'un texte imprimé. Il est composé de quatre parties:

- une description bibliographique du texte lisible par la machine,
- une description de la manière dont il a été codé,
- une description non bibliographique du texte,
- un historique de ses révisions.

Un texte TEI peut être *unitaire* (concernant un seul travail) ou *composite* (ensemble de différents travaux comme, par exemple, une anthologie). Dans les deux cas, deux parties délimitent l'avant et l'après texte (**front** et **back**). Entre ces deux parties, on trouve le corps du texte (qui lorsqu'il est composite peut être formé de groupes, chacun contenant d'autres groupes ou textes).

Les exemples qui suivent sont extraits des TEI version Lite [réf.3].

La structure d'un texte unitaire se présentera comme suit:

```
<TEI.2>
  <teiHeader> [ TEI Header information ] </teiHeader>
  <text>
    <front> [ front matter ... ] </front>
    <body> [ body of text ... ] </body>
    <back> [ back matter ... ] </back>
  </text>
</TEI.2>
```

Un texte composite inclut la notion de groupe de textes, chaque groupe ayant une couverture et un dos:

```
<TEI.2>
  <teiHeader> [ header information for the composite ] </teiHeader>
  <text>
    <front> [ front matter for the composite ] </front>
    <group>
      <text>
        <front> [ front matter of first text ] </front>
        <body> [ body of first text ] </body>
```

```

                                <back> [ back matter of first text ] </back>
                            </text>
                            <text>
                                <front> [ front matter of second text ] </front>
                                <body> [ body of second text ] </body>
                                <back> [ back matter of second text ] </back>
                            </text>
                            [ more texts or groups of texts here ]
                        </group>
                    <back> [ back matter for the composite ] </back>
                </text>
            </TEI.2>

```

Il existe également une structure définissant une combinaison de textes TEI connus sous le nom de TEI corpus:

```

<teiCorpus>
    <teiHeader> [header information for the corpus]</teiHeader>
    <TEI.2>
        <teiHeader>[header information for first text]</teiHeader>
        <text> [first text in corpus] </text>
    </TEI.2>
    <TEI.2>
        <teiHeader>[header information for second text]</teiHeader>
        <text> [second text in corpus] </text>
    </TEI.2>
</teiCorpus>

```

Le but de ce projet n'est pas de donner une explication complète de ce qu'offrent les directives de la TEI. Par contre, les structures ci-dessus révèlent parfaitement l'usage que l'on peut faire du format SGML. En effet, l'idée des concepteurs de SGML, et par conséquent des créateurs des TEI Guidelines, est de proposer aux utilisateurs de texte électroniques un outil de "découpage" de tout document qui pourra par la suite être dépecé, étudié, etc. Comme il a été déjà observé dans la partie d'introduction (chap. 1), les documents à base textuelle, tels que les romans, pièces de théâtre, recueils de poèmes, font l'objet de directives quasi-complètes dans la TEI. La création des scanners digitaux notamment (cf. section 4.3.) et les possibilités qu'ils offrent d'insérer directement des images dans l'ordinateur ont ouvert de nouveaux horizons aux documents électroniques. Cependant, ces horizons se limitent encore au stockage des images, la représentation du contenu de l'image restant un sujet d'avenir. Les chapitres qui suivent ne veulent donc pas révolutionner la représentation cognitive de l'image, mais plutôt proposer un sens de recherche au moyen de l'outil que l'on vient de définir, i.e. SGML.

3. LES PROBLEMES POSES PAR LA BANDE DESSINEE EN VUE D'UN CODAGE SGML

La bande dessinée, de par son aspect très "visuel", est considérée comme un art à part entière. Elle est d'ailleurs classée en neuvième place de la liste des arts reconnues par la Culture avec un grand "C". Art à part entière et pourtant regroupant sur papier tout ce que peuvent proposer les huit autres arts: prose, dialogue, jeu de rôles, illustration recherchée, art graphique, usage du son, etc. Mais ceci n'est qu'un bref aperçu de ce qu'est la BD. Ce chapitre a donc pour but de faire un petit tour de cette dernière en explorant principalement son étendue et son contenu.

3.1. La BD en tant que document illustré

3.1.1. Définition

"La définition de la bande dessinée est simple: c'est une suite de dessins contenant une anecdote ou une histoire: les personnages s'y expriment par des textes inscrits dans des bulles." [Filippini 89].

Cette définition pourrait parfaitement convenir à un non puriste de la bande dessinée. Cependant, la réalité n'est pas aussi "simple". En fait, on considère non pas UNE bande dessinée, mais bien DES bandes dessinées. La narration par le dessin remonte en fait à la préhistoire, le plus lointain exemple étant le *Livre des Morts* égyptien (1500 av. J.-C.). Ce sont des histoires figuratives dont les scènes successives s'ordonnent en bandes superposées, le texte (idéographique) étant situé à l'intérieur de l'image. On retiendra également beaucoup plus tard la tapisserie de Bayeux (XIe s.) ou les images d'Epinal (XVIIIe s.) comme d'autres ancêtres du-dit "neuvième art".

C'est pourtant aux Etats-Unis que naît une nouvelle forme d'expression illustrée par l'utilisation de la bulle (connue sous les autres noms de ballon et phylactère).

3.1.2. Historique

Comme il a été déjà observé, il existe plusieurs sortes de bandes dessinées. Les séries américaines, dont la toute première *The Yellow Kid* de Richard Felton Outcault paraît dans le "New York World" en 1895 en introduisant la notion de phylactère dès 1896, apparaissent dans les pages dites de "comics" des suppléments dominicaux des grands

journaux. Cet intérêt que portent les journaux pour la BD tend à se développer rapidement pour donner naissance dès 1912 à deux types de comics:

- l'aventure quotidienne en rangée de trois ou quatre images appelée "daily strip",
- la planche hebdomadaire en neuf ou douze images en couleurs connue sous le nom de "Sunday Page" [Annexe 1].

L'engouement des éditeurs est tel que ceux-ci finissent par créer des organismes appelés "syndicates", organismes de protection de la propriété des héros de séries. Notamment, en 1916, William Randolph Hearst fondera la *King Features Syndicate*, première agence chargée de distribuer les bandes dessinées de ses journaux dans tous les Etats-Unis, puis dans tous les pays de langue anglaise, et enfin à partir de 1929, dans le monde entier. Ainsi, deviendront célèbres des séries telles que *Mandrake the Phantom* ou encore les *Peanuts*.

En 1936 naît une nouvelle forme de diffusion: les "comic-books". Ce ne sont plus de simples pages diffusées dans les journaux mais bien des fascicules spécialisés, aux couvertures fantastiques et dont les personnages sont catégorisés de "super-héros".

En Europe, les ballons utilisés sporadiquement dans *Les Pieds Nickelés* (1908) de Louis Forton, ne seront adoptés définitivement que dans les années 20 dans *Zig et Puce* (1925) d'Alain Saint-Ogan et *Tintin et Milou au pays des Soviets* (Bruxelles 1929) d'Hergé. Mais ce système ne se généralisera qu'après le succès du Journal de Mickey paru en 1934.

Pendant la guerre, le matériel américain faisant défaut, les auteurs français s'adaptent aux phylactères et s'inspirent des séries américaines. En Italie sont créés les premiers fascicules de huit à seize pages contenant des histoires complètes appartenant à une série ou un récit totalement indépendant. Ce marché, dominé par les éditions Sage, Elan, Marcel Daubin, Artima, perdure après la guerre avec les titres: "*Alain la Foudre*" (1949-1950), "*L'Assaut du Ciel*" (1946-1948), "*Amok*" (1949-1950), "*L'Appel de la Jungle*" (1949-1952), "*La Série 2000*", "*Les Sélections Vaillance*" (1944-1949), "*Une aventure de ...*" (1949-1950), etc. Ces publications disparaissent avec l'arrivée des magazines en format de poche dans les années 50.

Après la guerre, la production de la bande dessinée en France s'oriente vers 3 types d'édition: les journaux réalisés par des auteurs et éditeurs français, les magazines en format de poche et les nouveaux hebdomadaires en provenance de Belgique. Du côté américain, une place semble leur être réservée en matière d'un autre style, le photo-roman, bandes sentimentales dont la plus connue est *The Heart of Juliet Jones* (*Juliette de mon coeur*).

L'édition d'albums est relativement récente. Elle a longtemps fait l'exclusivité aux seuls séries vedettes "*Spirou*" sous le label des éditions Dupuis et "*Tintin*" sous celui des éditions Lombard. A partir de 1960 cependant, les publications d'albums augmentent pour atteindre un chiffre dépassant les 600 albums par an, sous l'égide des éditions

Afin de récapituler les différents types de BD existants et pour faire le lien avec SGML, voici une proposition de balise "générale" d'une BD. En vue d'une éventuelle normalisation, nous représenterons la BD par l'élément **comic**.

- Etant donné la diversité des formes de diffusion des bandes dessinées, il est difficile de présenter une structure générale du genre. C'est pourquoi, nous considérerons uniquement le corps de la BD qui, lui, a peu de chance de changer d'un support à un autre.

"Il y a, selon moi, deux façons de lire un roman ou une bande dessinée. Premièrement, nous pouvons ne rien connaître de la "cuisine" qui préside à la création de toute oeuvre écrite ou dessinée (pourquoi les mots ou les images sont-ils agencés de telle façon plutôt que d'une autre, etc.) et nous comprendrons quand même très bien ce qu'a voulu dire l'auteur, nous tirerons peut-être beaucoup de satisfaction de notre lecture. Et cependant, cette façon de lire une oeuvre en restant à sa surface ne nous donnera jamais tout le plaisir que nous étions en droit d'attendre. Il y manque quelque chose... En revanche, pour peu que nous sachions suffisamment reconnaître les moyens qu'a utilisés l'auteur pour s'exprimer, notre satisfaction de lire une bonne histoire se trouvera tout naturellement augmentée du plaisir d'apprécier *comment celle-ci nous est racontée*. En d'autres termes, nous pourrions apprécier le style de son auteur, sa façon bien à lui d'exprimer ses idées, de tourner ses "phrases" avec plus ou moins d'originalité. De superficielle qu'elle était, notre lecture peut devenir alors un vrai régal!". [Duc 82]

page 25

principale difficulté à laquelle peut se trouver confronté un codeur est semblable à celle rencontrée par le lecteur: le "savoir-lire".

Bien que la norme SGML ait des buts focalisés sur les données textuelles d'un document, on peut imaginer que certaines données graphiques, telles que celles présentes dans une bande dessinée, doivent être détectées et comprises afin de poursuivre tout type de codage.

3.2.1. Les données graphiques

De manière similaire au théâtre où l'indication et la numérotation ordonnée des actes et des scènes (e.g. Acte II Scène 5), conjuguées à l'indication des personnages à chaque changement de locuteur (i.e. nom des personnages apparaissant en marge du dialogue), permettent de décrire le suivi d'une pièce, il existe 2 sortes de graphismes jouant le rôle de balise de l'histoire contée dans une BD: les **cadres** (ou cases) et les **bulles**. Les planches peuvent être également interrompues par des sous-titres, comparables à l'indication des actes ou des chapitres existant dans d'autres types de documents (roman, théâtre, etc.). Ces "coupures" peuvent apparaître sur une page seule, en séparation des différentes parties [Annexe 2, Figure 1] ou encore en entêtes indiquant la séparation entre différentes histoires courtes composées d'une ou deux planches [Annexe 2, Figure 2].

3.2.1.1. Les cadres

Le cadre est l'élément essentiel de la BD.

"Sans lui pas de récit; rien à raconter en un plan général. C'est la décomposition en images successives qui permet de créer, dès le premier changement d'angle, une histoire."
[Amiel 86]

Non seulement tout cadre contient un plan, mais, de par sa forme, il peut déterminer le type de dynamisme que l'on désire donner à ce plan. Ainsi, un cadre horizontal, s'il est de taille moyenne, représentera des scènes, décors ou paysages divers, tandis que s'il se caractérise par un étalement franchement horizontal, il représentera un vaste panorama, tel une plaine, une vue maritime, une course poursuite, une foule en cortège, etc. [Annexe 3, Figure 1].

Le cadre vertical représentera alternativement tout objet ayant une dynamisme vertical: une montagne, un building, une chute, un envol, etc. [Annexe 3, Figure 2].

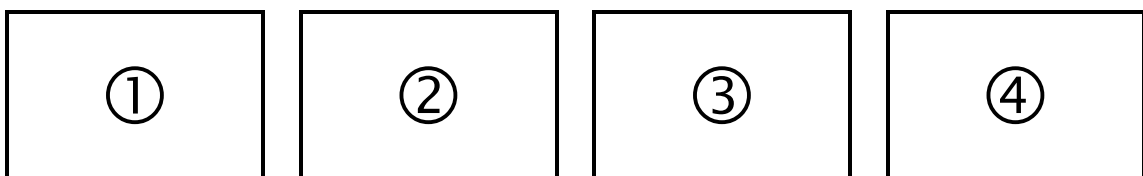
Le cadre rectangulaire ou le carré, représentation la plus courante, sera utilisé pour des plans d'ensembles ou des plans généraux.

Bien évidemment, des cadres de forme inattendue sont également utilisés, principalement afin d'attirer l'attention du lecteur. Ces types de cadres peuvent s'insérer directement dans un cadre principal [Annexe 3, Figure 3], ou encore faire le lien entre deux cases consécutives [Annexe 3, Figure 4]. Ces cadres peuvent donc prendre toute sorte de formes: ronds, carrés, rectangles aux angles arrondis, losanges, triangles, etc. Lorsqu'ils sont insérés dans un cadre, ils ne modifient pas la forme de celui-ci. En revanche, lorsqu'ils apparaissent sous forme de liaison, ils "cassent" le ou les cadres principaux. D'où la difficulté de leur donner une forme type. La seule particularité que l'on peut attribuer à un cadre est par conséquent ce trait plein marquant la limite entre le support papier et le dessin en lui-même. Ceci tout en sachant qu'il arrive que le cadre disparaisse pour ne laisser parler que le dessin!

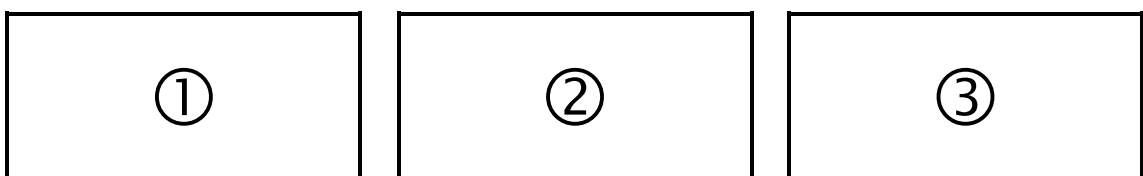
Outre les notions sémantiques que l'on peut attribuer aux différentes formes de cadres, ce qui intéressera le codeur SGML est leur position dans la page. En fait, le principal problème pour SGML sera de connaître le sens de lecture des cadres lorsqu'il les aura détectés. Le problème sera d'autant plus important que ces cases ne conservent pas toujours un alignement horizontal mais, par exemple, si l'on désire exprimer un certain mouvement, celles-ci peuvent être décalées par rapport à la page [Annexe 4].

Dans une BD classique, telle Lucky Luke ou Tintin, on trouvera rarement des "cassures" ou de gros changements morphiques dans la géométrie des cases. Dans ce cas, la lecture horizontale de la gauche vers la droite est assez apparente. Cependant cela n'est pas toujours aussi évident ce qui signifie que des règles particulières s'imposent:

(1) Lecture en rangées horizontales de gauche à droite (avec départ en haut de la page):



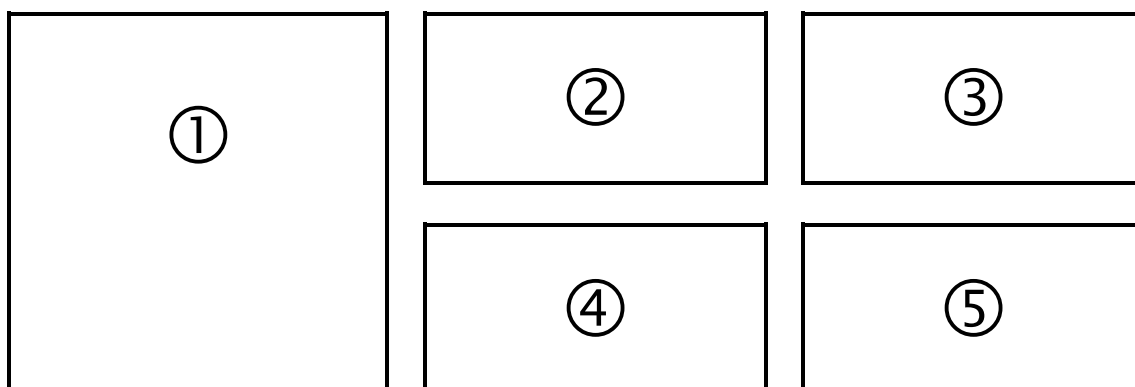
(2) (1) + retour en bas à gauche:



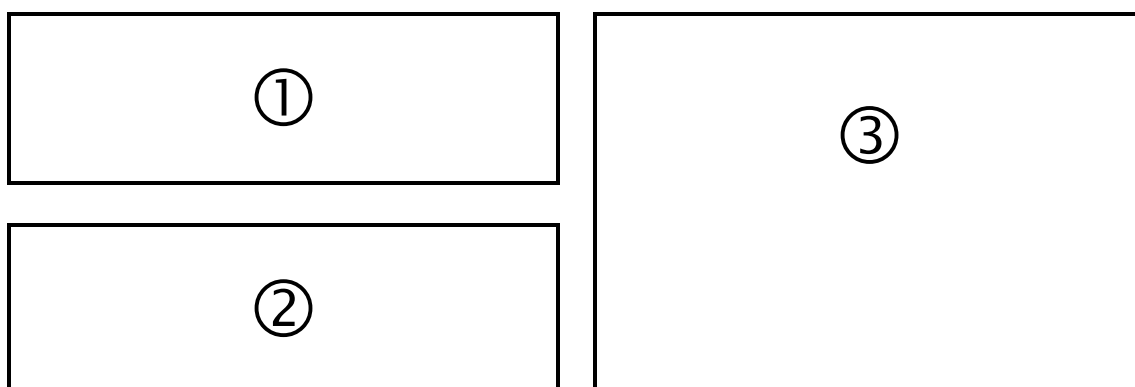
④

⑤

(3) (2) + retour vers la case non encore lue la plus à gauche:



(4) Direction donnée par la verticalité du plus grand cadre présent dans l'alignement (si ce dernier ne se trouve pas en tête de la bande)



En complément de ces règles, et dans le but de préciser ou imposer un sens de lecture, le dessinateur peut employer un système de fléchage, fléchage qui, bien entendu, varie en forme et en couleur selon les artistes [Annexe 5].

Dans une BD, une scène est généralement délimitée par un trait plein, néanmoins, d'autres types de traits peuvent être utilisés pour produire certains effets. En effet, le trait plein peut être substitué par un trait en "vagues" inspirant la narration de faits antérieurs ou inspirant le rêve [Annexe 6]. Bien sûr, d'autres truchements peuvent compléter ou remplacer ce système, notamment l'utilisation de couleurs particulières.

3.2.1.2. Les phylactères

Le phylactère (ou bulle ou ballon) est la boîte de dialogue des personnages. On peut dire que les textes se présentant en-dehors des bulles ne peuvent être directement attribués à un personnage en particulier. Alors qu'au théâtre le nom du personnage est indiqué en marge du texte:

DON DIEGUE Rodrigue, as-tu du coeur?, [Corneille 70]

en BD, rien ne précise *qui parle*, excepté la pointe de la bulle dirigée vers le locuteur.

La bulle est le fil conducteur du récit. Non seulement elle attribue un texte à un personnage ou à un objet en particulier, mais elle donne des indices de lectures. Comme on a vu que les cases se lisaient de gauche à droite et de haut en bas, il en est généralement de même pour la lecture des bulles. On remarquera que les bulles partent en haut et à gauche du cadre pour terminer vers la droite et en bas, toujours dans la même idée de donner un dynamisme au récit.

La bulle est en fait une borne temporelle du récit. Parfois, afin de montrer qu'une case est en prolongement d'une case précédente, il arrive qu'une bulle apparaisse dans la première case et que son locuteur ne soit désigné que dans la case suivante par la pointe de cette même bulle [Annexe 7, Figure 1]. Vice versa, un locuteur apparaissant dans une case n'aura la parole que dans la case suivante [Annexe 7, Figure 2].

Evidemment, le bon suivi des bulles dépend directement de la scène dessinée qui, poussée à l'extrême peut devenir le seul fil conducteur du récit [Annexe 8].

D'autre part, certains auteurs moins classiques feront complètement abstraction de la bulle pour ne garder qu'un trait plein dirigé vers le locuteur.

Outre l'attribution de la parole et de manière similaire aux cadres, le changement morphique du contour de la bulle suggère différents états d'esprit:

- le trait plein indique la parole entendue de tout le monde,
- le trait en pointillés ou constitué d'une suite de petites bulles, ou encore une bulle en forme de nuage exprime généralement la pensée ou l'aparté,
- le trait en dents de scie suggère souvent une voix retransmise (téléphone, radio, télévision) ou enregistrée (magnétophone),
- le trait en "pics" exprime une voix forte et/ou stridente,
- une bulle ornée de stalactites révèle la froideur du ton, l'hostilité,
- etc., etc.

On a dit précédemment que la bulle servait de borne temporelle du récit. D'autre part, ses différentes utilisations permettent de donner un rythme, un dynamisme aux dialogues, monologues ou apartés:

- des bulles étagées révèlent un dialogue très discuté,
- des bulles en vrac expriment la simultanéité des interventions orales,
- des bulles juxtaposées peuvent être attribuées à un seul personnage, souvent pour éviter de trop longs monologues ou encore exprimer un récit tourmenté,
- inversement, une même bulle peut pointer vers plusieurs personnages.

Dans le cas des bulles en vrac, il serait intéressant d'ajouter au niveau du codage un texte de commentaire similaire aux didascalies* utilisées au théâtre, comme "*En même temps...*". Une bulle pointant vers plusieurs personnages inspirerait un commentaire tel "*En chœur...*". Ce genre de commentaire éviterait la répétition de la même phrase attribuée à plusieurs personnages à l'aide des pointes de la bulle. Ainsi, on ne prendra en compte que le texte unique de la bulle et non le nombre de personnages vers qui elle se dirige. En ce qui concerne les bulles en vrac, le commentaire rendrait possible la lecture traditionnelle (partant en haut à gauche et finissant en bas à droite), dans le cas où il n'existe aucune priorité temporelle d'une bulle sur une autre.

3.2.2. Les données textuelles

Dans une BD, les données textuelles sont nombreuses et de caractère varié. Comme au théâtre, on a la partie des "acteurs" et la description temporelle et scénique de l'auteur. Comme dans les livres, on a des appels de notes. Comme au cinéma, on a des systèmes de bruitage. La BD est en fait un agglomérat d'autres arts, documentaires ET graphiques, d'où la difficulté de déterminer distinctement les éléments qui la compose. Pourtant, si l'on veut tenter un codage SGML, certains éléments doivent obligatoirement être pris en compte.

3.2.2.1. Les indications temporelles et scéniques

Pour soutenir la temporalité d'un récit, il n'est pas rare que l'auteur ajoute des phrases hors bulles indiquant le moment ou l'endroit dans lequel se situe l'action. Ces phrases se présentent souvent, mais pas toujours, sous la forme d'encadrés apparaissant dans l'un ou l'autre des angles du cadre. Parfois, si cela semble important aux yeux de l'auteur, c'est une case entière qui leur est destinée. Une telle indication intervient rarement en simultanéité avec l'histoire. Elle peut apparaître soit avant, soit après les autres composants de la case. Par contre, un tel encadré peut relier deux cases [Annexe 9, Figure 1].

Il ne faut pas confondre les indications purement scéniques et les commentaires de l'auteur ou du narrateur [Annexe 9, Figure 2]. Dans ce cas, il est utile de reconnaître certains mots-clés ou symboles permettant de différencier un commentaire ou une narration d'une donnée temporelle ou scénique: les dates, les noms de ville ou de pays, les expressions toutes faites telles que "*Le lendemain...*", "*Plus tard...*", etc., indiquant généralement le lieu ou l'instant de la scène. En revanche, les points d'interrogations, d'exclamations, les interjections (Oh!, Ah!, etc.), seront réservées aux commentaires.

* Didascalie: indication donnée à un acteur par l'auteur, sur son manuscrit, dans le théâtre grec (terme didactique utilisé dans le même sens pour le théâtre moderne).

3.2.2.2. Les onomatopées

Les onomatopées existant dans la langue française sont assez rares. Les auteurs de bandes dessinées se sont donc trouvés dans l'obligation de créer de toute pièce des onomatopées susceptibles de reproduire phonétiquement à peu près tous les bruits de la nature [Annexe 10].

Les voyelles "O" et "U", redoublées ou non, évoqueront toujours des sons plutôt graves, étouffés, creux, mous (POF, BOOM, OOOH, etc.), tandis que la lettre "I" suggérera au contraire des sons aigus, métalliques (ZIIP), sinon stridents (HIIII ou DRIIIN). Quant au "A", il produit des sons intermédiaires secs, non métalliques (CLAP, CALP, CRAAAAAC, etc.).

Dans le cas des onomatopées utilisées en BD, il est important de préciser le type de graphisme qui est employé afin de déterminer l'intensité du son produit. Des caractères plus ou moins grossissant expriment ainsi le crescendo ou le decrescendo du bruit. Dans ce cas et dans la mesure du possible, le codage devra rendre la taille des caractères. On a vu dans la partie consacrée au balisage SGML qu'il existait des étiquettes marquant des mots ou des phrases se distinguant par leur différence graphique. On peut ainsi affirmer que la plupart des onomatopées trouveront un signe caractéristique dans ce type de marquage, tout en considérant certaines limites. En effet, les onomatopées en BD sont rarement alignées horizontalement par rapport au support papier. Elles peuvent apparaître à n'importe quel endroit de la case, voire en-dehors. Elles peuvent être volontairement mal alignées et avec un graphisme de caractères plus ou moins "tremblé", pour exprimer des vibrations sonores. Pour les distinguer du reste du texte, elles peuvent être en italique, en gras, en caractères creux, etc. Elles peuvent parfois être difficilement lisibles, incomplètes ou carrément illisibles.

Le second intérêt est de savoir à quel moment précis de la scène intervient l'onomatopée:

- avant: elle provoque parfois une réaction du personnage; si celui-ci doit parler, alors la bulle se présente après l'onomatopée.
- simultanée: dans ce cas, il est aussi simple de respecter la lecture de gauche à droite et de haut en bas, sachant que, comme pour les bulles, on ajouterait un commentaire,
- après: la bulle apparaît logiquement à gauche du bruit provoqué.

3.2.2.3. Les symboles

Plus fréquemment utilisés dans les BD comiques, les symboles et graffitis utilisés dans les bulles expriment la pensée profonde d'un personnage [Annexe 11]. Cependant, même s'ils font partie entière du récit, il semble difficile de les décrire artificiellement. Retenons tout de même les symboles évoquant les émotions élémentaires, issus de caractères typographiques courants: les points d'interrogation et d'exclamation.

Ces caractères subissent le même sort que les onomatopées. Plus leur graphisme est épais et plus l'émotion ressentie sera forte. Isolés dans la bulle, ils sont le seul moyen d'expression de la pensée du personnage. En conjonction avec le texte, ils relèvent le dynamisme de la parole. Il ne faut pas oublier que la BD veut rendre un récit vivant, plus proche encore du cinéma que du théâtre. L'intonation y est donc primordiale.

Il arrive que ces symboles soient utilisés pour amplifier la valeur d'un autre graphisme remplaçant la parole du personnage, de la même façon qu'un rébus.

Un autre caractère typographique est souvent utilisé à outrance: les "...". On peut comparer les "..." aux points d'orgues utilisés en musique pour suspendre la mélodie ou le chant, pour mettre en alerte l'auditeur. En BD, cela va plus loin. La parole ou la pensée du personnage peut prendre plusieurs sens:

- une phrase non terminée, suivie de "..." peut exprimer le doute, la menace ("sinon..."),
- les "..." peuvent aussi être employés comme système de liaison; une phrase trouvera sa finalité dans une autre bulle. C'est alors un moyen de couper les monologues qui, sans quoi, pourraient devenir rébarbatifs,
- dans une conversation téléphonique, ils marquent les pauses pendant lesquelles c'est l'interlocuteur qui a la parole.
- utilisés tous seuls, les "..." incitent le lecteur à lire entre les lignes, à détecter tous les sentiments que les graphismes peuvent susciter, à se remémorer l'histoire, les raisons de la scène présente, etc. Ici, ils ne provoquent pas des scènes futures, mais au contraire, plongent dans les souvenirs du personnage et du lecteur. Les "..." peuvent dans ce cas avoir pour référent toutes les données antérieures qui font que la pensée du personnage est telle. Si l'on se reporte à SGML, on peut dire que les "..." isolés peuvent être remplacés par un référent unique qui engloberait toute l'histoire qui a précédé et provoqué la scène.

Parmi les autres symboles utilisés, on peut également retenir les notes de musique permettant de distinguer les parties parlées des parties chantées. Dans les parties "chantées", on distingue le chant réel et les voix mélodieuses telles la voix d'une hôtesse de l'air ou la voix perfide d'un personnage maléfique (expression de l'ironie, de l'hypocrisie).

3.2.2.4. Le contenu des dialogues dans les bulles

Un texte de BD se présente généralement toujours sous la même forme manuscrite du début à la fin de la BD. Cependant, comme on l'a observé pour les onomatopées ou les symboles, les variations graphiques sont multiples. Ces variations sont principalement dues aux changements de ton des personnages: l'étonnement, la colère, la peur, etc. On observe également la présence de nombreuses onomatopées s'immisçant dans le texte-même.

Le changement de police de caractères a des buts bien précis en BD. En général, il suggère un changement de "prononciation" ou d'intonation du personnage. Souvent, de système permet d'introduire un langage différent ou un accent typique [Annexe 12, Figure 1].

Selon le genre de la BD (comique, policière, etc.), les langages utilisés varient du plus littéraire au plus imagé. Plus simplement, on considérera:

- les langues issues des langues réelles: latin, anglais, etc.,
- la langue du lecteur,
- les paroles illisibles ou imaginaires représentées par un graphisme quelconque et indéchiffrable.

En ce qui concerne la langue du lecteur, elle peut être exprimée au moyen de n'importe quelle calligraphie ou type caractère. Ainsi, on ne balisera pas de la même manière la phrase latine "*sic transit gloria mundi*" [Annexe 12, Figure 2] prononcée par le pirate d'Astérix et le "*De diou*" lancé en cyrillique, mais *pas en russe*, par un personnage dans Spirou [Annexe 12, Figure 3].

Il est utile de pouvoir différencier les langages sachant que leur compréhension ou non compréhension par le lecteur a un but bien précis dans l'esprit de l'auteur. Par exemple, dans Lucky Luke, une partie du journal sera écrite dans la langue du lecteur (ce qui signifie qu'il en est le destinataire) et le reste dans la langue du lieu où se situe l'action (en l'occurrence, en anglais) [Annexe 12, Figure 4].

Il faut savoir que les langues imaginaires ressemblent rarement à des caractères existants. Cependant, que deviendrait cette page de Rosy [Annexe 12, Figure 5] si on "retirait" aux personnages le don de la parole. D'autre part, un auteur peut utiliser des caractères illisibles dans d'autres fins que d'exprimer une langue particulière. Dans *Spirou à New York*, par exemple, on aperçoit au loin un cimetière et une bulle contenant des traits en "vagues" exprimant le sermon inaudible (et incompréhensible) du prêtre [Annexe 12, Figure 6].

Il est vrai que les auteurs de Spirou semblent se délecter de ces jeux de langages comme on peut le constater dans *Z comme Zorglub* où les "Zorglhommes" communiquent en inversant les mots [Annexe 12, Figure 7] ou encore dans *Spirou à Moscou* où les lettres N et R inversées se veulent être une marque de la langue russe [Annexe 12, Figure 8].

3.2.2.5. Autres données textuelles

Par ce titre, on entend tous les mots, expressions ou phrases apparaissant à n'importe quel endroit du dessin mais ayant une inférence sur le cours de l'histoire. En vrac, cela peut être le texte d'un article de journal, une enseigne, une lettre, c'est-à-dire tout ce qui peut être lisible par le lecteur. On les distingue bien évidemment des onomatopées, mais aussi des commentaires et indications scéniques. Le problème semble être le même que

celui des onomatopées; il va falloir déterminer l'instant où la lecture de ce texte doit intervenir dans le récit. Si l'on estime que la case doit être lue de la gauche vers la droite, cela simplifie bien la tâche. Cependant, ce sens de lecture peut être controversé pour des raisons de temporalité.

En effet, la relation entre les expressions méta-texte et le récit peut être ou très étroite, ou, inversement, très lointaine. Si elle est très étroite, le sens du texte (en tant que relation sémantique), joue un rôle important dans l'ordonnancement de toutes les données. Par contre, si c'est uniquement un complément à une indication scénique, il peut être préférable de le faire apparaître avant le reste de la case (mais après les didascalies et autres commentaires inscrits en marge de la case - si ceux-ci sont exposés en haut à gauche).

3.2.2.6. Les notes infrapaginales

Les appels de notes en BD sont comparables à ceux de la littérature classique. En revanche, les caractères utilisés pour les renvois sont beaucoup plus diversifiés. Les plus courants sont bien sûr ceux de la littérature classique: astérisque ou étoile "*", numérotation (avec ou sans parenthèses), etc. D'autres peuvent être simplement des exercices de style en art graphique.

L'idée est donc de pouvoir repérer deux caractères semblables (sachant qu'il existe des marges de différences puisque la BD est manuscrite), dont l'un apparaît en exposant d'un texte et l'autre en note de bas de case ou de bas de page. Etrangement, on a déjà vu des appels de notes inachevés [Annexe 13, Figure 1], dus simplement à un oubli du dessinateur.

Il arrive pourtant que la note résultante disperse encore plus la compréhension de ce qui est annoté. Par exemple, dans Spirou, l'appel de note laisse espérer une traduction du chinois, et quelle n'est pas la surprise de retrouver en bas de page d'autres symboles dans cette même langue [Annexe 13, Figure 2]!!!. Effet comique, évidemment...

Autant de significations et de composants appartenant à un art révèle une grande richesse dans son contenu. Sans parler de l'art graphique utilisé par le dessinateur, souvent comparé à l'art pictural. Sans parler non plus du langage employé par le scénariste, style riche par excellence de par sa composition multiple (emploi des langues étrangères, usage et création d'onomatopées, style littéraire varié, etc.). D'où l'intérêt de l'intervention des TEI Guidelines pour mettre à disposition un tel puits d'informations sous format électronique.

4. PROPOSITION DE CODAGE D'UNE BD SOUS FORMAT SGML

Après avoir défini les différents composants du corps d'une bande dessinée, nous pouvons aborder librement le problème de codage sous format SGML. Nous verrons dans un premier temps comment s'effectue le codage des différentes parties d'une BD. Ensuite, nous effleurons le problème de l'écriture manuscrite et du codage des caractères et, enfin, les moyens mis en oeuvre pour l'implémentation des BD sous format électronique.

4.1. Codage des différentes parties d'une BD

Le corps de certains textes consiste simplement en une courte séquence d'éléments, comme par exemple des suites de paragraphes dans les textes prosodiques, les parties de discours et les didascalies dans les représentations théâtrales, les entrées de dictionnaire dans les dictionnaires. Dans d'autres cas, ces suites seront découpées de manière hiérarchisée en parties et sous-parties (e.g. chapitres et sous-chapitres). Les noms de ces parties varient selon le genre, la période ou encore la volonté de l'auteur, de l'éditeur ou du publieur: une sous-division d'un texte épique ou de la Bible est généralement appelé "book", celle d'un rapport "part" ou "section", celle d'une nouvelle "chapter", etc.

Dans un but de normalisation, les TEI Guidelines ont instauré un seul et même élément, **<div>**, dont la neutralité permet de s'adapter à toute sorte de découpage. Une BD se divise principalement en pages, en planches numérotées et en une succession de cases. Grâce à cette première assertion, on peut effectuer un premier découpage à partir d'une aventure complète de Tintin, *L'Ile Noire* [Hergé 66]. On représentera le titre en entête par l'élément **<head>**.

```
<body>
<div type='page' n=1>
  <head>L'Ile Noire</head>
  <div type='frame' n=1>
    </div>
  <div n=2>
    <sound> RRRRRRR</sound>
    <sp who='Tintin'>Tiens?... Un avion...</sp>
    <!--more divisions go here -->
  </div>
<!--more divisions go here -->
```

<trailer>FIN</trailer>

</body>

On peut d'ores et déjà observer la distinction du texte présent dans la bulle marqué par l'élément **<sp>** et la représentation de l'onomatopée par l'élément **<sound>**.

Après avoir réalisé ce premier découpage, le plus dur reste encore à faire: marquer par une balise normalisée les différents éléments qui composent une case. Tout au long de ce rapport, nous avons fait des comparaisons du genre de la BD avec le théâtre. La TEI propose déjà un ensemble de balises adaptables à toute forme de représentation de ce type (théâtrale, radiophonique ou cinématographique). Parmi ces balises, on retiendra:

- **<set>** contient une description du lieu de l'action, du moment, etc.
- **<sp>** pour speech, peut être utilisé en BD pour représenter le texte inclus dans les bulles. Dans ce cas, certains attributs peuvent être utilisés en complément d'information:

type donne une valeur propre au type de discours, i.e. *speech* pour discours parlé, *thought* pour inspirer la pensée, *music* pour les parties musicales,

who indique le nom du ou des personnages, s'ils sont connus,

id indique un référent arbitraire si le personnage n'est pas reconnu.

- **<stage>** contient toute sorte de direction scénique. L'attribut **type** est utilisé principalement pour décrire le jeu des personnages, leur apparition dans la scène, leurs mouvements, leur état d'esprit, etc. (e.g. *type=exit*).

D'autres types de représentations scéniques, comme les séries radiophoniques ou télévisées, introduisent d'autres éléments pouvant être utiles au codage de la BD:

- **<view>** décrit le contexte visuel d'une pièce télévisée, i.e. ce que le spectateur voit et qui est généralement indépendant du dialogue. En BD, on distinguera cet élément de l'élément **<stage>** pour en faire une balise de tout ce qui est mot, phrase ou texte plus ou moins indépendant du récit (présence d'enseignes, articles de journaux, lettres, etc.).
- **<sound>** indique la présence d'un son ou d'une séquence musicale. En BD, on attribuera cet élément à toutes les onomatopées utilisées à l'intérieur et à l'extérieur des bulles. L'intérêt est que les attributs déjà existants peuvent être adaptés au genre de la BD:

type catégorise le son, e.g. *music*, *special effect*, etc.,

discrete précise si le son chevauche le reste du texte ou s'il l'interrompt. Les valeurs qui lui sont associées sont:

y le son intervient entre les textes,

n le son chevauche le texte (notion de simultanéité),
u inconnu.

Outre ces balises empruntées des représentations scéniques, il existe des balises plus générales, i.e. applicables à plusieurs types de documents. On retiendra notamment l'élément **<note>** indiquant la présence d'appels de notes. En BD, les appels de notes sont réalisés de manière similaire; seule leur position varie (bas de case, bas de page, etc.).

Afin d'illustrer l'emploi de ces différents éléments, nous allons réaliser le codage des 4 premières cases de la page 51 de l'album *Le Mystère de la grande pyramide* [Annexe 14].

```
<body>
<!-- more divisions -->
<div type='page' n=51>
  <div type='frame' n=1>
    <stage>MORTIMER, S'ARRACHANT A SES PROFONDES
    REFLEXIONS, REPREND LENTEMENT LE CHEMIN DU
    RETOUR...
    </stage>
  </div>
  <div n=2>
    <stage>MAIS A PEINE A-T-IL FAIT QUELQUES PAS, QU'IL
    S'ARRETE BRUSQUEMENT...
    </stage>
    <sp who='Mortimer'>Tiens, qu'est-ce qui se passe là-bas?...
    </sp>
  </div>
  <div n=3>
    <sp who='Mortimer'>Cela m'a tout l'air d'une rixe...
    </sp>
    <stage>ET PIQUANT DES DEUX, IL LANCE SA MONTURE AU
    GALOP...
    </stage>
  </div>
  <div n=4>
    <sp who='Blake'>Quand j'interdis que l'on aille fouiner dans les coins, je
    désire être obéi, entends-tu?...
    </sp>
    <sp id=X1>Mais, effendi, le moudir <note id=X2>(1)</note>avait dit...
    </sp>
  <!-- more divisions -->
  <note place=foot target=X2>(1)DIRECTEUR DES TRAVAUX
</note>
```


4.2. Codage des caractères spéciaux

Les caractères de type textuel se différencient par leur typographie manuscrite (au lieu de dactylographique), leur taille, le type de graphisme utilisé (gras, souligné, etc.), leur sens (lettres inversées avec effet de miroir, mot inversé de la gauche vers la droite, etc.), leur origine linguistique (langue, type d'alphabet, etc.).

Il faut savoir que pratiquement tous les systèmes informatiques fournissent les 52 lettres de l'alphabet latin,

a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

ainsi que les 10 chiffres de l'écriture décimale,

0 1 2 3 4 5 6 7 8 9

et quelques symboles de base,

" % & ' () * + , - . / : ; < = > ? _ (espace)

D'autres caractères latins, tels que les diacritiques (e.g. ä ou é) ou les caractères d'origine autre que latine (e.g. grec, hébreu, arabe, cyrillique, etc.) sont moins courants. Encore moins fréquents sont les caractères scripts orientaux de par le problème posé par leur taille.

Cela signifie que pour avoir droit au plein emploi de ces caractères, il faudra modifier les déclarations SGML pour que tous les caractères soient définis en tant que caractères légaux. Ceci sans oublier que, la plupart du temps, ces caractères sont manuscrits, ce qui complique leur reconnaissance par un quelconque parseur. Et en supposant que les systèmes informatiques reconnaissent l'écriture manuscrite, ils devront également reconnaître les différentes calligraphies (similaire à la police de caractères en typographie).

Le changement de calligraphie est l'une des caractéristique des textes des albums d'Astérix. Si l'on désire marquer ce changement, le seul élément existant en TEI et qui semble le plus approprié à cet effet est l'élément **<distinct>**. Cependant, cet élément est employé pour représenter un langage linguistiquement distinct du reste du texte. Dans le cas qui nous préoccupe, la distinction se fera au niveau calligraphique:

```

<sp><distinct type=calligraphic>BIENVENUE A BORD! JE M'APPELLE KAREDA
ET JE SUIS MARCHAND GREC</distinct>
</sp>
<sp who='Astérix'>ASTERIX. GAULOIS. VOUS N'AURIEZ PAS VU TOMBER UN
FAKIR PAR HASARD?
</sp>
[Annexe 12, Figure 1]

```

Les changements de taille et d'épaisseur des lettres ont souvent en BD pour but de mesurer une intensité sonore, une variation vocale. A cet effet, la TEI propose l'élément **<shift>** auquel sont associés différents traits para linguistiques. L'attribut **feature** (pour trait para linguistique) contient des valeurs correspondant au type de trait (e.g. *loud*, *rhythm*, *voice*, etc.). L'attribut **new** intervient en complément pour indiquer la valeur du trait spécifié dans l'attribut **feature**. Parmi ces dernières valeurs associées aux traits, on peut retenir pour le codage de la BD:

- *loud* (pour intensité du son)

f	forte (fort)
ff	fortissimo (très fort)
cresc	crescendo (en grandissant)
p	piano (doux)
pp	pianissimo (très doux)
dimin	diminuendo (en diminuant)

- *voice* (pour voix)

whisp	whisper (murmure)
breath	breathy (voix haletante)
husk	husky (voix enrouée)
creak	creaky (qui grince, qui craque)
fals	falsetto (voix de fausset)
reson	resonant (voix sonore)
giggle	giggle (fou rire)
laugh	laugh (rire)
trem	tremulous (voix tremblante, chevrotante)
sob	sobbing (voix sanglotante)
yawn	yawning (bâillement)
sigh	sighing (soupir)

Par exemple, un son semblant se rapprocher pourra prendre pour valeur *cresc*.

ex: <sound feature=loud new=cresc>RRRRRRRR</sound>

[Hergé 66, case 2, page 1]

Le trait formant une bulle peut également inspirer de tels changements.

ex: `<sound feature=voice new=trem>Bê-ê-ê-ê...</sound>` [Hergé 66, case 3, page 9]

4.3. Les moyens nécessaires à l'implémentation de la BD sous format électronique

On ne peut parler de codage d'après le standard SGML si l'on n'aborde pas le problème des technologies nécessaires à l'implémentation de documents électroniques

"The main task of electronic document management is the creation, storage, processing, maintenance, distribution and eventual destruction of documents in electronic form in order to provide improved working efficiency and support for people and organisations who need or have to rely upon such documents in the course of their business". [Van Hoof 94]

En "surfant" sur Internet du côté des bandes dessinées, on remarquera qu'il existe déjà des moyens de stockage de ces dernières. En effet, SGML propose quelques balises permettant de les entrer dans les systèmes informatiques [réf.4].

La balise **<figure>** indique la position d'un graphisme, d'une illustration ou d'une figure. Le nom de fichier correspondant à l'image digitale est appelé par la valeur de l'attribut **entity**:

`entity="NOM DE FICHIER".`

Déjà ici, on peut noter la limite d'un tel système. Le genre de la BD est principalement formé d'images au sein desquelles est présent un texte. Cela signifie que la balise **<figure>** est très limitée dans l'espace et ne réfère qu'à la BD en général, c'est-à-dire sans distinction des éléments (graphiques ou textuels) qui la composent.

Mais poursuivons la démonstration. Considérons l'illustration extraite de *The Flirt* de Booth Tarkington [Annexe 15]. En contenu de la balise **<figure>**, on a les éléments:

- **<head>** peut être utilisé pour retranscrire l'entête ou le titre d'un graphisme:

```
<figure entity="TarFl122">
  <head>"Kiss me some more darl----"</head>
</figure>
```

- **<figDesc>** contient une courte description du contenu de l'illustration:

```
<figure entity="TarFl122">
  <head>"Kiss me some more darl----"</head>
  <figDesc>Grayscale illustration of a young girl trying to kiss a boy, under
  moonlight.</figDesc></figure>
```

Comme on l'a vu dans les parties 4.1. et 4.2., la reconnaissance du contenu de la BD requiert d'autres éléments. Apparemment, aucun outil actuel ne peut effectuer une telle reconnaissance. Il existe cependant des outils tels que ImageMagick (disponible sous W3 à l'adresse <ftp.x.org/contrib/applications/ImageMagick>) qui permet de convertir l'image d'un format à un autre (e.g. TIFF vers JPEG) et d'ajouter les balises délimitant les textes en entête et le reste de l'image, comme nous l'avons vu précédemment. Ce système laisse espérer que dans les prochaines années, les systèmes distingueront le texte du graphisme dans un document illustré quelconque.

Le premier pas vers l'informatisation des documents consiste à insérer ces documents sous forme papier à l'intérieur de l'ordinateur. C'est le rôle des scanners digitaux. Ceux-ci reproduisent des copies digitales des pages du document qui sont ensuite stockées en images *bitmap*. Une image *bitmap* est constituée de *pixels* (suite de petits points) et non de caractères textuels lisibles. Cela signifie que ces images peuvent être visibles à l'écran ou imprimées, mais sûrement pas "comprises" par l'ordinateur.

Un système complémentaire permet la conversion des images *bitmap* en texte *ASCII*, qui lui est lisible par l'ordinateur. Ce système est connu sous le sigle OCR (abréviation de *Optical Character Recognition*). Evidemment, ce système n'est pas encore paré à toute épreuve, et certainement pas à l'écriture manuscrite des textes de BD.

Les premiers systèmes OCR ne pouvaient reconnaître que des caractères spécifiques. Il arrivait même qu'ils reconnaissent des "non-caractères". Les systèmes modernes usent d'une nouvelle technologie, appelée *omnifont* (tout caractère) car elle permet de reconnaître une grande variété de caractères en utilisant les techniques d'intelligence artificielle qui extraient les caractéristiques des lettres. Parce que cette technologie est empruntée à l'intelligence artificielle, ces systèmes sont souvent appelés *Intelligent Character Recognition* (ICR). Un système permettant la conversion des images *bitmap* en un système rendant possible la reconnaissance du texte manuscrit reste donc encore à être développé.

Comme on peut le constater, loin s'en faut d'avoir découvert un système réellement performant quant à la reconnaissance visuelle des différents types de graphismes. Qu'un système puisse reconnaître Tintin et sa houppette ou Spirou et son costume de groom paraît être une idée quelque peu extravagante de nos jours. Cependant, si des propositions du type de ce projet et l'intérêt sont présents, cela ne peut qu'accélérer les efforts en matière de science informatique et surtout en matière de reconnaissance des documents illustrés.

5. CONCLUSION

Techniquement parlant, on a vu que le codage de la bande dessinée, en passant par son stockage sous la forme d'un document lisible par l'ordinateur, peut sembler véritablement un travail ardu et hors de portée des systèmes existants. Linguistiquement parlant, cela peut représenter un atout fort intéressant. Les buts du développement des documents électroniques sont multiples. Actuellement, on remarque une effervescence en matière de nouveaux moyens de communication inclus sous le terme générique de multimédia.

La BD, tout comme les autres documents imprimés mérite son qualificatif de "moyen d'expression". De par son aspect populaire et ludique, elle a même été adoptée par d'autres médias à des fins didactiques et publicitaires. On peut donc considérer celle-ci comme un média à part entière qui a donc droit à une place d'honneur dans le monde de l'informatique.

Le but de formats de codage tels que SGML est de pouvoir mettre à disposition sur support informatique des documents susceptibles d'être transmis d'un point du globe à un autre ou d'un siècle à un autre, et ceci en vue de n'importe quel post-traitement (stockage, extraction ou échange des données, analyse de documents, etc.).

Multimédias en tête, de par leur aspect également ludique (bien qu'à but culturel) peuvent exprimer un certain intérêt à s'approprier le genre de la BD. De plus, les albums de Tintin ont déjà fait l'objet d'analyses linguistiques et psychologiques. Le langage employé en BD peut aussi faire l'objet d'autres sortes d'utilisations à but principalement linguistique: compréhension du texte, alignement multilingue, établissement de dictionnaires, etc.

De plus, ce développement des systèmes de traduction informatique peut aider à l'élargissement du public de la BD qui bien souvent reste confiné à l'intérieur d'un seul et même territoire. Tintin et ses 42 versions en langues étrangères (dont le féroïcien, l'asturien, le romanche, l'espéranto, le luxembourgeois, etc.) est l'un des rares exemples contredisant cette règle.

SOURCES DOCUMENTAIRES

(1) Ouvrages consultés

- [Amiel 86] Amiel V., *Bande dessinée et architecture: l'espace encadré*. In: Les Cahiers de la Bande Dessinée, bimestriel n° 69, éd. Glénat, pp 24-25, Mai-Juin 1986.
- [Baudelaire 89] Baudelaire C., *L'Albatros*. In: Les Fleurs du Mal, éd. Presses Pocket, Usine de la Flèche, France, p. 32, 1989.
- [Corneille 70] Corneille, Le Cid. Coll. Classiques Larousse, éd. Larousse, Paris, France, 1970.
- [Duc 82] Duc B., L'Art de la B.D., éd. Glénat et B. Duc, Luçon, France, 1982.
- [Filippini 89] Filippini H., Dictionnaire de la Bande Dessinée, éd. Bordas, Poitiers, France, 1989.
- [Van Hoof 94] Van Hoof T., Smude P., Document Management Report, éd. Cambridge Market Intelligence, Londres, Grande-Bretagne, 1994.

(2) Albums de Bandes Dessinées

- [De Gieter 96] De Gieter, coll. Papyrus, *Les Momies Maléfiques*. In: [Spirou 96a], p. 30, 1996.
- [Fournier 76] Fournier, coll. Spirou et Fantasio, n° 21, *Du Glucose pour Noémie*, éd. Dupuis, Charleroi, Belgique, p. 40, 1976.
- [Franq-Jid 77] Franquin, Jidéhem, Greg, coll. Spirou et Fantasio, n° 15, *Z comme Zorglub*, éd. Dupuis, Charleroi, Belgique, p. 52, 1977.
- [Francq-Van H. 90] Francq P., Van Hamme J., coll. Largo Winch, *L'Héritier*, éd. Dupuis, Belgique, p. 34, 1990.

- [Gos 95] Gos, coll. Le Scrasmeustache, *Les Naufragés du Chastang*. In: [Spirou 95c], p. 16, 1995.
- [Graton 83] Graton J., coll. Michel Vaillant, n° 43, *Rendez-vous à Macao*, éd. Graton, Bruxelles, Belgique, p. 33, 1983.
- [Hergé 66] Hergé, coll. Les Aventures de Tintin, *L'Ile Noire*, éd. Casterman, Belgique, 1966.
- [Jacobs 86] Jacobs E. P., coll. Les Aventures de Blake et Mortimer, *Le Mystère de la Grande Pyramide*, tome 1, éd. Blake & Mortimer, Bruxelles, Belgique, p. 51, 1986.
- [Lambil-Cauvin 84] Lambil W., Cauvin R., coll. Les Tuniques Bleues, n° 21, *Les Cinq Salopards*, éd. Dupuis, Marcinelle, Belgique, p. 11, 1984.
- [Malet-Tardi 88] Malet L., Tardi, *120, rue de la Gare*, éd. Casterman, Tournai, Belgique, p. 155, 1988.
- [Michetz-Bosse 92] Michetz, Bosse, coll. Kogaratsu, n° 4, *Le Dos du Tigre*, p. 29, 1992.
- [Michetz-Bosse 94] Michetz, Bosse, coll. Kogaratsu, n° 5, *Par-delà les Cendres*, éd. Dupuis, Belgique, p. 16, 1994.
- [Morris-Fauche 84] Morris, Fauche X., Léturgie J., coll. Lucky Luke, *Le Daily Star*, éd. Dargaud, Turin, Italie, p. 5, 1984.
- [Olis-Gilson 95] Olis, Gilson, *Garage Isidore*. In: [Spirou 95a], p. 29, 1995.
- [Ortiz-Segura 93] Ortiz J., Segura, coll. Morgan, *Tome 5 - Petits Nègres et Négriers*, éd. Soleil, Strasbourg, France, 1982.
- [Roba 76] Roba, coll. Boule et Bill, n° 5, *60 Gags de Boule et Bill*, éd. Dupuis, Marcinelle-Charleroi, Belgique, p. 3, 1976.
- [Rosinski-Van H. 84] Rosinski G., Van Hamme J., coll. Thorgal, n° 2, *L'Ile des Mers Gelées*, éd. du Lombard, Bruxelles, Belgique, 1984.
- [Rosy 95] Rosy. In: [Spirou 95b], p. 5, 1995.
- [Séron 95] Séron, coll. Les Petits Hommes, *Melting Pot*. In: [Spirou 95c], p. 34, 1995.

- [Tabary 89] Tabary, coll. Les Aventures du Grand Vizir Iznogoud, *Iznogoud enfin Calife!*, éd. Tabary, Pont l'Abbé d'Arnoult, France, p. 9, 1989.
- [Tome-Janry 84] Tome, Janry, coll. Spirou et Fantasio, n° 33, *Virus*, éd. Dupuis, Belgique, p. 20, 1984.
- [Tome-Janry 87] Tome, Janry, coll. Spirou et Fantasio, n° 39, *Spirou à New York*, éd. Dupuis, Charleroi, Belgique, p. 5, 1987.
- [Tome-Janry 90] Tome, Janry, coll. Spirou et Fantasio, n° 42, *Spirou à Moscou*, éd. Dupuis, Belgique, p. 4, 1990.
- [Uderzo 87] Uderzo R., coll. Astérix, *Astérix chez Rahāzade*, éd. Albert René, Bergamo, Italie, 1987.

(3) Hebdomadaires

- [Spirou 95a] Spirou, hebdomadaire n° 2992, éd. Dupuis, 16 août 1995.
- [Spirou 95b] Spirou, hebdomadaire n° 2993, éd. Dupuis, 23 août 1995.
- [Spirou 95c] Spirou, hebdomadaire n° 2997, éd. Dupuis, 20 septembre 1995.
- [Spirou 96a] Spirou, hebdomadaire n° 3014, éd. Dupuis, 17 janvier 1996.

(4) Références WEB

- [réf.1] <http://www.sil.org/sgml/edw25.html>
Burnard L., *TEI EDW25: What is SGML and How Does It Help?*, 1991.
- [réf.2] <http://etext.virginia.edu/TEI.html>
TEI Guidelines for Electronic Text Encoding and Interchange.
- [réf.3] <http://www.uic.orgs/tei/intros/teiu5.html>
Burnard L., Sperberg-McQueen C. M., *TEI Lite: An Introduction to Text Encoding for Interchange*, 1995.
- [réf.4] <http://www.lib.virginia.edu/etext/tei/uvatei11.html>
Seaman D., *Illustrations*.

TABLE DES ANNEXES

ANNEXE 1: Sunday Page du Chicago Tribune [Filippini 89]	i
ANNEXE 2: L'utilisation des sous-titres en BD	ii
Figure 1: Planches textuelles partageant le récit en 4 sous-parties [Ortiz-Segura 93]	ii
Figure 2: Histoire courte avec entête [Roba 76]	iii
ANNEXE 3: Les différents types de cadres	iv
Figure 1: Cadre horizontal montrant une course automobile [Graton 83]	iv
Figure 2: Cadre vertical soulignant une chute [Francq-Van H. 90]	iv
Figure 3: Cadre intérieur [Michetz-Bosse 94]	v
Figure 4: Cadre reliant d'autres cadres [Michetz-Bosse 92]	v
ANNEXE 4: Exemple de cadres non alignés [Olis-Gilson 95]	vi
ANNEXE 5: Exemple de système de fléchage [Gos 95]	vii
ANNEXE 6: Narration de faits antérieurs au moyen de traits en "vagues" [Tabary 89]	viii
ANNEXE 7: La bulle en tant que borne temporelle	ix
Figure 1: Bulle pointant vers un personnage apparaissant dans la case suivante [Rosinski-Van H. 80, page 45]	ix
Figure 2: Bulle pointant vers un personnage d'une case précédente [Rosinski-Van H. 80, page 47]	ix
ANNEXE 8: Suivi des bulles dépendant du suivi du dessin [Séron 95]	x
ANNEXE 9: Indications scéniques	xi
Figure 1: Encadré reliant deux cases juxtaposées [De Gieter 96]	xi
Figure 2: Encadrés à buts narratifs [Malet-Tardi 88]	xi
ANNEXE 10: Petit dictionnaire des onomatopées (extrait de [Duc 82])	xii
ANNEXE 11: Petit dictionnaire des symboles et graffitis [Duc 82]	xv

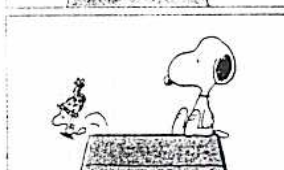
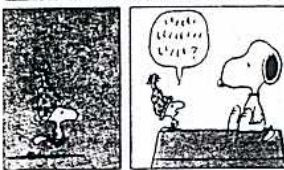
<u>ANNEXE 12: Les langues "étrangères" en bande dessinée</u>	<u>xvii</u>
Figure 1: L'accent grec dans Astérix [Uderzo 87, page 19]	xvii
Figure 2: Usage du latin en BD [Uderzo 87, page 16]	xvii
Figure 3: "De Diou" en cyrillique [Tome-Janry 84]	xviii
Figure 4: Article de journal en français et en anglais [Morris-Fauche 84]	xviii
Figure 5: Page de Rosy [Rosy 95]	xix
Figure 6: Caractères illisibles exprimant le sermon d'un prêtre [Tome-Janry 87]	xx
Figure 7: Inversion des mots [Franq-Jid 77]	xx
Figure 8: Inversion des lettres [Tome-Janry 90]	xx
<u>ANNEXE 13: Les appels de note</u>	<u>xxi</u>
Figure 1: Appel de note inachevé [Lambil-Cauvin 84]	xxi
Figure 2: Appel de note à but humoristique [Fournier 76]	xxi
<u>ANNEXE 14: Planche à coder [Jacobs 86]</u>	<u>xxii</u>
<u>ANNEXE 15: The Flirt, Booth Tarkington [réf.4]</u>	<u>xxiii</u>

ANNEXE 1: Sunday Page du Chicago Tribune [Filippini 89]

That fabulous Onassis: his life and loves
—series starts in Lifestyle



PEANUTS by SCHULZ



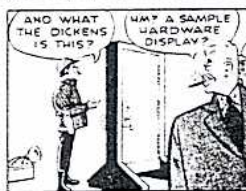
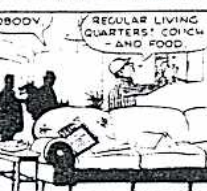
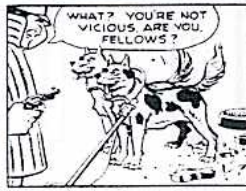
Chicago Tribune COMICS

SUNDAY • APRIL 30, 1972 • SECTION 8

CRIMESTOPPERS

BEWARE OF FLY-BY-NIGHT CUNNINGS
AND POOF REPAIR MEN WHO BY DECEIT
TAKE ADVANTAGE OF THE ELDERLY
AND THE GULLIBLE. CHECK!

DICK TRACY



FUN FACTS

fun things to know and tell.

JACA COURIER SYSTEMS IN USE 700 YEARS AGO OFTEN CARRIED MESSAGES 150 MILES A DAY... BY FOOT.

THE COLLOSSEUM IN ROME HELD MORE THAN FIFTY THOUSAND SPECTATORS.

A BIRD'S EYES OFTEN WEIGH MORE THAN ITS BRAIN.

YOU GET MORE TOP-QUALITY FULL-SIZE STICKS OF WHOLEY'S SPEARMINT GUM IN THE HANDY DIME PACK. CARRY IT WITH YOU.

IN THE 1860S, A \$10,000 PRIZE WAS OFFERED TO ANYONE WHO COULD DISCOVER A SUBSTITUTE FOR IVORY IN BILLIARD BALLS.

IF A MAN WERE AS STRONG FOR HIS SIZE AS MANY INSECTS, HE COULD EASILY CARRY A CAR ON HIS BACK.

ANNEXE 2: L'utilisation des sous-titres en BD

Figure 1: Planches textuelles partageant le récit en 4 sous-parties [Ortiz-Segura 93]

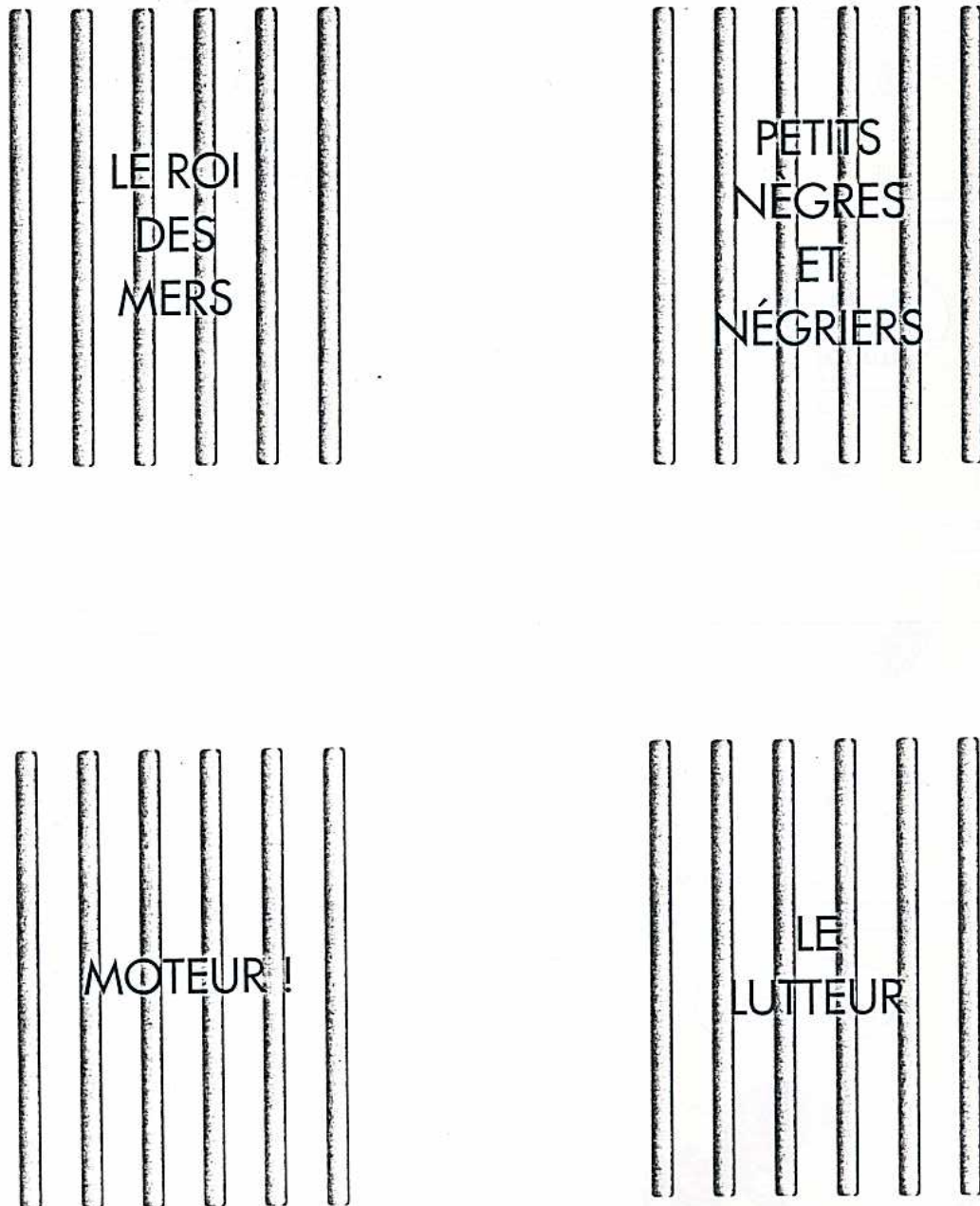
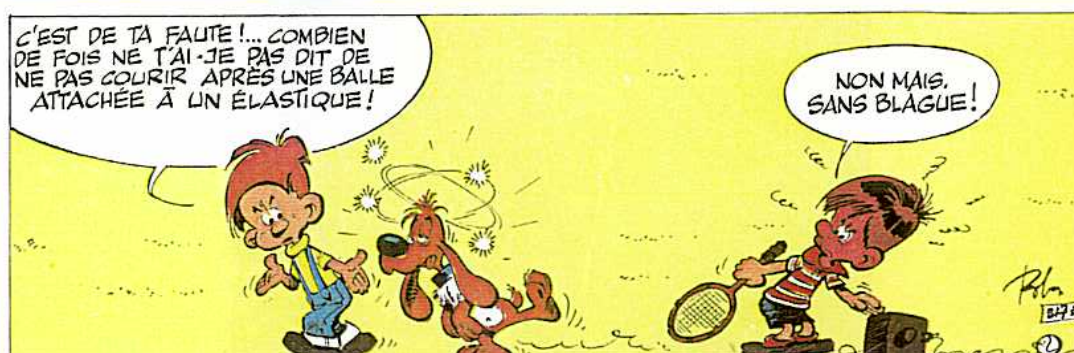
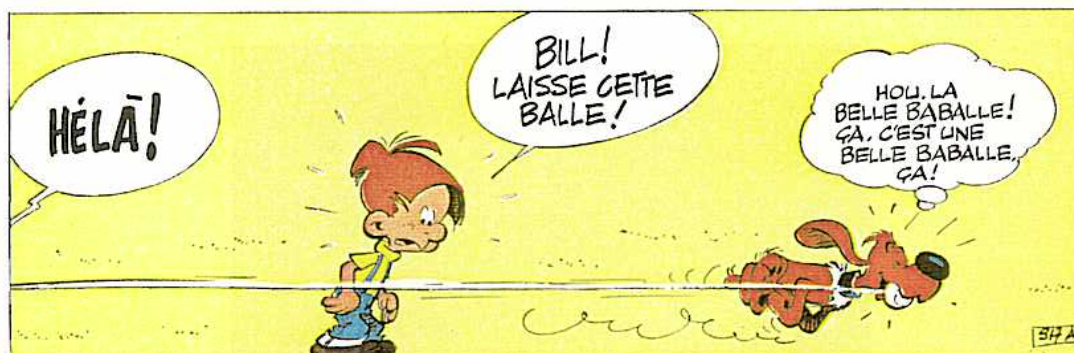
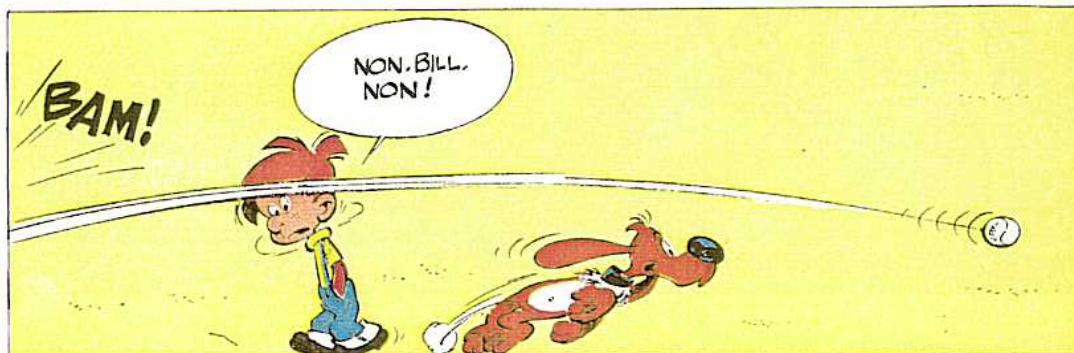


Figure 2: Histoire courte avec entête [Roba 76]



IL FAIT DES BONDS, IL FAIT DES BONDS...



ANNEXE 3: Les différents types de cadres

Figure 1: Cadre horizontal montrant une course automobile [Graton 83]

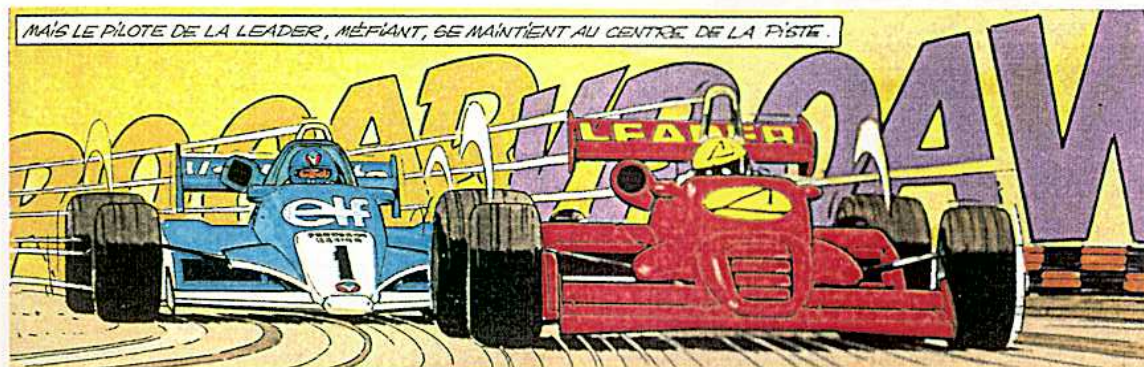


Figure 2: Cadre vertical soulignant une chute [Francq-Van H. 90]

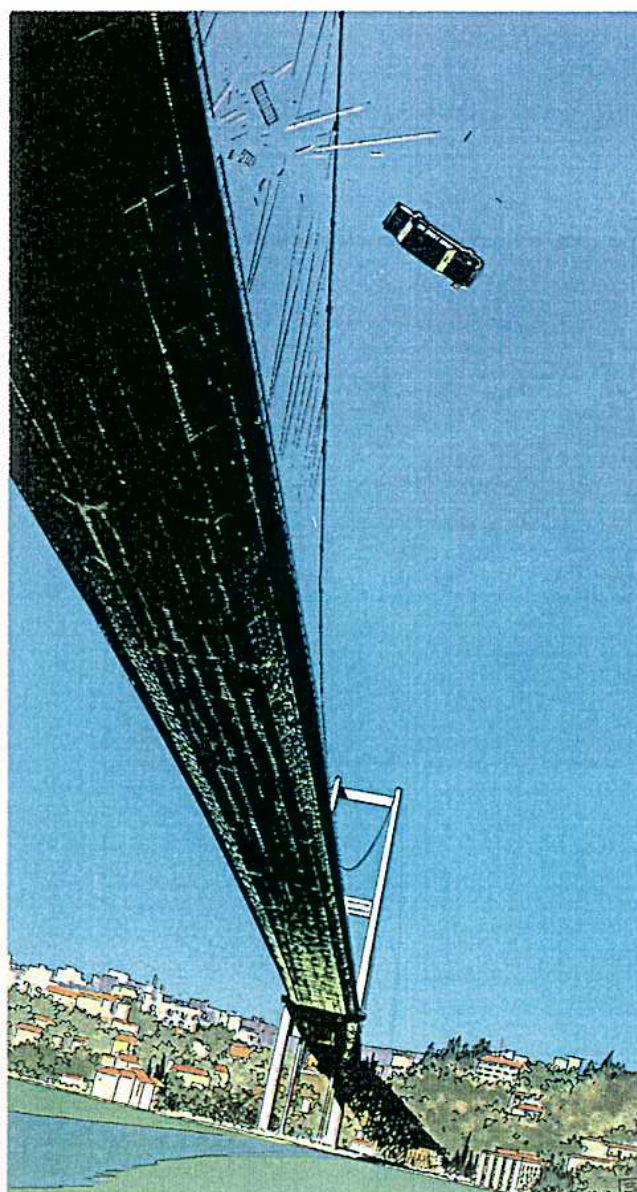


Figure 3: Cadre intérieur [Michetz-Bosse 94]

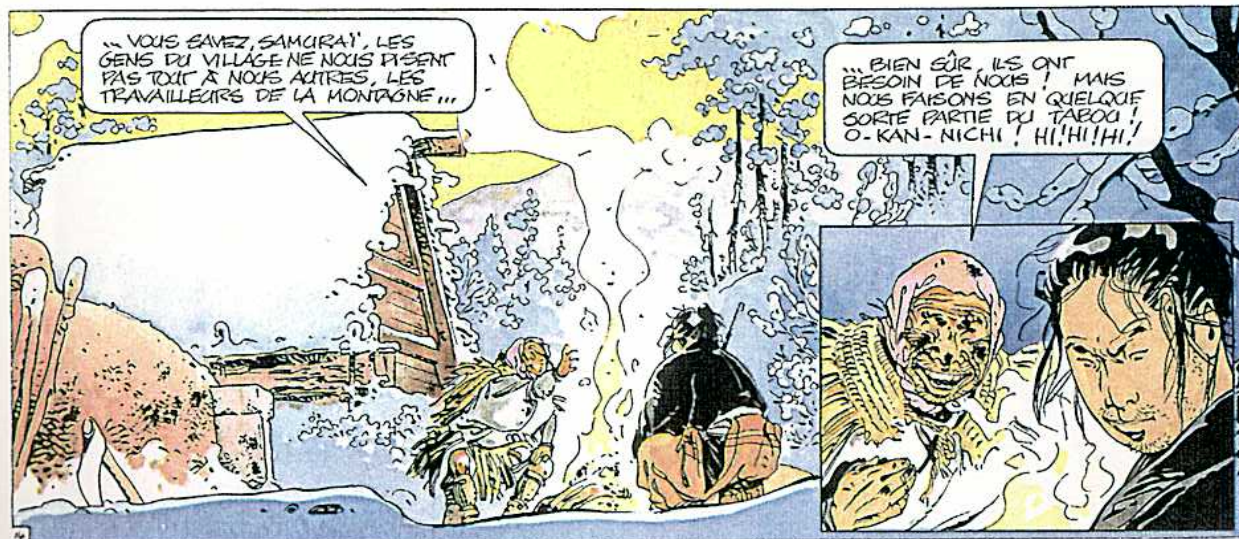
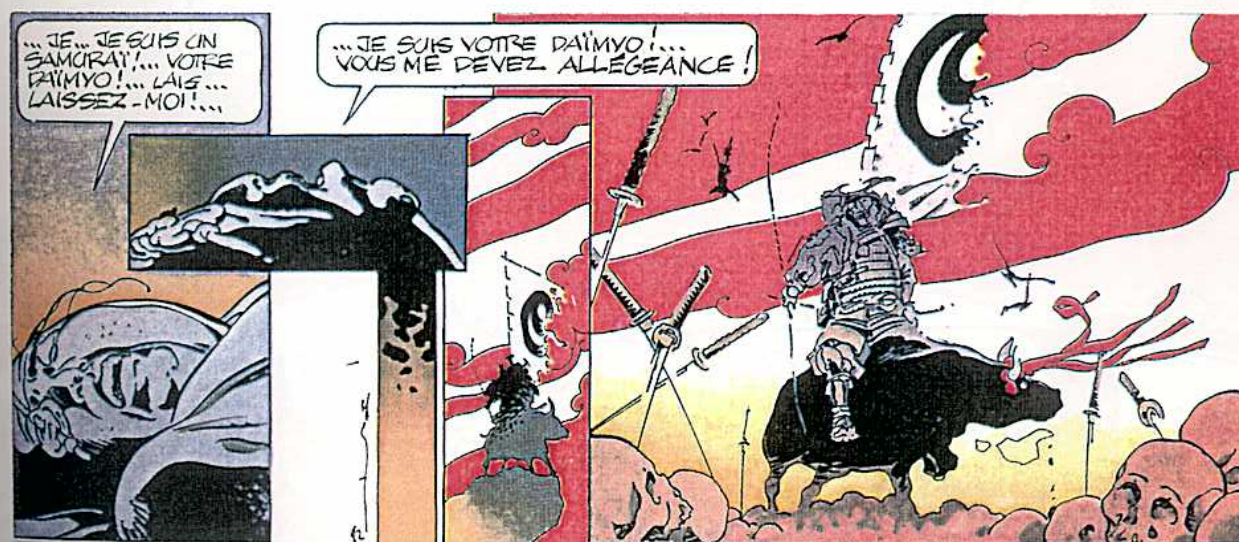


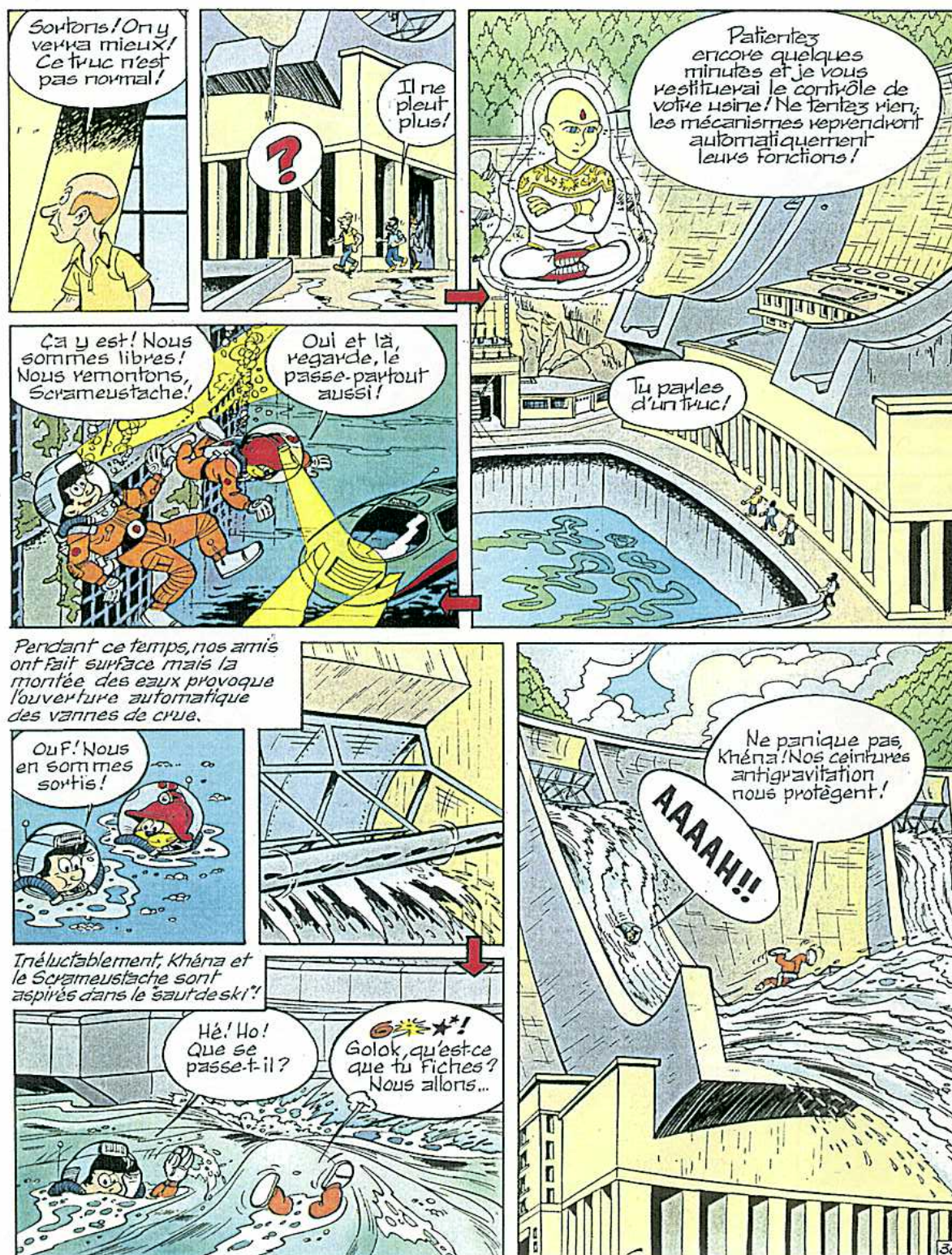
Figure 4: Cadre reliant d'autres cadres [Michetz-Bosse 92]



ANNEXE 4: Exemple de cadres non alignés [Olis-Gilson 95]



ANNEXE 5: Exemple de système de fléchage [Gos 95]



ANNEXE 6: Narration de faits antérieurs au moyen de traits en "vagues" [Tabary 89]



...EN MA FAVEUR IL S'EST MIS À ME FAIRE DES M. AGUES MORTELLES! LE COUP DE LA GAMELLE AU DESSUS DE LA PORTE, PAR EXEMPLE! - C'EST PAS DANGEREUX ÇA, PATRON!



SAUF QUE LA GAMELLE ÉTAIT EN FONTE!!

...ET TOUT DE SUITE APRÈS LA PEAU DE BANANE PLACÉE EN HAUT DU GRAND ESCALIER! BOOM MARCHES SUR TOUS LES MEMBRES!



LE JOUR MÊME, POUR SE FAIRE PARDONNER IL VOUS OFFRAIT UNE MAGNIFIQUE PÂTÈQUE! LA PLUS BELLE QUE LE SOL EST PRODUIT. VOUS COURÂTES VERS ELLE, FLATTÉ, HEUREUX, MORMAND! HA! HA! HA! - NE RIS PAS !!!



...IL L'AVAIT PLACÉE AU MILIEU DE L'ENDROIT OÙ L'ON ENTERRE LES DÉCHETS. LES DÉTRITUS, LES IMMONDICES, LES ORDURES "ÉVACUÉS" DU PALAIS...



ANNEXE 7: La bulle en tant que borne temporelle

Figure 1: Bulle pointant vers un personnage apparaissant dans la case suivante [Rosinski-Van H. 80, page 45]



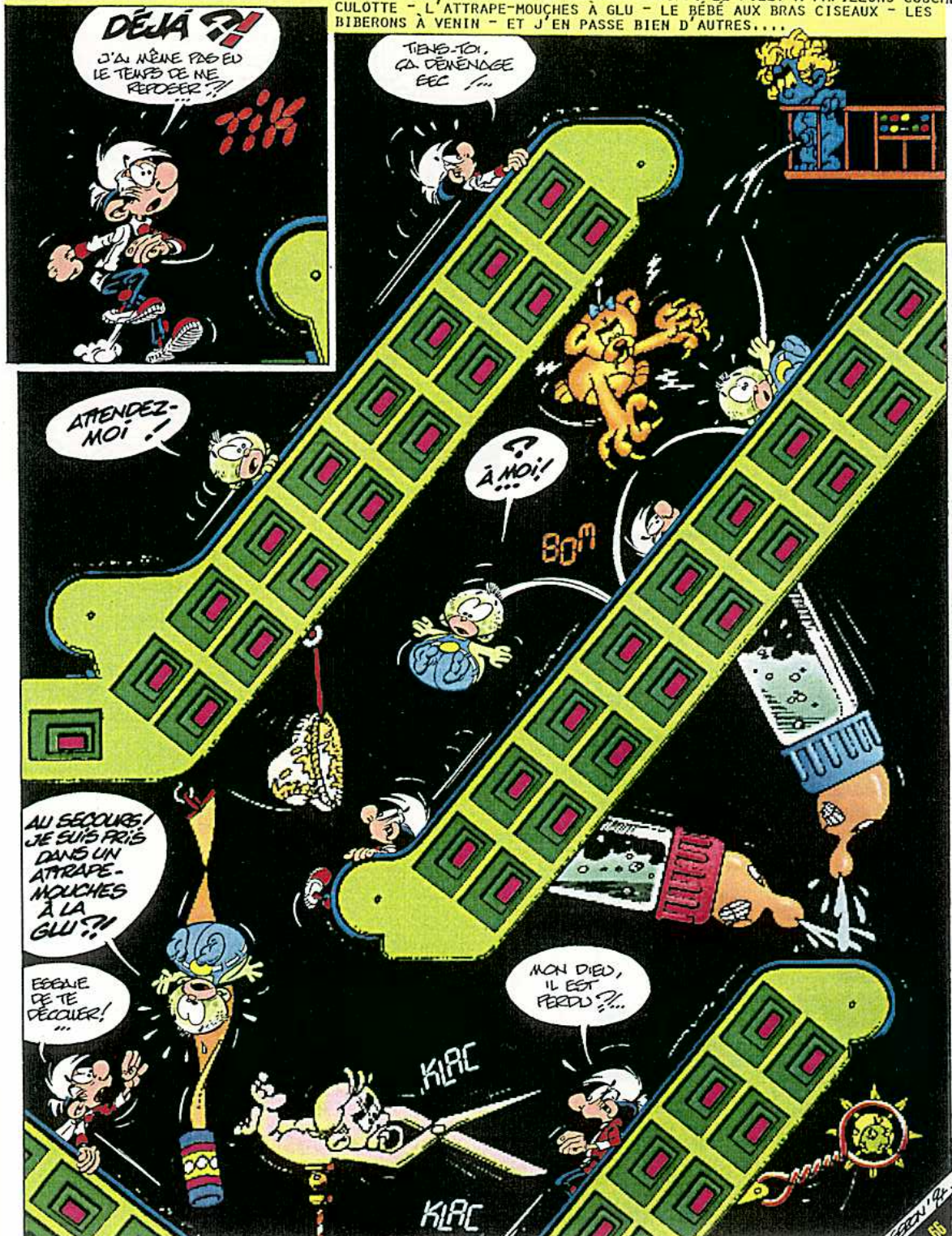
Figure 2: Bulle pointant vers un personnage d'une case précédente [Rosinski-Van H. 80, page 47]



ANNEXE 8: Suivi des bulles dépendant du suivi du dessin [Séron 95]

Renaud est toujours prisonnier du monde de jeux vidéo de Tchakakahn.

TIK...C'EST REPARTI...JE VAIS VOUS FAIRE GÔTER À UN TOUT NOUVEAU JEU QUE J'AI INTITULÉ "POUPONNIÈRE"... VOUS ALLEZ Y RENCONTRER LE MANNEKEN PIS DE L'ACIDE SULFURIQUE-LE NOUVEAU GRIFFU-LE FILET À PAPILLONS COUCHE-CULOTTE - L'ATTRAPE-MOUCHES À GLU - LE BÉBÉ AUX BRAS CISEAUX - LES BIBERONS À VENIN - ET J'EN PASSE BIEN D'AUTRES....



ANNEXE 9: Indications scéniques

Figure 1: Encadré reliant deux cases juxtaposées [De Gieter 96]

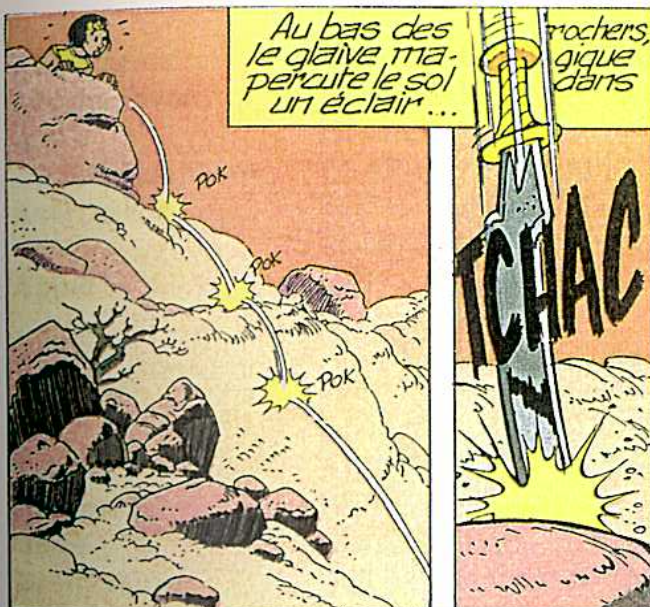
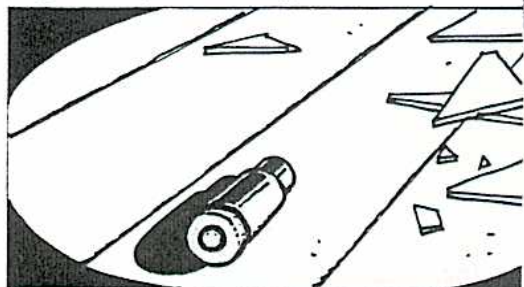


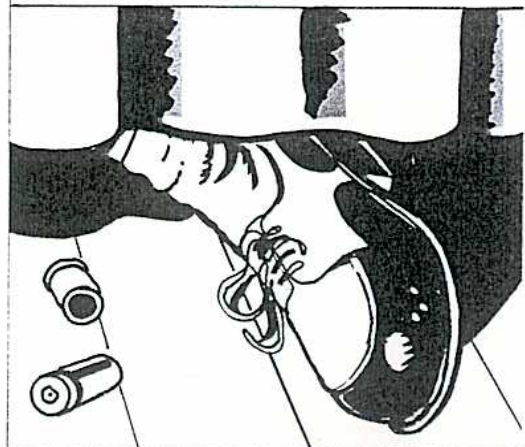
Figure 2: Encadrés à buts narratifs [Malet-Tardi 88]



Charles venait de trouver une douille éjectée par un Luger.



Un peu plus loin, nous découvrîmes deux autres douilles d'un calibre différent et un pied qui pointait sous les plis d'un rideau...



ANNEXE 10: Petit dictionnaire des onomatopées (extrait de [Duc 82])

A

Aboiements:	OUAH! OUAH! - WHAF! - WHOUAF!
Acclamations	HOURRA!
Ane	HI-HAN
Applaudissements	CLAP-CLAP! KLAP!
Avertisseur	TUT! - TUUUT! - T-U-U-U-U-T!; (vieux klaxon) COIN-COIN! - POUET-POUET!; (voiture de pompier) PIN-PON!
Appel	HEP! - HELP! - HE! - OHÉ! - OOOHÉÉÉ!

B

Bébé	AREU-AREU
Bêlement	BÊÊÊÊÊÊÊÊ
Boire	GLOU-GLOU! - GOUP-GLOUP!
Bruits sourds	BOUM! - BADAMOUM! - RAMDAM!

C

Chat	(miaulement) MIAOU!; (lapant du lait) LAP-LAP
Chute	POUF - PAF - PATATRAS
Claquement	CLAC - SPLATCH - CLAC-CLAC
Cloche	DING-DONG
Coq	COCORICÔÔÔÔÔ
Coup	(sec) PAF - CRAC; (coup de feu) BANG - PAN - PAW; (mitrailleuse) TATAATATA - TACATACATA
Cri	(de douleur) OUAÏE! - AÏE!; (de joie) YÉÉÉÉÉH!; (d'étonnement) OOOH!

D

Déchirement	CRAAATCH - SCRAAATCH
Déclat	CLIC - CLAC - FLIP - PLIP

E

Ecrasement	POUF - PLOUF - CRAAAC - CRASH
Eternuement	HAT...HAT...HATCHOUM!
Etranglement	GARGL - ARGL

ANNEXE [Duc 82]

des symboles et graffiti

S

Scie Z Z Z Z Z Z Z
Sifflet TRIT
Sonnerie DRIIIING - DRIIIIN
Sonnette DRELIN-DRELIN

T

Tremblement AGLAGLA

V

Vent WHOU-OU-OU-OU...

ANNEXE 11: Petit dictionnaire des symboles et graffitis [Duc 82]

Symboles évoquant les émotions élémentaires (interrogation, exclamation, perplexité...):

Le point d'interrogation et le point d'exclamation. Naturellement, plus l'exclamation (ou l'interrogation) sera forte, plus leur graphisme sera épais. Souvent, on verra aussi les points d'exclamation ou d'interrogation se multiplier ou se dédoubler à la mesure de l'émotion ressentie.

Symboles évoquant les états d'âme des personnages:

Les états d'âme des personnages seront plus subtilement suggérés, soit par un seul symbole venant s'inscrire dans une "bulle", soit par une association de différents symboles dans la même "bulle", venant très précisément illustrer le sentiment qui anime le personnage. Ce sera, par exemple:

- *Les symboles d'agressivité:*

Crâne, tibias entrecroisés, pavillon de pirate, gants de boxe, pistolet ou canon faisant feu, grenade, bombe dont la mèche brûle, couteau planté dans un dos, marteau, explosions diverses, éclairs, etc.

- *Les symboles d'évasion (farniente, repos, paresse, etc.):*

Soleil, palmiers, bord de mer, plage, parasol, yacht, hamac, barreaux de prison tordus, etc.

- *Symboles d'optimisme (euphorie, joie, bonheur, etc.):*

Fleurs, papillons virevoltants, cloches sonnant à toute volée, etc.
Souvent associés à des symboles d'évasion (voir ci-dessus).

- *Symboles de réussite:*

Couronne de lauriers, auréole de Saint, statue, etc.

- *Symboles de réussite matérielle (richesse, etc.):*

Coffre bien rempli, tas de pièces d'or, diamant scintillant, gros cigare, etc. (voir aussi symboles d'optimisme).

- *Symboles de pauvreté (faim, etc.):*

Arêtes de poisson dans une assiette, ceinture au dernier cran, poches vides, corde de pendu, ossements, etc.

- *Symboles de pessimisme (idées noires, etc.):*

Corbillard, corde de pendu, tombe, barreaux de prison, guillotine, etc.

- *Symboles sentimentaux:*

Coeur, coeur brisé, guirlande de coeurs, coeur transpercé d'une flèche, "amour" mythologique tirant à l'arc, lèvres, jolie fille souriante, etc.

- *Symboles de trahison:*

Couteau planté dans le dos, coup de massue, visage masqué, etc. Souvent associés à des symboles de pessimisme (voir ci-dessus).

- *Symboles de duplicité:*

Personnage enveloppé dans une cape, visage masqué, etc.

- *Symbole de sommeil:*

Une scie coupant une bûche évoquera souvent les ronflements du dormeur.

- *Symboles musicaux:*

Notes de musique, portée musicale, chef d'orchestre, violoniste, etc.

- *Idee lumineuse:*

Une ampoule allumée ou une bougie (si l'idée est moins bonne), des points d'exclamation, etc.

ANNEXE 12: Les langues "étrangères" en bande dessinée

Figure 1: L'accent grec dans Astérix [Uderzo 87, page 19]



Figure 2: Usage du latin en BD [Uderzo 87, page 16]



Figure 3: "De Diou" en cyrillique [Tome-Janry 84]



Figure 4: Article de journal en français et en anglais [Morris-Fauché 84]

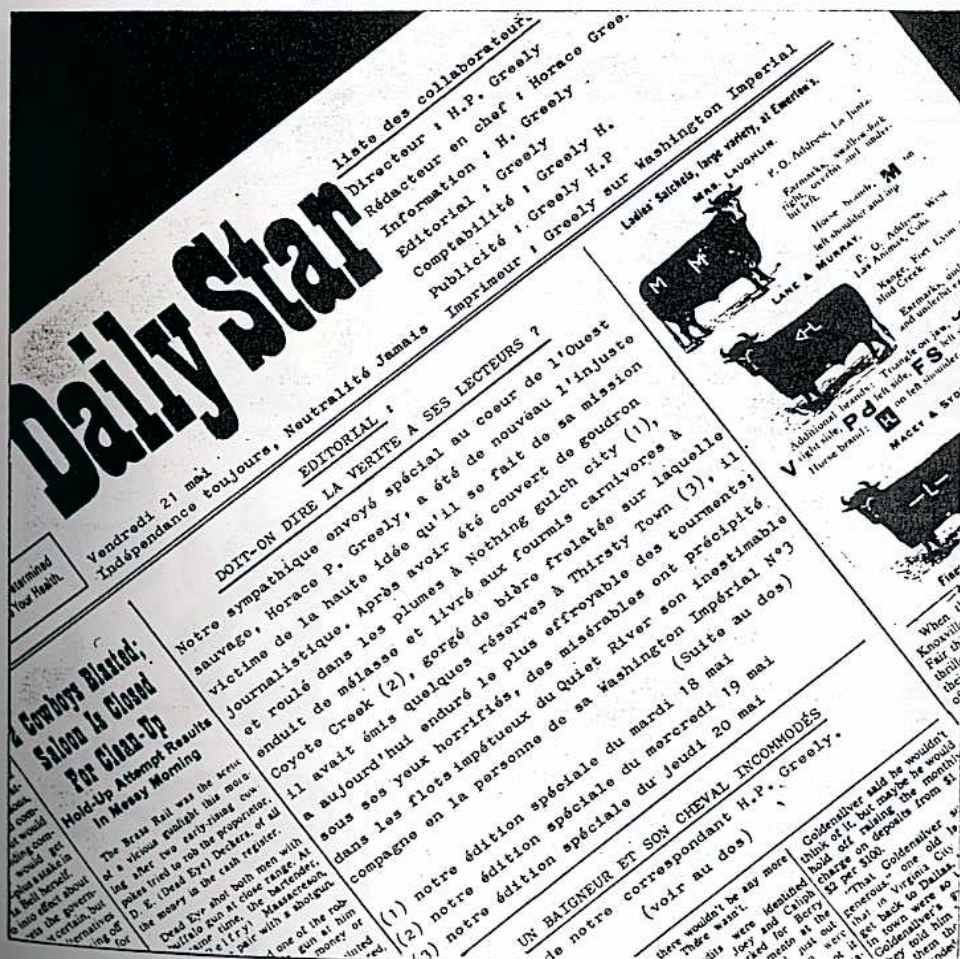


Figure 5: Page de Rosy [Rosy 95]

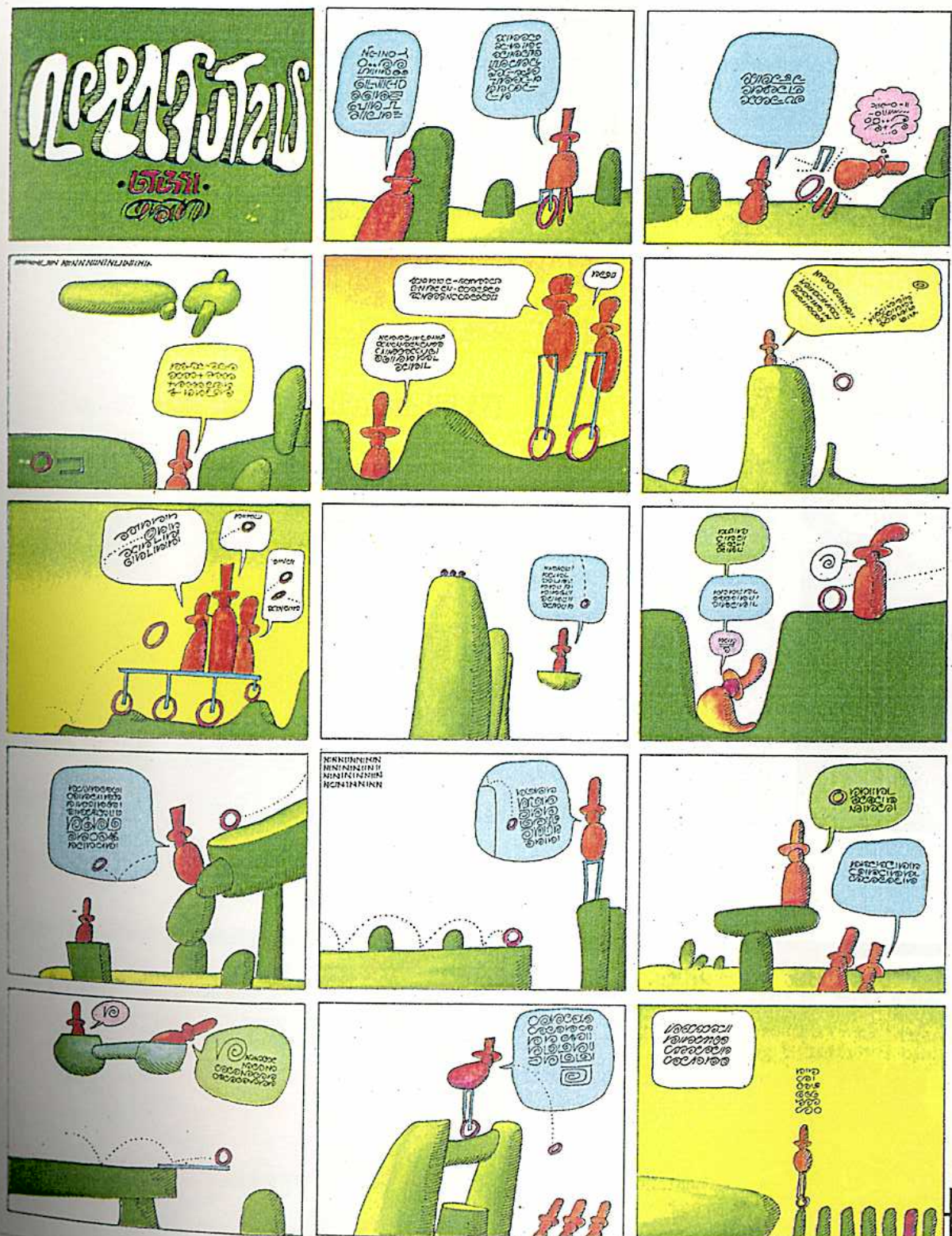


Figure 6: Caractères illisibles exprimant le sermon d'un prêtre [Tome-Janry 87]



Figure 7: Inversion des mots [Franq-Jid 77]

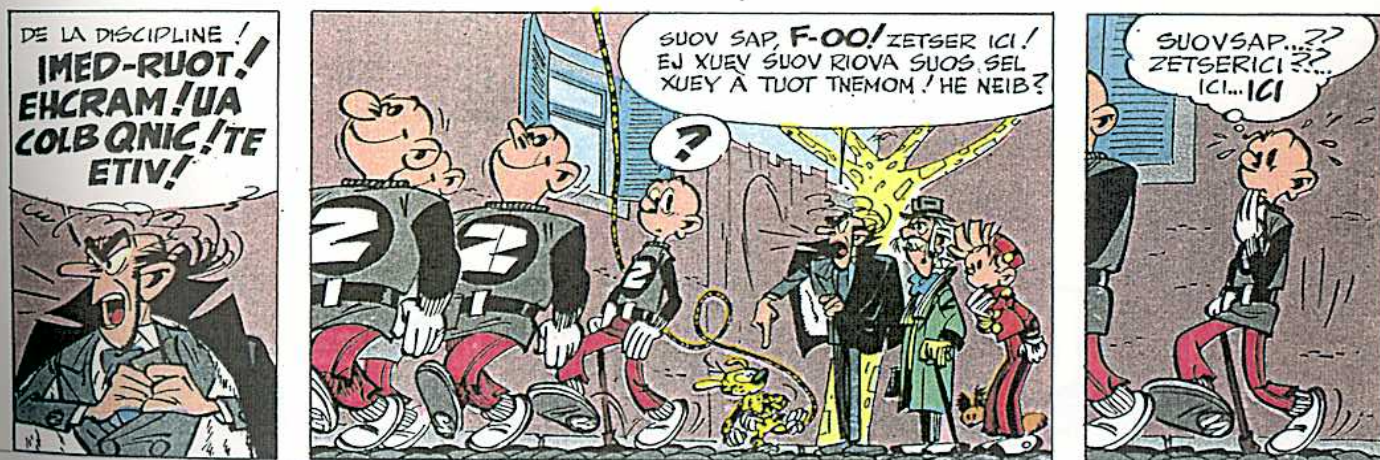


Figure 8: Inversion des lettres [Tome-Janry 90]



ANNEXE 13: Les appels de note

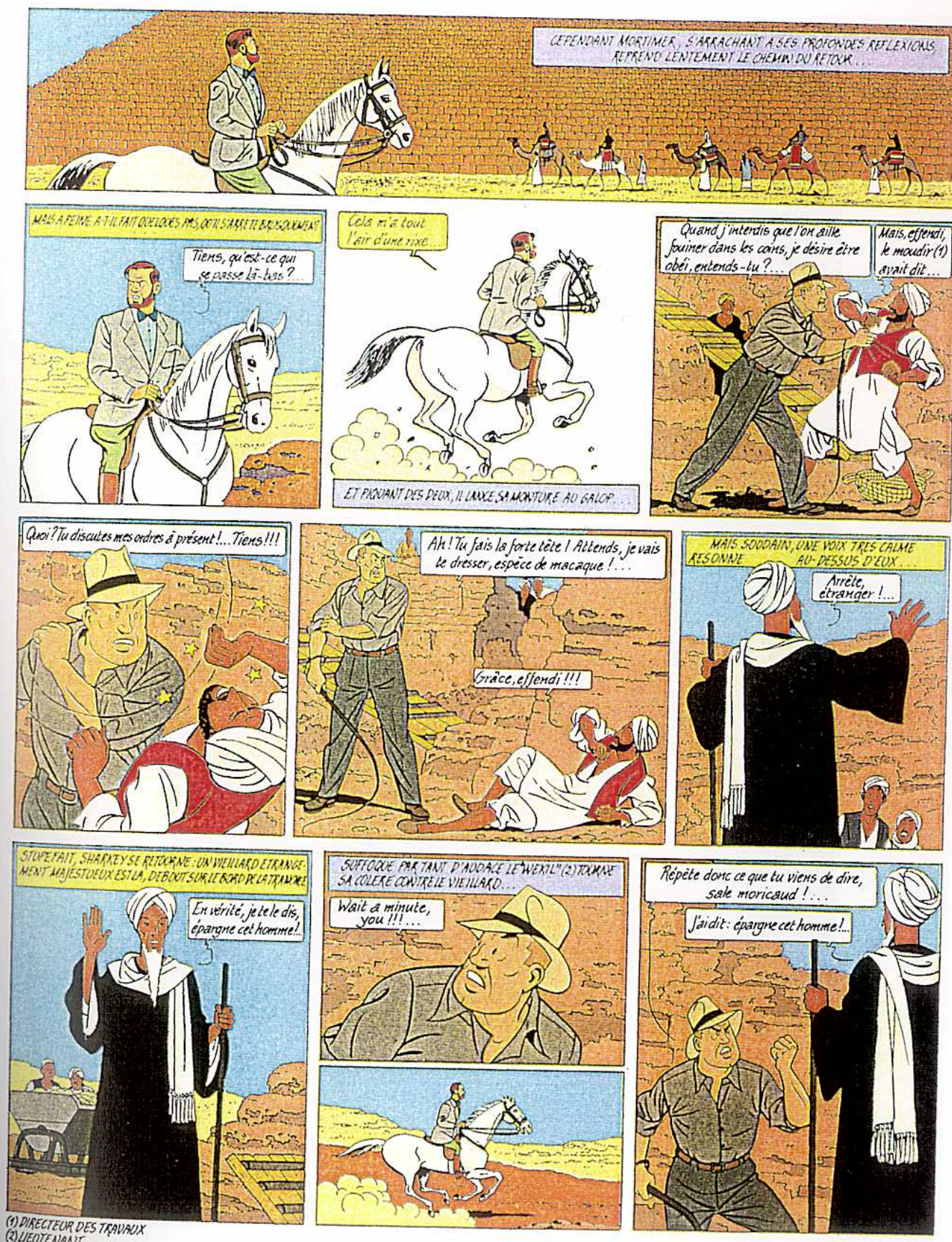
Figure 1: Appel de note inachevé [Lambil-Cauvin 84]



Figure 2: Appel de note à but humoristique [Fournier 76]



ANNEXE 14: Planche à coder [Jacobs 86]



(1) DIRECTEUR DES TRAVAUX
(2) JEOTENANT

ANNEXE 15: *The Flirt*, Booth Tarkington [réf.4]



"Kiss me some more, darl——"