

# *Cahiers* **GUT** *enberg*

## ☞ ASPECTS DE LA PROBLÉMATIQUE DE LA CONFECTION D'UNE FONTE POUR LES MATHÉMATIQUES ARABES

☞ Azzeddine LAZREK

*Cahiers GUTenberg*, n° 39-40 (2001), p. 51-62.

<[http://cahiers.gutenberg.eu.org/fitem?id=CG\\_2001\\_\\_39-40\\_51\\_0](http://cahiers.gutenberg.eu.org/fitem?id=CG_2001__39-40_51_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.



---

# Aspects de la problématique de la confection d'une fonte pour les mathématiques arabes

---

Azzeddine LAZREK

*Département d'Informatique, Faculté des Sciences, Université Cadi Ayyad,  
B.P. 2390, Marrakech, Maroc  
lazrek@ucam.ac.ma*

**Résumé.** Ce papier relate quelques difficultés rencontrées lors de la construction d'une fonte pour les textes mathématiques arabes avec expressions symboliques composées de symboles spécifiques et orientées de droite à gauche. Ces difficultés témoignent des limites des systèmes composés pour les besoins d'une langue déterminée lorsqu'on tente de les adapter à des contextes relativement lointains.

**Abstract.** *A great deal of Arabic texts of mathematics contains formulas composed with specific symbols in a writing running from right to left. Up till now, as far as we know, there is no system that allows typesetting such documents. Millions of scholars throughout Arabic countries use handbooks where symbols are still written by hand.*

*The system presented in this paper is an attempt to provide the possibility of typesetting such documents. The capabilities of the system T<sub>E</sub>X and of the package ArabT<sub>E</sub>X will be extended to suit with this situation.*

*The Arabic font Naskh designed by K. Lagally for his package ArabT<sub>E</sub>X and the family Computer Modern designed by D. E. Knuth in Metafont in use in his system T<sub>E</sub>X serve as basic fonts. In mathematical mode, the font Naskh will be adapted to various sizes according to the different positions (normal, superscript or subscript, superscript of superscript, etc.). This font is used to design new signs in different shapes and for abbreviations. The family Computer Modern, especially Computer Modern Math Symbols and Computer Modern Math Extension, will help to build some special symbols via glyphs inversions. Many difficulties come up afterwards. Heterogeneity of size, bold face level, the position of symbols with respect to the baseline, are some of such difficulties in symbolic writing. Shapes changes of Arabic characters between text mode and mathematical mode are further difficulties in text mode. These difficulties attest of the limits of systems composed for the needs of typesetting mathematics in a Latin Language whenever these systems are to be adapted to foreign lingual contexts.*

## 1. Introduction

Les textes mathématiques nécessitent *une police mathématique* propre. Le lecteur peut consulter [5] pour une discussion sur cette question. Ceci est encore plus vrai pour les textes mathématiques arabes. La calligraphie arabe étant un art millénaire qui a développé des règles raffinées, les textes mathématiques arabes requièrent encore plus de soins. Lorsqu'on utilise une fonte arabe disponible pour écrire des expressions symboliques, on se rend rapidement à l'évidence qu'il y a beaucoup d'imperfections et d'hétérogénéités : dénivellement de signes, non-homogénéité des tailles, etc.

Pour tenter de répondre aux besoins de la situation, nous avons entrepris d'étendre les capacités du système ArabTeX développé par K. Lagally [9]. Le système ArabTeX permet d'écrire du texte arabe mais n'offre pas la possibilité d'écrire des expressions symboliques arabes. Il comporte une fonte arabe en style Naskh en plus des fontes latine et grecque offertes par T<sub>E</sub>X. La description des commandes de cette extension est faite dans [10]. Un certain nombre de questions afférentes au sujet ne sont pas abordées ici parce qu'elles dépassent le cadre de cet article. Nous proposons à la fin de cet article un échantillon d'exemples d'écritures d'expressions mathématiques arabes réalisées avec le système obtenu.

Par abus de langage, on parlera de documents ou textes « mathématiques arabes » (respectivement latines) pour référer à une *présentation* en langue arabe (respectivement latine) de la mathématique qui est une et universelle et n'a rien à voir avec un quelconque nationalisme.

## 2. Ingrédients d'une fonte mathématique arabe

Les textes mathématiques arabes, comme tous textes mathématiques, nécessitent un grand nombre et une grande diversité de signes [2], [1] :

- les lettres de plusieurs *alphabets* (arabe, latin, grec, ...), en plusieurs *styles* (Naskh, Thuluth, Rouqa', Koufi, ... pour l'arabe, romain, italique, penché, calligraphique, ... pour l'alphabet latin). Ces alphabets doivent figurer en différents *traits de caractères* (minuscule et majuscule, gras, ...). Les formes de *glyphes* des lettres arabes varient selon la position de la lettre dans le mot (initiale, médiane, finale et isolée : م م م م). Par contre, certaines caractéristiques comme les majuscules n'ont pas d'analogue en écriture arabe. De plus l'alphabet arabe utilisé dans les expressions symboliques se présente sans points ou signes diacritiques, utilisés pour marquer les voyelles ;
- les accents en plusieurs formes et tailles ;

- les chiffres en plusieurs formes : deux formes latines (style normal et ancien), chiffres du Maghreb et du Machrek arabe pour l’arabe, ... ;
- les signes de ponctuation en plusieurs formes (en écriture arabe, la virgule est orientée vers le haut pour un couple de termes, elle est orientée vers le bas comme en français dans un nombre décimal) ;
- les délimiteurs en plusieurs formes et tailles ;
- les signes d’opérations arithmétiques, relationnels, logiques, ensemblistes, ... ces signes seront spécifiques ;
- les caractères composés à partir d’autres signes ( $\prime\lrcorner$ ,  $\leftarrow\Rightarrow$ , ... ) ;
- et d’autres symboles en plusieurs formes et tailles. La taille des symboles extensibles est fonction du contexte.

En écriture mathématique, excepté la taille, toute modification du signe est pertinente et donne lieu à une composante de sens. L’absence de signe est un signe. Le signe romain n’est pas celui italique, la position du signe sur la ligne courante, en indice ou en exposant, ... tout cela est sémantiquement pertinent [11].

Les signes utilisés en mode mathématique ne sont pas les mêmes que ceux utilisés en mode texte (ex. : l’écriture de la préposition a diffère de celle du symbole de désignation d’une variable ou d’une constante  $a$ ). Cela en dit long sur la diversité des signes requis pour les besoins de l’écriture mathématique. Ce grand nombre de symboles requiert un nombre important de codages variés et de polices de caractères également variées. D’un autre côté, la saisie du texte doit demeurer aisée. L’utilisation des commandes doit être directe et naturelle. La mise en page du texte doit fournir une page agréable à la lecture. L’esthétique du texte requiert parfois l’utilisation d’une combinaison de polices relativement voisines. Tout cela nécessite un haut degré d’homogénéité des caractéristiques des fontes (taille, position sur la ligne de base, trait de gras, ...). On s’en rend bien compte lors de la confection des caractères composés ou extensibles ou dans une liste de symboles [4].

### 3. Polices de caractères disponibles

Quelles sont les familles de polices de caractères susceptibles d’être utilisées pour écrire les expressions symboliques arabes ? On peut, pour éviter d’avoir à confectionner de nouvelles fontes, adopter la fonte arabe Naskh (nash ou xnsh) développée par K. Lagally pour Arab $\TeX$  [9] ainsi que la famille des fontes Computer Modern développée par D. E. Knuth en Metafont pour  $\TeX$  [7]. Les glyphes des symboles peuvent être réfléchis comme dans un miroir. D. E. Knuth et P. MacKay [8] proposent la possibilité d’une telle transformation.

Les familles de fontes susceptibles d'être proposées pourraient être :

- une famille nommée **amnash** générée à partir de la fonte Naskh. Cette famille de fontes contient les différentes tailles correspondant aux différentes positions (normale, en indice, en indice d'indice, ...) utilisées en mode mathématique. La famille **amnash** sera utilisée pour l'alphabet arabe, les chiffres arabes du Machrek et les signes de la ponctuation arabe ;  
Symboles en exposant avec variation de taille :

$2^2 2^2 2$      
  $۲۲۲$      
 ۲۲۲     
 ررر     
 ۲۲۲     
 جتا جتا جتا جتا

Symboles en indice avec variation de taille :

$2_2 2_2$      
 ۲۲۲     
 ۲۲۲     
 ۲۲۲     
 ۲۲۲     
 جتا جتا جتا جتا

- toute la famille de fontes Computer Modern. Elle sera utilisée pour les alphabets latin et grec (romain et italique), les chiffres arabes du Maghreb, etc. ;
- une famille nommée **amcmsy** générée à partir de la fonte Computer Modern Math Symbols, après réflexion de glyphes. Cette fonte sera utilisée pour les symboles non extensibles ( $>$ ,  $<$ ,  $\sqrt{\quad}$ , ...);
- une famille nommée **amcmex** générée à partir de la fonte Computer Modern Math Extension, après réflexion de glyphes. Cette fonte est utilisée pour les symboles extensibles ( $\int$ ,  $\lambda$ , ...).

Pour augmenter la diversité de tailles des caractères disponibles, il faudrait soit disposer d'une police distincte pour chaque taille souhaitée (ce qui nécessite un traitement préalable important et un espace mémoire de stockage également conséquent) soit se contenter d'une police unique avec la possibilité d'appliquer des homothéties sur les glyphes pour obtenir les différentes tailles souhaitées (ce qui nécessite un temps de traitement plus important et une correction optique, par exemple pour les petites tailles, pour rendre certains caractères lisibles ou pour atténuer l'effet de l'italique sur certains caractères)[4].

Un nombre important de possibilités de mise en page pourrait être offert en fonction des caractéristiques techniques et esthétiques des familles de fontes disponibles. En particulier, la composition d'une expression symbolique interagit avec :

- les espacements entre les termes de l'expression. Les espacements autour de la virgule dans un couple de termes, dans un nombre décimal ou encore entre deux termes d'une liste ne sont pas les mêmes. Les espacements autour du signe + de la positivité ou du signe + de l'opération de l'addition ne sont également pas les mêmes, ... ;

- la justification du paragraphe dans lequel l’expression symbolique est insérée ;
- la coupure entre les lignes du paragraphe ;
- l’interligne qui varie avec la hauteur et la profondeur de l’expression ;
- les marges de la page ;
- la largeur, la hauteur et la profondeur des symboles qui composent l’expression.

## 4. Éléments du système développé

Une expression symbolique composée de symboles classiques pourra être réalisée sur le système proposé au moyen de la commande :

$$\frac{\left(\sum_{i=1}^p \cos x_i\right) + \sqrt{2x}}{3^{p-4}} + 9$$

```

 $\frac{\left(\sum_{i=1}^p \cos x_i\right) + \sqrt{2x}}{3^{p-4}} + 9$ 

```

Pour écrire une formulation symbolique se déroulant de droite à gauche et composée de symboles arabes, on utilisera la commande :

$$9 + \frac{\sqrt{2} + \left(\sum_{i=1}^p \cos x_i\right)}{4 - 3^p}$$

```

 $9 + \frac{\sqrt{2} + \left(\sum_{i=1}^p \cos x_i\right)}{4 - 3^p}$ 

```

et de la même manière, avec des chiffres du Machrek :

$$9 + \frac{\sqrt{2} + \left(\sum_{i=1}^p \cos x_i\right)}{\xi - \tau^p}$$

```

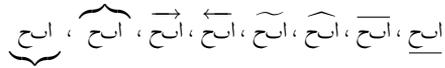
 $9 + \frac{\sqrt{2} + \left(\sum_{i=1}^p \cos x_i\right)}{\xi - \tau^p}$ 

```

### 4.1. Sens de déroulement de l’écriture

Les expressions symboliques arabes se déroulent de droite à gauche conformément au sens de déroulement de l’écriture arabe. Les principes d’inversion de déplacement, proposés par D. E. Knuth et P. MacKay [8], peuvent être aisément adaptés à cette fin.





#### 4.4. Chiffres

Sont proposées différentes formes de chiffres : les chiffres arabes du Maghreb en style usuel (9 , 8 , 7 , 6 , 5 , 4 , 3 , 2 , 1 , 0) et en style ancien (9 , 8 , 7 , 6 , 5 , 4 , 3 , 2 , 1 , 0) et les chiffres arabes du Machrek (٩ , ٨ , ٧ , ٦ , ٥ , ٤ , ٣ , ٢ , ١ , ٠). Les chiffres arabes du Maghreb sont empruntés à la fonte Computer Modern et les chiffres arabes du Machrek proviennent de la fonte Naskh. Les différentes formes de chiffres conservent les mêmes positions que dans leurs polices d’origines, ainsi le changement de la forme des chiffres peut être obtenu par un simple changement de la fonte dans un environnement déterminé. On peut observer que les chiffres arabes du Machrek se situent bien en dessous des signes de ponctuation arabe même s’ils proviennent de la même fonte. Les chiffres arabes du Maghreb, en style usuel, se situent également un peu plus bas que les signes de ponctuation arabe.

#### 4.5. Signes de ponctuation

On utilisera les différents signes de ponctuation arabe disponibles dans la fonte Naskh (à savoir : . . . ! : ; …)

mais aussi la virgule latine. En effet, la virgule arabe, orientée vers le haut, sera utilisée par exemple dans un couple de termes (ex. : (ب , ص)) mais la virgule latine, orientée vers le bas, sera utilisée par exemple dans un nombre décimal en notation française (ex. : 3,14). On peut remarquer que les signes de ponctuation apparaissent en haut de la ligne de base.

#### 4.6. Délimiteurs

Le système propose différentes formes de délimiteurs (ex. : les parenthèses ( et )). Il y a un problème d’appellation et de direction des délimiteurs. En effet, la parenthèse qui se trouve à gauche (respectivement à droite) d’une expression est la parenthèse ouvrante (respectivement fermante) en latin mais c’est la parenthèse fermante (respectivement ouvrante) en arabe. Cela a des conséquences néfastes sur la gestion automatique des espaces en fonction des types des symboles [7]. Les parenthèses et les crochets de la fonte Naskh (ex. : [ ( )]) n’ont pas la même forme ni le même trait de gras que ceux de la famille courante. Pour avoir un délimiteur extensible selon le contexte, dans une expression mathématique arabe qui sera inversée, il faut commencer par `\right`

puis ensuite `\left` à l'inverse de ce qu'on ferait pour le latin sauf si on inverse la définition de ces deux commandes.

#### 4.7. Symboles

On peut distinguer deux types de symboles : les symboles symétriques par rapport à l'axe vertical passant par leur milieu, ...  $=, \times, *, -, +$  et ceux qui ne le sont pas, ...  $<, >, /$ .

Les symboles symétriques seront pris sans changement. Certains symboles non symétriques comme  $/, %$  seront pris également sans changement.

Pour les symboles non symétriques qui possèdent, au sein de la fonte, des symboles symétriques, on a le choix entre l'interversion des positions des deux symboles (ex. : le symbole  $>$  devient le symbole  $<$  et vice-versa) ou bien l'interversion des noms des deux symboles (ex. : la commande `\in` devient `\ni` et inversement).

Pour les symboles non symétriques qui n'ont pas de symétrique dans la fonte, les glyphes seront inversés. Des problèmes surgissent alors pour certains symboles :

Le symbole  $\angle$ , généré par la commande `\angle`, n'est pas un glyphe unique ; il est composé du symbole  $/$  et d'un trait horizontal tel qu'il est défini dans Plain TeX [7]. On sera obligé de le redéfinir pour l'arabe pour avoir le symbole  $\sphericalangle$ .

Le symbole de l'intégrale en Computer Modern Math Extension est penché vers la droite. Une symétrie par rapport à l'axe vertical donne un symbole penché vers la gauche mais alors la position de la borne supérieure de l'intégrale devient très éloignée du symbole vers la droite  $\int_1^{\tau}$ . Pour remédier à cette situation, un redressement de la position de la borne supérieure à gauche devient nécessaire. Un autre symbole pour l'intégrale, plus droit peut être proposé  $\int_1^{\tau}$ .

Le symbole de racine a une hauteur qui est déterminée en fonction de la hauteur de l'expression sur laquelle il porte et de la position de la ligne de surlignement. La détermination automatique de la longueur du surlignement est fonction de la longueur de cette expression. Le problème de l'extension de la hauteur du symbole de racine en fonction de la hauteur de l'expression est résolu mais pas en fonction de la position de la ligne du surlignement  $(9 * \surd^3 ; 2 \surd ; \surd)$ . On obtient une expression symbolique de faible qualité :

$\sqrt[2]{s} \sqrt{*} \sqrt{j}$ . Il faudrait alors modifier la primitive `\radical` utilisée pour la racine d'un terme.

Par ailleurs, pour les symboles littéraux tels ceux de la somme, du produit ou de la limite, le système propose de nouveaux symboles qui épousent la forme d'abréviations

نہا، جد et مج

respectivement. L'extension de la khashidah (allongement de la dernière lettre de façon curviligne) de ces trois symboles d'opérateurs est fonction de la taille de la plus grande borne inférieure et supérieure pour la somme et le produit, ou de la taille de l'expression de la borne pour la limite. Généralement, l'extension de la khashidah de ces trois symboles se fait simplement par l'allongement du trait horizontal de la dernière lettre, de façon linéaire ; cela viole les règles de la calligraphie arabe.

$$\sum_{1=s}^s \quad ; \quad \text{مَج} \text{ }_{1=s}^s \quad ; \quad \text{مَج} \text{ }_1^s$$

$$\prod_{1=s}^s \quad ; \quad \text{جَد} \text{ }_{1=s}^s \quad ; \quad \text{جَد} \text{ }_1^s$$

$$\text{نَہَا} \text{ }_{0 \leftarrow s}^2 \quad ; \quad \text{نَہَا} \text{ }_{\infty + \leftarrow s}^2$$

La khashidah serait ici à revoir en fonction des règles de la calligraphie en s'inspirant des techniques utilisées dans [3].

#### 4.8. Fonctions

Nous définissons un ensemble d'abréviations de fonctions élémentaires utilisées en mathématiques arabes. On adoptera une écriture cursive pour les abréviations des fonctions comme en mode texte. Par contre, les noms de ces fonctions ne porteront pas de signes diacritiques même lorsque le texte arabe sera vocalisé. La possibilité de définir, de la même manière, de nouveaux noms de fonction sera également offerte.

#### 4.9. Gestion des espaces à l'intérieur des symboles

Il y a un problème de gestion automatique des espaces comme sous  $\text{\TeX}$  standard. En effet :

- en arabe, l'indice ou l'exposant d'un terme se situe à gauche du terme. Il faut éliminer l'espace généré avant un indice ou un exposant par l'adjonction du terme  $\{ \}$  comme cela est proposé par D. E. Knuth [7] lors de l'introduction d'un indice ou d'un exposant avant un terme en latin (symboles chimiques) ;
- il faut éliminer l'espace dans certains cas comme celui introduit après la virgule dans un nombre décimal en notation française ;
- pour d'autres cas, il faut introduire un espace comme celui entre les lettres qui se situent entre d'autres lettres et des chiffres, avant la virgule dans un couple de termes, séparés par une virgule, etc.

### 5. Exemples

Nous présentons dans ce qui suit quelques exemples d'expressions symboliques écrites avec le nouveau système.

Définition de la puissance d'un nombre dans un groupe multiplicatif :

$$\underbrace{ب \times \dots \times ب}_s = ب^s \text{تعرف}$$

Une liste de nombres se déroulant de droite à gauche. Les nombres sont alignés en colonnes :

$$\begin{array}{rcccccccc} \dots & 7 & 6 & 5 & 4 & 3 & 2 & 1 & 0 & = & ب \\ \dots & 124 & 64 & 32 & 16 & 8 & 4 & 2 & 1 & = & ب^{-2} \end{array}$$

Une matrice dont les termes sont doublement indexés de manière alphanumérique :

$$\begin{pmatrix} 1ب & \dots & 12ب & 11ب \\ 2ب & \dots & 22ب & 21ب \\ \vdots & \vdots & \vdots & \vdots \\ مب & \dots & م2ب & م1ب \end{pmatrix} = ك'ا$$

Un tableau récapitulatif représentant l'étude d'une fonction d'une variable réelle :

$\infty+$	0	$\infty-$	س
9-	-	0 +	5
2			(س)د
↙	↘		
7-		$\infty-$	

## 6. Conclusion

Plusieurs difficultés surgissent lorsqu'on utilise les éléments de familles de caractères différentes pour confectionner une fonte mathématique arabe. On peut citer entre autres :

- l'hétérogénéité du système symbolique résultant de l'importation de diverses familles de caractères : tailles, niveau de gras et surtout position du signe par rapport à la ligne de base, même si cela n'est pas propre à l'arabe ;
- contrairement au cas de l'alphabet latin avec ses traits d'écriture (romain, italique, penché, gras) pouvant être utilisés pour distinguer les lettres, la forme des lettres de l'alphabet arabe est la même qu'il s'agisse d'un caractère de texte ou d'un symbole de désignation mathématique, excepté que ce dernier est dépourvu de points diacritiques.

On voit donc que les fontes disponibles ne répondent pas tout à fait aux besoins.

L'utilisation d'autres fontes, fontes PostScript ou fontes virtuelles ou encore celles utilisées dans le système  $\Omega$ [6] pourrait constituer une alternative pour obtenir d'autres symboles tout en utilisant Unicode.

Nous projetons d'étendre le puissant système de traitement de documents multilingues  $\Omega$  pour l'adapter davantage aux spécificités des documents mathématiques arabes.

## Bibliographie

- [1] عزالدين لزرقي وخالد سامي، نظام مبتكر لمعالجة النصوص الرياضية العربية، المؤتمر السابع لتعريب العلوم، تعريب العلوم في منظومة التنمية القومية، مصر (Azzeddine LAZREK & Khalid SAMI, « Vers un système de traitement de document mathématique arabe », 7<sup>e</sup> congrès d'arabisation des sciences, *Arabisation des sciences dans le système de développement*, (2001) Egypte).
- [2] عزالدين لزرقي وخالد سامي، آفاق معالجة النصوص الرياضية باللغة العربية، مؤتمر المعلوماتية والصناعة البرمجية ودورها المستقبلي، اتحاد مجالس البحث العلمي العربية وكلية الحدباء الجامعة، العراق (Azzeddine LAZREK & Khalid SAMI, « Perspectives du traitement de document mathématique arabe », *Congrès Informatique et développement de logiciels et son futur impact*, Fédération des conseils de la recherche scientifique arabes et Faculté AlHadba' Universitaire, (2000) Irak).
- [3] Jacques ANDRÉ & Irène VATTON, « Dynamic optical scaling and variable-sized characters », *EPODD*, vol. 7 (4), (1994) p. 231–250.
- [4] Thierry BOUCHE, « Sur la diversité des fontes mathématiques », *Cahiers GUTenberg*, vol. 25, (1996) p. 1–24.
- [5] Yannis HARALAMBOUS, « Une police mathématique pour la Société mathématique de France : le SMF Baskerville », *Cahier GUTenberg*, vol. 32, (1999) p. 5–19.
- [6] Yannis HARALAMBOUS & John PLAICE, « Multilingual Typesetting with  $\Omega$  a Case Study: Arabic », in *Proceedings of the International Symposium on Multilingual Information Processing*, p. 137–154 (1997).
- [7] Donald E. KNUTH, *The TeXbook* (Addison-Wesley, 1984).
- [8] Donald E. KNUTH & Pierre MACKAY, « Mixing right-to-left texts with left-to-right texts », *TUGboat*, vol. 8 (1), (1987) p. 14–25.
- [9] Klaus LAGALLY, « ArabTeX - Typesetting Arabic with Vowels and Ligatures », congrès EuroTeX'92 (1992).
- [10] Azzeddine LAZREK, « A package for typesetting Arabic mathematical formulas », DANTE (2001).
- [11] Khalid SAMI, « Sur certains aspects de la formulation et de l'écriture de la mathématique en langue arabe », Thèse de doctorat, Université Catholique de Louvain, Louvain-La-Neuve (1992).