



## Tour du monde des ligatures

Yannis Haralambous

### ► To cite this version:

Yannis Haralambous. Tour du monde des ligatures. Cahiers Gutenberg, 1995, Ligatures & caractères contextuels, 22, pp.87-99. <hal-02101585>

**HAL Id: hal-02101585**

**<https://hal.science/hal-02101585v1>**

Submitted on 23 Apr 2019

**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.



HAL Authorization

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

§ TOUR DU MONDE DES LIGATURES

¶ Yannis HARALAMBOUS

*Cahiers GUTenberg*, n° 22 (1995), p. 87-99.

<[http://cahiers.gutenberg.eu.org/fitem?id=CG\\_1995\\_\\_22\\_87\\_0](http://cahiers.gutenberg.eu.org/fitem?id=CG_1995__22_87_0)>

© Association GUTenberg, 1995, 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.



---

# Tour du monde des ligatures

---

Yannis Haralambous

187, rue Nationale

59800 Lille

haralambous@univ-lille1.fr

## 1. Introduction

Le terme « ligature » désigne une quantité de choses différentes et est souvent source de confusions. Dans le monde T<sub>E</sub>X<sup>1</sup> la situation s'aggrave davantage, puisque D.E. KNUTH a donné ce nom à une partie des informations métriques des polices T<sub>E</sub>X.

Pour éclaircir ce sujet fort passionnant, nous allons tenter de classer les différents types de ligatures qu'on trouve dans les langues naturelles et ensuite situer T<sub>E</sub>X par rapport à elles.

On peut classer les ligatures *typographiques* en trois<sup>2</sup> classes :

1. les ligatures *linguistiques* ; ce sont celles qui sont indispensables pour l'écriture d'une langue donnée, et obéissent à des règles grammaticales (la grammaire étant liée à la *langue* et non pas à l'*écriture*). Souvent elles ont un statut de lettre et parfois même une place à part dans le dictionnaire, ce qui les rapproche des *digraphes*<sup>3</sup> ;
2. les ligatures *esthétiques* ; ce sont des graphies optionnelles qui existent pour des raisons esthétiques, de lisibilité, et/ou de tradition ;
3. les ligatures *contextuelles* ; ce sont des graphies qui se produisent sous certaines conditions, par exemple en début de mot. Elles sont obligatoires dans le contexte d'une écriture donnée, indépendamment des langues utilisant cette écriture.

---

<sup>1</sup>Voir l'index de l'éditorial de ce *Cahier*.

<sup>2</sup>Qui, comme les trois mousquetaires, en fait sont *quatre*, comme on le verra en page 92. Nous ne traitons pas ici des « ligatures » au sens de liaison entre lettres manuscrites.

<sup>3</sup>Un digraphe est une paire de lettres (*trigraphes* pour un triplet, etc.) qui dénote un seul son. Souvent les digraphes sont traités à part lors d'un tri alphabétique ; ainsi, en espagnol, « chacal » vient après « cucaracha », car « ch » est un digraphe, classé après « c ».

Dans le tableau suivant, nous classifions ces trois types de ligature suivant le critère obligatoire/optionnel, dans le contexte d'une langue donnée, ou celui d'une écriture donnée (c'est-à-dire de toutes les langues utilisant cette même écriture) :

Contexte	Type de ligature		
	<i>Linguistique</i>	<i>Esthétique</i>	<i>Contextuelle</i>
Langue	obligatoire	optionnelle	obligatoire
Écriture	optionnelle	optionnelle	obligatoire

Nous allons maintenant passer en revue les différents types de ligatures typographiques.

## 2. Ligatures linguistiques

Le cas le plus proche de nous est sans doute la ligature « œ ». On dit que le mot « œil » a *quatre* lettres, pourtant on écrit trois signes. Écrire « oeil » est plus que de la mauvaise typographie, c'est une faute de français ; il en est de même pour « mœlleux » (faux) au lieu de « mœlleux » (correct) : cet exemple montre que la ligature « œ » n'est pas automatiquement introduite à chaque fois qu'un « o » est suivi d'un « e ». Le signe « œ » est indispensable pour l'écriture correcte de la langue française. Pourtant, lors d'un tri alphabétique, « œ » est assimilé à « oe ».

La situation est plus claire pour le « æ » scandinave. Non seulement cette ligature fait indiscutablement office de *lettre* mais elle a une place à part dans le dictionnaire suédois : elle est placée après « z » (et non pas après « ad » comme on pourrait s'y attendre). Cependant, le même signe est utilisé de manière plus arbitraire en anglais : la mer Égée s'écrit « Ægæan sea » (comme transcription du mot grec Αἰγαῖον), *sans* que « Æ » soit considéré comme une lettre de l'alphabet anglais...

Plusieurs ligatures linguistiques proviennent historiquement de ligatures esthétiques. On connaît l'évolution historique du « et » en « & » (l'« esperluette »), il en est de même pour le « ß » allemand, qui provient de la ligature « f » long + « s » court.

Certaines ligatures ont du mal à se faire respecter. C'est le cas du « ij » flamand et néerlandais. Peu de locuteurs du néerlandais (et encore moins de non-locuteurs) savent que « ij » constitue dans 99,9% des cas une ligature dans cette langue (les exceptions sont des mots importés, comme « bijou », « bijektie » etc.). La preuve de l'existence de cette ligature est pourtant très simple : il suffit de lire dans une carte de Hollande le nom de la ville « IJmegen » où on retrouve évidemment la version majuscule de « ij »<sup>4</sup>.

<sup>4</sup>Il existe un rapport obscur entre cette ligature « ij » et la lettre diacritée « ÿ ». Cette dernière n'est — à la connaissance de l'auteur — utilisée qu'en gallois moderne et en français du moyen âge. Le rapport entre « ij » et « ÿ » doit être cherché ailleurs que dans les grammaires. À ce propos, une anecdote : un des grands jeunes

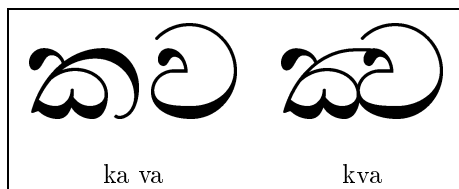


Figure 1: Exemple de ligature cingalaïse

Une ligature peut être considérée *indispensable*, même si elle n'est pas considérée comme *une* lettre unique. C'est le cas, par exemple, du « ٧ » arabe : il est impensable d'écrire « ٧ », que ce soit en typographie, en dactylographie, ou même en écriture manuscrite. Pourtant il s'agit toujours de la lettre « ٧ » en forme initiale « ٧ » suivie de la lettre « ٧ » en forme finale « ٧ ». Comme nous le verrons dans la section 4, toute autre ligature arabe est soit esthétique, soit contextuelle.

Ce phénomène est assez courant dans les langues orientales. Dans les langues indiennes, les ligatures jouent un rôle très important : elles indiquent l'absence de voyelles entre deux consonnes. Ainsi, en cingalais<sup>5</sup>, on voit sur la fig. 1 les caractères « ka » et « va », suivis de la ligature « kva » (dans le signe à droite la voyelle implicite « a » entre les deux consonnes n'est pas prononcée).

Dans certaines écritures, les voyelles sont indiquées par des traits supplémentaires qui s'ajoutent aux consonnes et n'existent pas isolés : en cambodgien<sup>6</sup>, ក est le caractère « ka » et កា, prononcé « kâ », est la ligature de ce dernier avec ា « â », un caractère qui n'existe pas isolé. Pour noter une syllabe consistant uniquement en une voyelle, on utilise le signe អ (sans prononciation) qui sert comme « porteur » de voyelle : « â » s'écrit អា.

On est donc à la limite de la définition de ligature, puisque on lie un caractère existant à un caractère non existant (sous forme isolée). En amharique<sup>7</sup>, on va encore plus loin : l'adjonction d'une voyelle peut entraîner la réduction de la forme de la lettre. Ainsi, (cf. fig. 2) « ze » peut être considéré comme la forme de base. On obtient « zu » « zē », « zē »,

mathématiciens français s'appelle « Ghys ». D'où vient ce nom? Les ancêtres de M. Ghys, s'appelaient « Ghijs » et vivaient en Flandre flamande. Quand ils sont venus s'installer en France, leur nom était écrit dans leur passeport — à l'époque ceux-ci étaient remplis à la main — en belle écriture manuscrite liée comme on l'apprenait à l'école. Mais si vous écrivez à la main « ij », les deux lettres bien liées, il est impossible de les distinguer d'un « ÿ ». C'est en effet ce que l'employé du bureau des naturalisations a cru voir. Et, fonctionnaire consciencieux, il a décidé d'office que « ÿ » n'était pas du bon français et a retiré le ...tréma, et le nom est devenu « Ghys ».

<sup>5</sup>Le système T<sub>E</sub>X cingalais a été créé par l'auteur, suite à une commande du *Wellcome Institute for the History of Medicine* de Londres. Il sera bientôt disponible sur serveur CTAN. Le lecteur en trouvera la description dans [7].

<sup>6</sup>Le système T<sub>E</sub>X cambodgien a été créé par l'auteur sous la supervision d'Alain Daniel, de l'Institut National des Langues et Civilisations Orientales. Il est décrit dans [6].

<sup>7</sup>Le système T<sub>E</sub>X amharique a été créé par Abass Amnulehe. Il est décrit dans [2] et [1].

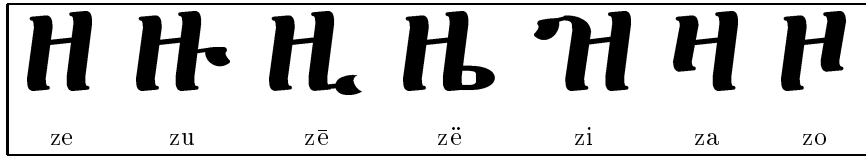


Figure 2: Quelques syllabes amhariques

« zi », en ajoutant des petits traits diacritiques ; par contre, pour obtenir « za », « zo », on réduit la forme de la lettre de base.

On trouvera le plus bel exemple de ligature orientale dans l'écriture coréenne Hangoul : une syllabe peut être composée d'une à quatre lettres qui sont alors mises en échelle et placées dans un carré invisible. Sur la fig. 3, le lecteur trouvera un exemple utilisant trois lettres : le mot « chang » (chapitre) est composé des lettres « ch », « a » et « ng ». Les deux premières forment la ligature « cha » (en bas à gauche). Cette ligature suivie de la troisième lettre forme de nouveau une ligature : « chang » (en bas à droite), le mot « chapitre ». En composant ainsi les 10 voyelles coréennes (+ 11 diphtongues) avec les 14 consonnes (+ 16 consonnes doubles) de cette manière, on obtient quelques milliers de caractères Hangoul<sup>8</sup>.

### 3. Les ligatures esthétiques

Elles sont très faciles à caractériser : ce sont celles qui ne sont pas indispensables (celles qu'on peut remplacer par leurs composantes, non-liées, sans changer la validité grammaticale, ou le sens du texte. Les exemples les plus courants (en typographie occidentale) sont ceux des ligatures « ff », « fi », « fl », « ffi », « ffl »<sup>9</sup>. Mais même pour celles-ci il y a parfois des règles : on ne peut, certes, imposer leur utilisation, mais souvent on l'interdit catégoriquement.

C'est le cas par exemple de la langue turque : l'alphabet turc dispose de caractères distincts « i » (prononcé « i ») et « ı » (prononcé comme le ع arabe ou le и russe. Il serait pour le moins déroutant d'écrire « fi » en turc (est-ce « f » + « i », ou « f » + « ı » ?) ; les Turcs sont d'autant plus sensibles à ce point que l'ancêtre de la langue turque moderne, la langue *ottomane*, écrite en caractères arabes, battait tous les records de confusion possible entre lettres écrites de la même manière et désignant des sons tout à fait différents<sup>10</sup>. On évite donc les problèmes en interdisant ces ligatures.

<sup>8</sup> Ceci n'est qu'une des deux écritures utilisées en Corée du sud. À celle-ci s'ajoute l'écriture *Hancha* c'est à dire *chinoise* : les idéogrammes chinois. Le régime de Corée du Nord a interdit l'utilisation des caractères Hancha.

<sup>9</sup> Auxquelles s'ajoute la ligature « fj », utilisée par exemple dans le mot norvégien « fjörd », et ses dérivés.

<sup>10</sup> Sans doute un artifice de l'aristocratie lettrée pour décourager le peuple à apprendre la lecture ; du moins ce fût l'interprétation de Kemal Atatürk, qui introduisit l'alphabet latin dans les années trente.

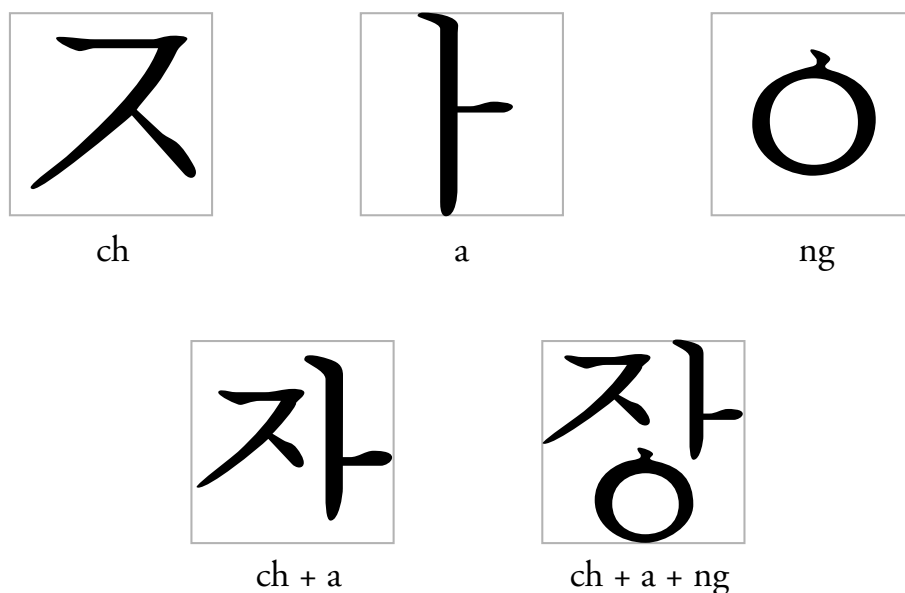


Figure 3: L'écriture coréenne Hangoul

Même interdiction pour certains mots composés allemands : par exemple, le mot „Auflage“ qui est composé des mots „auf“ et „Lage“ ; on ne pourrait former une ligature avec des lettres appartenant à des mots différents, même si ces mots sont concaténés. De même pour „Schiffahrt“ (ATTENTION ! dans „Schiffahrt“, PAS de ligature « ff »), où on voit re-apparaître la ligature « ff » lors de la césure : „Schiff-fahrt“ (la première composante du mot se terminant par deux « f », et la seconde commençant par cette même lettre, on n'écrit que deux « f »<sup>11</sup>).

Certaines traditions typographiques évitent les ligatures du genre « fi », sans raison particulière. C'est le cas de l'italien et du portugais.

Un exemple très frappant de ligatures esthétiques est celui de l'écriture arménienne. Comment faire quand on a une lettre comme *մ* « m » suivie d'une lettre comme *ն* « n » ? Il serait donc inacceptable d'écrire *մն*, qui donne l'impression de deux mots séparés. La solution traditionnelle arménienne est d'écrire *մն* : les deux lettres entrelacées ! Le lecteur trouvera les principales ligatures arméniennes sur la fig. 4.

<sup>11</sup>Cette règle montre l'économie de la grammaire allemande : la distinction entre « consonne double » et « consonne simple » est significative au niveau de la prononciation : en effet, une voyelle sera longue devant une consonne simple, et courte devant une double. Cette règle ne mentionne pas les consonnes *triples*, qui sont traitées comme des doubles ; elles sont alors inutiles et on les remplace donc par des doubles. La langue française ne peut pas, malheureusement, en dire autant : comparer « vile » et « ville », qui se prononcent de la même manière.



Մ	+	Է	→	ՄԷ	Մ	+	Ի	→	ՄԻ
Մ	+	Ի	→	ՄԻ	Մ	+	Ե	→	ՄԵ
Մ	+	Խ	→	ՄԽ	Վ	+	Ն	→	ՎՆ
Մ	+	Լ	→	ՄԼ	Մ	+	Ն	→	ՄՆ

Figure 4: Les principales ligatures arméniennes

#### 4. Les ligatures contextuelles

Une ligature contextuelle est une chaîne de caractères prenant des formes spéciales suivant leur position dans le mot en obéissant à des règles grammaticales strictes et liées uniquement à l'écriture. L'écriture latine n'utilise aucune ligature contextuelle : on a classifié « œ » comme étant une ligature *linguistique*, puisque la chaîne de caractères « oe » n'est pas *toujours* remplacé par la ligature « œ » : les critères sont d'ordre grammatical, et ne sont pas des invariants de l'écriture latine, mais plutôt de la langue française : en allemand cette même chaîne de caractères ne produit jamais de ligatures (mais peut par contre être remplacé dans certains cas par le caractère diacrité « ö »). Par contre les ligatures contextuelles ne dépendent que de l'écriture et non pas de la langue qui se sert de cette écriture.

Le cas par excellence d'écriture riche en ligatures contextuelles est l'écriture arabe<sup>12</sup>. Ainsi, en arabe, aura-t-on :

$$\begin{aligned}
 & \text{ت} + \text{غ} \rightarrow \text{تغ} \\
 & \text{ت} + \text{غ} + \text{ه} \rightarrow \text{تغه} \\
 & \text{ت} + \text{غ} + \text{ه} + \text{ك} \rightarrow \text{تغكه}
 \end{aligned}$$

et ainsi de suite...

<sup>12</sup>Qui comme nous le savons est utilisé pour un grand nombre de langues tout à fait différentes : l'arabe, le persan, l'ourdou, le sindhi, le pachto, le kirghiz, l'ouïgour, le malais, et j'en passe. Par définition, ces ligatures sont strictement les mêmes pour toutes ces langues.

Chaque lettre peut être écrite de quatre manières différentes suivant sa position dans le mot : initiale, médiane, finale, isolée. Si l'on veut être minimaliste, pour écrire en arabe il suffit de connaître les quatre formes de chaque lettre et de savoir qu'on a une ligature linguistique (voir section 2) : la paire « lām-alif » ne produit pas  $\text{ﻻ}$  qui est la concaténation du  $\text{ﻻ}$  initial et du  $\text{ﻻ}$  final, mais  $\text{ﻻ}$ . L'écriture arabe exige donc, comme ensemble minimal de règles, l'utilisation d'une ligature linguistique et de plusieurs ligatures contextuelles.

Heureusement tout le monde n'est pas minimaliste, et l'écriture arabe va beaucoup plus loin que cela ! À part une ligature *linguistique* et les ligatures *contextuelles* (qui font partie de la grammaire de base de l'arabe), on a aussi des ligatures *esthétiques* : en bonne typographie égyptienne (l'Égypte est considérée comme un des centres de la culture arabe), au lieu d'avoir  $\text{ﻻ} + \text{ﻻ} \rightarrow \text{ﻻﻻ}$  on aura  $\text{ﻻﻻ}$ , les lettres  $\text{ﻻ}$ ,  $\text{ﻻ}$ ,  $\text{ﻻ}$  vont former la triple ligature  $\text{ﻻﻻﻻ}$  au lieu de l'écriture linéaire  $\text{ﻻﻻﻻ}$ , le nom « Mohammed » s'écrira  $\text{ﻣﻮﻫﻤﻢ}$  au lieu de  $\text{ﻣﻮﻫﻤﻢ}$ , etc.

Nous avons fait une étude détaillée des ligatures esthétiques (typographiques<sup>13</sup>) de l'écriture arabe, [5]. Un cas similaire à celui de  $\text{ﻻ}$  est  $\text{ﷲ}$ , ligature utilisée *uniquement* dans le mot  $\text{ﷲ}$  (Dieu).

D.E. Knuth généralise la notion de ligature contextuelle, en considérant le début et la fin de mot comme des caractères : notons-les  $\langle \text{bow} \rangle$  et  $\langle \text{eow} \rangle$ . Ainsi, la transformation du sigma grec en fin de mot en sigma final, est-elle une ligature contextuelle :  $\Gamma\iota\acute{\alpha}\nu\nu\eta\sigma\langle \text{eow} \rangle \rightarrow \Gamma\iota\acute{\alpha}\nu\nu\eta\varsigma$ .

On rencontre le phénomène de ligature (au sens étendu) de début/fin de mot également en vieil allemand, et même en vieux français : un contemporain de La Fontaine écrirait

Le bon chasseur fachant chaffer fait chaffer sans son chien<sup>14</sup>

où les « s » sont courts en fin de mot et sinon longs. En allemand la situation est plus compliquée : on aura un « s » court même à l'intérieur d'un mot, si ce mot est composé<sup>15</sup> et que la lettre « s » se trouve en fin de composante :

Haus + Meister = Hausmeister

ou mieux :

Nihilistendynamittheaterkältchenfprengungssattentätsverfuchungen,

exemple tiré du *TeXbook*.

Le lecteur pourra trouver une étude de la typographie vieille-allemande dans [4].

<sup>13</sup>Et non pas *calligraphiques*, dont le nombre et l'envergure ne dépend que de l'imagination et du savoir-faire du calligraphe...

<sup>14</sup>Nous avons assemblé les polices utilisant automatiquement un « f » long en début ou à l'intérieur d'un mot et un « s » court à la fin d'un mot.

<sup>15</sup>De nouveau l'ergonomie allemande : le « s » a un rôle grammatical bien précis, il dénote la limite d'une composante : „Anfchrift“ est clairement composé de „an“ et „Schrift“, tandis que „Ausgang“, de „aus“ et „Gang“.

Pour clore la partie « non-TeX » de ce tour du monde des ligatures, mentionnons une quatrième catégorie de ligatures, tout à fait récente : les ligatures *pictographiques*. Le lecteur ayant accès au réseau l'a sans doute deviné, il s'agit bien évidemment des « *smilies* » (sourires), utilisés couramment pour exprimer certains sentiments ou états d'âme : :- ) pour désigner la joie, le bonheur, la remarque ironique ou à ne pas prendre au sérieux, :- ( la tristesse, ; - ) le clin d'œil, : - o la surprise, 8 - | le suspense, :- | et X - ( des états proches de l'apopléxie, | - o le baillement, et ainsi de suite...

Le lecteur intéressé trouvera une liste assez longue de *smilies* dans le livre récemment paru de Seth Godin [3] (on y trouve aussi des « créations » en matière de *smilies*, comme c | :- = ) (Charlot), \* < : o ) (Bozo le clown) = ) :- ) (Oncle Sam), etc.). Que dire de cette catégorie bien particulière de ligatures, imprégnées d'humour américain ? Tout d'abord soulignons qu'il est faux de dire que ces pictogrammes sont portables « parce qu'ils sont composés en ASCII 7-bit ». En fait, ils dépendent de la police de caractères utilisée : ils sont le mieux représentés par une police écran de console UNIX, ou par une police de type machine à écrire. Affichés ou composés dans une autre police, ils seraient difficilement reconnaissables.

Mais le plus important est peut-être le fait que ces ligatures pictographiques ne sont que la représentation ASCII 7-bit (la « transcription ») de pictogrammes réels, qu'on a intérêt à utiliser si l'environnement le permet. Certains *smilies* ont même été « institutionnalisés » : il s'agit des deux *smilies* les plus classiques (datant de l'époque des Beatles), :- ) et :- ( (le second étant aussi connu comme *frownie*), ou des pictogrammes ☺ et ☹, qui font bel et bien partie du code Unicode/ISO 10646-1 (positions 0x2639 et 0x263A), dans la même table que les signes météorologiques, les symboles religieux ou idéologiques, les hexagrammes du Yi-King, les signes du zodiaque, les pièces du jeu d'échecs, du jeu de cartes, et les notes de musique. Le lecteur trouvera ces deux *smilies* dans la police *smilie* (au format METAFONT ou PostScript) sur <ftp.ens.fr>, dans `pub/tex/yannis/sourires`. L'auteur trouve que ces signes ont leur place même dans un texte scientifique sérieux et favorisent son assimilation<sup>16</sup>. Après tout, la philosophie de TeX est basée sur la convivialité, même dans des conditions difficiles.

## 5. Du côté de TeX

Après avoir fait le tour des principales catégories de ligatures, voyons comment TeX parvient à les composer efficacement. Pour lui la situation est beaucoup plus simple : les ligatures sont des caractères d'une police et peuvent être accédées

<sup>16</sup>Un effet de décompression psychologique analogue à celui de la plaisanterie devenue presque obligatoire au milieu d'un cours magistral difficile.

- soit directement, par une commande (`\oe{ }` ou `\Mm{oe}` pour « œ », `\ae{ }` ou `\Mm{ae}` pour « æ » etc., la commande `\Mm` servant à déjouer les pièges de la commande `\uppercase`)<sup>17</sup> ;
- soit par une suite de caractères (`f i` donnera « fi », `f f i` donnera « ffi » ; en arabe, `l A` donnera  $\text{لأ}$ , etc.)

Dans le deuxième cas, on parle de *ligature*  $\text{T}_{\text{E}}\text{X}$  ; en effet, il s'agit d'une opération inhérente à la police. Cette opération consiste à remplacer une paire de caractères  $(A, B)$  par un troisième caractère  $C : A + B \mapsto C$ . Ainsi pour  $A = \text{« f »}$ ,  $B = \text{« f »}$ , on aura  $C = \text{« ff »}$  ; pour  $A = \text{« ff »}$ ,  $B = \text{« i »}$ , on aura  $C = \text{« ffi »}$ , etc. Depuis la version 3 de  $\text{T}_{\text{E}}\text{X}$  on dispose aussi de ligatures spéciales (connues généralement sous le nom de « ligatures intelligentes », traduction de *smart ligatures*). Celles-ci se présentent de la manière suivante :

$$\begin{aligned} A + B &\mapsto AC \\ A + B &\mapsto CB \\ A + B &\mapsto ACB \end{aligned}$$

On a également la possibilité de considérer le début de mot ou la fin de mot comme caractère (au sens de  $\text{T}_{\text{E}}\text{X}$ ) et définir les ligatures du genre

$$\begin{aligned} \text{<bow>} + B &\mapsto C \\ A + \text{<eow>} &\mapsto C \end{aligned}$$

Une application intéressante de ces outils est la contextualisation de l'écriture arabe : en écrivant `AHmd`,  $\text{T}_{\text{E}}\text{X}$  écrirait en l'absence de ligatures :  $\text{أحمد}$  ; par contre, si on introduit les ligatures (intelligentes)

$$\begin{aligned} \text{<bow>} + \text{ل} &\mapsto \text{ل} \\ \text{ل} + \text{ا} &\mapsto \text{لا} \end{aligned}$$

on retrouve bien le nom  $\text{أحمد}$  (au fait, on va encore plus loin en introduisant la ligature esthétique  $\text{ا} + \text{ا} \mapsto \text{اا}$ , pour avoir comme résultat final  $\text{أحمدا}$ ).

## 5.1. La syntaxe des ligatures

On peut définir des ligatures  $\text{T}_{\text{E}}\text{X}$ , soit en `METAFONT`, au moment de la création de la police, soit dans le fichier `PL` (qui est le fichier `TFM` rendu lisible à l'aide de `TFtOPL`). La syntaxe est très simple dans les deux cas.

<sup>17</sup>Cette commande est inutile quand on utilise  $\Omega$  à la place de  $\text{T}_{\text{E}}\text{X}$  :  $\Omega$  est une extension de  $\text{T}_{\text{E}}\text{X}$  développée par John Plaice et l'auteur, cf. [8].

### 5.1.1. Définir des ligatures dans le fichier MF

On peut utiliser les opérateurs suivants : `ligtable`, `boundarychar`, `=:`, `|=:`, `=:|`, `|=:|`, `|=:>`, `=:|>`, `|=:|>`, et `|`. Le cas le plus simple (celui de la ligature « fi ») est :

```
ligtable "f"\colon{} "i" =\colon{} hex"1C";
```

qui signifie : « la table de ligatures de la lettre « f » est la suivante : quand cette lettre est suivie par « i », remplacer les deux lettres par le caractère de position hexadécimale 1C de la police ».

Les opérateurs `|=:`, `=:|`, `|=:|`, signifient « garder la première lettre et remplacer la deuxième par... », « garder la deuxième lettre et remplacer la première par... », « garder les deux lettres et insérer ce qui va suivre entre elles », respectivement. À quoi servent alors les `|=:>`, `=:|>` et `|=:|>` ?

En fait, quand METAFONT va garder une ou deux lettres et remplacer ou insérer une autre, il re-examine ensuite la situation et applique, le cas échéant, de nouveau des ligatures ou crénages. Ainsi, supposons qu'on ait défini une part de crénage entre « A » et « C » et une ligature

```
ligtable "A"\colon{} "B" |=: "C";
```

alors T<sub>E</sub>X remplace d'abord « AB » par « AC » et effectue ensuite le crénage. Par contre, si la ligature était définie par

```
ligtable "A"\colon{} "B" |=:> "C";
```

il aurait remplacé « AB » par « AC » et serait passé immédiatement au reste du texte. Le signe `>` signifie donc « passer à l'étape suivante ».

Enfin, comment définir des ligatures de début ou de fin de mot ? On nomme un caractère (qui peut être invisible) « caractère de frontière » (*boundarychar*). Ce caractère fait office de « fin de mot » dans la table de ligatures. On peut bien sûr l'utiliser normalement dans le texte, l'« utilisation abusive » d'un caractère normal restant interne à la table de ligatures. À noter que la « fin de mot » au sens T<sub>E</sub>X, signifie « caractère suivi d'espace blanc ». Si l'on veut obtenir les mêmes transformations devant les signes de ponctuation, il faut les définir séparément. Ainsi, dans nos polices grecques, nous demandons un sigma final devant un point : « σ » + « . »  $\mapsto$  « ζ. », mais PAS devant une barre oblique : on écrit souvent Θες/νίκη comme abréviation de Θεσσαλονίκη. Pour éviter une ligature de ce type, on peut utiliser la commande `\noboundary`, au niveau T<sub>E</sub>X : ainsi `f'ilos\noboundary` donne-t-il « φίλος. » (abréviation de φιλοσοφία), au lieu de « φίλος. ».

Pour le début de mot, on dispose de l'opérateur `|`. La syntaxe est la même que pour un caractère ordinaire :

---

```
ligtable ||\colon{ } "B" =: "C";
```

signifie qu'en début de mot, B sera remplacé par C.

### 5.1.2. Définir des ligatures dans le fichier PL

Les opérateurs sont les mêmes qu'en 5.1.1, écrits en syntaxe PL : LIGTABLE, BOUNDARY-CHAR, LIG, /LIG, LIG/, /LIG/, /LIG>, LIG/>, /LIG/>>, et | |. Le premier exemple de la section précédente sera écrit

```
(LIGTABLE C f
  (LIG C i H lC)
)
```

Il n'y a donc que la présentation qui change ; pensez quand même à laisser *trois* blancs pour chaque retrait, sinon PLtoTF va se plaindre.

## 5.2. Exemple

Voici le problème le plus vicieux auquel nous avons été confronté, et comment nous l'avons résolu :

- supposez que vous êtes en train d'appliquer des ligatures (intelligentes) de T<sub>E</sub>X à une police codée en T1 (codage de Cork), pour obtenir des « f » longs automatiques. Pour être complets il va falloir dessiner des ligatures du type « fi », « fl », etc. utilisant le « f » long : « fi », « fl » et ainsi de suite. Il y en a une plus méchante que les autres : c'est la ligature « ff ». Voyons d'abord pourquoi elle est méchante.
- supposez que vous soyez T<sub>E</sub>X. Vous venez de lire un f, et vous tombez sur un s. Vous savez que fs doit produire la ligature « ff », vous remplacez donc fs par ce signe, et continuez la lecture. Et là vous découvrez avec horreur que vous êtes en fin de mot : on n'a pas prévu de ligature « f + s court ».

Deux solutions possibles : bidouiller au niveau de ligatures T<sub>E</sub>X, ou dessiner une telle ligature hypothétique.

### 5.2.1. Bidouillage au niveau des ligatures T<sub>E</sub>X

La ligature « f + s court » n'existe pas, il FAUT donc retourner en arrière, l'effacer et à sa place composer un « f » normal, suivi d'un « s » court. Mais T<sub>E</sub>X ne peut pas faire marche arrière. Le caractère « ff » est déjà emmagasinée, on est pris au piège.

Voici la solution : on va utiliser un caractère auxiliaire qui n'apparaîtra jamais à l'intérieur d'un mot, par exemple le caractère <14> (guillemet simple ouvrant , qui est censé n'apparaître que devant un mot). On applique les ligatures suivantes :

1. « f » + « s » = : « ff » (« f » et « s » donnent la ligature « ff ») ;
2. « ff » + <eow> | = : | <14> (« ff » en fin de mot donne « ff<14><fdm> ») ;
3. « ff » + <14> = : | « f » (« ff » suivi de <14> donne « f » suivi de <14>) ;
4. « f » + <14> | = : > « ff » (« f » suivi du caractère auxiliaire <14>, donne de nouveau « f », suivi d'un « f » long et PASSE À L'ÉTAPE SUIVANTE !) ;
5. « f » + <eow> = : | « s » (le cas classique : « f » long en fin de mot, donne « s » court).

Le point crucial est la quatrième étape : dans les trois premières étapes on s'arrange pour avoir, à partir de la ligature « ff » en fin de mot, le caractère « f » suivi du caractère « bidon » <14>. On remplace cette paire de caractères par « f » suivi de « f » long. Le lecteur voit le danger imminent : on est à la case-départ ; si on ne fait pas attention, le processus va recommencer *ad infinitum*. D.E. Knuth n'a pas prévu de mécanisme de détection des cercles vicieux ; on risque donc de condamner notre magnifique machine à remplacer des paires « f<14> » par des paires « ff » à l'infini...

Voici donc l'utilité de l'opérateur > : « f<14> » est remplacé par la paire non ligaturée « ff » et on passe tout de suite à l'étape suivante, qui va remplacer le « f » long par un « s » court, et on est enfin « sorti de l'auberge ».

### 5.2.2. Dessiner une ligature hypothétique « f » + « s court »

Sachant que le *es-zet* allemand « ß » provient de la ligature « s long » + « s court », on pourrait appliquer le même processus graphique à la combinaison de lettres « f » + « s court » et dessiner une ligature « ß ». Sauf que cette ligature n'a jamais été gravée (à la connaissance de l'auteur) et demeure donc une invention de notre part. Est-ce bien raisonnable d'inventer une telle ligature aujourd'hui, alors que le « f long » n'est plus utilisé depuis quelques siècles ? La question reste ouverte. Nous présentons une telle tentative au lecteur sur la figure 5.

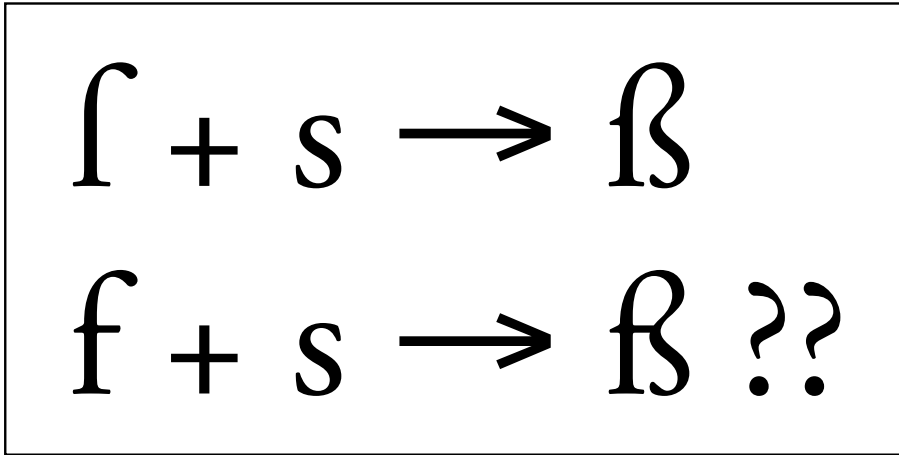


Figure 5: Ligature hypothétique « f » + « s court »

## Bibliographie

- [1] በኮምፒውተር ፕሮሰሰ የመልክል-ፊደል ነጠፋ መፅሀፈ ኢላተክዩሰነድ ዝግጅት መመሪያ አባስበላይ አላምነህ, 1993.
- [2] A. ANDULEM (AMNULEHE). « The road to Ethiopic T<sub>E</sub>X ». *TUGboat*, 10(3):352–354, 1989.
- [3] S. GODIN. *The Smiley Dictionary*. Peachpit Press, Berkeley, 1993.
- [4] Y. HARALAMBOUS. « Typesetting Old German: Fraktur, Schwabacher, Gotisch and initials ». *TUGboat*, 12(1):129–138, 1991.
- [5] Y. HARALAMBOUS. « Typesetting the Holy Qur’ān with T<sub>E</sub>X ». In *Proceedings of the 2nd International Conference on Multilingual Computing—Arabic and Latin script (Durham)*, 1992.
- [6] Y. HARALAMBOUS. « The Khmer Script tamed by the Lion (of T<sub>E</sub>X) ». In *Proceedings of the 14th T<sub>E</sub>X Users Groups Annual Meeting (Aston, Birmingham)*, 1993.
- [7] Y. HARALAMBOUS. « Indica, a Preprocessor for Indic Languages—Sinhalese T<sub>E</sub>X ». In *Proceedings of the 15th T<sub>E</sub>X Users Groups Annual Meeting (Santa Barbara)*, 1994.
- [8] Y. HARALAMBOUS and J. PLAICE. « First Applications of Ω: Greek, Arabic, Khmer, Poetica, ISO 10646/UNICODE, etc. ». In *Proceedings of the 15th T<sub>E</sub>X Users Groups Annual Meeting (Santa Barbara)*, 1994.