Cahiers GUTenberg

MOTIFS FRANÇAIS DE CÉSURE TYPOGRAPHIQUE

Daniel Flipo, Bernard Gaulle, Karine Vancauwenberghe

Cahiers GUTenberg, nº 18 (1994), p. 35-60.

<http://cahiers.gutenberg.eu.org/fitem?id=CG_1994___18_35_0>

Motifs français de césure typographique

Daniel FLIPO [α]
Bernard GAULLE [β]
Karine VANCAUWENBERGHE [γ]

Résumé. Cet article a pour objectif de faire un bilan des différentes versions de fichiers de motifs de césure français et de proposer une nouvelle version entièrement revue et corrigée. Une introduction à la division des mots y est faite ainsi qu'au processus de division employé par TEX.

Abstract. *The aim of this article is to compare the various current versions of the French hyphenation files and to propose a completely new updated and corrected version. A short introduction is given to French hyphenation as well as to TEX word-splitting mechanisms.*

1. La méthode LIANG

Avant les années 80, la division des mots se faisait souvent encore *à la main* selon les habitudes acquises de l'expérience ou de la sagesse des anciens. L'auteur d'un document n'intervenait que rarement dans ces décisions qui regardaient surtout les typographes. L'informatisation de la chaîne éditoriale a changé progressivement les méthodes et les professions. Les meilleurs programmes de division de mots ne s'intéressaient qu'aux règles classiques de syllabation du français, sur lesquelles nous reviendrons par la suite. Il était alors souvent nécessaire de se constituer des dictionnaires d'exceptions personnels. Par ailleurs, l'aptitude de ces programmes à traiter des documents multilingues était quasiment nulle. C'est de l'université de Stanford (EUA) que nous est arrivé encore une fois, la solution à ce problème.

F.M. LIANG [18] a publié une thèse en 1983 dans laquelle il propose un algorithme capable de prendre en compte plusieurs langues, pourvu que

α. Maître de conférences, université de Lille 1.
β. Ingénieur de recherche, IDRIS-CNRS, Orsay.
γ. Agrégée de lettres modernes, Bordeaux.

lui soient fourni des données linguistiques adaptées. C'est cet algorithme que Donald KNUTH a justement choisi d'implémenter dans TEX et c'est grâce à lui que J. DÉSARMÉNIEN a pu commencer en 1983 son travail [4] sur les césures françaises dans TEX (toujours a Stanford). Nous verrons que cet algorithme n'utilise pas de dictionnaire mais des *motifs* spécifiques à chaque langue. Cette méthode, efficace et économique, a malgré tout sa contre-partie : toute altération même minime d'un seul motif peut *casser* entièrement la validité de l'algorithme pour la langue en question. C'est pourquoi nous souhaitons que les motifs français continuent d'être *certifiés* par l'association GUTenberg, en faveur de laquelle nous avons réalisé ce travail.

Pour la petite histoire il faut dire que cet algorithme a été *repris* par Ronald MCINTOSH[1] qui fournit le monde entier et en exclusivité, en algorithmes de coupure de mots pour photocomposeuses. Ses fichiers de motifs sont bien entendu très confidentiels, nous ne les avons pas approchés ; mais les milieux bien informés disent qu'une révision est déjà annoncée pour le français...

2. La division des mots : une opération délicate

La division des mots a lieu normalement aux frontières des syllabes graphiques : la coupe en syllabes graphiques répond à un certain nombre de règles (voir plus loin page 38), la syllabe graphique se distinguant par quelques points de la syllabe phonique. En français, la syllabe est une unité phonique comprenant obligatoirement une voyelle, accompagnée éventuellement de consonnes ou de semi-consonnes (qui la précèdent et/ou la suivent).

Les principales divergences entre syllabe graphique et syllabe phonique sont les suivantes :

- la syllabe graphique conserve tout « *e* » muet placé entre une consonne et une autre consonne ou une fin de mot ;
- la syllabe graphique sépare les consonnes doubles même si elles sont prononcées simples.

1. Voir 1° R. MCINTOSH, *Hyphenation*, Ronald MCINTOSH publication, ISSBN 1 872757-00-6, London, 1990 ; 2° D. FAWTHROP, *Hyphenation by Algorithm of English/American and other Languages*, Computer Hyphenation Ltd, 1990, ISSBN 1 872757-02-2.

On peut voir, par exemple, que l'on distingue graphiquement trois syllabes dans `pu-re-té`, même si l'on prononce `[pyr-te]` (deux syllabes phoniques).

Il est coutume de dire que la division des mots se fait normalement entre syllabes mais aussi conformément à l'étymologie lorsque celle-ci est « *nettement perceptible* » [12]. Il n'existe malheureusement pas de traité ni de dictionnaire français sur ce sujet. Contrairement au domaine orthographique géré par l'Académie, il n'existe pas d'organisme détenteur de droits sur la typographie française. Du coté de la linguistique, domaine qui nous occupe aussi tout naturellement et où les *traités, grammaires* ou *dictionnaires* foisonnent, rien n'est établi non plus concernant la division des mots. Seuls quelques auteurs, en nombre restreint, ont publié des éléments de réponse en donnant des exemples, mais jamais on ne trouvera une liste vraiment exhaustive des règles à appliquer. Notre objectif n'est toutefois pas d'y remédier car nous pensons qu'il faut laisser la place aux différentes écoles, tout comme à l'innovation. Les règles que nous proposons plus loin permettront à chacun de disposer d'éléments de décision (lorsque l'une d'elles est difficilement mise en œuvre dans TEX nous l'indiquons en note de bas de page).

Pour poser clairement le problème de la division des mots en français, citons Jacques Désarménien [5] :

« Comme l'écrit Thimonnier [19] : « ceux dont le français est la langue « maternelle, appliquent spontanément les règles de la syllabation parlée. « Ils n'ont qu'à se laisser guider par la prononciation usuelle. » dans la pratique typographique, les choses ne sont pas aussi simples : on sait bien que, en français, la correspondance n'est pas parfaite entre l'écriture et la prononciation. De fait, les problèmes rencontrés pour la détermination de la coupure d'un mot proviennent de la confusion faite généralement entre syllabation parlée et syllabation écrite. »

On ne saurait donc être trop prudent avant d'énoncer quelques règles complémentaires. Pourtant, il nous est apparu qu'il était important de diffuser quelques rudiments que l'on trouve épars dans quelques livres d'auteurs [2, 3, 6, 12, 19, 15, 10], ou dans le *Code typographique* [7]. Ces références divergent parfois sur quelques points de vue ; nous avons fait notre propre choix.

3. Définitions

Le terme de *césure* (*typographique*) est souvent jugé équivalent aux termes plus simples de *coupe*, *coupure*, ou *division*. Il nous faut, cependant, expliquer les distinctions que nous allons leur appliquer par la suite :

- la *division d'un mot* est une sorte d'algorithme, non formel, qui doit être employé lorsqu'un mot arrive en bout de ligne et déborde dans la marge, de façon à obtenir l'endroit approprié de sa coupure (marquée par un trait d'union en fin de ligne) ;
- la *césure* est une autre forme d'algorithme, qui peut être tout à fait formel comme dans le cas de TEX et qui sert à déterminer *tous* les points possibles de coupure dans un mot donné ; il faut alors parler de *points de césure* ; l'algorithme de TEX est basé sur un ensemble de *motifs de césure* dont nous parlerons plus loin ;
- le terme *coupure* sera souvent employé par la suite en lieu et place de *division*.

Dans ce qui suit nous allons aussi adopter les notations suivantes :

`-`	:	trait d'union ou point de césure
`/`	:	division interdite à cet endroit
`-/`	:	division interdite après le trait d'union
`-[`	:	division proposée
`[ ]`	:	notation phonétique (API [2])

4. Quelques règles

Avant de faire état des règles de coupure syllabique et étymologique, voyons d'abord quelques règles générales puis typographiques. Le terme de *règle* employé ici et par la suite est probablement un peu fort s'il est lu avec sa connotation de *règlement*. Il s'agit plutôt pour nous de règles de conduite, voire de conseils, d'autant que ces règles évoluent dans le temps, parfois très fortement, nous en verrons des exemples par la suite : aussi ce que nous écrivons aujourd'hui doit être considéré comme un recueil (non exhaustif) de ce que nous considérons comme de *bonnes habitudes* actuelles.

2. Alphabet Phonétique International.

Les points de césure figurant dans les exemples à venir sont aussi ceux exactement produits par TEX après application des éléments fournis dans cet article.

4.1. Règles générales

g-a) Il faut éviter au maximum de couper les mots. Cela étant dit, nous devrions nous en tenir là, mais comme rien n'est parfait... et qu'il faut, malgré tout, parfois couper, voyons les autres règles.

g-b) Il est important de suivre les mêmes règles tout au long du document.

g-c) On n'isole pas en fin de ligne une seule lettre et on en rejette au moins trois (ex. : `obéi` n'est pas coupé, tout comme `agréé`, voir [6]).

g-d) « Les mots étrangers employés dans un texte en français doivent être divisés, le cas échéant, suivant les règles de la langue étrangère » (extrait de [7]). Il faut donc que l'utilisateur de logiciel puisse indiquer de quelle langue il s'agit, et ce au niveau le plus bas c'est-à-dire au niveau du mot.

Il est précisé, par exemple, dans [2] que : « Les règles d'emploi du trait d'union varient d'une langue à l'autre. En français, on ne coupe pas un mot avant ou après un *x* ou un *y* placé entre deux voyelles. En anglais, on peut le faire. »

Cette idée est reprise par [3] qui donne les exemples suivants : en français, *fixer*, *moyen* ne peuvent pas être coupés, à l'inverse de *mixture*, *tex-tile* et *cy-près*.

g-e) « On évitera les coupures malsonnantes :

```
con/trôle    cul/ture
```

« [...] De ce fait on ne peut couper certains mots comme `connexion` » (extrait de [7]).

g-f) On ne coupe pas les noms propres (on ne coupe pas non plus entre initiale et nom [3]), les abréviations, les acronymes, les nombres et les dates exprimés en chiffres [4].

g-g) On ne sépare pas un nombre de l'unité qui lui est rattachée [5] (m, kg, %, F, *etc.*).

3. Nous n'abordons pas, volontairement, le problème de la coupure entre les mots qui n'est pas l'objet de cet article mais qui, à lui tout seul, mériterait une longue discussion (on en trouvera un apperçu avec la règle suivante).

4. En (LA)TEX l'option de style `french` apporte quelques commandes pour effectuer ce travail.

5. Pour éviter ce problème, les logiciels disposent en général d'un espace insécable (~ en TEX).

g-h) On coupe les mots composés au trait d'union qu'ils possèdent déjà ;

- dans le cas du *t* euphonique, on coupe naturellement au premier trait d'union [6] : `dira-[t-il`
- par contre, « pour la même raison d'euphonie, on divisera : `c'est-à-[dire` et *non* `c'est-/à-dire` » (extrait de [7]).

4.2. Règles typographiques

Traitant, ici encore, de règles très générales, nous n'entrerons pas dans le débat *pour* ou *contre* la justification à droite dans certaines conditions, ceci étant un préalable à la décision de couper ou de ne pas couper les mots. Nous ne parlerons pas non plus de la taille des justifications qui modifient parfois les données de la coupure ; on se reportera à [7], par exemple, pour de plus amples détails. On retiendra toutefois les quelques points suivants, applicables dans la grande majorité des cas :

t-a) Aucune césure ne doit intervenir en bas de page (et spécialement sur les pages impaires) pour ne pas rendre la lecture plus difficile[6].

t-b) Un paragraphe ne doit pas être terminé, du fait d'une division de mot, par une ligne de longueur inférieure au renfoncement[7].

t-c) On ne doit jamais trouver 3 lignes consécutives terminées par un trait d'union[8]. Mais au temps des DIDOT on n'hésitait pas à couper 6 ou 7 fois de suite !

4.3. Règles de coupure en syllabes graphiques

Les règles générales et typographiques étant énoncées, il est maintenant possible de savoir si un mot donné, placé à un endroit précis du document, peut être candidat à la division ou non. En supposant qu'il le soit, il faut

6. TEX choisit l'endroit où il doit changer de page en maximisant une « note d'aspect » pour la page en cours. Malheureusement aucun paramètre n'est prévu pour pénaliser de telles coupures de mots, et il arrive que TEX produise effectivement des pages dont le dernier mot est coupé (voir [17] et [13]). Dans ce cas il est nécessaire d'intervenir manuellement.

7. Cette règle n'est pas mise en œuvre directement en TEX sous cette forme. Dans son calcul de note d'aspect TEX prend en compte le paramètre `\finalhyphendemerits` dont la valeur standard (5 000) le dissuade fortement de terminer par une césure l'avant dernière ligne d'un paragraphe (cf. [17] p. 98).

8. TEX prend en compte un paramètre `\doublehyphendemerits` dont la valeur standard (10 000) pénalise très fortement toute paire de lignes consécutives terminées par une césure (cf. [17] p. 98). Dans le cas d'un triplet de telles lignes la pénalité serait doublée, rendant la disposition du paragraphe inacceptable.

alors s'attacher, en premier lieu, à le diviser « normalement » aux frontières des syllabes graphiques. Voici quelques règles à cet effet :

s-a) la coupure est possible entre voyelle et consonne unique [6] :
`co-[lon`

s-b) la coupure peut être réalisée entre les consonnes lorsqu'il y en a deux :
`col-[lant` `fac-[teur`
`al-[bi-nos` `par-[tir`

La coupure entre deux consonnes identiques a lieu même si ces consonnes doublées dans l'écriture sont presque toujours prononcées simples. Ainsi on coupe `souf-[fler` et `col-[lè-gue`, bien que l'on prononce généralement `[su-fle]` et `[ko-lɛg]`.

- la coupure s'effectue devant la plus forte de deux consonnes formant un groupe décroissant. Un tel groupe est en général constitué d'une consonne que l'on peut qualifier de forte – en particulier une occlusive, sourde (`[p]`, `[t]`, `[k]`) ou sonore (`[b]`, `[d]`, `[g]`), ou parfois une constrictive, notamment `[f]` et `[v]` – suivie d'un phonème plus *faible* : une semi-consonne[9] ou le plus souvent une sonante, telle la latérale `[l]` et la vibrante `[r]`. D'où de nombreux groupes indissociables notés `cr`, `dr`, `bl`, `fr`... :
`câ-[bler` `nau-[frage` `ta-[bleau`
- mais on peut aussi tenir compte de l'étymologie (voir 4.4) :
`sub-[stan-tif`[10]
- *ch, gn, ph, th,* ne sont jamais coupés lorsqu'ils ne transcrivent qu'un son, ce qui n'est pas toujours le cas. En effet, on relève quelques termes, qui font figure d'exceptions, pour lesquels les deux consonnes transcrivent deux phonèmes, et sont donc susceptibles d'être coupés. J. DÉSARMÉNIEN avait parfaitement noté cela pour les mots suivants :
`stag-[na-tion` `wag-[né-rien`.

s-c) Des voyelles consécutives en hiatus (ex. `thé/âtre`) constituent des syllabes entre lesquelles on ne va pas à la ligne, sauf s'il s'agit de préfixes [6] :
`pré-[avis` `pro-[émi-nent`

9. Les trois semi-consonnes du français sont :
`[j]`, comme dans *ciel, faille*
`[ɥ]`, comme dans *lui, nuit*
`[w]`, comme dans *loi, Louis*

10. [15] préfère la coupure syllabique ordinaire : `subs-tan-tif`.

s-d) On ne rejette pas à la ligne, pour la commodité de la lecture, une syllabe muette [6] :
`es-[pè/rent`

s-e) On ne divise jamais après une apostrophe, ni immédiatement après une élision (comme dans `s'en-tr'/égor-ger`).

4.4. Règles de coupure étymologique

Le *Code typographique* [7] fait grandement abstraction des conditions à appliquer pour respecter l'étymologie :
« La division étymologique est si rarement et si difficilement observée, même par l'Académie et les lexicographes, qu'il est préférable de suivre l'avis des techniciens du Livre tels que TASSIS, FREY, Théotiste LEFÈVRE, etc. qui recommandent la division d'après l'épellation française.

« Nous reconnaissons néanmoins que certains auteurs de travaux scientifiques préfèrent la division étymologique, qui fait ressortir la racine grecque ou latine. »

Nous avions vu en g-h qu'il fallait couper les mots composés au trait d'union qu'ils possèdent déjà.

e-a) Il en est de même pour les mots composés sans trait d'union[11] :
`baise-[main  chèvre-[feuille  contre-[maître`

Par ailleurs, J. DÉSARMÉNIEN donne *ses* bons exemples mais aussi les cas difficiles, voire insolubles de façon informatique usuelle. Pour schématiser à l'extrême, disons :

e-b) les préfixes tels que `re-`, `ré-`, `co-`, `trans-` (mais attention ici aux faux amis, on coupera `trans-[at-lan-tique` mais `tran-[sac-tion`) et `inter-` peuvent être considérés comme étymologiques (on peut alors couper `inter-[action`[12]) ;

e-c) ceux tels que `sur-`, `psy-`, `pro-`, `ex-` et `sub-` sont aussi à examiner avec attention, mais à l'exception de ce dernier la coupure syllabique coïncide en général avec la séparation préfixe/base ;

11. Le *Code typographique* [7] dit improprement « les mots composés dont le trait d'union a disparu ». En effet, le mot *chèvrefeuille* existait déjà en un seul mot au Moyen Age ; il semble donc qu'il n'ait jamais comporté de trait d'union.

12. Cette coupure n'est pas effectuée automatiquement par TEX avec les motifs actuels.

e-d) d'autres comme `para-` sont plus difficilement décidables : *parabole*, *paravent* et *parachever* n'ont pas la même racine étymologique, ainsi on devrait couper `pa/ra-[chro-nisme` [12] mais aussi `par-[ache-ver`.

e-f) et parfois on s'interroge, sans fin, sur des préfixes comme `poly-` ou `télé-` : faut-il couper ou conserver intact ces préfixes ?

J. DÉSARMÉNIEN précise dans [5] que 96,8 % des mots peuvent être traités de façon syllabique classique. Alors on est porté à croire que la méthode et les quelques règles que nous venons d'exposer conviendront dans la grande majorité des cas pour les textes usuels.

5. Un peu de TEXnique

Le procédé utilisé par TEX pour choisir les points de césure est exposé dans le TEXbook [17] appendice H p. 449.

Rappelons ici quelques notions nécessaires à la lecture d'un fichier de césures. Un tel fichier se compose essentiellement d'une macro `\patterns{}` dont le contenu est une suite de lignes contenant chacune un ou plusieurs motifs séparés par un espace.

Ce fichier principal peut éventuellement être accompagné d'un fichier d'exceptions (par exemple `enhyphex.tex` contient les exceptions au fichier de motifs anglo-américains, `f7hyphex.tex` [13] contient les exceptions françaises en format 7-bits).

5.1. `\patterns` : un motif de discussion...

Pour préciser l'interprétation des motifs de césure, nous allons considérer quelques lignes extraites du fichier `f7hyph.tex` [14] :

```
\patterns{...
           il2l
          vil3l
        chevil4l
```

13. Il existe aussi `f8hyphex.tex` qui lui est équivalent mais sous une forme 8-bits. Dans la pratique c'est le fichier `frhyphex.tex` qui est appelé et qui décide quel est le fichier à charger en mémoire, soit le 7-bits, soit le 8-bits. Tous ces fichiers font partie intégrante de la distribution des *fichiers du style french* sur le serveur d'archives GUTenberg : `ftp.univ-rennes1.fr` dans le répertoire `pub/GUTenberg/french`.

14. C'est le code contenu dans le fichier `frhyph.tex` qui est exécuté et qui décide quel est le fichier à charger en mémoire, soit le 7-bits `f7hyph.tex`, soit le 8-bits `f8hyph.tex`.

```
...
4che.
4ches.
.ch4
...}
```

Le principe est le suivant : les chiffres impairs indiquent les coupures possibles, les chiffres pairs non nuls les coupures interdites. Considérons les mots en *ill* : lorsque ce groupe se prononce `[ij]` (grillage, fille...), la coupure `il-l` est interdite par le motif `il2l`, mais cette même coupure entre les deux *l* est permise dans *vil-lage* pour le son `[il]` : on ajoute pour ce faire le motif `vil3l` ; ensuite il faut apporter une correction pour que le mot *cheviller* (phonème `[j]`) ne soit pas coupé entre les deux *l* : c'est l'objet du motif `chevil4l`. TEX prend en compte comme valeur paire ou impaire le maximum des chiffres rencontrés dans tous les motifs concernés : si une coupure est interdite par un `2` on peut l'autoriser dans certains cas particuliers par un `3`, ensuite l'interdire de nouveau par un `4` pour certaines exceptions et ainsi de suite.

On remarque dans la macro `\patterns` des lignes commençant ou se terminant par des points : en début de ligne un point signifie que le motif de la ligne ne s'applique qu'aux mots *commençant* par lui, le point en fin de ligne indique que le motif concerne seulement les mots *se terminant* par lui. En français, comme nous l'avons rappelé précédemment, on ne renvoie jamais à la ligne suivant une terminaison de mot muette : par exemple les coupures devant la syllabe terminale *che(s)* des mots *pervenche(s) cravache(s) fantoche(s)...* sont interdites par les motifs `4che.` et `4ches.`. De même, sans le motif `.ch4` la coupure `ch/rysanthème` serait possible...

5.2. Mettre l'accent : comment ?

Pour les langues utilisant des caractères accentués se pose le problème de la représentation de ces caractères dans les motifs de césure. Il existe deux représentations possibles, le choix dépend du moteur TEX utilisé :

– si on travaille avec un moteur MlTEX (cas de la distribution UNIX de GUTenberg) il est possible d'utiliser une représentation des caractères accentués sous la forme `\'e` pour `é` comme cela est fait dans les fichiers `f7hyph.tex` et `f7hyphex.tex` pour les césures françaises ;

– si on travaille avec TEX V3.xx *et des caractères 8-bits*, soit avec des fontes DC/EC (cas de la distribution MAC de GUTenberg), soit avec des fontes virtuelles (cas de la distribution PC de GUTenberg, à base emTEX) il *faut* utiliser une représentation des caractères accentués sous la forme hexadécimale `^^xx` décrite dans [17] p.73 (`^^e9` pour `é`) ; c'est ce qui est fait dans les fichiers `f8hyph.tex` et `f8hyphex.tex` pour les césures françaises.

Rappelons enfin que pour obtenir toutes les coupures possibles des mots accentués, il faut[15] impérativement utiliser soit MlTEX, soit un TEX V3.xx *associé à des fontes 8-bits* (virtuelles ou réelles comme les DC). En effet, si les caractères accentués sont fabriqués à partir de la macro `\accent` (qui superpose le caractère et l'accent), toute coupure est inhibée après le premier caractère accentué du mot.

6. Petite histoire de la césure française dans TEX

Les premiers travaux sur l'adaptation de TEX aux césures françaises ont été faits par J. DÉSARMÉNIEN qui a commencé son travail à Stanford au début des années 80. Une première publication a été faite dans le *TUGboat* en 1984 [4]. Mais l'essentiel du travail de J. DÉSARMÉNIEN sur les césures est exposé dans la revue *TSI* [5].

La solution technique utilisée par J. DÉSARMÉNIEN pour parvenir à des coupures correctes des mots accentués reposait sur l'utilisation de fontes francisées : les caractères spécifiques du français (â, ê, î, ô, û, é, ç, ë, è, î) remplaçaient les caractères grecs placés par Don KNUTH en position 01 à 10 des fontes CMR[16].

Dans le même temps M.-J. FERGUSON apportait une solution intéressante aux problèmes multilingues avec son MlTEX. Il devenait possible de couper des mots contenant des lettres accentuées sans recourir à des fontes francisées. Le fichier des motifs de césure de J. DÉSARMÉNIEN a donc été revu et adapté au codage 7-bits standard de TEX. MlTEX a commencé à cette époque à être très largement diffusé dans les milieux francophones (aidé en cela par les distributions PCTEX de Personal TEX Inc.). Les der-

15. Ces conditions ne sont pas entièrement suffisantes, les `\lccode` et `\catcode` des caractères utilisés dans le texte étant, notamment, très importants.

16. Le « *à* » n'a pas pu trouver place dans les dix positions réservées aux caractères grecs, mais cette lettre n'intervient qu'en fin de mot (*voilà*) ou dans des mots composés : (*tout-à-l'égout*).

nières « *marques* » de M.-J. FERGUSON dans les fichiers semblent dater de juin 88.

Ensuite est sorti TEX V3, reprenant l'essentiel (mais pas la totalité) des idées de M.-J. FERGUSON. Cette différence a amené M.-J. FERGUSON à faire son MlTEX V3.

Une autre façon de traiter le problème des césures de mots accentués, qui généralise l'approche de J. DÉSARMÉNIEN, consiste à disposer dans chaque fonte de tous les caractères accentués nécessaires. L'adoption à Cork en 1990 [8] du codage de fontes EC[17] à 256 caractères résout donc définitivement le problème pour le français comme pour toutes les autres langues européennes, à condition toutefois de disposer de fontes réelles (DC[18] par exemple) ou virtuelles.

L'apparition de cette norme a motivé le transcodage des fichiers de césures prévus pour MlTEX en fichiers 8-bits. Le transcodage naïf consiste à remplacer les séquences du type `\'e` par le caractère 8-bits correspondant... Mais le code de ce caractère dépend bien sûr de la machine utilisée (les codages PC, Mac, ISO-Latin1 et PostScript sont *tous* différents). Il faudrait donc un fichier 8-bits spécial pour chaque machine... Il existe malheureusement encore aujourd'hui sur les réseaux, y compris dans les toutes nouvelles archives CTAN, des versions utilisant ce type de codage (cf. `f8hyph.tex` récupéré le 01/12/93 sur `ftp.tex.ac.uk`).

La seule façon raisonnable de coder un fichier de motifs de césure en 8-bits est d'utiliser la notation hexadécimale `^^xy` de TEX V3 et de respecter le codage le plus standard, c'est-à-dire le codage DC. Cette notation interne à TEX a deux qualités essentielles : elle est indépendante de la machine utilisée[19] et elle résiste (en principe) aux transferts par les réseaux puisqu'elle ne fait appel à aucun caractère de code supérieur à 127.

C'est ce codage qu'a fort judicieusement choisi Y. HARALAMBOUS pour créer son fichier de motifs adapté aux fontes DC[20] ; malheureusement la liste des motifs de césure dont il est parti était altérée. Ce fichier est encore présent dans les archives sous le nom de `fr8hyph.dc` ou `fr8hyph.tex` (version 1.0 du 10 juillet 1991)...

17. Ce codage est désormais appelé `T1` avec LATEX2ε.

18. Le créateur de ces fontes considère que leur nom est temporaire ; le nom définitif doit être EC.

19. Sous réserve d'utiliser des fontes avec le même codage.

20. Il est nécessaire de disposer du fichier de motifs adapté au codage de la police utilisée à la composition du document.

Au moment où nous avons effectivement commencé notre travail (fin 1991) de nombreuses versions différentes des motifs de césure français étaient disponibles sur les réseaux. Celles que nous possédions nous-mêmes étaient-elles correctes ? Les fichiers originaux de J. DÉSARMÉNIEN avaient de toute évidence subi des ajouts et des modifications, certaines judicieuses et d'autres moins : comment faire le tri ? Dans un premier temps nous avons décidé de travailler uniquement sur les fichiers de motifs prévus pour MlTEX et de transcoder ceux-ci, lorsqu'ils seraient au point, en version 8-bits.

7. Notre travail

Nous avons commencé par une comparaison, à la main, des différents fichiers de césure française à notre disposition, dans le domaine public, en vue de déterminer lequel pouvait être considéré comme le meilleur point de départ pour notre étude. Très vite il a fallu se rendre à l'évidence et renoncer devant l'énormité du travail. Nous disposions de 5 fichiers d'origines diverses mais tous étaient vraiment différents.

Dans le doute, nous avons pris comme référence le fichier de césure distribué par GUTenberg depuis près de trois ans, pour le comparer avec les motifs publiés par J. DÉSARMÉNIEN. Nos comptages manuels font apparaître 259 différences (tableau 1).

TABLE 1 - *Divergences survenues depuis le travail de J.* DÉSARMÉNIEN

Motifs de césure	modifiés	disparus	nouveaux	total
étymologiques	25	63	69	157
phonétiques	0	60	42	102
	25	123	111	259

Notre fichier de référence n'était visiblement pas bon. Il fallait donc refaire le décompte sur d'autres fichiers mais sans avoir la garantie que celui qui se rapproche le plus de celui de J. DÉSARMÉNIEN soit vraiment le meilleur. Alors nous avons décidé, non plus de comparer les motifs de césure, mais de comparer les résultats obtenus avec la liste des 499 mots publiée en annexe de l'article de J. DÉSARMÉNIEN [5] (nous avons ajouté en fait un 500e mot, *correctement*, souvent mal coupé : `cor-re/cte-ment` au lieu de `cor-rec-[te-ment`). Il fallait alors automatiser ce processus de comparaison pour pouvoir le répéter souvent.

Nous avons donc créé les outils permettant d'effectuer les comparaisons : un premier fichier de code TEX lit le fichier des 500 mots tests correctement *hyphénés*[21] en ignorant les points de césure présents et crée un fichier `.log` comportant toutes les coupures proposées par le système. Un deuxième relit ce fichier `.log` et compare les points de césure trouvés à ceux du fichier test.

Dès les premiers essais, des anomalies graves apparurent dans la majorité des fichiers de motifs de césure disponibles :

- les points en début de ligne étaient souvent remplacés par des doubles points, ce qui a pour effet pratique d'invalider totalement le motif correspondant ; probablement est-ce dû à des problèmes de transmission sur les réseaux...
- le plus souvent, tous les motifs concernant *ill* (liste allant des motifs `il2l` à `xil3l` chez J. DÉSARMÉNIEN) avaient disparu... Est-ce le résultat d'une fausse manœuvre d'un contributeur anonyme ?
- les motifs `.con4` et `.cons4` (règle g-e) présents chez J. DÉSARMÉNIEN étaient absents dans plusieurs fichiers.
- un motif `1ct` a été ajouté, par qui ? Pourquoi ? Il a pour effet fâcheux de faire couper *correctement* en cor-re/cte-ment !

Inversement, d'autres personnes semblent être intervenues (judicieusement) sur les transcriptions du fichier original en vue de son utilisation par MlTEX, comme celles de M.-J. FERGUSON (juin 88) pour ajouter un 0 derrière les `\i` et les `\oe` de façon à permettre effectivement la coupure après les motifs concernés.

Comme on le voit, ces fichiers de motifs mis à la disposition de tous, dans le domaine public, ont pu être utilisés par tous mais aussi modifiés avec plus ou moins de bonheur. Nous étions alors encore plus confortés dans l'idée qu'il était indispensable de mettre un peu d'ordre et de laisser ensuite l'association GUTenberg seul maître et garant de leur pérénité.

Après toutes ces corrections, notre même fichier de motifs (qui comportait 259 anomalies) donna, *seulement*, 118 mots mal divisés sur les 500. Ce n'était pas encore un exploit !

21. Cet anglicisme signifie que chaque mot comportait les points de césure (sous forme de trait d'union) proposés par J. DÉSARMÉNIEN.

Le meilleur fichier de motifs disponible s'est avéré être celui de la machine Gould de l'ex-CICB de Rennes[22] qui donnait seulement 7 différences significatives[23] par rapport à la référence. Le tableau 2 récapitule les 7 différences observées. Nous avons choisi d'en faire notre référence et d'essayer

TABLE 2 - *Dernières différences observées.*

Original de J.D.	Notre résultat	Index
désami-don-ner	dés-ami-don-ner	(1)
désen-flure	dés-en-flure	(2)
déshy-dro-gé-ner	dés-hy-dro-gé-ner	(3)
déso-rien-ter	dés-orien-ter	(4)
pé-réqua-tion	pér-équa-tion	(5)
phy-toph-thora	phy-to-phthora	(6)
pontet	pon-tet	(7)

de supprimer les différences observées en le modifiant le moins possible. Le premier type de différences concerne les coupures étymologiques des mots commençant par `d\'es`. Les motifs concernant ces mots semblent avoir été profondément modifiés par rapport à la version de J. DÉSARMÉNIEN. Nous ne disposons d'aucune indication ni sur l'auteur de ses modifications ni sur ses intentions.

Nous avons réintroduit le motif `.d\'e2s` qui était présent dans le fichier original de J. DÉSARMÉNIEN, et supprimé les motifs suivants :

- `.d\'e2s1a2` responsable de la différence (1)
- `.d\'e2s1e2` responsable de la différence (2)
- `.d\'e2s3h` responsable de la différence (3)
- `.d\'e2s1o2` responsable de la différence (4)

De plus, nous avons ajouté le motif `.d\'es2a3m` pour obtenir la division `dés-a-mi-don-ner` et non `désami-don-ner`. Ce choix n'est pas celui de J. DÉSARMÉNIEN, mais il nous semble préférable de traiter de la même

22. Actuellement appelé CRI de l'université de Rennes 1.

23. Le premier essai nous donna en fait 6 erreurs de plus avec MlTEX V3, ce qui était dû à des fautes de frappe. Mais nous avons surtout été troublés un temps par les 33 erreurs produites par MlTEX V2. M.-J. FERGUSON nous expliqua alors que les accents perturbaient le décompte des caractères en fin de mot. De fait, tous les mots supplémentaires en erreur se terminaient par un e accent aigu.

manière `dés-a-mi-don-ner`, `dés-a-mor-cer` et `dés-a-bu-ser`. Enfin, la réintroduction du motif `.d\'e2s` nous a obligés à augmenter la force de la césure des motifs présents dans le fichier original `.d\'e1s2c` `.d\'e1s2p` et `.d\'e1s2t` en `.d\'e3s2c` `.d\'e3s2p` et `.d\'e3s2t`.

La différence (5) provenait vraisemblablement d'une faute de frappe (`p\'e2r1\'e2q` au lieu de `p\'e1r2\'e2q`) d'un auteur inconnu car ce motif n'existait pas chez J. DÉSARMÉNIEN.

Le motif `.phyto3ph2` (ajouté par qui ?) n'était pas bon, il donnait `phy-to-phthora` (6) au lieu de `phy-toph-[thora`; nous l'avons supprimé.

Enfin le cas du mot `pon/tet` (7) peut être réglé en ajoutant le motif `.pon2tet`.

Nous avons remplacé le motif `p\'e2n1ul` par `p\'e2nul` conformément au fichier original de J. DÉSARMÉNIEN (pour `an-té-pénul-[tième`).

Autres modifications introduites :

- rétablissement des motifs `.re2s3cisi` et `.re2s3ciso`, présents chez J. DÉSARMÉNIEN mais absents du fichier de Rennes ;
- les exceptions précisées en e-a ont aussi été ajoutées[24] ;
- rétablissement du motif `1\c c` présent chez J. DÉSARMÉNIEN mais absent du fichier de Rennes ;
- ajout du motif `.cul4` (même raison que `.con4` et `.cons4`) ;
- le mot `ré/union` était mal coupé (règle s-c) ; nous avons remplacé le motif en cause (`.r\'e1u2`) par `.r\'eu2` ;
- nous n'avons pas voulu retirer les motifs de début de mot à 1 caractère ainsi que ceux de fin de mot à 2 caractères qui n'ont pas de sens avec TEX, puisque par défaut il ne coupera jamais de la sorte ; mais les deux paramètres `\righthyphenmin` et `\lefthyphenmin` qui permettent, depuis la version 3.00 de TEX, d'imposer le minimum de lettres en fin de ligne et le minimum à rejeter sur la ligne suivante, peuvent avoir été malencontreusement changées ; et puis il reste aussi quelques installations MlTEX V2 pour lesquelles l'algorithme de dénombrement des caractères de début et de fin de mot n'est par sûr à 100 % ;

24. Nous les avions introduites dans un premier temps dans les fichiers d'exceptions, mais il est beaucoup plus performant, en terme de gestion mémoire, de les mettre sous forme de motifs de césure.

– nous avons complètement revu l'ordre des motifs de césure afin d'essayer de faciliter la lecture et la maintenance ultérieure du fichier : les motifs étymologiques regroupés en deuxième partie du fichier par J. DÉSARMÉNIEN ont été interclassés avec les motifs phonétiques[25]. L'ordre retenu est l'ordre alphabétique, à ceci près que nous avons regroupé les motifs agissant sur le même groupe de lettres. Par exemple, le motif générique `il2l` est à sa place alphabétique et les motifs complémentaires concernant le *ill* (`cil3l`, `rcil4l`, ... `xil3l`) sont rassemblés à la suite de celui-ci et indentés de manière à aligner verticalement le groupe *ill* afin d'améliorer la lisibilité. De même, nous avons préféré ne pas séparer les terminaisons muettes : par exemple les motifs `4de.`, `4des.`, `2dent.` sont dans cet ordre, les exceptions étymologiques `.d\'e1a2`, `.d\'e1io`, `.d\'e1o2`, étant reléguées après, en (légère) violation de l'ordre alphabétique.

Le test des 500 mots demeurant pour nous trop réduit par rapport à la richesse de notre vocabulaire français, nous avons souhaité vérifier les césures obtenues sur plusieurs milliers de mots. Il aurait été souhaitable de reprendre le fichier du « vocabulaire informatisé du laboratoire d'automatique et de linguistique » de Paris VII fourni par le professeur GROSS comme base de travail à J. DÉSARMÉNIEN mais nous ne l'avions pas. Par contre nous avons pu bénéficier d'une convention de recherche avec le GDR-PRC « Communication Homme-Machine » du CNRS, ce qui nous a permis d'utiliser la *base de données lexicales du français écrit et parlé* (BDLEX), soit environ 330 000 formes fléchies du français.

De ce volumineux fichier nous avons extrait aléatoirement 3 000 mots et produit la liste des mots avec les césures proposées par TEX. Nous avons relevé quelques problèmes, notamment les terminaisons muettes de pluriel de verbes (règle s-d), telles que :

`at-té-nuè/rent` `fric-tion/nent` `se-ri-nas/sent`

Aussi avons-nous décidé d'essayer de régler le problème des terminaisons en `-ent`, celles-ci n'ayant pas été traitées par J. DÉSARMÉNIEN. Le problème n'est pas simple car ces terminaisons peuvent être de trois types :

– muettes : les verbes à la troisième personne du pluriel (cf. ci-dessus), les plus nombreux et de loin, étant en `-sent` et `-rent` (imparfait du subjonctif et passé simple...),

25. Dans les fichiers `f7hyph.tex` et `f8hyph.tex`, les motifs étymologiques sont indentés de 20 caractères de façon à les distinguer des motifs phonétiques. Dans la liste publiée en appendice, ils sont précédés du signe |.

– sonores : principalement les adverbes en `-ment` (*vraiment, sûrement*) et d'autres mots (*comment, élément, parent, présent*),

– mixtes, sonores ou muettes selon le cas : `président` (*ils président, le président*), `pressent` (*ils pressent, il pressent*), `équivalent`, `ferment` etc.

TEX n'analyse pas le contexte d'un mot et ne sait pas couper *ils président* dans un cas et *le pré-si-dent* dans l'autre. Nous avons donc choisi de traiter les terminaisons mixtes comme si elles étaient muettes et donc *insécables* : nous avons préféré interdire la coupure possible *le prési-dent* plutôt que d'accepter la césure incorrecte *ils prési-dent* ; ce mot sera dans tous les cas coupé par TEX en `pré-[si/dent`. L'auteur a toujours la possibilité d'infléchir *localement* ce choix en codant dans le texte `pré\-si\-dent` au lieu de `président`.

La base BDLEX nous a été d'une grande utilité pour faire le tri des terminaisons : grâce aux utilitaires standard d'UNIX, nous avons extrait *tous* les mots se terminant par `ent`, puis nous les avons triés par ordre alphabétique mais *de droite à gauche* afin de dégager des règles de classement en motifs sonores ou muets. Une lecture attentive du fichier produit nous a permis d'établir la liste de motifs à ajouter. Le fichier BDLEX nous a semblé assez complet pour permettre une classification satisfaisante.

Le tableau 3 présente le bilan de toutes les terminaisons en `ent.` présentes dans le fichier BDLEX.

L'ajout des 182 motifs proposés permet de corriger 18 331 coupures incorrectes parmi les mots du fichier BDLEX.

Prenons, pour expliquer notre démarche, l'exemple des terminaisons en `-cent` : elles sont en général muettes (115 contre 47, cf. tableau 3), on ajoute donc le motif de base `2cent.` pour interdire en général la coupure devant le `c`. Mais lorsque la syllabe `-cent` est précédée de `-ja-` (*adjacent, subjacent*) ou de `-é-` (*décent, indécent, récent*) elle est sonore et on corrige en ajoutant `ja3cent.` et `\'e3cent.` pour autoriser la coupure devant le `c` dans ce cas. Les terminaisons en `-escent` (*évanescent, incandescent, sénescent...*) sont également sonores sauf *acquiescent*, on ajoute donc `es3cent.` et `acquies4cent.`

On comprend sur cet exemple que l'ordre pertinent pour l'analyse des *terminaisons* de mots est l'ordre alphabétique *de droite à gauche*. C'est dans cet ordre qu'on été rangés les motifs corrigeant chaque motif de base

TABLE 3 - *Fréquences des terminaisons en* `-ent` *dans* BDLEX.

	muettes	sonores	motifs ajoutés		muettes	sonores	motifs ajoutés
-bent	55	0	1	-kent	1	0	1
-blent	35	0	1	-lent	647	22	18
-brent	29	0	1	-ment	164	2247	58
-cent	115	47	12	-nent	698	14	8
-chent	179	0	1	-pent	140	3	4
-ckent	2	0	1	-phent	5	0	1
-clent	16	0	1	-plent	14	0	1
-crent	9	0	1	-prent	3	0	1
-dent	313	16	12	-quent	183	6	4
-dlent	4	0	1	-rent	6767	21	7
-drent	12	0	1	-sent	7307	4	4
-fent	44	0	1	-shent	1	0	1
-flent	27	0	1	-tent	877	16	9
-frent	16	0	1	-trent	70	0	1
-gent	203	18	14	-vent	162	2	3
-glent	18	0	1	-vrent	21	0	1
-gnent	86	0	1	-went	1	0	1
-grent	12	0	1	-xent	19	0	1
-guent	70	1	2	-zent	5	1	2
-jent	1	0	1	Total	18 331	2 418	182

dans les fichiers `f7hyph.tex` et `f8hyph.tex`: ainsi `2cent.` est placé à la lettre `c`, après les deux motifs `4ce.` et `4ces.` qui ont la même fonction que lui, et les exceptions (`ja3cent` ... `is3cent.`, `immis4cent.`) sont classées par ordre alphabétique *de droite à gauche* à la suite, tous les motifs du groupe ayant été alignés sur le point final afin d'améliorer la lisibilité.

8. Conclusion provisoire et perspectives

Le présent travail constitue une remise en ordre des fichiers de césures françaises de J. DÉSARMÉNIEN. Nous avons essayé de faire le tri des améliorations apportées et des détériorations survenues. Nous ne prétendons pas que les fichiers de motifs (`f7hyph.tex`, `f8hyph.tex`) soient parfaits, il y a encore beaucoup de travail à faire sur ce sujet: si les motifs phonétiques nous semblent bien au point, les motifs étymologiques mériteraient eux d'être testés plus avant, mais aucun des trois auteurs n'a eu le temps, à ce jour, de se mettre à ce travail. Toutes les propositions de collaboration dans ce domaine sont évidemment les bienvenues.

Les nouveaux fichiers de césure se trouvent dans la distribution des *fichiers de style french*, ils seront inclus dans les distributions GUTenberg. Des versions bien antérieures qui sont encore actuellement diffusées sur les réseaux, sont loin d'être satisfaisantes. Nous comptons faire en sorte que toutes les versions altérées soit retirées des archives CTAN et remplacées par les seuls fichiers :

- `frhyph.tex` (V1.06 du 20/4/94) : code TEX effectuant le chargement en mémoire des motifs de césure sous leur forme 7 ou 8-bits selon le cas le plus approprié ;
- `f7hyph.tex` (V2.0 du 20/5/94) motifs de césure 7-bits ;
- `f8hyph.tex` (V2.0 du 20/5/94) motifs de césure 8-bits (codage DC).

Toutes les suggestions et améliorations peuvent nous être transmises à l'adresse électronique suivante : `cesure-l@ens.fr` ou directement aux auteurs.

Remerciements Nous tenons à remercier les diverses personnes qui ont bien voulu critiquer une version provisoire de cet article, et en particulier Jacques ANDRÉ.

Références bibliographiques

[1] Association suisse des compositeurs à la machine : *Guide du typographe romand*, 5e édition, 1994

[2] M. ARRIVE, F. GADET et M. GALMICHE, *La grammaire d'aujourd'hui : guide alphabétique de linguistique française*, Flammarion, Paris, 1986.

[3] H. BONNARD : *Code du français courant*, Magnard, 1981.

[4] J. DÉSARMÉNIEN : « How to run TEX in a French environment: Hyphenation, Fonts, Typography », *TUGboat*, vol. 5 no. 2, 1984, pp. 91–102.

[5] J. DÉSARMÉNIEN : « La division par ordinateur des mots français : application à TEX», Techniques et Sciences Informatiques, vol. 5 no. 4, 1986, pp. 251–265.

[6] B. DUPRIEZ, *Gradus : les procédés littéraires (dictionnaire)*, Union générale d'Éditions, Paris, "10-18", 1984.

[7] Fédération CGC de la communication : *Code typographique, Choix des règles à l'usage des auteurs et professionnels du livre*, 16e édition, 1989.

[8] M.-J. Ferguson : « Fontes latines européennes et TEX 3.0 », *Cahiers* GUT*enberg* N° 7, 1990.

[9] D. Flipo and L. Siebenmann : « Hyphenation in presence of accents and diacritics. An easy and low-cost solution » *Cahiers* GUT*enberg* N° 15 (EuroTEX 92), 1992.

[10] A. Frey : *Manuel nouveau de typographie*, réed. Léonce Laget, 1979, *épuisé*.

[11] B. Gaulle : *Manuel d'utilisation du style* `french`, document électronique fourni avec la distribution logicielle, 6e édition, 1994.

[12] J. Girodet : *Dictionnaire du bon français*, Bordas, 1981.

[13] M. Goossens, F. Mittelbach and A. Samarin : *The* LATEX *Companion*, Addison-Wesley, New York, 1994.

[14] C. Gouriou : *Mémento typographique*, Hachette, 1973.

[15] M. Grévisse : *Le bon usage*, 13e édition Duculot, 1994.

[16] Imprimerie nationale : *Lexique des règles typographiques en usage à l'Imprimerie nationale*, 3e édition, 1990.

[17] D. E. Knuth : *The TEXbook*, Addison Wesley, 1991.

[18] F. M. Liang : *Word Hy-phen-a-tion by Comput-er*, Ph. D. Thesis, Departement of Computer Science, Stanford University, report N° STANCS-83-977, 1983.

[19] R. Thimonnier : *Code orthographique et grammatical*, Marabout, 1978, *épuisé*.

Annexe Liste des motifs de césure français

Les lettres accentuées des motifs de césure sont exprimées en notation TEX de base (ASCII c'est-à-dire 7-bits). Chaque macro-instruction terminale est suivie du chiffre 0 pour clore effectivement le motif. Rappelons que les chiffres impairs à l'intéreur des motifs autorisent la coupure des mots calqués sur ce motif (à condition qu'aucun autre motif ne le contredise, voir à ce sujet le paragraphe 5.1). Les motifs étymologiques sont ici précédés du caractère |. Le classement est d'ordre alphabétique, de gauche à droite pour l'essentiel, mais aussi de droite à gauche pour les exceptions (on les remarquera par leur alignement à droite).

```
\patterns{
  2'2
  .a4
  'a4
  .\^a4
  '\^a4
| .ab3r\'ea
| 'ab3r\'ea
| a1\`e2dre
| .ae3s4ch
| 'ae3s4ch
| 1alcool
| a2l1algi
| .amino1a2c
| 'amino1a2c
| .ana3s4tr
| 'ana3s4tr
| 1a2nesth\'esi
| .anti1a2
| 'anti1a2
| .anti1e2
| 'anti1e2
| .anti1\'e2
| .anti2enne
| 'anti2enne
| 'anti1\'e2
| .anti1s2
| 'anti1s2
| .apo2s3ta
| 'apo2s3ta
| apo2s3tr
| archi1\'e2pis
| .as2ta
| 'as2ta
| a2s3tro
  1ba
  1b\^a
| .bai2se3main
  1be
  1b\'e
  1b\`e
  1b\^e
  4be.
  4bes.
  2bent.
  1bi
  1b\^\i0
| .bi1a2c
| .bi1a2t
| .bi1au
| .bio1a2
| .bi2s1a2
| .bi1u2
  1b2l
  4ble.
  4bles.
  2blent.
  1bo
  1b\^o
  1b2r
  4bre.
  4bres.
  2brent.
  1bu
  1b\^u
  1by
  1\c c
  1ca
  1c\^a
  ca3ou3t2
  1ce
  1c\'e
  1c\`e
  1c\^e
  4ce.
  4ces.
          2cent.
        ja3cent.
        ac3cent.
       \'e3cent.
    munifi3cent.
    r\'eti3cent.
  privatdo3cent.
      inno3cent.
         es3cent.
   acquies4cent.
        is3cent.
     immis4cent.
  .ch4
  1c2h
  4ch.
  2chb
  4che.
  4ches.
  2chent.
| .ch\`e2vre3feuille
  2chg
  ch2l
  4chle.
  4chles.
| chlo2r3a2c
| chlo2r3\'e2t
  2chm
  2chn
  2chp
  ch2r
  4chre.
  4chres.
  2chs
  2cht
  2chw
  1ci
  1c\^\i0
| .ci2s1alp
  1c2k
  4ck.
  2ckb
  4cke.
  4ckes.
  2ckent.
  2ckf
  2ckg
  2ck3h
  2ckp
  2cks
  2ckt
  1c2l
  4cle.
  4cles.
  2clent.
  1co
  1c\^o
| co1acc
| co1acq
| co1a2d
| co1ap
| co1ar
| co1assoc
| co1assur
| co1au
| co1ax
  1c\oe0
| co1\'e2
| co1ef
| co1en
| co1ex
| .con4
| .cons4
| .contre1s2c
| .contre3ma\^\i0tre
| co2nurb
| .co1o2
| .co2o3lie
  1c2r
  4cre.
  4cres.
  2crent.
  1cu
  1c\^u
  1cy
  .cul4
  1d'
  1da
  1d\^a
| .dacryo1a2
  d1d2h
  1de
  1d\'e
  1d\`e
  1d\^e
  4de.
  4des.
        2dent.
  d\'eca3dent.
     \'e3dent.
     cci3dent.
    inci3dent.
   confi3dent.
     tri3dent.
   dissi3dent.
   chien3dent.
     .ar3dent.
    impu3dent.
     pru3dent.
```

```
| .d\'e1a2
| .d\'e1io
| .d\'e1o2
| .d\'e2s
| .d\'e3s2a3cr
| .d\'es2a3m
| .d\'e3s2a3tell
| .d\'e3s2astr
| .d\'e3s2c
| .d\'e2s1\'e2
| .d\'e3s2\'e3gr
| .d\'e3s2ensib
| .d\'e3s2ert
| .d\'e3s2exu
| .d\'e2s1i2
| .d\'e3s2i3d
| .d\'e3s2i3gn
| .d\'e3s2i3li
| .d\'e3s2i3nen
| .d\'e3s2invo
| .d\'e3s2i3r
| .d\'e3s2ist
| .d\'e3s2o3d\'e
| .d\'e2s1\oe0
| .d\'e3s2o3l
| .d\'e3s2o3pil
| .d\'e3s2orm
| .d\'e3s2orp
| .d\'e3s2oufr
| .d\'e3s2p
| .d\'e3s2t
| .d\'e2s1u2n
| 3d2hal
| 3d2houd
  1di
  1d\^\i0
| di2s3cop
| .di1a2c\'e
| .di1a2cid
| dia2g3n
| .di1ald
| .di1a2mi
| .di1a2tom
| .di1e2n
| .di2s3h
  2dlent.
  1do
  1d\^o
  1d2r
  4dre.
  4dres.
  2drent.
  d1s2
  1du
  1d\^u
  1dy
| .dy2s3
| .dy2s1a2
| .dy2s1i2
| .dy2s1o2
| .dy2s1u2
  .e4
  'e4
  .\^e4
  '\^e4
  .\'e4
  '\'e4
  .\`e4
  '\`e4
| 1\'e2drie
| 1\'e2drique
| 1\'e2lectr
| 1\'e2l\'ement
| .en1a2
| 'en1a2
| 1\'e2nerg
| e2n1i2vr
| .en1o2
| 'en1o2
| \'epi2s3cop
| \'epi3s4cope
| e2s3cop
| .eu2r1a2
| 'eu2r1a2
| eu1s2tat
| extra1
| extra2c
| extra2i
  1fa
  1f\^a
  1fe
  1f\'e
  1f\`e
  1f\^e
  4fe.
  4fes.
  2fent.
  1fi
  1f\^\i0
  1f2l
  4fle.
  4fles.
  2flent.
  1fo
  1f\^o
  1f2r
  4fre.
  4fres.
  2frent.
  f1s2
  1fu
  1f\^u
  1fy
  1ga
  1g\^a
  1ge
  1g\'e
  1g\`e
  1g\^e
  4ge.
  4ges.
         2gent.
     r\'e3gent.
    entre3gent.
     indi3gent.
     dili3gent.
  intelli3gent.
    indul3gent.
      tan3gent.
      rin3gent.
   contin3gent.
      .ar3gent.
      'ar3gent.
      ser3gent.
      ter3gent.
  r\'esur3gent.
  1g2ha
  1g2he
  1g2hi
  1g2ho
  1g2hy
  1gi
  1g\^\i0
  1g2l
  4gle.
  4gles.
  2glent.
      1g2n
  .sta2g3n
    wa2g3n
  4gne.
  4gnes.
  2gnent.
  1go
  1g\^o
  1g2r
  4gre.
  4gres.
  2grent.
  1gu
  1g\^u
  g1s2
  4gue.
  4gues.
     2guent.
  .on3guent.
  'on3guent.
  1gy
  1ha
  1h\^a
  1he
  1h\'e
  1h\`e
  1h\^e
| h\'emi1\'e
| h\'emo1p2t
  4he.
  4hes.
  1hi
  1h\^\i0
  1ho
  1h\^o
  1hu
  1h\^u
  1hy
| hypera2
| hypere2
| hyper\'e2
| hyperi2
| hypero2
| hypers2
| hype4r1
| hyperu2
| hypo1a2
| hypo1e2
| hypo1\'e2
| hypo1i2
| hypo1o2
| hypo1s2
| hypo1u2
  .i4
  'i4
  .\^\i4
  '\^\i4
| i1algi
| i1arthr
| i1\`e2dre
       il2l
     cil3l
    rcil4l
    ucil4l
   vacil4l
     gil3l
     hil3l
     lil3l
       l3lion
     mil3l
     mil4let
  \'emil4l
   semil4l
    rmil4l
   armil5l
   capil3l
   papil3la
   papil3le
   papil3li
   papil3lom
   pupil3l
   piril3l
   thril3l
   cyril3l
   ibril3l
   pusil3l
   .stil3l
  distil3l
  instil3l
  fritil3l
  boutil3l
   vanil3lin
   vanil3lis
     vil3l
    avil4l
  chevil4l
   uevil4l
    uvil4l
     xil3l
| 1informat
| .in1a2
| 'in1a2
| .in2a3nit
| 'in2a3nit
| .in2augur
| 'in2augur
| .in1e2
| 'in1e2
| .in1\'e2
```

```
| 'in1\'e2
| .in2effab
| 'in2effab
| .in2\'e3lucta
| 'in2\'e3lucta
| .in2\'e3narra
| 'in2\'e3narra
| .in2ept
| 'in2ept
| .in2er
| 'in2er
| .in2exora
| 'in2exora
| .in1i2
| 'in1i2
| .in2i3miti
| 'in2i3miti
| .in2i3q
| 'in2i3q
| .in2i3t
| 'in2i3t
| .in1o2
| 'in1o2
| .in2o3cul
| 'in2o3cul
| .in2ond
| 'in2ond
| .in1s2tab
| 'in1s2tab
| 'inte4r3
| .intera2
| 'intera2
| .intere2
| 'intere2
| .inter\'e2
| 'inter\'e2
| .interi2
| 'interi2
| .intero2
| 'intero2
| .inte4r3
| .interu2
| 'interu2
| .inters2
| 'inters2
| .in1u2
| 'in1u2
| .in2uit
| 'in2uit
| .in2u3l
| 'in2u3l
| io1a2ct
| i1oxy
| i1s2tat
  1j
  2jk
  4je.
  4jes.
  2jent.
  1ka
  1k\^a
  1ke
  1k\'e
  1k\`e
  1k\^e
  4ke.
  4kes.
  2kent.
  1k2h
  4kh.
  .kh4
  1ki
  1k\^\i0
  1ko
  1k\^o
  1k2r
  1ku
  1k\^u
  1ky
  1la
  1l\^a
  1l\`a
| la2w3re
  1le
  1l\'e
  1l\`e
  1l\^e
  4le.
  4les.
          2lent.
       .ta3lent.
       iva3lent.
  \'equiva4lent.
    monova3lent.
    polyva3lent.
        re3lent.
       .do3lent.
      indo3lent.
      inso3lent.
     turbu3lent.
     succu3lent.
    f\'ecu3lent.
     trucu3lent.
       opu3lent.
     corpu3lent.
        ru3lent.
     sporu4lent.
  1li
  1l\^\i0
  1lo
  1l\^o
  l1s2t
  1lu
  1l\^u
  1ly
  1ma
  1m\^a
| .ma2c3k
| .macro1s2c
| .ma2l1a2dres
| .ma2l1a2dro
| .ma2l1ais\'e
| .ma2l1ap
| .ma2l1a2v
| .ma2l1en
| .ma2l1int
| .ma2l1oc
| .ma2l1o2d
| .ma2r1x
  1me
  1m\'e
  1m\`e
  1m\^e
| .m\'e2g1oh
| .m\'e2sa
| .m\'e3san
| .m\'e2s1es
| .m\'e2s1i
| .m\'e2s1u2s
| .m\'eta1s2ta
  4me.
  4mes.
        \^a2ment.
         da2ment.
         fa2ment.
     amalga2ment.
        cla2ment.
         ra2ment.
  temp\'era3ment.
         ta2ment.
      testa3ment.
        qua2ment.
        \`e2ment.
     car\^e2ment.
   diaphrag2ment.
       ryth2ment.
         ai2ment.
        rai3ment.
      ab\^\i2ment.
       \'eci2ment.
        vidi2ment.
       subli2ment.
       \'eli2ment.
        reli2ment.
          mi2ment.
         ani2ment.
        veni2ment.
          ri2ment.
    d\'etri3ment.
      nutri3ment.
        inti2ment.
        esti2ment.
           l2ment.
        flam2ment.
        gram2ment.
        .gem2ment.
          om2ment.
        .com3ment.
         \^o2ment.
      slalo2ment.
        chro2ment.
          to2ment.
          ar2ment.
        .sar3ment.
          er2ment.
    antifer3ment.
        .ser3ment.
         fir2ment.
          or2ment.
          as2ment.
          au2ment.
      \'ecu2ment.
           fu2ment.
           hu2ment.
      fichu3ment.
        llu2ment.
        plu2ment.
        bou2ment.
        bru2ment.
         su2ment.
         tu2ment.
  1mi
  1m\^\i0
| .milli1am
| 1m2n\'emo
| 1m2n\`es
| 1m2n\'esi
  1mo
  1m\^o
  1m\oe0
| .mono1a2
| .mono1e2
| .mono1\'e2
| .mono1i2
| .mono1\"\i2d\'e
| .mono1o2
| .mono1u2
| .mono1s2
| mon2t3r\'eal
  m1s2
  1mu
  1m\^u
  1my
| moye2n1\^a2g
  1na
  1n\^a
  1ne
  1n\'e
  1n\`e
  1n\^e
  4ne.
  4nes.
             2nent.
        r\'ema3nent.
          imma3nent.
         perma3nent.
        .\'emi3nent.
  pr\'e\'emi3nent.
    pro\'emi3nent.
    sur\'emi3nent.
        immi3nent.
       conti3nent.
       perti3nent.
       absti3nent.
  1ni
  1n\^\i0
  1no
  1n\^o
  1n\oe0
| .no2n1obs
  1nu
  1n\^u
| n3s2at.
| n3s2ats.
  n1x
  1ny
```

```
  .o4
  'o4
  '\^o4
  .\^o4
| o2b3long
| 1octet
| o1d2l
| o1\'e2dre
| o1ioni
| ombud2s3
| omni1s2
| o1s2tas
| o1s2tat
| o1s2t\'ero
| o1s2tim
| o1s2tom
| o1s2trad
| o1s2tratu
| o1s2triction
| .oua1ou
| 'oua1ou
| .ovi1s2c
| 'ovi1s2c
| oxy1a2
  1pa
  1p\^a
| pal\'eo1\'e2
| .pa2n1a2f
| .pa2n1a2m\'e
| .pa2n1a2ra
| .pa2n1is
| .pa2n1o2ph
| .pa2n1opt
| .pa2r1a2che
| .pa2r1a2ch\'e
| .para1s2
| .pa2r3h\'e
  1pe
  1p\'e
  1p\'e
  1p\^e
  4pe.
  4pes.
     2pent.
   re3pent.
  .ar3pent.
  'ar3pent.
  ser3pent.
| .pen2ta
| per3h
| p\'e2nul
| .pe4r
| .per1a2
| .per1e2
| .per1\'e2
| .per1i2
| .per1o2
| .per1u2
| p\'e1r2\'e2q
| .p\'eri1os
| .p\'eri1s2
| .p\'eri2s3s
| .p\'eri2s3ta
| .p\'eri1u2
  1p2h
  .ph4
  4ph.
| .phalan3s2t
  4phe.
  4phes.
  2phent.
  ph2l
  4phle.
  4phles.
  2phn
| photo1s2
  ph2r
  4phre.
  4phres.
  2phs
  2pht
| 3ph2tal\'e
| 3ph2tis
  1pi
  1p\^\i0
  1p2l
  4ple.
  4ples.
  2plent.
| .pluri1a
| 1p2n\'e
| 1p2neu
  1po
  1p\^o
| po1astre
| poly1a2
| poly1e2
| poly1\'e2
| poly1\'e2
| poly1i2
| poly1o2
| poly1s2
| poly1u2
| .pon2tet
| .pos2t3h
| .pos2t1in
| .pos2t1o2
| .pos2t3r
| .post1s2
  1p2r
  4pre.
  4pres.
  2prent.
| .pr\'e1a2
| .pr\'e2a3la
| .pr\'e2au
| .pr\'e1\'e2
| .pr\'e1e2
| .pr\'e1i2
| .pr\'e1o2
| .pr\'e1u2
| .pr\'e1s2
| .pro1\'e2
| .pro1s2c\'e
| pro2s3tat
| .prou3d2h
| 1p2sych
| .psycho1a2n
| 1p2t\'er
| 1p2t\'er
  1pu
| .pud1d2l
  1p\^u
  1py
  1q
  4que.
  4ques.
          2quent.
       \'e3quent.
     \'elo3quent.
  grandilo3quent.
  1ra
  1r\^a
| radio1a2
  1re
  1r\'e
  1r\'e
  1r\^e
| .r\'e1a2
| .r\'e2a3le
| .r\'e2a3lis
| .r\'e2a3lit
| .r\'e2aux
| .r\'e1\'e2
| .r\'e1e2
| .r\'e2el
| .r\'e2er
| .r\'e2\'er
| .r\'e1i2
| .r\'e2i3fi
| .r\'e1o2
| .re1s2
| .re2s3cap
| .re2s3cisi
| .re2s3ciso
| .re2s3cou
| .re2s3cri
| .re2s3pect
| .re2s3pir
| .re2s3plend
| .re2s3pons
| .re2s3quil
| .re2s3s
| .re2s3t
| .re3s4tab
| .re3s4tag
| .re3s4tand
| .re3s4tat
| .re3s4t\'en
| .re3s4t\'er
| .re3s4tim
| .re3s4tip
| .re3s4toc
| .re3s4top
| .re3s4tr
| .re4s5trein
| .re4s5trict
| .re4s5trin
| .re3s4tu
| .re3s4ty
| .r\'eu2
| .r\'e2uss
| .r\'etro1a2
  4re.
  4res.
         2rent.
      .pa3rent.
    appa3rent.
  transpa3rent.
      \'e3rent.
      tor3rent.
      cur3rent.
  1r2h
  4rhe.
  4rhes.
| 2r3heur
| 2r3hydr
  1ri
  1r\^\i0
  1ro
  1r\^o
  1ru
  1r\^u
  1ry
  1sa
  1s\^a
  .sch4
| 1s2caph
| 1s2cl\'er
| 1s2cop
   1s2ch
  e2s3ch
  i2s3ch\'e
  i2s3chia
  i2s3chio
  4sch.
  4sche.
  4sches.
  2schs
  1se
  1s\'e
  1s\'e
  1s\^e
| sesqui1a2
  4se.
  4ses.
       2sent.
    ab3sent.
  pr\'e3sent.
   .res3sent.
  .seu2le
  .sh4
  1s2h
  4sh.
  4she.
  4shes.
  2shent.
  2shm
| 2s3hom
  2shr
  2shs
  1si
  1s\^\i0
| 1s2lav
| 1s2lov
  1so
```

```
  1s\^o                | .su2r1inf            4thre.                    2vent.
  1s\oe0               | .su2r1int            4thres.                conni3vent.
| 1s2patia             | .su2r1of             2ths                    .sou3vent.
| 1s2perm              | .su2r1ox             1ti                    1vi
| 1s2por                 1sy                  1t\^\i0                1v\^\i0
| 1s2ph\‘er              1ta                  1to                    1vo
| 1s2ph\’er              1t\^a                1t\^o                  1v\^o
| 1s2piel                1t\‘a                1t2r                 | vol2t1amp
| 1s2piros             | tachy1a2           | tran2s1a2              1v2r
| 1s2tandard           | tchin3t2           | tran3s2act             4vre.
| 1s2tein                1te                | tran3s2ats             4vres.
| st\’er\’eo1s2          1t\’e              | tran2s3h               2vrent.
| 1s2tigm                1t\‘e              | tran2s1o2              1vu
| 1s2tock                1t\^e              | tran2s3p               1v\^u
| 1s2tomos             | t\’el\’e1e2        | tran2s1u2              1vy
| 1s2troph             | t\’el\’e1i2            4tre.                  1wa
| 1s2tructu            | t\’el\’e1o2b           4tres.                 1we
| 1s2tyle              | t\’el\’e1o2p           2trent.                4we.
  1su                  | t\’el\’e1s2        | .tri1a2c               4wes.
  1s\^u                  4te.                 | .tri1a2n               2went.
| .su2b1a2               4tes.                | .tri1a2t               1wi
| .su3b2alt                      2tent.       | .tri1o2n               1wo
| .su2b1\’e2                  .la3tent.       | t1t2l                  1wu
| .su3b2\’e3r                 .pa3tent.         1tu                    1w2r
| .su2b1in                 comp\’e3tent.        1t\^u                  2xent.
| .su2b3limin                \’eni3tent.       tung2s3                .y4
| .su2b3lin                m\’econ3tent.       1ty                    ’y4
| .su2b3lu                  omnipo3tent.       .u4                  | y1asth
| sub1s2                 ventripo3tent.         ’u4                  | y1s2tom
| .su2b1ur               \’equipo3tent.         .\^u4                | y1algi
| supero2                    impo3tent.         ’\^u4                  1za
| supe4r1                     mit3tent.       | uni1o2v                1ze
| supers2                .th4                 | uni1a2x                1z\’e
| .su2r1a2               1t2h                 | u2s3tr                 1z\‘e
| su3r2ah                4th.                   1va                    4ze.
| .su3r2a3t              4the.                  1v\^a                  4zes.
| .su2r1e2               4thes.                 1ve                           2zent.
| .su3r2eau            | thermo1s2              1v\’e                  privatdo3zent.
| .su3r2ell            | 2t3heur                1v\‘e                  1zi
| .su3r2et               2thl                   1v\^e                  1zo
| .su2r1\’e2             2thm                 | v\’elo1s2ki            1zu
| .su2r3h                2thn                   4ve.                   1zy
| .su2r1i2m              th2r                   4ves.                }
```