

Cahiers **GUT** *enberg*

☞ VERS UNE EXTENSION MULTILINGUE DE BIBTEX

☞ Jean-Michel HUFFLEN

Cahiers GUTenberg, n° 39-40 (2001), p. 23-38.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_2001__39-40_23_0>

© Association GUTenberg, 2001, tous droits réservés.

L'accès aux articles des *Cahiers GUTenberg*

(<http://cahiers.gutenberg.eu.org/>),

implique l'accord avec les conditions générales

d'utilisation (<http://cahiers.gutenberg.eu.org/legal.html>).

Toute utilisation commerciale ou impression systématique

est constitutive d'une infraction pénale. Toute copie ou impression

de ce fichier doit contenir la présente mention de copyright.

Vers une extension multilingue de BIB_TE_X

Jean-Michel HUFFLEN

LIFC — Université de Franche-Comté
16, route de Gray
25030 BESANÇON CEDEX
FRANCE
hufflen@lifc.univ-fcomte.fr

Résumé. Cet article décrit une extension de BIB_TE_X dans le cadre du multilinguisme et les étapes prévues pour sa réalisation. Nous montrons comment nous tentons de satisfaire à la fois les auteurs souhaitant que la langue utilisée pour la bibliographie d'un ouvrage soit celle de l'ouvrage et ceux qui préfèrent que la langue utilisée pour décrire chaque référence soit celle de la référence elle-même. L'outil que nous réalisons sera adaptable à tout paquetage multilingue de L^AT_EX 2_ε; et nos conventions assurent une compatibilité aussi bien avec les fichiers bibliographiques .bib existants qu'avec les fichiers de style .bst existants.

Mots-clés : Bibliographies, multilinguisme, L^AT_EX, BIB_TE_X.

Abstract. *This paper describes a multilingual extension to BIB_TE_X, and the steps of its implementation. We meet the requirements of users who wish the language used for the bibliographical references of a printed work to be the work's language, as well as users who prefer the language of each reference to be the reference's. Our tool will be able to work with any multilingual package of L^AT_EX 2_ε. Besides, our extension allows both existing bibliographical files (.bib files) and bibliography style files (.bst files) to be processed compatibly.*

Keywords: *Bibliographies, multilingual features, L^AT_EX, BIB_TE_X.*

Introduction

L'un des domaines dans lequel le logiciel L^AT_EX a connu de très fortes améliorations depuis ses débuts [30] est sans conteste le multilinguisme. Même si T_EX fournissait les commandes nécessaires à la production des accents et autres signes diacritiques des langues utilisant l'alphabet latin [29, ch. 9], il était fastidieux, dans la première version de L^AT_EX, d'écrire dans des langues autres que

l'anglais, sa langue « natale ». Un pas important a été franchi avec M \TeX (pour *Multilingual T \TeX*) de Michael J. Ferguson [17], qui permettait la division en fin de ligne de mots comportant des lettres accentuées. Par la suite, Donald E. Knuth a repris quelques-unes des idées de M \TeX dans la version 3 de T \TeX , version qui a permis la gestion de plusieurs ensembles de divisions possibles, en fonction de la langue considérée. En outre, la définition de polices dans lesquelles les lettres accentuées sont directement codées — plutôt qu'obtenues par superposition d'un accent à une lettre¹ — a contribué à homogénéiser les applications de l'algorithme de division utilisé par T \TeX ². Puis sont apparues, avec L \TeX 2 ϵ [21], les premières versions des paquetages³ multilingues french [20] et german [39], facilitant l'écriture en français ou en allemand. Par « facilitant l'écriture », nous entendons que L \TeX s'adapte aux conventions typographiques de la langue concernée. À titre d'exemple, les utilisateurs familiers avec le paquetage french savent qu'ils peuvent taper le texte :

N'est-ce pas ?

et le résultat sera « N'est-ce pas ? » où, devant le point d'interrogation, ce n'est pas une espace inter-mots qui est laissée, mais une espace fine, suivant les conventions typographiques de la langue française. Les adaptations réalisées par ces paquetages french et german incluent donc la sélection d'un fichier adéquat pour la division des mots et la redéfinition des commandes produisant les mots-clés tels que « chapitre » ou « figure ». Plus multilingue encore est le paquetage babel [5], dans la mesure où il tente d'homogénéiser le traitement de toutes les langues qu'il connaît, sans que les possibilités pour une langue particulière soient *a priori* mises en avant. En particulier, le paquetage babel est très adapté pour réaliser des *mélanges* de plusieurs langues à l'intérieur d'un même document⁴. Pour une introduction aux commandes de base de paquetages french et babel, voir [24, § 2] ; des informations complémentaires peuvent être trouvées sur les sites du réseau CTAN⁵. Mentionnons également un panorama des possibilités multilingues actuelles de L \TeX 2 ϵ dans [16, ch. 8].

Face à ces perfectionnements, force est de reconnaître que la version actuelle de B \TeX [36], l'outil associé à L \TeX pour la gestion des références bibliographiques, ne présente pas de facilités analogues. Quelques tentatives d'intro-

1. Ce qui s'effectue par la commande `\accent` de T \TeX [29, ch. 25].

2. Cet algorithme est expliqué dans [29, annexe H].

3. Beaucoup d'auteurs francophones traduisent le mot « *package* » de la terminologie anglophone de L \TeX 2 ϵ par « extension », ou encore par « option de style ». Il nous semble quant à nous que le mot « paquetage » traduit mieux cette notion reliée aux modules des langages de programmation.

4. Quoiqu'il soit possible, si l'on utilise les paquetages french et german, de repasser dans la configuration initiale de L \TeX , ce qui permet l'insertion aisée de passages écrits en anglais.

5. *Comprehensive T \TeX Archive Network*. Le site du réseau CTAN le plus accessible en France est <ftp://ftp.loria.fr/pub/unix/ctan>.

duction de multilinguisme dans les fichiers bibliographiques .bib traités par BIB_{TEX} ont été proposées, mais restent rares. Quant aux *styles bibliographiques* utilisés par BIB_{TEX} ⁶ (les fichiers .bst) et permettant de contrôler la présentation des bibliographies, il n'existe pour l'instant que des versions liées à une langue (surtout l'anglais, mais aussi le français⁷) mais peu de versions réellement multilingues au sens de permettre le mélange de plusieurs langues dans une bibliographie. Au demeurant, comme nous le verrons dans la première section, deux philosophies s'affrontent quant à cette fonctionnalité.

Assisté par deux étudiants dans le cadre d'un projet de DESS⁸, nous avons débuté en octobre 2000 une réalisation que nous avons appelée $\text{MIBIB}_{\text{TEX}}$ (pour *Multilingual BIB_{TEX}*). Le but de ce prototype n'est pas de se substituer au « vrai » BIB_{TEX} , mais de permettre la conduite d'expérimentations dans le cadre des bibliographies multilingues. Indépendamment de ce que sera la version « définitive » de BIB_{TEX} — la version 1.0 annoncée par son auteur dans [37] —, nous pensons d'ici peu avoir un recul par rapport à l'ergonomie des solutions que nous proposons, après quelques essais sur des bibliographies en vraie grandeur.

La lecture de cet article ne présuppose qu'une connaissance de base de BIB_{TEX} , telle que l'on peut l'acquérir dans la plupart des manuels d'introduction à L_{TEX} : [31, Appendix B], [40, ch. 8] et [16, § 6.5.1]. D'autre part, nous ne préoccuperons pas non plus des conventions typographiques relatives à la présentation des bibliographies ; une courte introduction à ce sujet figure en [24, § 1.8], des références plus détaillées étant :

- [12, § 94], [22, p. 37-39] et [33, p. 31-36] pour la langue française (la référence de base étant la norme [1] établie par l'AFNOR⁹) ;
- [8, § 10], [11, § 15.54-15.76] et [23, p. 50-54] pour la langue anglaise (les normes publiées par *British Standards* étant [6, 7], la norme américaine ANSI¹⁰ [2]) ;

mais nous étudierons uniquement l'aspect multilingue dans les bibliographies. Dans la première section, nous décrivons les besoins à satisfaire quant à l'introduction du multilinguisme dans une bibliographie. Nous donnons aussi les moyens actuels *ad hoc* de les résoudre. Ensuite, nous décrivons dans la section 2 nos extensions multilingues telles qu'elles seront traitées par $\text{MIBIB}_{\text{TEX}}$.

6. [35] donne les conventions du langage utilisé dans ces fichiers.

7. Quelques exemples de styles bibliographiques francisés — leur nom commence en général par « f . . . » — figurent dans la liste donnée en [25, § 4.3].

8. Diplôme d'Études Supérieures Spécialisées. Il s'agit du DESS *Génie informatique, systèmes distribués et multimedia* de l'Université de Franche-Comté.

9. Association Française de NORmalisation.

10. *American National Standards Institute*.

Enfin, la section 3 décrit l'état d'avancement de notre projet et les différentes phases que nous prévoyons.

1. BIB_TE_X et les besoins multilingues

1.1. Point de vue « général »

Deux options sont possibles pour choisir la ou les langues d'une bibliographie.

- La première consiste à considérer qu'une référence bibliographique n'est réellement exploitable que par quelqu'un qui connaît la langue dans laquelle l'œuvre correspondante a été écrite. Adoptant cette optique, on spécifiera sans complexe toutes les informations dans la langue de la référence : s'il s'agit d'un ouvrage paru lors d'un mois de mars, on indiquera « mars » pour un ouvrage en français, « *March* » pour un ouvrage en anglais, « *März* » pour un ouvrage en allemand, etc. Ce choix, que nous avons adopté pour la bibliographie du présent article, présente l'avantage de l'homogénéité si l'on examine référence par référence les bibliographies qui en résultent. En outre, les informations à porter dans la bibliographie se recopient directement à partir de l'ouvrage : par exemple, le champ SCHOOL d'une thèse pourra être « Institut National Polytechnique de Grenoble », en français, ou « *University of Stirling* », en anglais, selon les informations qui apparaissent sur le document de thèse considéré.
- L'autre choix tente de donner le maximum d'informations dans la langue de l'ouvrage dont on établit la bibliographie. Reprenons notre référence parue lors d'un mois de mars : si elle apparaît dans un ouvrage écrit en français, en fait, pour *tous* les ouvrages de la bibliographie parus en mars, les références feront apparaître « mars » comme mois de parution. Ou « *March* » si l'ouvrage est écrit en anglais, « *März* » si l'ouvrage est écrit en allemand. Ce choix produit des bibliographies dont l'aspect visuel est beaucoup plus homogène ; ainsi, une bibliographie en français pour un ouvrage en français utilisera systématiquement des guillemets à la française¹¹ :

« À la poursuite de Quâr »

et non tantôt des guillemets à la française, tantôt des guillemets à l'anglaise :

« Les mystères de New-York »
'World's Fair Goblin'

11. Produits par les commandes `\leftguillemets` et `\rightguillemets` du paquetage french [20, § 1.7], ou par les commandes `\og` et `\fg` dans l'option frenchb du paquetage babel [19, p. 2].

Poussée à l'extrême, cette convention oblige cependant à traduire certaines informations : par exemple, indiquer « Université de Stirling » pour le champ `SCHOOL` d'une thèse soutenue à Stirling, en Écosse, et non « *University of Stirling* » si cette référence doit apparaître dans la bibliographie d'un ouvrage en français.

Signalons qu'il existe des paquetages permettant de réaliser l'une ou l'autre approche : `mlbib` [34] qui repose sur l'ajout d'un champ `LANGUAGE` pour écrire une référence particulière dans une langue particulière¹² et `oxford` [3] qui permet de choisir globalement une langue pour toute une bibliographie.

Toutefois, après cet exposé des orientations globales, nous pouvons remarquer que ces approches sont en réalité complémentaires ; une version multilingue de $\text{BIB}\text{T}_\text{E}\text{X}$ devrait être capable de les mettre en œuvre toutes les deux. D'ailleurs, si l'on considère un livre dont les bibliographies sont à donner chapitre par chapitre¹³ et dont les chapitres sont écrits dans des langues différentes, on voit bien que le second choix conduit à exprimer la bibliographie de chaque chapitre dans la langue du chapitre et que la notion de langue globale pour toute la bibliographie de l'ouvrage disparaît.

1.2. Étude technique

Considérons, à titre d'exemple, la référence donnée dans la figure 1. Un simple coup d'œil nous convainc que cette entrée de $\text{BIB}\text{T}_\text{E}\text{X}$ référence un ouvrage écrit en anglais et qu'elle a été établie pour une utilisation dans un document écrit en anglais. Tentons à présent de l'adapter pour qu'elle soit exploitable aussi bien dans des documents anglais que dans des documents écrits dans une autre langue que l'anglais, par exemple le français. Dans ce qui va suivre, nous supposons que le paquetage multilingue employé pour le document est `babel`, ce qui est certes restrictif, mais les manipulations que nous décrivons ci-après — passage de l'anglais au français et *vice versa* — peuvent aussi être réalisées avec le paquetage `french`, au moyen des commandes `\french` et `\english`.

Adopter le premier choix et présenter cette référence en anglais même si elle doit apparaître à l'intérieur d'un document écrit en français

```
\usepackage[... ,frenchb]{babel}
...
```

se résout en générant la commande `\bibitem` suivante dans le fichier `.bbl` résultant de $\text{BIB}\text{T}_\text{E}\text{X}$

12. Voir aussi [28, § 284].

13. Ce qui est actuellement possible avec les paquetages `chapterbib` ou `bibunits` [21, § 13.3].

```
@BOOK{robeson1968i,
  AUTHOR = {Kenneth Robeson},
  TITLE = {The Flaming Falcons},
  PUBLISHER = {Bantam},
  SERIES = {Doc Savage Series},
  NUMBER = 30,
  NOTE = {Originally published June 1939 in \emph{Doc
Savage Magazine}},
  YEAR = 1968,
  MONTH = nov}
```

FIGURE 1 – Exemple de référence anglaise en $\text{BIB}_{\text{T}}\text{E}_{\text{X}}$.

```
\bibitem{robeson1968i}
\begin{otherlanguage}{english}Kenneth Robeson. ...
...
\end{otherlanguage}
```

Ceci peut se réaliser en ajoutant un champ supplémentaire à la référence `robeson1968i`

```
LANGUAGE = english
```

et en définissant un nouveau style bibliographique (fichier `.bst`) comme nous le montrons dans [24, § 3.2]. Ce style, couplé avec `babel`¹⁴, utilisera les commandes données dans la table 1 pour les mots-clés et les informations telles que les noms des mois, etc.

Venons-en à présent au second choix possible : l’insertion de cette référence dans une bibliographie écrite, autant que faire se peut, en français. Là encore, l’utilisation du mnémotechnique `nov` pour le champ `MONTH` et d’un style bibliographique couplé avec `babel` nous tire d’affaire, pourvu que la langue sélectionnée pour la bibliographie soit le français. Par ailleurs, il est judicieux de se protéger contre les mauvaises coupures dans les parties qui ne peuvent être citées qu’en anglais : dans notre exemple, les champs concernés sont `TITLE`, `PUBLISHER` et `SERIES`. Nous indiquons pour ces champs un changement temporaire de langue au moyen de la commande `\foreignlanguage` de `babel` [5, § 1]. Cette manipulation n’est pas envisageable avec le champ `AUTHOR`, à cause des conventions très précises de $\text{BIB}_{\text{T}}\text{E}_{\text{X}}$ quant à la formulation de ce champ

14. Précisons que l’utilitaire `makebst` (voir [21, § 13.9] pour une introduction et [15] pour une présentation plus approfondie) permet de générer des styles bibliographiques couplés avec le paquetage `babel`.

Noms de mois	Nombres ordinaux	Autres mots-clés
<code>\bbljan</code>	<code>\bblfirsto</code>	<code>\bbland</code>
<code>\bblfeb</code>	<code>\bblsecondo</code>	<code>\bblchap</code>
<code>\bblmar</code>	<code>\bblthirdo</code>	<code>\bbled</code>
<code>\bblapr</code>	<code>\bblfourtho</code>	<code>\bbledby</code>
<code>\bblmay</code>	<code>\bblfiftho</code>	<code>\bbledn</code>
<code>\bbljun</code>	<code>\bblst</code>	<code>\bbleds</code>
<code>\bbljul</code>	<code>\bblnd</code>	<code>\bblin</code>
<code>\bblaug</code>	<code>\bblrd</code>	<code>\bblmasterthesis</code>
<code>\bblsep</code>	<code>\bblth</code>	<code>\bblno</code>
<code>\bbloct</code>		<code>\bblof</code>
<code>\bblnov</code>		<code>\bblp</code>
<code>\bbldec</code>		<code>\bblphdthesis</code>
		<code>\bblpp</code>
		<code>\bbltechrep</code>
		<code>\bblvol</code>

Mentionnons que la bibliographie de cet article a été obtenue avec un style bibliographique utilisant ces commandes ainsi que l’environnement `bblquoteditle`, permettant d’obtenir des guillemets à la française, à l’anglaise ou à l’allemande, selon la langue de la référence.

TABLE 1 – Mots-clés des bibliographies en interface avec le *package* `babel`.

[36, § 4]. Quant au champ `NOTE`, la meilleure adaptation consiste à en donner à la fois une version française et une version anglaise au moyen de la commande `\iflanguage` de `babel`. Le résultat est donné à la figure 2 : nous pouvons remarquer que la référence obtenue est utilisable aussi bien dans une bibliographie « totalement » francophone

Kenneth ROBESON : *The Flaming Falcons*. N° 30 in Doc Savage Series. Bantam. Première édition en juin 1939 dans *Doc Savage Magazine*. Novembre 1968.

que dans une bibliographie « totalement » anglophone

Kenneth Robeson. *The Flaming Falcons*. No. 30 in Doc Savage Series. Bantam. Originally published June 1939 in *Doc Savage Magazine*. November 1968.

Dans le cas où plus de deux langues sont en lice, nous arrivons, certes, à des emboîtages importants de la commande `\iflanguage` de `babel`, mais sans limitation théorique.

```

@BOOK{robesson1968i,
  AUTHOR = {Kenneth Robeson},
  TITLE = {\foreignlanguage{english}{The Flaming Falcons}},
  PUBLISHER = {\foreignlanguage{english}{Bantam}},
  SERIES = {\foreignlanguage{english}{Doc Savage Series}},
  NUMBER = 30,
  NOTE = {\iflanguage{frenchb}{Premi\`{e}re \`{e}dition en
juin 1939 dans}{Originally published June 1939 in} \emph{Doc
Savage Magazine}},
  YEAR = 1968,
  MONTH = nov}

```

FIGURE 2 – Réécriture multilingue de la référence de la figure 1.

Cela dit, outre sa faible lisibilité, cette solution *ad hoc* ne peut pas être mise en œuvre telle quelle si nous souhaitons cacher à l'utilisateur de base de $\text{BIB}\text{T}\text{E}\text{X}$ l'utilisation du paquetage `babel`, utilisation qui ne doit d'ailleurs pas être ressentie comme une obligation.

2. Extensions proposées

Étant donné l'énorme nombre de fichiers bibliographiques `.bib` déjà établis, il est certain que notre extension doit assurer une compatibilité avec ce qui existe. D'autre part, vu l'utilisation de $\text{L}\text{A}\text{T}\text{E}\text{X}$ et $\text{BIB}\text{T}\text{E}\text{X}$ principalement dans le domaine scientifique où la plupart des ouvrages sont écrits en anglais, il ne nous paraît pas aberrant de supposer qu'en l'absence de toute indication contraire, les références sont à générer suivant les conventions anglaises. Ces indications contraires fournies par nos extensions concernent soit une référence, soit un champ d'une référence.

Spécifier la langue d'une référence particulière s'effectue au moyen du champ `LANGUAGE`, comme nous l'avons déjà montré en § 1.2. Les valeurs permises pour ce champ seront tout préfixe non ambigu désignant une langue connue : le nom d'un paquetage multilingue ou d'une option de `babel`¹⁵. Alors,

15. La convention d'utilisation d'un préfixe non ambigu permettra, en cas de traitement par `babel` de la bibliographie générée, d'utiliser l'identificateur « `french` » pour caractériser un passage écrit en français, que l'option choisie pour la langue française soit `french` — le paquetage `french` [20] — ou `frenchb` [19].

- en cas de langue liée à une référence, la référence toute entière sera générée suivant les conventions de la langue correspondante ;
- en cas de langue choisie globalement pour la bibliographie, l'information fournie par LANGUAGE servira à traiter les champs :

ADDRESS	JOURNAL	TITLE
AUTHOR	PUBLISHER	TYPE
BOOKTITLE	SCHOOL	
EDITOR	SERIES	

MIBib_TE_X fournit également des moyens de spécifier des informations relatives aux langues à l'intérieur de la valeur associée à un champ. L'écriture

$$[string_0] * idf_0 [string_1] * idf_1 \dots [string_n] * idf_n \quad (1)$$

où $string_0, string_1, \dots, string_n$ sont des suites de caractères et $idf_0, idf_1, \dots, idf_n$ des identificateurs représentant des langues, signifie

- (i) s'il existe i , avec $0 \leq i \leq n$, tel que la langue de la référence — choisie globalement ou pour chaque référence — est égale à idf_i , alors la valeur spécifiée par (1) se ramène à la valeur de $string_i$;
- (ii) sinon — la langue de la référence n'est égale à aucune des langues données par idf_0, \dots, idf_n — la valeur choisie est celle associée à la langue indiquée par le champ LANGUAGE : si cette valeur n'existe pas, (1) est remplacée par la chaîne vide.

Les parties du champ en dehors de cette notation en crochets sont considérées comme des parties communes quelle que soit la langue à considérer et sont donc traitées de façon standard. Par exemple

$$\begin{aligned} & [University of] * english \\ & [Universit'\{e\} de] * french \\ & Stirling \end{aligned} \quad (2)$$

sera compris comme « University of Sirling » si la référence est à traiter en anglais, et comme « Universit'\{e\} de Stirling » si elle est à traiter en français. En outre, la mention « * english » est superflue pour une référence dont le champ LANGUAGE vaut english ; si nous nous trouvons dans ce cas, l'exemple (2) peut se simplifier en :

$$[University of] [Universit'\{e\} de] * french Stirling$$

Une autre écriture est également proposée :

$$[string_0] ! idf_0 [string_1] ! idf_1 \dots [string_n] ! idf_n \quad (3)$$

```
@BOOK{robesson1968i,
  AUTHOR = {Kenneth Robeson},
  TITLE = {The Flaming Falcons},
  PUBLISHER = {Bantam},
  SERIES = {Doc Savage Series},
  NUMBER = 30,
  NOTE = {[Première édition en juin 1939 dans] * french
[Originally published June 1939 in] * english \emph{Doc Savage
Magazine}},
  YEAR = 1968,
  MONTH = nov
  LANGUAGE = english}
```

FIGURE 3 – La référence de la figure 2, pleinement utilisable par MIBIB_TE_X.

Elle a le même effet que la notation (1) dans le cas (i). Dans le cas (ii), la chaîne (3) est remplacée par la chaîne vide, et un message d'avertissement (*warning*) est émis par MIBIB_TE_X.

Remarquons que compte tenu de ces conventions, la notation

$$s_0 \dots [s_i] \dots s_n$$

est équivalente à « $s_0 \dots [s_i] * l \dots s_n$ » où l est la langue indiquée par le champ LANGUAGE, donc, en fait, à « $s_0 \dots s_i \dots s_n$ ».

Si l'on désire que, dans la valeur d'un champ, les crochets droits soient interprétés comme de « vrais » crochets droits qui doivent apparaître dans le texte imprimé, les faire suivre de la chaîne vide.

Consulter [chapitre~3] {}

produira « Consulter [chapitre 3] ». Cet emploi de la chaîne vide permet également la mention d'un crochet fermant suivi du caractère « * » dans le texte imprimé

[comme] {} * ceci

Nous donnons dans la figure 3 la ré-écriture, telle qu'elle sera pleinement utilisable par MIBIB_TE_X, de la référence de la figure 1.

3. Le projet MIBibTeX

Nous avons adopté dès le départ l'idée d'une ré-écriture complète de BibTeX, même si nous réutilisons l'expérience des développements précédents en consultant les textes sources actuels. Cette ré-écriture s'est effectuée dans le langage C [27], dans le but de pouvoir largement diffuser ce produit, dans la mesure où C est très répandu tant dans le monde UNIX que dans le monde GNU¹⁶. En outre, ce choix nous permet de tirer pleinement parti d'outils tels que *flex* et *bison* pour les analyses lexicale et syntaxique¹⁷, outils précieux lorsqu'on cherche à étendre la syntaxe d'un langage, à travailler sur la grammaire qui le caractérise. En outre, le travail de ré-implantation proprement dit a été précédé d'une analyse fine des besoins qui a débouché sur une décomposition claire de notre nouvelle version en modules, ainsi que sur la définition d'une terminologie précise pour nommer les données et fonctions qui composent chaque module. Tous ces détails techniques sont consignés dans [14].

La ré-implantation « brute » du BibTeX actuel a produit en janvier 2001 une version « zéro » de MIBibTeX. Cette version préliminaire a été enrichie en une version 1.0 à présent en phase de tests et, vu son avancement, nous pensons qu'elle devrait être disponible à l'été 2001. Cette version 1.0 inclut le traitement de la syntaxe étendue proposée en § 2 pour tous les champs d'une référence bibliographique excepté `AUTHOR` et `EDITOR`. Ce traitement s'effectue en interface avec les options de *babel* correspondant aux langues écrites avec l'alphabet latin, ainsi qu'avec les paquetages multilingues *french* et *german*. Plus précisément, lorsque le fichier `.bbl` — regroupant les citations de l'environnement `thebibliography` à partir des citations présentes dans le fichier `.aux` et des références présentes dans les fichiers `.bib` sélectionnés — est traité par \LaTeX , nous savons détecter aussi bien l'éventuel chargement d'un paquetage multilingue que les options utilisées dans le cas du chargement de *babel*, et agir en conséquence.

Comme nous l'avons dit, le projet MIBibTeX cherche surtout à fournir un outil pour l'expérimentation, car il est difficile pour l'instant d'avoir du recul et de juger de notre ergonomie. En outre, il est préférable d'avancer à petits pas afin de faciliter les éventuelles mises à jour progressives d'outils fondés sur les fichiers `.bib`¹⁸ Mais nous pensons que la compatibilité *absolue* de notre prototype avec tous les fichiers `.bib` et `.bst` existants devrait l'ouvrir à nombre d'utilisateurs qui pourront expérimenter nos solutions, ce qui devrait nous permettre

16. Acronyme récursif : *GNU's Not Unix*. Pour plus de détails, se reporter à [13, p. 19-20] ou à la *Web page* <http://www.gnu.org>.

17. *flex* et *bison* sont les outils GNU correspondant respectivement aux utilitaires « natifs » d'UNIX *lex* et *yacc*. Une bonne introduction à ces outils et à leur utilité est [32].

18. Par exemple, les outils `bib2bib` et `bibtex2html`, présentés dans [18].

de « prendre la température » avant des réalisations plus ambitieuses. Réalisations plus ambitieuses qui pourraient inclure dans de futures versions

- le traitement multilingue des valeurs des champs AUTHOR et EDITOR des références ;
- l'utilisation possible d'Unicode, permettant la représentation de tous les caractères du monde sur 16 bits¹⁹ et ouvrant la voie à une utilisation conjointe de MIB \TeX avec les moteurs Ω et Λ [38], fondés sur ce codage²⁰ ;
- la possibilité de spécifier finement les tris des références bibliographiques : si l'on choisit l'ordre alphabétique (en adoptant le style bibliographique alpha par exemple), cet ordre alphabétique dépend des langues²¹ ; dans ce but, nous étudions à présent le générateur d'index $x\text{indy}$ ²² qui propose de semblables fonctionnalités ;
- l'amélioration du langage des fichiers .bst dans le sens de la modularité : permettre des inclusions de fichiers afin d'éviter les duplications de parties identiques d'un style bibliographique à l'autre ; cette question est également à l'étude en ce moment ;
- le langage des fichiers bibliographiques .bib pourrait lui aussi être rendu plus modulaire.

Cette liste n'est bien entendu pas limitative ; d'autres idées ne manqueront pas de nous venir au fur et à mesure de notre travail sur les versions successives de MIB \TeX .

Conclusion

La réalisation de cette extension multilingue de B \TeX constitue pour nous une tâche très stimulante. D'abord parce ce genre d'outil nous a souvent personnellement manqué. Ensuite parce ce type de réalisation pose des problèmes intéressants : par exemple, la nécessité de définir une syntaxe très précise pour tenir compte des imbrications des accolades auxquels viennent s'ajouter la présence de crochets droits²³. Tout utilisateur un tant soit peu avancé de B \TeX sait que, pour un titre, les valeurs :

19. Une introduction aux principes d'Unicode est [9], le document de référence étant [41].

20. Actuellement, MIB \TeX utilise le code ASCII (*American Standard Code for Information Interchange*) sur 7 bits, de même que le B \TeX « classique ».

21. À titre de variante par rapport à nos habitudes, considérons l'espagnol, où les mots commençant par « 11... » sont classés à part, après les mots en « 1a... », ..., « 1y... », juste avant les mots en « m... ».

22. *flexible INDEXing sYstem* [26]. C'est le successeur multilingue de *makeindex* [10] [21, § 12].

23. Encore que l'utilisation de crochets droits pour étendre la syntaxe de B \TeX ne soit pas réellement une invention de notre part. Elle figure dans le paquetage CAMEL [4].

{The Thousand-Headed Man} et {{The Thousand-Headed Man}}

ne sont pas traitées de façon équivalente. Pour « coller » à ce comportement, il nous a fallu prendre en compte les cas suivants :

1. {[The Thousand-Headed Man] * english
[L'homme aux mille t[^]{e}tes] * french}
2. {{{The Thousand-Headed Man}}} * english
[L'homme aux mille t[^]{e}tes]} * french}
3. {{{[The Thousand-Headed Man] * english
[L'homme aux mille t[^]{e}tes] * french}}}

et traiter de façon analogue les cas 2 et 3, différemment du cas 1.

Cette remarque étant faite, force nous est de reconnaître, d'un point de vue de génie logiciel, que nous avons pu mesurer toutes les difficultés d'étendre un produit en vue d'une utilisation qui n'avait pas été pensée au départ, tout en gardant le souci d'assurer la compatibilité avec l'existant.

Remerciements

Je remercie Laurent Comas et Éric Langlet, pour leur très active participation à l'implantation du projet MIBibT_EX. Merci également à Denis B. Roegel — ainsi qu'aux lecteurs du comité de programme de GUT' 2001 — pour leurs relectures de cet article.

Bibliographie

- [1] AFNOR Z44-045: *Documentation — Références bibliographiques — Contenu, forme et structure*. Norme disponible auprès de l'AFNOR, voir <http://www.afnor.fr>. décembre 1987.
- [2] ANSI/NISO Z39.71-1999: *Holding Statements for Bibliographic Items*. See <http://www.ansi.org>. 1999.
- [3] Peter ANTMAN: *Oxford Style Package*. Version 0.4. 1997. See CTAN: [biblio/bibtex/contrib/oxford/](http://ctan.org/tex-archive/biblio/bibtex/contrib/oxford/).
- [4] Frank G. BENNETT, JR: *User's Guide to the CAMEL Citator*. July 1995. See <http://www.loria.fr/services/tex/english/bibdex.html>.
- [5] Joannes BRAAMS: *Babel, a Multilingual Package for Use with L^AT_EX's Standard Document Classes*. Version 3.7. February 2001. See CTAN: [macros/latex/required/babel/babel.dvi](http://ctan.org/tex-archive/macros/latex/required/babel/babel.dvi).

-
- [6] British Standards 1629:1989: *Recommendation for References to Published Materials*. See <http://bsonline.techindex.co.uk>. 1989.
- [7] British Standards 5605:1990: *Recommendations for Citing and Referencing Published Materials*. See <http://bsonline.techindex.co.uk>. 1990.
- [8] Judith BUTCHER: *Copy-Editing. The Cambridge Handbook for Editors, Authors, Publishers*. 3rd edition. Cambridge University Press. 1992.
- [9] François CHAHUNEAU: "The Unicode Standard. A Global Solution to Localization Problems in Electronic Documents". *Document numérique*, Vol. 1, no. 4, p. 385–401. December 1997.
- [10] Pehong CHEN and Michael A. HARRISON: "Index Preparation and Processing". *Software—Practice and Experience*, Vol. 19, no. 9, p. 897–915. September 1988.
- [11] *The Chicago Manual of Style*. The University of Chicago Press. The 14th edition of a manual of style revised and expanded. 1993.
- [12] *Code typographique. Choix de règles à l'usage des professionnels du livre*. Fédération nationale du personnel d'encadrement des industries polygraphiques et de la communication, Paris. 17^e édition. 1993.
- [13] COLLECTIF: *Le logiciel libre*. Collection Précis et concis. Éditions O'Reilly France. Traduction française de *Open Source — kurz & gut* par James Guérin. 2001.
- [14] Laurent COMAS et Éric LANGLET: *Ré-implantation des commandes de formatage d'une bibliographie construite avec Bib_TE_X*. Rapport de fin de projet DESS. février 2001.
- [15] Patrick W. DALY: *Customizing Bibliographic Style Files*. Version 3.2. February 1999. Part of L_AT_EX' distribution.
- [16] Bernard DESGRAUPES: *L_AT_EX. Apprentissage, guide et référence*. Vuibert Informatique, Paris. 2000.
- [17] Michael J. FERGUSON: "A Multilingual TeX". *TUGboat*, Vol. 6, no. 2, p. 57–58. July 1985.
- [18] Jean-Christophe FILLIÂTRE and Claude MARCHÉ: *Bib_TE_X2HTML: A Translator of Bib_TE_X bibliographies into HTML*. February 2001. See <http://www.loria.fr/services/tex/english/outils.html>.
- [19] Daniel FLIPO: *A babel Language Definition File for French*. Version v1.5e. March 2001. See <http://www.tex.ac.uk>.
- [20] Bernard GAULLE: *Notice d'utilisation du style french multilingue pour L_AT_EX*. Version pro V5.01. janvier 2001. Voir sur CTAN : loria/language/french/pro/french/ALIRE.pdf.

-
- [21] Michel GOOSSENS, Frank MITTELBACH and Alexander SAMARIN: *The \LaTeX Companion*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1994.
- [22] Louis GUÉRY: *Dictionnaire des règles typographiques*. Techniques rédactionnelles, 2^e édition. CFPJ-éditions. janvier 2000.
- [23] *Hart's Rules for Composers and Readers at the University Press*. Oxford University Press. 39th edition. 1999.
- [24] Jean-Michel HUFFLEN: « Typographie : les conventions, la tradition, les goûts, ... et \LaTeX ». *Cahiers GUTenberg*, vol. 35-36, p. 169–214. mai 2000.
- [25] Jean-Michel HUFFLEN, Denis RÈGEL et Karl TOMBRE: *Guide local (\LaTeX) du LORIA*. Millésime 1998. rapport technique 98–R–214, LORIA. septembre 1998.
- [26] Roger KEHR: “xindy—A Flexible Indexing System”. *Cahiers GUTenberg*, Vol. 28-29, p. 223–230. March 1998.
- [27] Brian W. KERNIGHAN and Denis M. RITCHIE: *The C Programming Language*. 2nd edition. Prentice Hall. 1988.
- [28] Marie-Paule KLUTH: *FAQ \LaTeX française pour débutants et confirmés*. Vuibert Informatique, Paris. Également disponible sur CTAN : help/LaTeX-FAQ-francaise/. janvier 1999.
- [29] Donald Ervin KNUTH: *Computers & Typesetting. Vol. A: the \TeX book*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1984.
- [30] Leslie LAMPORT: *\LaTeX . A Document Preparation System*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1986.
- [31] Leslie LAMPORT: *\LaTeX . A Document Preparation System. User's Guide and Reference Manual*. Addison-Wesley Publishing Company, Reading, Massachusetts. 1994.
- [32] John LEVINE, Tony MASON and Doug BROWN: *lex & yacc*. 2nd edition. O'Reilly. October 1992.
- [33] *Lexique des règles typographiques en usage à l'Imprimerie Nationale*. Imprimerie Nationale. 1990.
- [34] Wenzel MATIASKE: *Multilinguale Zitierformate*. Oktober 1995. Siehe CTAN: macros/latex/contrib/supported/mlbib/.
- [35] Oren PATASHNIK: *Designing Bib \TeX styles*. February 1988. Part of \LaTeX' distribution.
- [36] Oren PATASHNIK: *Bib \TeX ing*. February 1988. Part of \LaTeX' distribution.
- [37] Oren PATASHNIK: “Bib \TeX 1.0”. *TUGboat*, Vol. 15, no. 3, p. 269–273. September 1994.

- [38] John PLAICE and Yannis HARALAMBOUS: *Draft Documentation for the Ω System*. March 1998. See <http://www.loria.fr/services/tex/english/moteurs.html>.
- [39] Bernd RAICHLE: *Die Makropakete „german“ und „ngerman“ für \LaTeX 2 ϵ , \LaTeX 2.09, Plain- $T_{\text{E}}X$ and andere darauf Basierende Formate*. Version 2.5. Juli 1998. Im Software \LaTeX .
- [40] Christian ROLLAND: *\LaTeX par la pratique*. Éditions O'Reilly. octobre 1999.
- [41] *The Unicode Standard Version 3.0*. Addison-Wesley. February 2000.