

Francais-GUTenberg v1.0
Manuel de référence

Christophe Pythoud

7 juin 1999

Introduction

De quoi s'agit-il?

Francais-GUTenberg est un dictionnaire spécialement composé pour le correcteur ISPELL¹. Il a été développé à l'instigation de l'association GUTenberg, à partir de mon travail de diplôme qui portait précisément sur ISPELL et sur un autre dictionnaire pour le français, Francais-IREQ. Reportez-vous au chapitre 5 pour en savoir plus.

Copyright

Ce programme est distribué sous la licence GPL-2 (voir les fichiers `GPL.fr` et `GPL.en`) par Christophe Pythoud (`Christophe.Pythoud@ling.unil.ch`) et l'Association GUTenberg (<http://www.gutenberg.eu.org/>).

Comment lire cette documentation

La documentation de Francais-GUTenberg se veut la plus complète possible. Néanmoins, si tout se passe bien, vous ne devriez pas avoir à lire plus de quelques lignes par chapitre. Si Francais-GUTenberg a déjà été installé par l'administrateur de votre système, tout ce qui vous intéresse est le chapitre 4.

La documentation est divisée en trois parties :

- la première (chapitres 1 à 3) concerne l'installation de Francais-GUTenberg ; lisez-la si vous avez la responsabilité d'installer le correcteur ; vous y apprendrez comment (re)compiler ISPELL si nécessaire (chapitre 1), comment installer Francais-GUTenberg (chapitre 2) et comment faire coexister le correcteur avec (X)Emacs ;
- la seconde (chapitre 4) concerne l'utilisation de Francais-GUTenberg ;

1. <http://ficus-www.cs.ucla.edu/ficus-members/geoff/ispell.html>

- la troisième présente des informations complémentaires sur les principes qui ont présidé à l’élaboration de Français-GUTenberg (chapitre 5), sur le fichier d’affixes (chapitre 6) et sur l’ajout de dictionnaires au système (chapitre 7).

Les chapitres 5 et 7 sont, pour l’essentiel, tirés d’un article publié dans les Cahiers GUTenberg (Pythoud 1998). Reportez-vous aux pages du présent document plutôt qu’à cet article, car de nombreux éléments ont été revus depuis.

Tout commentaire sur ces pages sera le bienvenu.

Comment signaler anomalies et problèmes

Bien que Français-GUTenberg se trouve maintenant dans la phase finale de son développement, il est fort possible que des anomalies de fonctionnement subsistent. Les problèmes que vous pourriez rencontrer seront de deux types :

- un mot du dictionnaire est mal orthographié ;
- un mot que vous utilisez est absent du dictionnaire.

Dans le premier cas, il faut absolument me signaler le problème ! Bien que des efforts importants aient été consacrés à vérifier la validité des entrées du dictionnaire, certaines erreurs ont pu subsister.

Dans le second cas, il faut bien distinguer deux états de fait :

- le mot appartient au vocabulaire courant et il convient de me le signaler ;
- le mot appartient au vocabulaire spécialisé d’une profession ou d’un domaine scientifique, auquel cas il serait préférable d’envisager la création d’un dictionnaire spécialisé (cf. chapitre 7).

Vous pouvez m’envoyer un courrier privé, décrivant le problème que vous rencontrez, à l’adresse suivante : `Christophe.Pythoud@ling.unil.ch`.

Si vous souhaitez participer activement à l’élaboration de Français-GUTenberg, vous pouvez vous abonner à la liste de discussions `dico-francais-gutenberg` en envoyant un message contenant :

```
sub dico-francais-gutenberg <Prénom> <Nom>  
à l’adresse sympa@cru.fr.
```

Hommage à Martin Boyer

Je tiens à saluer l’immense travail accompli par Martin Boyer sur Français-IREQ, sans lequel je ne me serais peut-être jamais intéressé au problème de la correction de l’orthographe française à l’aide de ISPELL, et sans lequel Français-GUTenberg n’aurait probablement pas vu le jour.

Sommaire

Introduction	iii
I Installation et configuration	1
1 Installation/configuration de ISPELL	3
1.1 Que faut-il faire?	3
1.1.1 Test 1: ISPELL est-il présent?	3
1.1.2 Test 2: l'utilitaire buildhash est-il présent?	4
1.1.3 Test 3: utilisez-vous la version 3.1.20 de ISPELL?	4
1.1.4 Test 4: ISPELL est-il convenablement configuré?	4
1.2 Télécharger les sources de ISPELL	5
1.2.1 Depuis la Belgique...	5
1.2.2 Depuis le Canada...	5
1.2.3 Depuis la France...	5
1.2.4 Depuis la Suisse...	6
1.3 (Re)compiler ISPELL	6
1.3.1 Décompresser l'archive	6
1.3.2 Le fichier <code>local.h</code>	6
1.3.3 Et pour disposer d'un dictionnaire anglais...	7
1.4 Compilation et installation	8
1.4.1 En cas d'erreur...	8
1.4.2 Installation manuelle des fichiers	8
1.4.3 Installation manuelle de dictionnaires	9
2 Installation de Francais-GUTenberg	11
2.1 Dernière version	11
2.2 L'utilitaire makehash	11
2.2.1 La bonne version de ISPELL?	12
2.2.2 Autres erreurs possibles	12
2.2.3 Continuer ou interrompre les opérations?	13

2.2.4	Création du dictionnaire	13
2.2.5	Caractères accentués dans les fichiers T _E X	14
2.3	Installation manuelle	15
2.3.1	Choix des listes de mots	15
2.3.2	Utilisation de buildhash	15
2.3.3	Installation des fichiers	16
3	Configuration de Emacs et XEmacs	17
3.1	Généralités	17
3.2	Emacs	17
3.3	XEmacs	18
3.4	FlySpell	18
II	Utilisation	19
4	Utilisation de Francais-GUTenberg	21
4.1	Appeler ISPELL depuis la ligne de commande	21
4.2	Commandes de ISPELL	21
4.3	Utilisation depuis (X)Emacs	22
4.3.1	Avec Emacs...	22
4.3.2	Avec XEmacs...	23
4.3.3	Avec FlySpell...	23
4.4	Fonctionnement du système	24
4.4.1	Qu'est-ce qu'un correcteur orthographique?	24
4.4.2	Le format du dictionnaire	24
4.4.3	Vérification	25
4.4.4	Suggestion	25
4.5	Éléments pour résoudre quelques problèmes	26
4.5.1	Absence de suggestions	26
4.5.2	ISPELL vous propose de valider une affixation	26
4.5.3	Élision du pronom <i>que</i>	27
4.5.4	Élision de <i>le</i> et <i>de</i> devant un mot avec une capitale	27
4.5.5	Le caractère <i>æ</i>	27
III	Fonctionnement et création de dictionnaires	29
5	Élaboration du système	31
5.1	Bref historique de ce projet	31
5.2	Choix des lemmes	31

5.3	Le problème de l'élision	35
5.3.1	Vérification et élision	35
5.3.2	Suggestion et élision	35
5.4	Points divers	36
5.4.1	Écrire les nombres en toutes lettres	36
5.4.2	Le vocabulaire spécialisé	36
6	Le fichier d'affixes	39
6.1	Options de base	39
6.1.1	La commande <code>allaffixes</code>	39
6.1.2	Les caractères composant les mots	39
6.1.3	Les caractères <code>TEX</code>	41
6.2	Description des drapeaux	42
6.2.1	Les préfixes	42
6.2.2	Les suffixes (noms et adjectifs)	48
6.2.3	Les suffixes (verbes)	50
7	Ajout de nouveaux dictionnaires	63
7.1	Ajouter des noms et des adjectifs	63
7.1.1	Le cas trivial	63
7.1.2	Féminin et masculin	63
7.1.3	Le problème de l'apostrophe	64
7.1.4	Utilisation des autres préfixes	65
7.2	Ajouter des verbes	66
7.2.1	Éléments de base	66
7.2.2	Infinitif, participe passé et adjectif verbal formé sur le parti- cipe présent	67
7.2.3	Le problème de l'élision	69
7.2.4	Premières personnes du conditionnel et des subjonctifs présent et imparfait avec élision de <i>je</i>	73
7.2.5	Élision et emplois réflexifs	73
7.2.6	Élision et verbes <i>essentiellement pronominaux</i>	73
7.2.7	Les verbes du troisième groupe	75
7.3	Entrer des abréviations	77
7.4	Vérification de la validité des entrées	77
7.5	Dernières étapes	77

Répertoire des tables

7.1	Les pronoms du français et l'élision	69
7.2	Verbes et élision : quelques exemples	72
7.3	Drapeaux et conjugaisons du <i>Bescherelle</i>	75
7.3	Drapeaux et conjugaisons du <i>Bescherelle</i>	76
7.3	Drapeaux et conjugaisons du <i>Bescherelle</i>	77

Première partie

Installation et configuration

Chapitre 1

Installation/configuration de ISPELL

Ce chapitre présente les paramètres à prendre en compte pour s'assurer que l'on dispose d'une configuration de ISPELL susceptible d'être utilisée avec Francais-GUTenberg.

1.1 Que faut-il faire ?

Cette section présente une série de tests, vous permettant de savoir quelles tâches vous allez devoir accomplir en ce qui concerne ISPELL.

Si vous avez l'impression que votre version de ISPELL va de toute façon remplir les critères, vous pouvez essayer de lancer immédiatement l'utilitaire `makehash` (cf. 2.2) et voir ce qui se passe.

1.1.1 Test 1 : ISPELL est-il présent ?

Le plus simple pour s'assurer que ISPELL est disponible est de taper la commande :

```
ispell -v
```

Si vous n'obtenez pas de message d'erreur indiquant que le fichier est introuvable, vous êtes sur la bonne voie. Sinon ISPELL est absent de votre système et vous devez vous le procurer (cf. 1.2) ou vous adresser à l'administrateur de votre système. (Il est également possible que ISPELL soit installé, mais que vous n'y ayez pas accès : vérifiez la variable `PATH` de votre environnement ou adressez-vous à votre administrateur.)

1.1.2 Test 2 : l'utilitaire buildhash est-il présent ?

Cet utilitaire est indispensable pour compiler un nouveau dictionnaire. Essayez la commande :

```
buildhash
```

Si vous obtenez un message d'erreur indiquant que le fichier est introuvable, il faut absolument l'installer. Il est livré avec ISPELL (cf. 1.2).

1.1.3 Test 3 : utilisez-vous la version 3.1.20 de ISPELL ?

Lorsque vous tapez la commande :

```
ispell -v
```

la première ligne de la sortie devrait être :

```
@(#) International Ispell Version 3.1.20 10/10/95
```

Il est possible que Francais-GUTenberg fonctionne avec une autre version de ISPELL, mais ce n'est en aucun cas garanti. Il est à peu près certain que Francais-GUTenberg ne fonctionnera pas avec ISPELL 3.0.x ou antérieur.

1.1.4 Test 4 : ISPELL est-il convenablement configuré ?

Lors de la compilation de ISPELL, il est possible de choisir de nombreuses options qui auront une influence sur le comportement du programme. Pour que Francais-GUTenberg fonctionne correctement, il est indispensable que ISPELL accepte :

- soixante-quatre drapeaux (à propos de cette notion, cf. 4.4.2) ou davantage ;
- les caractères accentués.

Pour savoir si ISPELL remplit ces conditions, tapez :

```
ispell -vv
```

Parmi les lignes qui sont imprimées, vous devez impérativement trouver :

```
MASKBITS = 64
!NO8BIT (8BIT)
```

La valeur de MASKBITS doit être 64, 128 ou 256, mais pas 32. Le point d'exclamation devant NO8BIT est fondamental : il indique que cette variable n'est pas définie. Si le point d'exclamation n'apparaît pas, c'est-à-dire si la variable est définie, ISPELL sera incapable de traiter les caractères accentués (les caractères dits « huit bits »).

Si ISPELL ne remplit pas l'une de ces conditions, ou les deux, il doit être recompilé (cf. 1.3).

1.2 Télécharger les sources de ISPELL

À moins que vous ne disposiez déjà des sources de ISPELL, vous devrez les télécharger. Comme l'indique la liste ci-dessous, vous pouvez les trouver en de nombreux endroits¹.

1.2.1 Depuis la Belgique...

```
1 ftp.rug.ac.be /pub/gnu/ispell-3.1.20.tar.gz
2 ftp.glo.be /pub/mirrors/slackware-3.4/source/ap/ispell/ispell-3.1.20.tar.gz
3 ftp.interpac.be /mirror/linux/distributions/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
4 ftp.skynet.be /mirror/linux/distributions/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
5 ftp.z03.glo.be /pub/mirrors/slackware-3.4/source/ap/ispell/ispell-3.1.20.tar.gz
```

1.2.2 Depuis le Canada...

```
1 ftp.srv.ualberta.ca /pub/OpenBSD/distfiles/ispell-3.1.20.tar.gz
2 ftp.trentu.ca /unix/gnu/ispell-3.1.20.tar.gz
3 ftp.dgim.crc.ca /pub/solaris/ispell-3.1.20.tar.gz
4 ftp.sscl.uwo.ca /pub/linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
5 ftp.uwinnipeg.ca /pub/mirror/sunsite-linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
6 ftp.sun.ca /pub/freeware/SOURCES/ispell-3.1.20.tar.gz
7 ftp.tec.ualberta.ca /pub/Gnu/ispell-3.1.20.tar.gz
8 ftp.ucs.ubc.ca /pub/gnu/ispell-3.1.20.tar.gz
9 ftp.dgim.crc.ca /pub/unix/text/ispell-3.1.20.tar.gz
```

1.2.3 Depuis la France...

```
1 ftp.recherche.enac.fr /pub/SRC/ispell-3.1.20.tar.gz
2 ftp.supelec.fr /pub/sys/unix/GNU/ispell-3.1.20.tar.gz
3 ftp.ens-cachan.fr /pub/dico/ispell-3.1.20.tar.gz
4 ftp.cru.fr /pub/linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
5 ftp.iut-bm.univ-fcomte.fr /pub3/linux/distributions/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
6 ftp.lip6.fr /pub6/linux/distributions/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
7 ftp.univ-st-etienne.fr /pub/linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
8 ftp.esse-metz.fr /pub4/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
9 ftp.iut-bm.univ-fcomte.fr /pub1/linux/distributions/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
10 ftp.ujf-grenoble.fr /mirror/linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
11 ftp.crihan.fr /pub/mirror/Linux/source/ap/ispell/ispell-3.1.20.tar.gz
12 ftp.eudil.fr /pub2/linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
13 ftp.imag.fr /pub2/linux/slackware/source/ap/ispell/ispell-3.1.20.tar.gz
14 ftp.rez-gif.supelec.fr /pub/linux/slackware.3.2/source/ap/ispell/ispell-3.1.20.tar.gz
15 ftp.rez-gif.supelec.fr /pub/linux/slackware.3.4/source/ap/ispell/ispell-3.1.20.tar.gz
16 ftp.ujf-grenoble.fr /mirror/solaris/SOURCES/ispell-3.1.20.tar.gz
17 ftp.ens.fr /pub/unix/gnu/ispell-3.1.20.tar.gz
18 ftp.efrei.fr /pub/gnu/ispell-3.1.20.tar.gz
19 ftp.inria.fr /gnu/ispell-3.1.20.tar.gz
20 ftp.cenatls.cena.dgac.fr /pub/gnu/prep/ispell-3.1.20.tar.gz
21 ftp.dgac.fr /mirror/gnu/ispell-3.1.20.tar.gz
22 ftp.esse-metz.fr /pub/Gnu/ispell-3.1.20.tar.gz
23 ftp.esse-metz.fr /pub5/Gnu/ispell-3.1.20.tar.gz
24 ftp.esse-metz.fr /pub9/susix/source/apps/spell-checkers/ispell-3.1.20.tar.gz
25 ftp.espci.fr /pub/Unix/gnu/ispell-3.1.20.tar.gz
26 ftp.imag.fr /pub1/gnu/ispell-3.1.20.tar.gz
27 ftp.iut-bm.univ-fcomte.fr /pub1/gnu/ispell-3.1.20.tar.gz
28 ftp.lip6.fr /pub8/gnu/ispell-3.1.20.tar.gz
29 ftp.netplus.fr /pub3/gnu/prep/ispell-3.1.20.tar.gz
30 ftp.obspm.fr /pub1/computing/gnu/ispell-3.1.20.tar.gz
31 ftp.pasteur.fr /pub/Gnu/ispell-3.1.20.tar.gz
32 ftp.univ-lyon1.fr /pub/mirrors/gnu/ispell-3.1.20.tar.gz
33 ftp.cmla.ens-cachan.fr /pub/gnu/ispell-3.1.20.tar.gz
34 ftp.ens-cachan.fr /pub/gnu/ispell-3.1.20.tar.gz
35 ftp.eunet.fr /.01/security/pub/gnu.old/ispell-3.1.20.tar.gz
36 ftp.eurobretagne.fr /pub/unix/packages/gnu/ispell-3.1.20.tar.gz
37 ftp.univ-reunion.fr /pub/mirrors/gnu/ispell-3.1.20.tar.gz
38 ftp.utbm.fr /.pub2/unix/gnu/ispell-3.1.20.tar.gz
39 ftp.via.ecp.fr /pub/gnu/ispell-3.1.20.tar.gz
```

1. Source: <http://ftpsearch.com/index.html>.

```

40 ftp.enpc.fr          /pub/cermics/TeX/ispell-3.1.20.tar.gz
41 ftp.grolier.fr      /pub/gnu/ispell-3.1.20.tar.gz
42 ftp.mgn.fr          /pub/gnu/ispell-3.1.20.tar.gz
43 ftp.imag.fr         /pub1/ispell/ispell-3.1.20.tar.gz
44 ftp.lip6.fr         /pub8/ispell/ispell-3.1.20.tar.gz
45 ftp.esse-metz.fr    /pub7/ispell/ispell-3.1.20.tar.gz

```

1.2.4 Depuis la Suisse...

```

1 ftp.eunet.ch          /mirrors4/gnu/ispell-3.1.20.tar.gz
2 sunsite.cnlab-switch.ch /mirror/gnu/ispell-3.1.20.tar.gz

```

1.3 (Re)compiler ISPELL

Cette section aborde les détails de la compilation de ISPELL.

1.3.1 Décompresser l'archive

La première chose à faire est de décompresser l'archive dans un répertoire approprié, par exemple `/usr/local/src` si vous disposez des privilèges du « super-utilisateur », ou dans votre arborescence personnelle si ce n'est pas le cas.

Une commande du type :

```

zcat ispell-3.1.20.tar.gz | tar xvf -

```

devrait faire l'affaire. Alternativement, si vous disposez de GNU tar, vous pouvez économiser quelques caractères :

```

tar zxvf ispell-3.1.20.tar.gz

```

1.3.2 Le fichier local.h

Toutes les informations de configuration doivent figurer dans `local.h`. Il faut créer ce fichier en exécutant simplement :

```

cp local.h.samp local.h

```

1.3.2.1 Caractères huit bits

Ouvrez `local.h` dans votre éditeur favori. Sautez les quelques lignes de commentaires en début de fichier jusqu'à :

```

#define MINIMENU /* Display a mini-menu at the bottom of the screen */
#define NO8BIT   /* Remove this if you use ISO character sets */
#undef USG       /* Define this on System V */

```

En français, on a besoin des caractères accentués. Il faut donc remplacer `#define NO8BIT` par `#undef NO8BIT`.

1.3.2.2 Type de système (System V ou BSD)

Pour compiler ISPELL avec succès, il faut savoir si notre Unix est du type « System V » ou « BSD »². Dans le premier cas, il faut remplacer `#undef USG` par `#define USG`. (Il faut le faire pour Linux, par exemple.)

1.3.2.3 Les chemins

Les lignes suivantes de `local.h` méritent un peu d'attention :

```
/*
 * Important directory paths
 */
#define BINDIR "/usr/local/bin"
#define LIBDIR "/usr/local/lib"
#define ELISPDIR "/usr/local/lib/emacs/site-lisp"
#define TEXINFODIR "/usr/local/info"
#define MAN1DIR "/usr/local/man/man1"
#define MAN4DIR "/usr/local/man/man4"
```

Les deux variables les plus importantes sont `BINDIR` et `LIBDIR`. La première permet à ISPELL de trouver certains programmes auxiliaires dont il peut avoir besoin. La seconde est capitale : elle indique où ISPELL va aller chercher ses dictionnaires.

En principe, vous n'avez pas de raison de modifier ces variables, sauf si vous n'avez pas accès aux répertoires indiqués (si vous ne disposez pas des privilèges du super-utilisateur).

1.3.2.4 Augmenter le nombre des drapeaux disponibles

Français-GUTenberg utilise un grand nombre de drapeaux. La plupart du temps, ISPELL est compilé avec une limite de trente-deux drapeaux, ce qui est insuffisant. Pour corriger cela, il faut ajouter la ligne suivante à `local.h` :

```
#define MASKBITS 64
```

1.3.3 Et pour disposer d'un dictionnaire anglais...

Si vous souhaitez disposer d'un dictionnaire anglais, vous pouvez ajouter une ligne ressemblant à :

```
#define LANGUAGES "{american,MASTERDICTS=american.med,HASHFILES=americanmed.hash}"
```

(Je vous suggère d'ajouter cette ligne dans tous les cas, car bien souvent les options par défaut pour les dictionnaires engendrent une erreur lors de la compilation.)

2. Ces distinguos n'ont plus tellement court, mais ISPELL est un vieux programme.

1.4 Compilation et installation

Une fois la configuration terminée, tapez :

```
make
```

Si la compilation se déroule sans problème, vous pouvez immédiatement taper :

```
make install
```

pour achever l'installation. Si tout s'est bien déroulé jusqu'ici, vous pouvez directement passer au chapitre 2.

1.4.1 En cas d'erreur...

Trois types d'erreurs peuvent se produire pendant la compilation :

- une erreur dans la compilation de ISPELL lui-même ou de l'un des utilitaires qui l'accompagnent, comme `buildhash` ;
- une erreur lors de la création de certains fichiers de documentation (pages de manuel et fichiers `texinfo`) ;
- une erreur lors de la création du dictionnaire anglais.

Dans le premier cas, la première chose à vérifier est que vous avez fait le bon choix en ce qui concerne la variable `USG` (cf. 1.3.2.2). Si le problème ne vient pas de là, vous allez malheureusement devoir consulter la documentation de ISPELL pour voir s'il s'y trouve des informations relatives à votre système d'exploitation.

Dans le deuxième cas, cela signifie que vous ne disposez pas des programmes nécessaires pour produire la documentation. Si vous tenez absolument à ce que la compilation se déroule sans accroc, vous pouvez installer ces programmes ou modifier le `Makefile` de ISPELL. Mais si vous ne voulez pas vous embêter avec ça, vous pouvez tout simplement procéder à une installation manuelle des fichiers (cf. 1.4.2).

Le troisième type d'erreurs provient généralement d'une faute de frappe dans `local.h`. Vérifiez que les indications que vous avez données correspondent bien à des fichiers existants. Notez qu'il n'est pas du tout nécessaire de créer les dictionnaires lors de la compilation de ISPELL, et que cela peut être fait manuellement plus tard (cf. 1.4.3).

1.4.2 Installation manuelle des fichiers

Regardez dans le répertoire où vous avez compilé ISPELL. Vous devriez y trouver les exécutable suivants :

```
buildhash  
findaffix  
icombine  
ijoin
```

```
ispell  
munchlist  
sq  
subset  
tryaffix  
unsq  
zapdups
```

Vous devez les copier dans le répertoire indiqué par la variable `BINDIR` telle que définie dans le fichier `local.h` (cf. 1.3.2.3). En fait, pour assurer le bon fonctionnement de Français-GUTenberg, seuls `ispell` et `buildhash` sont nécessaires. Libre à vous également de copier les pages de manuel aux endroits appropriés si vous le souhaitez.

1.4.3 Installation manuelle de dictionnaires

Pour créer un dictionnaire anglo-américain, tapez les commandes suivantes à partir du répertoire `languages/english`:

```
cat american.0 american.1 english.0 english.1 > american.med  
buildhash american.med english.aff american.hash
```

Si vous avez une préférence pour l'anglais d'Angleterre, essayez plutôt :

```
cat british.0 british.1 english.0 english.1 > english.med  
buildhash english.med english.aff english.hash
```

(Vous pouvez bien entendu utiliser d'autres noms pour certains fichiers.)

Il ne vous reste plus qu'à copier les fichiers `*.aff` et `*.hash` dans le répertoire précisé par la variable `LIBDIR` du fichier `local.h` (cf. 1.3.2.3).

Chapitre 2

Installation de Français-GUTenberg

Ce chapitre présente l'installation de Français-GUTenberg proprement dite, c'est-à-dire la création d'un dictionnaire français.

2.1 Dernière version

Tout d'abord, assurez-vous que vous disposez de la dernière version de Français-GUTenberg. Elle est disponible à l'adresse suivante :

<http://www.unil.ch/ling/cp/frgut.html>

Des erreurs contenues dans les dictionnaires sont parfois corrigées entre deux distributions. Vous trouverez la liste des dictionnaires mis à jour à l'adresse :

<http://www.unil.ch/ling/cp/frgut-dicos.html>

2.2 L'utilitaire `makehash`

L'utilitaire `makehash` est un programme en Perl 5 qui assure la création du dictionnaire français d'après certains paramètres fournis par l'utilisateur. Ses fonctions sont les suivantes :

- estimer si la version de ISPELL installée sur votre machine peut être utilisée avec Français-GUTenberg (cf. 2.2.1) ;
- décider quelles listes de mots vont être incluses dans le dictionnaire (cf. 2.2.4) ;
- décider s'il est nécessaire de prendre des mesures particulières pour corriger des fichiers (L^A)T_EX depuis la ligne de commande (cf. 2.2.5).

Si vous ne disposez pas de Perl 5, ou pour toute autre raison, vous pouvez également procéder à une installation manuelle (cf. 2.3).

2.2.1 La bonne version de ISPELL ?

`makehash` effectue avant toute chose un certain nombre de tests pour s'assurer que votre version de ISPELL est apte à fonctionner avec Francais-GUTenberg. Si aucun message d'erreur ne s'est affiché lorsque vous avez lancé `makehash`, vous pouvez passer à la section 2.2.4, sinon lisez ce qui suit. Consultez la section 2.2.2 si `makehash` affirme que des fichiers lui manquent.

Si le message suivant s'affiche :

```
AVERTISSEMENT : impossible d'accéder à ISPELL. Ou le programme n'est pas
                 installé sur cette machine, ou il ne figure pas dans votre
                 chemin d'accès (PATH).
```

cela signifie que `makehash` ne peut pas trouver ISPELL. Reportez-vous à la section 1.1.1.

Si le message suivant s'affiche :

```
AVERTISSEMENT : impossible d'accéder à l'utilitaire buildhash. Ou le
                 programme n'est pas installé sur cette machine, ou il ne
                 figure pas dans votre chemin d'accès (PATH). buildhash
                 est nécessaire à l'accomplissement des dernières
                 opérations de création du dictionnaire.
```

cela signifie que `makehash` ne peut pas trouver l'utilitaire `buildhash`, qui est un programme indispensable à la création de dictionnaires. Reportez-vous à la section 1.1.2.

Si le message suivant s'affiche :

```
AVERTISSEMENT : apparemment votre version de ISPELL ne permet pas l'usage
                 de caractères 8 bits.
```

cela signifie que votre version de ISPELL ne peut pas gérer les caractères accentués (cf. 1.1.4). Il faut recompiler le programme avec les paramètres adéquats (cf. 1.3.2.1).

Si le message suivant s'affiche :

```
AVERTISSEMENT : apparemment votre version de ISPELL n'est pas configurée
                 avec l'option MASKBITS=64. Cela est nécessaire pour que
                 Francais-GUTenberg puisse utiliser les lettres minuscules
                 comme drapeaux.
```

cela signifie que ISPELL ne dispose pas de suffisamment de drapeaux pour gérer toutes les règles d'affixation de Francais-GUTenberg (cf. 1.1.4). Il faut recompiler le programme avec les paramètres adéquats (cf. 1.3.2.4).

2.2.2 Autres erreurs possibles

Si le message suivant s'affiche :

```
ERREUR : le répertoire dicos n'est pas accessible.
```

Vérifiez son existence ou les permissions qui lui sont affectées.

cela signifie que `makehash` n'arrive pas à trouver le répertoire contenant les dictionnaires de Français-GUTenberg. Vous devez vérifier l'intégrité de vos fichiers et de la distribution que vous avez obtenue.

Si le message suivant s'affiche :

```
AVERTISSEMENT : le dictionnaire essentiel [nom du dico] est introuvable !
```

cela signifie qu'un des dictionnaires essentiels au bon fonctionnement de Français-GUTenberg est introuvable. Vous devez vérifier l'intégrité de vos fichiers et de la distribution que vous avez obtenue. (Le répertoire `dicos` contient à la fois des dictionnaires essentiels et des dictionnaires facultatifs. La présence de ces derniers est automatiquement détectée par `makehash`. Ce programme s'assure également que les fichiers essentiels sont présents avant de continuer les opérations.)

2.2.3 Continuer ou interrompre les opérations ?

Les erreurs évoquées ci-dessus ne sont pas fatales à `makehash` (sauf l'impossibilité d'accéder au répertoire contenant les dictionnaires) et l'utilisateur est toujours libre de poursuivre la composition du dictionnaire jusqu'à la création de `francais.dico`. Par contre les dernières opérations devront être effectuées à la main (cf. 2.3). (Cela peut être utile pour composer un dictionnaire sur une machine où ISPELL n'est pas installé. Mais après, il faudra, ou installer ISPELL, ou changer de machine.)

2.2.4 Création du dictionnaire

Il s'agit maintenant de choisir quels fichiers vont être inclus dans votre dictionnaire. La première question que `makehash` vous pose est la suivante :

```
Un certain nombre de dictionnaires sont considérés comme essentiels au bon
fonctionnement de ISPELL. Souhaitez-vous tous les inclure automatiquement
(si vous répondez non, vous pourrez en effectuer la sélection dictionnaire
par dictionnaire) (o/n) ?
```

En principe, vous répondrez toujours « oui » à moins de vraiment savoir ce que vous faites. Si l'un de ces dictionnaires n'est pas inclus, il est fort probable que Français-GUTenberg ne fonctionnera pas correctement. (Si vous répondez « non », vous pourrez choisir parmi ces dictionnaires ceux que vous souhaitez intégrer.)

`makehash` affichera ensuite un par un les noms des dictionnaires facultatifs disponibles, accompagnés (en principe) d'une brève description de leur contenu. Effectuez votre sélection avec sagesse : n'incluez pas les mots d'un domaine si vous savez que vous n'en aurez pas besoin ; certains mots spécialisés risquent de faire passer inaperçues certaines fautes. Pour la même raison, à moins que votre intuition orthographique soit très bonne, évitez d'inclure des listes de mots rares.

2.2.5 Caractères accentués dans les fichiers \TeX

Si tout s'est bien passé, en fin de parcours, `makehash` va vous proposer de créer un dictionnaire spécial pour corriger des fichiers $(\text{\LaTeX})\text{\TeX}$ contenant des caractères accentués depuis la ligne de commande. Pourquoi cela ?

Le problème est le suivant : lorsque ISPELL est invoqué depuis la ligne de commande, comme dans :

```
ispell -d francais document.tex
```

le programme passe en mode \TeX , à cause de l'extension `.tex` du fichier¹. Cela a deux conséquences :

- tout mot précédé d'une barre oblique inversée est considéré comme une commande \TeX et n'est donc pas soumis au correcteur ; d'autre part, ISPELL reconnaît certaines commandes de mise en page et n'en analyse pas les arguments (comme `\label`, `\ref`, etc.) ;
- les caractères accentués « à la \TeX » sont reconnus (`\'e` à la place de `é`).

La première propriété est indispensable. La seconde peut rendre service, mais elle présente un caractère exclusif ennuyeux : toutes les équivalences sept bits définies dans `francais.aff` (vers la ligne 55) sont reconnues, par contre les caractères huit bits ne le sont plus. Il devient donc impossible de corriger les mots contenant des caractères huit bits. (Le problème ne se pose pas si ISPELL est invoqué depuis (X)Emacs.)

On peut envisager deux solutions peu satisfaisantes :

- débrayer le mode \TeX pour effectuer la correction de ce type de fichier² : le problème est alors que les commandes \TeX ne sont plus reconnues comme telles et sont donc soumises au correcteur (bref, ISPELL devient inutilisable) ;
- convertir le texte en sept bits avant d'en vérifier l'orthographe (et éventuellement le reconvertir en huit bits après) : c'est envisageable pour un long texte, mais très pénible pour une série de petits fichiers ; d'autre part, il est nettement plus agréable de voir s'afficher des caractères accentués à l'écran plutôt que les codes \TeX .

Heureusement, le problème peut être réglé dans `francais.aff` en supprimant purement et simplement la partie définissant les équivalences \TeX . C'est ce qui est fait dans le fichier `francais-TeX8b.aff`. Comment procéder ? Il est possible de créer un dictionnaire `francais-TeX8b.hash` en répondant oui à la question suivante (posée par `makehash`) :

```
Souhaitez-vous également produire un dictionnaire permettant de corriger
des fichiers TeX huit bits depuis la ligne de commande (pas nécessaire si
vous utilisez EMACS -- voir la section 2.2.5 du manuel) (o/n) ?
```

1. On peut également forcer ISPELL à passer en mode \TeX en lui donnant l'option `-t` ou `-T TeX`.

2. `ispell -n -d francais document.tex`

2.3 Installation manuelle

L'installation du dictionnaire Francais-GUTenberg se déroule en trois étapes :

- il faut rassembler les différents fichiers constituant le dictionnaire (cf. 2.3.1) ;
- il faut invoquer `buildhash` pour produire le dictionnaire proprement dit, dans un format qui sera « compris » par ISPELL (cf. 2.3.2) ;
- il faut placer les fichiers adéquats au bon endroit, pour que ISPELL les trouve (cf. 2.3.3).

2.3.1 Choix des listes de mots

Cette opération peut être accomplie par l'utilitaire `makehash`, même si vous ne disposez pas de ISPELL ou `buildhash` sur votre système. (D'autre part, `makehash` ne se contente pas d'assembler les fichiers : il trie le dictionnaire obtenu et en supprime d'éventuels doublons.) Pour faire les choses par vous-même, suivez les instructions ci-dessous.

Les listes de mots sont contenues dans le répertoire `dicos`. Vous y trouverez deux types de fichiers : les listes de mots proprement dites, dans des fichiers avec l'extension `.dico` et, correspondant à chacun de ces fichiers, une brève description stockée dans un fichier de même nom mais avec l'extension `.desc`. Vous pouvez consulter ces derniers pour savoir quelles listes de mots inclure dans votre dictionnaire.

Les fichiers suivants devraient en principe toujours être inclus dans le dictionnaire : `nonverbes.dico`, `series.dico`, `verbes-gp12.dico`, `verbes-gp3.dico` et `verbes-varia.dico`. Pour les autres listes de mots, c'est à vous de décider.

Pour composer votre dictionnaire, utilisez une commande du type :

```
cat nonverbes.dico series.dico verbes-gp12.dico [...] | sort > francais.dico
```

les crochets (`[...]`) indiquant un nombre quelconque de dictionnaires, selon vos besoins (en principe les dictionnaires essentiels plus quelques autres).

2.3.2 Utilisation de `buildhash`

Cette opération peut normalement être accomplie par l'utilitaire `makehash` si vous disposez sur votre système d'une version correctement configurée de ISPELL et de `buildhash`.

Utilisez la commande suivante pour effectuer l'opération manuellement (vous devez disposer d'un fichier `francais.dico` — cf. 2.3.1) :

```
buildhash francais.dico francais.aff francais.hash
```

Si vous comptez corriger des fichiers (L^A)T_EX depuis la ligne de commande, vous devez également créer `francais-TeX8b.hash` (cf. 2.2.5). Utilisez la commande suivante pour ce faire :

```
buildhash francais.dico francais-TeX8b.aff francais-TeX8.hash
```

2.3.3 Installation des fichiers

La dernière étape consiste à placer `francais.aff` et `francais.hash` (et, éventuellement, `francais-tex8b.aff` et `francais-tex8.hash`) à l'endroit où ISPELL s'attend à les trouver. Si vous avez utilisé `makehash` jusqu'ici, le programme vous indique où placer ces fichiers avant de s'arrêter. Sinon, vous pouvez obtenir cette précieuse information en utilisant la commande :

```
ispell -vv
```

et en cherchant une ligne du type :

```
LIBDIR = "/usr/local/lib"
```

Copiez les fichiers dans le répertoire indiqué par la variable LIBDIR.

Chapitre 3

Configuration de Emacs et XEmacs

Ce chapitre évoque brièvement les problèmes que vous pourriez rencontrer avec Emacs ou XEmacs lors de l'utilisation de Francais-GUTenberg. L'utilisation de FlySpell pour la correction « au vol » est également envisagée.

Il est fort probable que la version de (X)Emacs dont vous disposez fonctionne d'emblée sans problème avec Francais-GUTenberg. Dans ce cas, vous pouvez sauter ce chapitre. (Si vous n'utilisez pas FlySpell, jetez quand même un coup d'œil à la section 3.4.)

Pour savoir comment utiliser Francais-GUTenberg dans (X)Emacs, reportez-vous à la section 4.3.

3.1 Généralités

Il est possible d'invoquer ISPELL depuis Emacs ou XEmacs. Cela peut se faire grâce à la présence du fichier `ispell.el` (et/ou `ispell.elc`, le premier étant le code source Lisp, le second le résultat après compilation). Ce fichier est généralement livré avec (X)Emacs. Il en existe également une version qui accompagne les sources de ISPELL.

3.2 Emacs

Généralement, il est possible d'utiliser ISPELL sans problème avec les versions 19.x et 20.x de Emacs. Le fichier `ispell.el` qui accompagne Emacs 19.x est en principe identique à celui que vous trouverez dans les sources de ISPELL. Par contre, il faut utiliser le `ispell.el` livré avec Emacs 20.x si vous utilisez cette version de l'éditeur.

Si, par hasard, votre version de Emacs était livrée sans `ispell.el`, vous pouvez copier celui livré avec les sources de ISPELL dans le répertoire `site-lisp` de l'arborescence de Emacs¹. En principe, cela ne marchera pas avec Emacs 20.x, mais cette version devrait toujours être livrée avec son propre `ispell.el`.

3.3 XEmacs

Généralement, il est possible d'utiliser ISPELL sans problème avec les versions 19.x et 20.x de XEmacs. Le fichier `ispell.el` qui accompagne XEmacs est en principe identique à celui que vous trouverez dans les sources de ISPELL.

Si, par hasard, votre version de XEmacs était livrée sans `ispell.el`, vous pouvez copier celui livré avec les sources de ISPELL dans le répertoire `site-lisp` de l'arborescence de XEmacs².

Il arrive parfois que ISPELL refuse de fonctionner avec XEmacs après avoir été recompilé. J'ai toujours pu régler ce problème en supprimant les fichiers `ispell.el` et `ispell.elc` livrés avec XEmacs³ et en plaçant le fichier `ispell.el` livré avec ISPELL à l'endroit indiqué au paragraphe précédent.

3.4 FlySpell

FlySpell est un mode mineur de (X)Emacs développé par Manuel Serrano⁴ qui permet la correction « au vol » des fautes. Le fichier `flyspell.el` est inclus avec la distribution de Francais-GUTenberg.

Pour l'installer, il suffit de copier `flyspell.el` dans le répertoire `site-lisp` de l'arborescence de votre (X)Emacs (cf. 3.2 et 3.3). Ajoutez ensuite les deux lignes suivantes à votre fichier `.emacs` dans votre répertoire racine :

```
(autoload 'flyspell-mode "flyspell" "On-the-fly spelling checking" t)
(autoload 'global-flyspell-mode "flyspell" "On-the-fly spelling" t)
```

Pour savoir comment utiliser FlySpell, reportez-vous à la section 4.3.3.

1. Généralement : `/usr/share/emacs/site-lisp`, mais cela peut varier d'un système à l'autre.

2. Généralement : `/usr/lib/xemacs/site-lisp`, mais cela peut varier d'un système à l'autre.

3. Dans le répertoire `.../xemacs-[version]/lisp/packages` en général.

4. Manuel.Serrano@unice.fr

Deuxième partie

Utilisation

Chapitre 4

Utilisation de Francais-GUTenberg

Ce chapitre présente les éléments nécessaires à l'utilisation de ISPELL avec Francais-GUTenberg. Il ne constitue pas une présentation exhaustive du correcteur ISPELL ; pour cela reportez-vous aux pages de manuel de ce programme. Les principes théoriques ayant présidé à l'élaboration de Francais-GUTenberg ne seront pas davantage exposés ici, mais traités au chapitre 5.

4.1 Appeler ISPELL depuis la ligne de commande

Si vous corrigez un fichier texte normal ou un fichier (L^A)T_EX ne comportant pas de caractères accentués (c'est-à-dire utilisant les codes T_EX pour représenter les caractères accentués), utilisez la commande suivante :

```
ispell -d francais document.txt
```

(L'extension du fichier n'a pas d'importance pour un fichier texte, mais elle doit être `.tex` pour un fichier (L^A)T_EX.)

Si vous corrigez un fichier (L^A)T_EX contenant des caractères accentués (c'est-à-dire codés sur huit bits), utilisez la commande suivante :

```
ispell -d francais-TeX8b document.tex
```

(Pour comprendre pourquoi cela est nécessaire, reportez-vous à la section 2.2.5.)

Il est possible d'indiquer simultanément plusieurs fichiers à ISPELL.

4.2 Commandes de ISPELL

Une fois lancé, ISPELL affiche, dans leur contexte, les mots dont il considère la graphie comme erronée (appelés ci-après « mots erronés »). Il propose généralement quelques remplacements possibles en précisant sur quelle touche vous devez appuyer pour effectuer la substitution.

Les autres commandes disponibles sont les suivantes :

- touche **R** : permet de taper un mot pour remplacer complètement le mot affiché ;
- barre d'espace : accepte le mot pour cette fois seulement (le mot sera à nouveau signalé comme erroné la prochaine fois qu'il sera rencontré) ;
- touche **A** : accepte le mot tel quel (il ne sera plus signalé comme erroné jusqu'à l'arrêt de ISPELL) ;
- touche **I** : accepte le mot et l'inscrit dans votre dictionnaire personnel (ce mot ne sera plus jamais considéré comme erroné par ISPELL) ;
- touche **U** : accepte le mot et l'inscrit dans votre dictionnaire personnel, mais en minuscules (ce mot ne sera plus jamais considéré comme erroné par ISPELL) ;
- touche **L** : vous permet de taper un mot et ISPELL affiche les mots voisins dans le dictionnaire (selon l'ordre alphabétique) ; il est possible d'utiliser l'astérisque comme joker (comme remplacement d'un ou de plusieurs caractères) ;
- touche **X** : abandonne la correction du fichier tout en sauvegardant les corrections effectuées (ISPELL passe au fichier suivant si plusieurs fichiers ont été indiqués sur la ligne de commande) ;
- touche **Q** : abandonne la correction du fichier sans modifier celui-ci ;
- touche **!** : permet de taper des commandes de l'interpréteur (*shell*) ;
- touche **?** : affiche un petit texte d'aide, proche de ce que vous venez de lire.

4.3 Utilisation depuis (X)Emacs

Cette section présente les trucs et astuces liés à l'utilisation de (X)Emacs avec Francais-GUTenberg.

4.3.1 Avec Emacs...

Les commandes qui nous intéressent se situent dans le menu *Spell*. Il s'agit d'un sous-menu du menu *Edit*. La première chose à faire est de choisir la commande *Select Francais*. Ceci fait, vous devriez voir apparaître dans le « mini-buffer » (au bas de l'écran), le message suivant : « *(Next local Ispell command will use francais dictionary)* ».

Vous pouvez maintenant choisir une des commandes suivantes :

- *Check Buffer* : pour corriger la totalité du document que vous êtes en train d'éditer ;
- *Check Region* : pour corriger une zone de texte que vous aurez préalablement sélectionnée, soit avec la souris, soit avec le clavier ;

- *Check Word* : pour corriger le mot dans lequel, ou devant lequel, se trouve le curseur (peut aussi être invoqué par la combinaison de touches *Meta-\$*, ce qui est bien pratique).

Une fois la correction lancée, vous disposez des mêmes options que si ISPELL avait été invoqué depuis la ligne de commande. Reportez-vous à la section 4.2 pour en avoir une brève description.

4.3.2 Avec XEmacs...

XEmacs fonctionne comme Emacs, à la seule exception que le sous-menu *Spell* n'apparaît pas dans le menu *Edit*, ce qui est ennuyeux. Pour remédier à cet inconvénient, appuyez sur la combinaison de touches *Meta-\$*, ceci aura pour effet d'activer ISPELL. Si ISPELL vous envoie un message à propos d'un mot mal orthographié, appuyez simplement sur la barre d'espace pour vous en débarrasser.

Si vous cliquez maintenant sur le menu *Edit*, vous verrez que le sous-menu *Spell* est disponible. Reportez-vous à la section 4.3.1 pour la suite des opérations.

4.3.3 Avec FlySpell...

Si vous avez installé FlySpell (cf. 3.4), vous pouvez l'utiliser avec Francais-GUTenberg pour corriger vos fautes « au vol ». Son utilisation est très simple. Pour l'invoquer dans un document que vous êtes en train de rédiger, appuyez sur les touches *Meta-X*, puis tapez la commande :

```
flyspell-mode
```

dans le « mini-buffer ».

Dans le sous-menu *Spell* du menu *Edit*, choisissez la commande *Select Francais* (sinon FlySpell utilisera le dictionnaire anglais).

Vous pouvez maintenant taper votre texte : les mots qui ne sont pas reconnus apparaîtront soulignés et dans un rouge éclatant (si c'est la première fois que le terme apparaît) ou en jaune moutarde — officiellement couleur or — (si le terme est déjà présent plus haut dans le document).

Pour effectuer une correction, cliquez avec le bouton du milieu de la souris sur le mot souligné et un menu vous proposant des remplacements fera son apparition. Faites votre choix et le tour est joué.

Si vous aviez déjà tapé du texte avant de lancer FlySpell, passez sur celui-ci en utilisant les touches du curseur pour activer la correction.

4.4 Fonctionnement du système

Cette section présente brièvement les principes de fonctionnement de ISPELL qu'il peut vous être utile de connaître dans votre utilisation quotidienne (?) de Francais-GUTenberg. (Pour une présentation plus détaillée de certains mécanismes, reportez-vous au chapitre 5.)

Si cette cuisine interne ne vous intéresse pas, vous pouvez sauter immédiatement à la section suivante (4.5) et revenir ici si vous rencontrez alors une difficulté de compréhension.

4.4.1 Qu'est-ce qu'un correcteur orthographique ?

On appelle généralement « correcteur orthographique » tout programme qui compare les mots d'un texte à ceux présents dans une liste de mots, appelée « dictionnaire » (procédure de vérification, cf. 4.4.3). Lorsqu'un mot du texte en vérification ne figure pas dans le dictionnaire, un algorithme calcule à partir du mot (potentiellement) mal orthographié un certain nombre (qui peut être nul) de corrections possibles qui sont des mots graphiquement proches figurant dans le dictionnaire (procédure de suggestion, cf. 4.4.4).

Certains logiciels plus ambitieux tentent de corriger, en plus de l'orthographe d'usage, l'orthographe d'accord ou « grammaticale ». ISPELL n'entre pas dans cette seconde catégorie et fonctionne selon le premier modèle évoqué : vérification puis, éventuellement, suggestion de graphies alternatives.

4.4.2 Le format du dictionnaire

Un dictionnaire ISPELL prêt à l'emploi se présente sous la forme de deux fichiers : `langue.aff` et `langue.hash`. Ils s'appellent `francais.aff` et `francais.hash` dans Francais-GUTenberg (et sont complétés le cas échéant par `francais-TeX8b.aff` et `francais-TeX8b.hash` — cf. 2.2.5). Le premier contient la liste des caractères pouvant licitement apparaître dans un mot, ainsi qu'un ensemble de règles d'affixation dont il va être question un peu plus loin. Le second est une table de hachage obtenue à partir du premier et d'un troisième fichier, que l'on peut appeler `langue.dico`¹ (pour le français : `francais.dico`).

Un fichier `langue.dico` présente deux types d'entrées :

(4.1) `mot1`

(4.2) `mot2/drapeau(x)`

1. Il n'existe pas d'extension standard pour ce troisième fichier qui est d'ailleurs généralement le produit de l'agrégation de plusieurs listes de mots.

Dans le premier cas², on a simplement affaire à un mot faisant partie du dictionnaire. Dans le second, le mot est accompagné d'un ou plusieurs drapeaux renvoyant aux règles d'affixation définies dans `langue.aff` : à la ligne (4.2) vont correspondre plusieurs mots du dictionnaire. Par exemple, dans le fichier `francais.dico` de Francais-GUTenberg, on peut trouver :

(4.3) `coordinatrice/G`

(4.3) indique que le mot *coordinatrice* est présent dans le dictionnaire, de même que les mots *coordinatrices*, *coordinateur* et *coordinateurs* engendrés à partir de certaines règles d'affixation appartenant au drapeau `G` :

(4.4) `[^SXZ]` > `S`

(4.5) `T R I C E` > `-RICE, EUR`

(4.6) `T R I C E` > `-RICE, EURS`

La règle (4.4) engendre un pluriel en « s » à partir de n'importe quel mot qui ne s'achève ni par « s », ni par « x » et ni par « z ». La règle (4.5) transforme tout mot s'achevant en *-trice* en mot s'achevant en *-eur* en soustrayant les lettres « r », « i », « c » et « e » (dans cet ordre) à la fin du mot de départ et en lui ajoutant les lettres « e », « u » et « r » (dans cet ordre). La règle (4.6) fait de même, mais ajoute encore la lettre « s » pour créer le pluriel du masculin.

On voit que Francais-GUTenberg part du féminin pour aboutir au masculin, contrairement à une certaine tradition qui veut que l'on cherche un nom dans un dictionnaire d'après la forme du masculin singulier. Reportez-vous à la section 5.2 pour une justification de cette manière de faire (qui n'a que peu d'incidence sur ce qui nous intéresse ici). En lexicographie, le mot dont on part pour aboutir à d'autres mots qui sont ses variantes morphologiques s'appelle un « lemme ».

4.4.3 Vérification

ISPELL considère qu'un mot est correctement orthographié s'il s'agit d'un lemme du dictionnaire ou s'il peut être engendré par une règle d'affixation (cf. 4.4.2) : la présence de `coordinatrice/G` dans le dictionnaire implique que *coordinatrice*, *coordinatrices*, *coordinateur* et *coordinateurs* sont corrects).

4.4.4 Suggestion

Lorsque ISPELL considère un mot comme erroné, il va suggérer quelques graphies proches qui figurent dans son dictionnaire (comme lemmes ou engendrées par les

2. Les numéros entre parenthèses servent à identifier les différents exemples donnés et ne font pas partie des entrées du dictionnaire.

règles d’affixation). Une graphie sera considérée comme proche si elle remplit une (et une seule) des conditions suivantes :

- elle peut être obtenue en permutant deux lettres voisines (**cmaembert* → *camembert*) ;
- elle peut être obtenue en changeant une lettre (**cjaleur* → *chaleur*) ;
- elle peut être obtenue en supprimant une lettre (**beaucoup* → *beaucoup*) ;
- elle peut être obtenue en ajoutant une lettre (**corp* → *corps*) ;
- ou si l’insertion d’un espace dans la chaîne donne deux mots reconnus par le dictionnaire (**nonlinéaire* → *non linéaire*).

4.5 Éléments pour résoudre quelques problèmes

ISPELL n’est pas un programme parfait et l’orthographe française présente des particularités difficiles à gérer dans le cadre de ce logiciel. Bien que Français-GUTenberg ait été élaboré pour répondre à ces deux problèmes, certains dysfonctionnements subsistent. Cette section leur est consacrée.

4.5.1 Absence de suggestions

L’algorithme de suggestion de ISPELL n’est pas très évolué, comme on peut le voir à la section 4.4.4 : il ne permet la correction que de fautes portant sur une seule lettre. Il ne pourra pas suggérer *précisément* à partir de **précisemment*, ou *notamment* à partir de **nottament*. Dans ces deux cas, ISPELL ne fera aucune suggestion. Alors comment s’en sortir ?

Le plus simple est de tester différentes variantes : utilisez la touche **R** (cf. 4.2) pour essayer des graphies alternatives jusqu’à ce que vous tombiez sur la bonne ou que ISPELL parvienne à vous faire des suggestions sensées. Si vous utilisez FlySpell, vous pouvez taper quelques variantes à côté du mot considéré comme faux et voir ce qui se passe.

Le dictionnaire de Français-GUTenberg est assez vaste, aussi y a-t-il peu de chance pour qu’un mot courant en soit absent. Prenez donc bien garde de ne pas ajouter inconsidérément des mots mal orthographiés à votre dictionnaire personnel en utilisant la touche **I** dès que ISPELL est incapable de faire une suggestion.

4.5.2 ISPELL vous propose de valider une affixation

Parfois, plutôt que de vous faire des suggestions (cf. 4.4.4), ISPELL vous demande si un mot peut être obtenu par la combinaison d’un lemme et de suffixes. Dans la plupart des cas, il ne faut pas valider la suggestion : par exemple, si vous avez tapé **l’haricot*, ISPELL va vous demander s’il s’agit d’une combinaison valide de **l’** +

haricot (ce n'est pas le cas bien sûr). En théorie, il existe un moyen d'inhiber ce comportement (cf. 5.3.1), mais en pratique cela ne fonctionne pas.

4.5.3 Élision du pronom *que*

Dans certains cas, il faudrait que l'élision du pronom « que » soit admise (par exemple : « qu'honneurs », « qu'utile », « qu'irrespectueuse »), mais pas dans d'autres (« qu'Afrique », « qu'âme », etc.). Cela dépend du contexte et du sens des mots et n'a pas pu être spécifié dans le dictionnaire (ce sera peut-être réglé dans la prochaine version de Français-GUTenberg). Ce genre de construction, fort heureusement assez rare, va donc engendrer une fausse alerte.

Si vous avez un doute sur la graphie du mot qui suit le pronom « que » élidé et que vous utilisez FlySpell, vous pouvez taper un espace après l'apostrophe pour voir ce qui se passe. Sinon vous pouvez utiliser la touche **L**.

4.5.4 Élision de *le* et *de* devant un mot avec une capitale

Si vous devez écrire un mot avec une capitale initiale et le faire précéder d'un article élidé (par exemple : « l'Absence »), ISPELL ne va pas l'accepter. Cela provient de la manière dont ISPELL gère la capitalisation des mots de son dictionnaire qui ne prévoit pas ce genre de choses, typique du français (ISPELL a été essentiellement conçu pour l'anglais et l'allemand). Notez que le problème ne se pose pas avec les mots qui commencent toujours pas une majuscule.

Si vous avez un doute sur la graphie du mot qui suit le pronom « le » ou « de » élidé et que vous utilisez FlySpell, vous pouvez taper un espace après l'apostrophe pour voir ce qui se passe. Sinon vous pouvez utiliser la touche **L**.

4.5.5 Le caractère *œ*

ISPELL utilise l'encodage *Isolatin-1* qui ne comprend pas le caractère « œ ». En attendant l'apparition de *Isolatin-9* (ISO-8859-15), ce caractère est remplacé dans les mots du dictionnaire qui le contiennent par les lettres « oe ».

Lorsque vous tapez un mot contenant ce caractère (`{\oe}uvre` → œuvre), ISPELL va malheureusement produire une fausse alerte. Si vous avez un doute sur l'orthographe du mot considéré, vous pouvez utiliser la touche **L** pour vous rassurer ou, si vous utilisez FlySpell, tapez le mot à côté en utilisant les caractères « oe ».

La prochaine version de Français-GUTenberg utilisera *Isolatin-9* là où c'est possible afin de résoudre ce problème³. (Le noyau 2.2.x de Linux permet l'utilisation de

3. Pendant l'élaboration de Français-GUTenberg, j'ai envisagé plusieurs solutions « bricolées » pour résoudre le problème, mais aucune n'était très élégante et nous courrions le risque de les voir perdurer alors qu'elles auraient cessé d'être utiles avec l'introduction de *Isolatin-9*.

Isolatin-9, et la dernière version de `kbd` — 0.99 — contient des polices avec le jeu de caractères *Isolatin-9*. Espérons que des polices pour X-Windows seront également bientôt disponibles.)

Troisième partie

Fonctionnement et création de dictionnaires

Chapitre 5

Élaboration du système

Ce chapitre présente les éléments fondamentaux ayant présidé à l'élaboration de Francais-GUTenberg. Sa lecture vous permettra de mieux comprendre le fonctionnement général de Francais-GUTenberg (voir aussi 4.4) et vous sera utile si vous comptez créer un dictionnaire spécialisé en suivant les instructions du chapitre 7.

5.1 Bref historique de ce projet

Aux origines de Francais-GUTenberg se trouve mon travail de diplôme (Pythoud 1996). Je souhaitais que mon mémoire de linguistique porte sur un problème de traitement automatique du langage. Plus particulièrement, je m'intéressais au problème de l'orthographe, bien qu'ayant été le dernier des cancre dans ce domaine pendant mes années d'école.

L'existence de ISPELL et de Francais-IREQ m'a fourni un sujet de choix. Après un travail critique, j'ai abouti, entre autres, à un certain nombre de conclusions sur la manière de composer un dictionnaire pour le français dans le cadre de ISPELL. Francais-GUTenberg est le fruit de ces cogitations.

Les lignes qui suivent reprennent quelques points de ce travail, importants pour comprendre le fonctionnement de Francais-GUTenberg.

5.2 Choix des lemmes

Lors de la présentation du format du dictionnaire, à la section 4.4.2, j'avais attiré l'attention du lecteur sur le fait que pour traiter des noms et des adjectifs, Francais-GUTenberg partait du féminin singulier et non du masculin comme dans les dictionnaires sur papier auxquels nous sommes habitués. Ceci est extrêmement important.

Comme déjà indiqué en fin de section 4.4.2 (page 25), en lexicographie, on appelle « lemme »¹ un mot choisi pour représenter toute une famille de formes apparentées. Par exemple, si je trouve la forme « soliloquions » dans un texte et que je désire en connaître le sens précis à l'aide d'un dictionnaire courant, je vais chercher dans celui-ci la forme « soliloquer » et non « soliloquions ». Nous sommes en effet habitués à chercher les verbes sous leur forme infinitive dans les dictionnaires et les ouvrages de grammaire. L'infinitif constitue donc le lemme ou « mot-vedette », pour reprendre la terminologie de Nina Catach (1995), pour représenter les verbes. Parallèlement, pour un nom ou un adjectif, je me référerai à la forme du masculin singulier.

Par analogie, on peut dire que ISPELL fonctionne de cette manière : au travers des règles d'affixation dont il dispose, il va chercher à trouver le « mot-vedette », le lemme, correspondant à la graphie qu'il est en train d'analyser (cf. 4.4.3). Ici une question se pose : quels lemmes allons-nous utiliser dans la composition de notre dictionnaire ? l'infinitif pour les verbes et le masculin singulier pour les noms et les adjectifs ? ou autre chose ?

Même pour un dictionnaire sur papier, l'emploi de l'infinitif et du masculin singulier n'est qu'une convention et rien ne nous empêcherait, sinon la tradition, de choisir comme lemme, pour un verbe, la première personne du présent de l'indicatif. C'est d'ailleurs cette forme qui est utilisée dans la plupart des dictionnaires grecs et latins. Est conventionnel aussi le fait de n'avoir qu'un lemme pour une famille de mots, convention d'ailleurs très relative : lorsqu'un nom ou un adjectif présente un féminin ou un pluriel irrégulier, cela est généralement indiqué.

Or, le choix des lemmes n'est justement pas un problème trivial au niveau de la composition d'un dictionnaire français pour ISPELL. Français-IREQ adopte la solution traditionnelle : infinitif et masculin singulier. On peut faire mieux en étant plus audacieux.

Par exemple, comment traiter la féminisation² et la mise au pluriel des mots : « pleureur », « puériculteur », « chasseur » et « docteur » ? Si on choisit de partir du masculin singulier, il faudra quatre règles différentes pour obtenir les féminins (et il vaudrait mieux ne pas se tromper en attribuant les drapeaux) : « pleureuse », « puéricultrice », « chasserresse » et « doctoresse ». Cela donnerait quelque chose comme ça :

```
flag *A: # pleureur, pleureuse
          E U R                >          -R,SE
          E U R                >          -R,SES
```

1. Je préfère recourir à la terminologie des lexicographes plutôt que d'utiliser le terme de « racine », traduction de l'anglais « root », présent dans la documentation de ISPELL, et qui est contraire à la tradition francophone.

2. Les remarques qui suivent ne concernent bien entendu que les mots possédant une forme masculine *et* féminine. Le traitement des termes épiciens est beaucoup plus simple.

```

flag *B: # puériculteur, puéricultrice
    E U R > -EUR,RICE
    E U R > -EUR,RICES

flag *C # chasseur, chasseresse
    E U R > -UR,RESSE
    E U R > -UR,RESSES

flag *D # docteur, doctoresse
    E U R > -EUR,ORESSE
    E U R > -EUR,ORESSES

```

avec pleureur/AS³, puériculteur/BS, chasseur/CS et docteur/DS dans le dictionnaire.

Maintenant, si on observe les terminaisons des mots féminins, on s'aperçoit qu'elles sont toutes différentes et qu'il est donc possible, en utilisant le féminin singulier comme lemme, de n'utiliser qu'un seul drapeau. C'est ce qui est fait dans Francais-GUTenberg et le drapeau G est consacré aux mots en *-eur* :

```

flag *G: # mots en -eur
    E > S

    # pleureuse
    E U S E > -SE,R
    E U S E > -SE,RS

    # puéricultrice
    T R I C E > -RICE,EUR
    T R I C E > -RICE,EURS

    # chasseresse
    E R E S S E > -RESSE,UR
    E R E S S E > -RESSE,URS

    # doctoresse
    O R E S S E > -ORESSE,EUR
    O R E S S E > -ORESSE,EURS

```

avec pleureuse/G, puéricultrice/G, chasseresse/G et doctoresse/G dans le dictionnaire.

Le drapeau G est réservé aux mots en *-eur* (et aux mots en *-al* faisant leur pluriel

3. En partant du principe qu'il existe un drapeau S qui renvoie à un traitement général des pluriels.

en *-als*). Les autres termes non épiciens sont tous traités par le drapeau F, dont voici un extrait :

# Consonne(s) (t)t + e		
T T E	>	-TTE,T # coquette, sothe
T T E	>	-TE,S
[^èT] T E	>	-TE,T # idiote
[^èT] T E	>	-E,S
è T E	>	-èTE,ET # discrète
è T E	>	-èTE,ETS

qui présente les mêmes caractéristiques d'économie et dont je laisse l'analyse comme exercice au lecteur.

Quand on s'attaque aux verbes, un problème du même ordre, mais de plus grande ampleur, se présente : Français-IREQ utilise comme lemme l'infinitif. Or, une des difficultés de la conjugaison en français est qu'un verbe peut avoir plusieurs radicaux : *manger* en a deux, *mang-* et *mange-* ; *envoyer* en a trois, *envoi-*, *envoy-* et *enverr-* ; *tenir* en a cinq, *tien-*, *tienn-*, *ten-*, *tin-* et *tiend-*. Ceci a pour conséquences, si on veut traiter exhaustivement du verbe français dans le cadre de Français-IREQ, les problèmes suivants :

- une multiplication des drapeaux ;
- des règles d'affixation extrêmement complexes ;
- l'*impossibilité* de traiter de tous les verbes du français, d'où l'absence de nombreuses formes du troisième groupe dans Français-IREQ ;
- la *nécessité*, au vu de ce qui précède, d'entrer directement dans le dictionnaire les formes qui ne peuvent être traitées par le système d'affixation, ce qui n'en facilite pas la relecture ni la mise à jour (il est difficile de ne pas commettre d'erreurs en entrant « manuellement » de nombreuses formes verbales).

La solution consiste ici à augmenter le nombre de lemmes, ce qui amène une diminution, apparemment paradoxale (du même ordre que ce qui se passe pour les mots en *-eur*), des drapeaux, réduit la complexité des règles et permet un traitement exhaustif des verbes des trois groupes. Les cinq lemmes suivants sont utilisés par Français-GUTenberg : les premières personnes du singulier du présent, de l'imparfait, du passé simple et du futur de l'indicatif, ainsi que l'infinitif. Cette stratégie permet également de résoudre l'élosion du pronom personnel *je* (cf. 7.2.3).

La démonstration est un peu courte et incomplète sur ce dernier point : le lecteur intéressé pourra se reporter à Pythoud (1996).

5.3 Le problème de l'élision

Cette section présente brièvement ce qu'est le « problème de l'élision », familier des utilisateurs de Français-IREQ, et comment il se trouve résolu dans Français-GUTenberg.

5.3.1 Vérification et élision

Nous avons vu à la section 4.4.3 que ISPELL considérait comme correctement orthographié tout mot faisant partie des lemmes du dictionnaire ou pouvant être engendré par une règle d'affixation (cf. 4.4.2).

Maintenant, si le fichier `français.aff` contient la mention `allaffixes on` sur une ligne, ISPELL essayera alors *tous* les drapeaux sur le mot pour tenter d'aboutir à un lemme. Par exemple, supposons que vous tapiez *l'éléphant* et que le dictionnaire contienne `éléphant/S`, mais pas `éléphant/LS` (avec `S` gérant le pluriel et `L` l'élision du déterminant *le*) : ISPELL va s'apercevoir que *l'éléphant* serait correct si `éléphant/S` était accompagné du drapeau `L` ; il va donc le signaler à l'utilisateur et lui demander de valider (ou non) l'affixation `l' + éléphant`.

À priori, ce comportement peut sembler adéquat ; il ne l'est pas, car il ne sert qu'à récupérer une erreur commise dans la conception du dictionnaire : l'absence d'un drapeau. ISPELL produit donc une fausse alerte qui gêne l'utilisateur : si de nombreux drapeaux manquent, comme c'est malheureusement le cas dans Français-IREQ, les fausses alertes de ce type seront nombreuses et le confort d'utilisation s'en trouvera très diminué (cf. 5.3.2).

Le problème est résolu dans Français-GUTenberg en s'assurant que tous les mots du dictionnaire sont accompagnés des drapeaux adéquats et en précisant `allaffixes off` dans `français.aff`. En fait, je me suis rendu compte que même avec `allaffixes off`, ISPELL continue à chercher un affixe permettant d'accepter le mot, *i.e.*, si vous tapez **l'haricot*, il vous proposera de valider le mot comme la combinaison de `l' + haricot`. Ne le faites pas (cf. 4.5.2) !

5.3.2 Suggestion et élision

L'élision des pronoms personnels et des articles pose des problèmes complexes : il en sera continuellement question au fil du chapitre 7. Cela peut même engendrer des problèmes au niveau de la procédure de suggestion : reprenons notre exemple de *l'éléphant*. Supposons que le dictionnaire contienne l'entrée `éléphant/S`, qui engendre *éléphant* et *éléphants*, mais pas `éléphant/LS` qui engendrerait, en plus de ce qui précède, *l'éléphant*. Supposons encore que, victime d'une fatigue soudaine, je tape `éléphent` : ISPELL n'aura aucune peine à me suggérer de remplacer cette graphie fautive par *éléphant*. Mais, si j'ai tapé `l'éléphent`, ISPELL ne pourra que me

signaler que cette graphie est probablement mauvaise, mais en aucun cas il ne parviendra à suggérer *l'éléphant* comme remplacement. Pourquoi? Parce que le drapeau L manque et que par conséquent, la forme *l'éléphant* est absente du dictionnaire. Voilà une raison supplémentaire de veiller à ce que chaque entrée du dictionnaire soit accompagnée des drapeaux adéquats.

5.4 Points divers

5.4.1 Écrire les nombres en toutes lettres

Les règles de l'orthographe française exigent que, dans certains contextes, on écrive les nombres en toutes lettres. Or, beaucoup de nombres manquent dans Français-IREQ. Vous les trouverez, sous leurs formes cardinale et ordinale, dans Français-GUTenberg.

5.4.2 Le vocabulaire spécialisé

Un reproche souvent fait à Français-IREQ est de ne pas connaître le vocabulaire de telle ou telle discipline, en particulier les mathématiques. Idéalement, lorsqu'on corrige un texte relevant d'une discipline particulière, on devrait pouvoir invoquer, en plus du dictionnaire principal, un dictionnaire spécialisé contenant les termes propres à la discipline considérée. ISPELL ne dispose pas de ce mode de fonctionnement (contrairement aux correcteurs de certains traitements de textes commerciaux, il ne permet pas l'utilisation simultanée de plusieurs dictionnaires). La seule solution est donc d'inclure le vocabulaire propre à la discipline qui nous intéresse directement dans le dictionnaire. Mais si le dictionnaire contient des termes mathématiques, les physiciens vont alors tout à fait légitimement demander qu'y figure également le vocabulaire de la physique, les juristes celui du droit, les linguistes celui de la linguistique, etc. Inclure le jargon de tous ces différents domaines dans le dictionnaire n'est pas envisageable, cela pour deux raisons :

- la taille excessive du dictionnaire finirait par affecter les performances du correcteur ;
- certaines fautes risqueraient de passer inaperçues à cause de termes spécialisés qui les masqueraient ⁴.

4. Par exemple, lorsqu'on s'intéresse à la logique de Montague (Montague 1974), on est souvent amené à parler d'*intension* (par opposition à *extension*). On peut imaginer que ce terme soit contenu dans le fichier `logique.dico` inutilement inclus dans le dictionnaire principal. Si un utilisateur, qui n'a jamais entendu parler des théories de Montague, écrit **intension* à la place de *intention*, la faute ne sera pas détectée.

Au vu de ce qui précède, la solution envisagée est la suivante : l'utilisateur a la possibilité de choisir, lors de la compilation du dictionnaire, quelles listes de mots il compte y faire figurer. La procédure de compilation et la création de dictionnaires spécialisés sont décrites réciproquement aux chapitres 2 et 7.

Chapitre 6

Présentation du fichier d'uffixes de Francais-GUTenberg

Ce chapitre présente `francais.aff`, le fichier d'uffixes de Francais-GUTenberg.

6.1 Options de base

Cette section décrit les options de base spécifiées dans `francais.aff`. Chaque sous-section commence par citer *in extenso* la partie du fichier d'uffixes dont elle traite ; les commentaires suivent.

6.1.1 La commande `allaffixes`

```
# Pas d'uffixation automatique
```

```
allaffixes off
```

Lorsque ISPELL trouve un mot inconnu, mais que ce mot pourrait être admis si on l'analysait comme issu de la combinaison d'un lemme avec un affixe, ISPELL propose à l'utilisateur de valider la combinaison. Ce comportement n'est adéquat que lorsqu'un dictionnaire est en développement ou qu'il lui manque des drapeaux. Ce n'est pas le cas de Francais-GUTenberg : `allaffixes off` est donc spécifié pour débrayer cette fonction. (Malheureusement, dans la pratique, le débrayage n'est pas effectué. Il s'agit d'un bogue de ISPELL.)

Pour plus d'informations sur le sujet, voir la section 5.3.1.

6.1.2 Les caractères composant les mots

```
# Définition des caractères
```

```
defstringtype "list" "nroff" ".list"
```

```
boundarychars  [-]
boundarychars  '
boundarychars  [.]
boundarychars  @
wordchars      a      A
stringchar     à      À
stringchar     á      Á
stringchar     â      Â
stringchar     ä      Ä
wordchars      [b-c]  [B-C]
stringchar     ç      Ç
wordchars      [d-e]  [D-E]
stringchar     è      È
stringchar     é      É
stringchar     ê      Ê
stringchar     ë      Ë
wordchars      [f-i]  [F-I]
stringchar     í      Í
stringchar     î      Î
stringchar     ï      Ï
wordchars      [j-o]  [J-O]
stringchar     ó      Ó
stringchar     ô      Ô
stringchar     ö      Ö
wordchars      [p-u]  [P-U]
stringchar     ù      Ù
stringchar     û      Û
stringchar     ü      Ü
wordchars      [v-z]  [V-Z]
```

La liste ci-dessus énumère les caractères qui seront considérés par le système comme étant les lettres susceptibles de former des mots du français. Les caractères sont classés en trois catégories :

boundarychars : il s’agit de caractères qui sont considérés comme des lettres lorsqu’ils sont précédés et suivis par des *wordchars* ou des *stringchars* ; en fin ou en début de mot, ils sont ignorés (d’où la notion de « frontière » — *boundary* — dans leur dénomination) ; le tiret est là pour les mots à trait d’union, l’apostrophe pour l’élision (et quelques mots comme *aujourd’hui*), le point pour les URL et les adresses électroniques, l’arobas (@) pour les adresses électroniques ;

wordchars : il s’agit de caractères alphabétiques standards (dont le code décimal est compris entre 0 et 127) ; on indique la minuscule et la majuscule correspondante ; il est possible de préciser plusieurs lettres à la fois (la syntaxe est évidente ; je ne la commente pas) ;

stringchar : il s’agit de caractères (dont le code décimal est compris entre 128 et

255) susceptibles d'être codés par des commandes (*i.e.*, les codes $\text{T}_{\text{E}}\text{X}$) dans le fichier à vérifier (voir la section suivante pour les détails); certains de ces caractères n'existent en principe pas en français: ils sont là pour quelques noms propres d'origine étrangère.

6.1.3 Les caractères $\text{T}_{\text{E}}\text{X}$

```
#
# TeX
#
altstringtype "tex" "TeX" ".tex" ".bib"

altstringchar  \\'a           à
altstringchar  \\'A           À
altstringchar  \\'a           á
altstringchar  \\'A           Á
altstringchar  \\'^a         â
altstringchar  \\'^A         Â
altstringchar  "\\c c"       ç
altstringchar  "\\c C"       Ç
altstringchar  \\'e           è
altstringchar  \\'E           È
altstringchar  \\'e           é
altstringchar  \\'E           É
altstringchar  \\'^e         ê
altstringchar  \\'^E         Ê
altstringchar  \\'"e         ë
altstringchar  \\'"E         Ë
altstringchar  \\'\\i        í
altstringchar  \\'\\I        Í
altstringchar  \\'^\\i       î
altstringchar  \\'^\\I       Î
altstringchar  \\'"\\i       ï
altstringchar  \\'"\\I       Ï
altstringchar  \\'o           ó
altstringchar  \\'O           Ó
altstringchar  \\'^o         ô
altstringchar  \\'^O         Ô
altstringchar  \\'"o         ö
altstringchar  \\'"O         Ö
altstringchar  \\'u           ù
altstringchar  \\'U           Û
altstringchar  \\'^u         û
altstringchar  \\'^U         Û
altstringchar  \\'"u         ü
altstringchar  \\'"U         Ü
```

Cette section est absente du fichier `francais-TeX8b.aff` pour les raisons ex-

posées à la section 2.2.5. Chaque ligne indique la correspondance entre un « *string-char* » et la commande T_EX qui permet de l’obtenir.

6.2 Description des drapeaux

Cette section présente les différents drapeaux qui gèrent les règles d’affixation. Pour une introduction sommaire à la syntaxe de ces dernières reportez-vous à la section 4.4. Pour une présentation plus exhaustive, voyez mon mémoire (Pythoud 1996).

Le fichier `francais.aff` contient quelques indications sur la répartition des drapeaux :

```
# Les affixes dont le drapeau est une minuscule concernent, le plus souvent,
# les verbes ; les affixes dont le drapeau est une majuscule concernent
# essentiellement noms et adjectifs, sauf s’ils renvoient à certains
# préfixes.
```

```
# Drapeau      A B C D E F G H I J K L M N O P Q R S T U V W X Y Z
# Préfixe      * * * * *      * * * * * * * * * *      * * * *
# Suffixe      * * * * *      * * * * *      * * * *
# Libre        * * * * *      * * * * *      * * * * *      # #
```

```
# Drapeau      a b c d e f g h i j k l m n o p q r s t u v w x y z
# Préfixe      * * * * *      * * * * *      * * * * *      * *
# Suffixe      * * * * *      * * * * *      * * * * *      * * * * *
# Libre        * * * * *      * * * * *      * * * * *      # # #
```

6.2.1 Les préfixes

La syntaxe des règles de préfixation est un peu différente de celle des règles de suffixation. On opère ici, bien évidemment, sur les premières lettres du lemme et non sur les dernières. De plus, les règles ne peuvent pas être soustractives, c’est-à-dire qu’on ne peut pas retrancher des lettres en début de mot comme on peut le faire en fin de mot.

prefixes

noms et adjectifs (principalement)

```
flag *A:      # préfixe latin 'a(d)-'
               # verbes et non-verbes
               > AD
               > A
               > AC
               > AF
```

```

[L] > AL # exceptions !
[N] > AN # id.
[P] > AP # id.
[R] > AR
[S] > AS
[T] > AT # id.

flag *B: # préfixe latin 'a(d)-'
# non-verbés uniquement
[^BCDFGLMNPQRST] > D'AD
[BDGM] > D'A
[CQ] > D'AC
[F] > D'AF
[L] > D'AL # exceptions !
[N] > D'AN # id.
[P] > D'AP # id.
[R] > D'AR
[S] > D'AS
[T] > D'AT # id.

flag H: # préfixe latin 'a(d)-'
# non-verbés uniquement
[^BCDFGLMNPQRST] > L'AD
[BDGM] > L'A
[CQ] > L'AC
[F] > L'AF
[L] > L'AL # exceptions !
[N] > L'AN # id.
[P] > L'AP # id.
[R] > L'AR
[S] > L'AS
[T] > L'AT # id.

flag *C: # préfixe latin 'con-'
[^LMR] > CON
[L] > COL
[M] > COM
[R] > COR

flag *D: # préfixe 'dé-'
[^aâæèéêïïoôuh] > Dé
[aâæèéêïïoôuh] > DéS

flag *N: # préfixe latin 'dis-'
[^F] > DIS
[F] > DIF

flag *E: # préfixe latin 'ex-'
# verbés et non-verbés

```

```

    [^FLMNRV] > EX # exceptions pour C,P,T !
    [LMNRV] > é
    [F] > EF

flag *Y: # préfixe latin 'ex-'
          # non-verbés uniquement
    [^FLMNRV] > D'EX # exceptions pour C,P,T !
    [LMNRV] > D'é
    [F] > D'EF

flag Z: # préfixe latin 'ex-'
         # non-verbés uniquement
    [^FLMNRV] > L'EX # exceptions pour C,P,T !
    [LMNRV] > L'é
    [F] > L'EF

flag *I: # les deux préfixes latins 'in-'
          # (privatif et locatif/inchoatif)
          # verbés et non-verbés
    [^BLMPR] > IN
    [L] > IL
    [BMP] > IM
    [R] > IR

flag *J: # préfixes latins 'in-'
          # non-verbés uniquement
    [^BLMPR] > D'IN
    [L] > D'IL
    [BMP] > D'IM
    [IR] > D'IR

flag K: # préfixes latins 'in-'
         # non-verbés uniquement
    [^BLMPR] > L'IN
    [L] > L'IL
    [BMP] > L'IM
    [IR] > L'IR

flag *P: # préfixe 'pré-'
          . > PRé # exceptions !

flag *R: # préfixes 're-' et 'ré-'
          # ('é' devant V et 'e' devant C)
    [^aââeèèééiifoûhs] > RE
    [aââeèèééiifoûh] > Ré
    [S] > RES

flag *O: # préfixes 're-' et 'ré-'
          # ('é' devant C et élision devant V)

```

```

[˘aàâeèéêïïoðuh] > Ré
[aàâeèéêïïoðuh] > R

flag *W: # préfixe latin 'sub-'
[˘CFP] > SUB
[C] > SUC
[F] > SUF
[P] > SUP

# Selon le nombre de drapeaux restant, on traitera également des
# préfixes 'sur' et 'sous'

flag L:
[aàâeèéêïïoðuh] > L'

flag *M:
[aàâeèéêïïoðuh] > D'

flag *Q:
[aàâeèéêïïoðuh] > QU'
[aàâeèéêïïoðuh] > QUOIQU'
[aàâeèéêïïoðuh] > PUISQU'
[aàâeèéêïïoðuh] > LORSQU'

flag *U:
[aàâeèéêïïoðuh] > JUSQU'

# verbes (principalement)

flag j: # pour accompagner le système à 5 lemmes
[aàâeèéêïïoðuh] > J'

flag *n: # élision de la négation et de 'que'
[aàâeèéêïïoðuh] > N'
[aàâeèéêïïoðuh] > QU'

flag *m: # élision du pronom 'me'
[aàâeèéêïïoðuh] > M' # tu m'aides

flag *t: # élision du pronom 'te'
[aàâeèéêïïoðuh] > T' # je t'énerve
# pose problème : vous t'énervez ???
# moins ennuyeux que 'l'éléphants' cpdt
# et peu probable

flag *l: # élision d'un pronom complément d'objet
# direct uniquement, à n'employer que pour

```

```

# les verbes transitifs directs et surtout
# jamais pour les noms ou adj. (utiliser L)
[aââeèéëïïoôuh] > L' # je l'attends

flag s: # élision du pronom réfléchi 'se'
# à employer uniquement devant l'infinitif
# (sinon avec parcimonie)
[aââeèéëïïoôuh] > S' # s'attendre

flag *a: # préfixe latin 'a(d)-'
# verbes transitifs uniquement
[^BCDFGLMNPQRST] > L'AD
[BDGM] > L'A
[CQ] > L'AC
[F] > L'AF
[L] > L'AL # exceptions !
[N] > L'AN # id.
[P] > L'AP # id.
[R] > L'AR
[S] > L'AS
[T] > L'AT # id.
[^BCDFGLMNPQRST] > M'AD
[BDGM] > M'A
[CQ] > M'AC
[F] > M'AF
[L] > M'AL # exceptions !
[N] > M'AN # id.
[P] > M'AP # id.
[R] > M'AR
[S] > M'AS
[T] > M'AT # id.
[^BCDFGLMNPQRST] > T'AD
[BDGM] > T'A
[CQ] > T'AC
[F] > T'AF
[L] > T'AL # exceptions !
[N] > T'AN # id.
[P] > T'AP # id.
[R] > T'AR
[S] > T'AS
[T] > T'AT # id.

flag b: # préfixe latin 'a(d)-'
# verbes uniquement
[^BCDFGLMNPQRST] > J'AD
[BDGM] > J'A
[CQ] > J'AC
[F] > J'AF

```

```

[L] > J'AL # exceptions !
[N] > J'AN # id.
[P] > J'AP # id.
[R] > J'AR
[S] > J'AS
[T] > J'AT # id.

flag *e: # préfixe latin 'ex-'
          # verbes transitifs uniquement
[^FLMNRV] > L'EX # exceptions pour C,P,T !
[LMNRV] > L'é
[F] > L'EF
[^FLMNRV] > M'EX # exceptions pour C,P,T !
[LMNRV] > M'é
[F] > M'EF
[^FLMNRV] > T'EX # exceptions pour C,P,T !
[LMNRV] > T'é
[F] > T'EF

flag f: # préfixe latin 'ex-'
         # verbes uniquement
[^FLMNRV] > J'EX # exceptions pour C,P,T !
[LMNRV] > J'é
[F] > J'EF

flag *i: # préfixes latins 'in-'
         # verbes transitifs uniquement
[^BLMPR] > L'IN
[L] > L'IL
[BMP] > L'IM
[R] > L'IR
[^BLMPR] > M'IN
[L] > M'IL
[BMP] > M'IM
[R] > M'IR
[^BLMPR] > T'IN
[L] > T'IL
[BMP] > T'IM
[R] > T'IR

flag k: # préfixes latins 'in-'
        # verbes uniquement
[^BLMPR] > J'IN
[L] > J'IL
[BMP] > J'IM
[R] > J'IR

```

6.2.2 Les suffixes (noms et adjectifs)

Les règles qui suivent concernent les noms et les adjectifs. Elles permettent de régler l’accord en genre et en nombre. Pour une description du fonctionnement du système, reportez-vous à la section 7.1.

suffixes

noms et adjectifs (principalement)

```

flag *F:                # gestion du masculin et du pluriel à
                        # partir du féminin (sauf mots en -eur)
# règle générale du féminin pluriel
[^SXZ]                >      S

# Voyelle + e/è :
é E                    >      -éE,é  # = -E (employée)
I E                    >      -IE,I   # = -E (endolorie)
i E                    >      -iE,i   # = -E (ouïe)
[^GQ] U E             >      -UE,U   # = -E (bienvenue)
[^GQ] U E             >      -E,S
[éIi] E               >      -E,S
U è                    >      -Uè,U   # = -è (aiguë)
U è                    >      -è,S

# Finale -(c)que
[^C] Q U E           >      -QUE,C  # caduque
[^C] Q U E           >      -QUE,CS # caduque
C Q U E              >      -CQUE,C # = -QUE (grecque)
C Q U E              >      -QUE,S

# Finale -gue
G U E                >      -GUE,G  # = -UE (longue)
G U E                >      -UE,S   # = -UE (longue)

# Finale -che
[^è] C H E          >      -CHE,C  # = -HE (blanche)
[^è] C H E          >      -HE,S
è C H E              >      -èCHE,EC # sèche
è C H E              >      -èCHE,ECS

# Consonne c + e
C E                  >      -CE,X   # douce

# Consonne d + e
D E                  >      -DE,D   # = -E (lourde)
D E                  >      -E,S

# Consonnes ff + e

```

F F E	>	-FFE,F # = -FE (cheffe)
F F E	>	-FE,S
# Consonne(s) (l)l + e		
[EIU] L L E	>	-LLE,L # = -LE (éventuelle, nulle
[EIU] L L E	>	-LE,S # pareille, gentille)
A L E	>	-LE,L # = -E (buccale)
A L E	>	-LE,UX
è L E	>	-èLE,EL # fidèle
è L E	>	-èLE,ELS
[IOUâ] L E	>	-LE,L # = -LE (subtile, espagnole,
[IOUâ] L E	>	-E,S # veule, saoule)
# Consonne(s) (n)n + e		
[AEO] N N E	>	-NNE,N # = -NE (paysanne, citoyenne,
[AEO] N N E	>	-NE,S # canadienne, wallonne)
[AIU] N E	>	-NE,N # = -NE (américaine, voisine,
[AIU] N E	>	-E,S # opportune, romane)
I G N E	>	-GNE,N # bénigne, maligne
I G N E	>	-GNE,NS
# Consonne r + e		
è R E	>	-èRE,ER # dernière
è R E	>	-èRE,ERS
[IUâY] R E	>	-RE,R # = -E (supérieure, noire)
[IUâY] R E	>	-E,S
# Consonne(s) (s)s + e		
[EO] U S E	>	-SE,X # heureuse, jalouse
		# exception : andalouse/andalou
[^US] S E	>	-SE,S # = -E (précise, québécoise)
[^EO] U S E	>	-SE,S # recluse
O U S S E	>	-SSE,X # rousse
[^U] S S E	>	-SSE,S # = -SE (basse, épaisse,
		# métisse)
# Consonne(s) (t)t + e		
T T E	>	-TTE,T # coquette, sotté
T T E	>	-TE,S
[^èT] T E	>	-TE,T # idioté
[^èT] T E	>	-E,S
è T E	>	-èTE,ET # discrète
è T E	>	-èTE,ETS
# Consonne v + e		
è V E	>	-èVE,EF # brève
è V E	>	-èVE,EFS
[IiU] V E	>	-VE,F # vive, naïve, veuve
[IiU] V E	>	-VE,FS

```

flag *G:                                # mots en -eur et en -al qui font le pluriel
                                         # en -als (pas en -aux -> drapeau F)
    E                                    >          S

    # prometteuse
    E U S E                              >          -SE,R
    E U S E                              >          -SE,RS

    # opératrice, ambassadrice
    [DT] R I C E                          >          -RICE,EUR
    [DT] R I C E                          >          -RICE,EURS

    # chasseresse
    E R E S S E                          >          -RESSE,UR
    E R E S S E                          >          -RESSE,URS

    # doctoresse
    O R E S S E                          >          -ORESSE,EUR
    O R E S S E                          >          -ORESSE,EURS

    # australe
    A L E                                 >          -LE,L   # = -E
    A L E                                 >          -E,S

flag *S:                                # pluriels en 's'
    [^SXZ]                                >          S

flag *X:                                # pluriels en 'x'
    [AEO] U                               >          X
    A L                                   >          -L,UX
    A I L                                >          -IL,UX

```

6.2.3 Les suffixes (verbes)

Les règles qui suivent constituent l’implantation du système « à cinq lemmes », décrit à la section 7.2.

verbes (essentiellement)

Lemme = 1ère pers. sing. de l’IND PR

Pour : 1ère, 2e, 3e pers. sing de l’IND PR, 3e pers. plur. de l’IND PR

1ère, 2e, 3e pers. sing du SUBJ PR, 3e pers. plur. du SUB PR

flag *u:

```

    E                                    >          S
    E                                    >          NT

```

[^TD] S	>	-S,T	
T S	>	-TS,T	# = -S
D S	>	-DS,D	# = -S
[^O] I S	>	SENT	
[^O] I S	>	SE	
[^O] I S	>	SES	
I E N S	>	-S,NENT	
I E N S	>	-S,NE	
I E N S	>	-S,NES	
I E R S	>	-ERS,èRENT	
I E R S	>	-ERS,èRE	
I E R S	>	-ERS,èRES	
[^I] E N S	>	-S,TENT	
[^I] E N S	>	-S,TE	
[^I] E N S	>	-S,TES	
[TD] S	>	-S,ENT	
[TD] S	>	-S,E	
[TD] S	>	-S,ES	
[AEO] I N S	>	-NS,GNE	
[AEO] I N S	>	-NS,GNES	
[AEO] I N S	>	-NS,GNENT	
[^E] R S	>	-S,ENT	
[^E] R S	>	-S,E	
[^E] R S	>	-S,ES	
O I S	>	-S,VENT	# 'recevoir', 'boire',
O I S	>	-S,VE	# 'devoir'
O I S	>	-S,VES	
O î S	>	-OîS,OISSENT	# 'croître'
O î S	>	-OîS,OISSE	
O î S	>	-OîS,OISSES	
[PU] S	>	-S,ENT	# P pour la famille
[PU] S	>	-S,E	# de rompre
[PU] S	>	-S,ES	
[FVP] [AE] U X	>	-X,T	# sauf 'prévaloir'
V A U X	>	-UX,LENT	
V A U X	>	-UX,ILLE	
V A U X	>	-UX,ILLES	
V A U X	>	-UX,ILLENT	
V E U X	>	-X,LENT	# 'vouloir'
V E U X	>	-X,ILLE	
V E U X	>	-X,ILLES	

```

V E U X          >      -X,ILLENT

                                # verbe 'fallir' (désuet)
F A U X          >      -UX,ILLENT
F A U X          >      -UX,ILLE
F A U X          >      -UX,ILLES

                                # verbe 'pouvoir'
P E U X          >      -X,VENT
P E U X          >      -EUX,UIS
P E U X          >      -EUX,UISSE
P E U X          >      -EUX,UISSES
P E U X          >      -EUX,UISSENT

flag *g:         # Verbes de la famille de 'prendre'
                 # IND et SUBJ PR
P R E N D S      >      -DS,D   # = -S
P R E N D S      >      -DS,NE
P R E N D S      >      -DS,NES
P R E N D S      >      -DS,NENT

                 # Verbes de la famille de 'asseoir'
S I E D S        >      -DS,D   # = -S
S I E D S        >      -IEDS,EYENT
S I E D S        >      -IEDS,EYE
S I E D S        >      -IEDS,EYES

                 # Verbes de la famille de 'coudre'
C O U D S        >      -DS,D   # = -S
C O U D S        >      -DS,SENT
C O U D S        >      -DS,SE
C O U D S        >      -DS,SES

                 # Verbes de la famille de 'moudre'
M O U D S        >      -DS,D   # = -S
M O U D S        >      -DS,LENT
M O U D S        >      -DS,LE
M O U D S        >      -DS,LES

                 # Verbes de la famille de 'mettre' et de 'battre'
                 # IND et SUBJ PR
[AE] T S         >      -TS,T   # = -S
[AE] T S         >      -S,TE
[AE] T S         >      -S,TES
[AE] T S         >      -S,TENT

                 # Verbes de la famille de 'faire'
                 # IND et SUBJ PR
F A I S          >      -S,T

```

```

F A I S > -AIS,ONT
F A I S > -IS,SSE
F A I S > -IS,SSES
F A I S > -IS,SSIONS
F A I S > -IS,SSIEZ
F A I S > -IS,SSENT

# Verbes de la famille 'connaître', 'paraître', 'naître'
# IND et SUBJ PR
[~T] [NR] A I S > -IS,îT
[~T] [NR] A I S > SENT
[~T] [NR] A I S > SE
[~T] [NR] A I S > SES

# Verbes de la famille de 'traire'
T R A I S > -S,T
T R A I S > -S,ENT
T R A I S > -S,E
T R A I S > -S,ES

# Verbes de la famille de 'plaire'
P L A I S > -IS,îT
P L A I S > ENT
P L A I S > E
P L A I S > ES

# Verbe 'taire'
T A I S > -S,T
T A I S > ENT
T A I S > E
T A I S > ES

# 'paître' et 'repaître'
P A I S > -AIS,AîT
P A I S > SENT
P A I S > SE
P A I S > SES

# Verbe 'hair'
H A I S > -S,T
H A I S > -IS,ïSSENT
H A I S > -IS,ïSSE
H A I S > -IS,ïSSES

# Verbe 'savoir'
S A I S > -S,T
S A I S > -IS,VENT
S A I S > -IS,CHE
S A I S > -IS,CHES

```

```

S A I S                >      -IS,CHENT

      # Verbes en -SCRIRE + 'écrire' et 'décrire'
      # IND et SUBJ PR
[és] C R I S          >      -S,T
[és] C R I S          >      -S,VENT
[és] C R I S          >      -S,VE
[és] C R I S          >      -S,VES

      # Verbes de la famille de 'lire', 'dire' et 'confire'
[LDF] I S              >      -S,T
[LDF] I S              >      ENT
[LDF] I S              >      E
[LDF] I S              >      ES

      # Verbes de la famille de 'cuire'
U I S                  >      -S,T
U I S                  >      ENT
U I S                  >      E
U I S                  >      ES

      # Verbes de la famille de 'voir', 'croire', 'pouvoir'
O I S                  >      -S,T
O I S                  >      -S,ENT
O I S                  >      -S,E
O I S                  >      -S,ES

      # 'prévaloir' uniquement
V A U X                >      -X,T
V A U X                >      -UX,LENT
V A U X                >      -UX,LE
V A U X                >      -UX,LES

      # 'bouillir' et 'débouillir'
B O U S                >      -S,T
B O U S                >      -S,ILLENT
B O U S                >      -S,ILLE
B O U S                >      -S,ILLES

      # 'absoudre', 'dissoudre' et 'résoudre'
S O U S                >      -S,T
S O U S                >      -US,LVENT
S O U S                >      -US,LVE
S O U S                >      -US,LVES

      # 'dormir' et 'endormir'
D O R S                >      -S,T
D O R S                >      -S,MENT
D O R S                >      -S,ME

```

```

D O R S          >      -S,MES

      # Verbes de la famille de 'servir'
S E R S          >      -S,T
S E R S          >      -S,VENT
S E R S          >      -S,VE
S E R S          >      -S,VES

      # Verbes de la famille de 'mouvoir'
M E U S          >      -S,T
M E U S          >      -S,VENT
M E U S          >      -S,VE
M E U S          >      -S,VES

      # 'vaincre' et 'convaincre'
V A I N C S      >      -CS,C   # = -S
V A I N C S      >      -CS,QUENT
V A I N C S      >      -CS,QUE
V A I N C S      >      -CS,QUES

      # verbes de la famille de 'sortir'
S O R S          >      -S, T
S O R S          >      -S, TE
S O R S          >      -S, TES
S O R S          >      -S, TENT

# Lemme = 1ère pers. sing. de l'IND IMP
# Pour : toutes les pers. de l'IND IMP, les 1ère et 2e pers. plur. de l'IND PR,
#       les 1ère et 2e pers. plur. du SUBJ PR, le participe présent et la
#       grande majorité des adjectifs verbaux en -ant
flag *v:
A I S            >      -S,T
[˘çE] A I S      >      -AIS,IONS
[˘çE] A I S      >      -AIS,IEZ
ç A I S          >      -çAIS,CIONS   # 'placer'
ç A I S          >      -çAIS,CIEZ
E A I S          >      -E AIS,IONS   # 'manger'
E A I S          >      -E AIS,IEZ
A I S            >      -S,ENT
A I S            >      -AIS,ONS
[˘çE] A I S      >      -AIS,EZ
ç A I S          >      -çAIS,CEZ
E A I S          >      -AIS,Z
A I S            >      -IS,NT
flag *c:
A I S            >      -IS,NTS
A I S            >      -IS,NTE
A I S            >      -IS,NTES

```


[IUâ] S	>	SES	# idem
[IUâ] S	>	SIONS	# idem
[IUâ] S	>	SIEZ	# idem
[IUâ] S	>	SENT	# idem

U S	>	-US,âMES
U S	>	-US,âTES
U S	>	-US,âT

û S	>	-S,MES
û S	>	-S,TES

		# V du 3e gp
I N S	>	-INS,âNMES
I N S	>	-INS,âNTES
I N S	>	SE
I N S	>	SES
I N S	>	-INS,âNT
I N S	>	SIONS
I N S	>	SIEZ
I N S	>	SENT

H A î S	>	-S,MES
H A î S	>	-S,TES

Lemme = 1ère pers. sing. du FUT

Pour : toutes les pers. du FUT et du COND PR

flag *x:

A I	>	-I,S
A I	>	-AI,A # = -I
A I	>	-AI,ONS
A I	>	-AI,EZ
A I	>	-AI,ONT
A I	>	S
A I	>	T
A I	>	-AI,IONS
A I	>	-AI,IEZ
A I	>	ENT

Lemme = INF

Pour : toutes les formes du participe passé

flag *y:

E R	>	-ER,é
E R	>	-ER,és
E R	>	-ER,éE
E R	>	-ER,éES

		# V du 2e gp et certains du 3e gp
[^0] I R	>	-IR,I # = -R

```

[˘O] I R > -R,S
[˘O] I R > -R,E
[˘O] I R > -R,ES

# V du 3e gp sur le modèle de 'recevoir'
C E V O I R > -CEVOIR,çU
C E V O I R > -CEVOIR,çUS
C E V O I R > -CEVOIR,çUE
C E V O I R > -CEVOIR,çUES

# V du 3e gp sur le modèle de 'asseoir'
S E O I R > -EOIR,IS
S E O I R > -EOIR,ISE
S E O I R > -EOIR,ISES

# V du 3e gp sur le modèle de 'rendre'
# et 'rompre'
[PD] R E > -RE,U
[PD] R E > -RE,US
[PD] R E > -RE,UE
[PD] R E > -RE,UES

# V du 3e gp sur le modèle de 'cuire',
# 'faire' et 'traire'
[AU] I R E > -RE,TS
[AU] I R E > -RE,TE
[AU] I R E > -RE,TES

# V du 3e gp sur le modèle de 'lire'
L I R E > -IRE,U
L I R E > -IRE,UE
L I R E > -IRE,UES

# V du 3e gp sur le modèle de 'dire'
# et 'confire'
[DF] I R E > -RE,TS
[DF] I R E > -RE,TE
[DF] I R E > -RE,TES

# V du 3e gp sur le modèle de 'mettre'
M E T T R E > -ETTRE,IS
M E T T R E > -ETTRE,ISE
M E T T R E > -ETTRE,ISES

# V du 3e gp sur le modèle de 'battre'
B A T T R E > -RE,U
B A T T R E > -RE,UE
B A T T R E > -RE,US
B A T T R E > -RE,UES

```

```

# V du 3e gp sur le modèle de 'connaître',
# et 'paraître'
[NR] A î T R E > -AîTRE,U
[NR] A î T R E > -AîTRE,US
[NR] A î T R E > -AîTRE,UE
[NR] A î T R E > -AîTRE,UES

# V du 3e gp sur le modèle de 'clore'
C L O R E > -RE,S
C L O R E > -RE,SE
C L O R E > -RE,SES

# V du 3e gp : conclure, exclure, reclure
C L U R E > -URE,U # = -RE

# V du 3e gp sur le modèle de 'boire'
# et 'croire'
O I R E > -OIRE,U
O I R E > -OIRE,US
O I R E > -OIRE,UE
O I R E > -OIRE,UES

# 'croître' et 'recroître'
C R O î T R E > -OîTRE,û
C R O î T R E > -OîTRE,US
C R O î T R E > -OîTRE,UE
C R O î T R E > -OîTRE,UES

# 'haïr'
H A ï R > -ïR,ï # = -R
H A ï R > -R,S
H A ï R > -R,E
H A ï R > -R,ES

# 'vaincre' et 'convaincre'
V A I N C R E > -RE,U
V A I N C R E > -RE,US
V A I N C R E > -RE,UE
V A I N C R E > -RE,UES

# 'suivre' et 'poursuivre'
S U I V R E > -RE,I
S U I V R E > -RE,IS
S U I V R E > -RE,IE
S U I V R E > -RE,IES

# 'vivre', 'revivre' et 'survivre'
V I V R E > -IVRE,éCU

```

```

V I V R E > -IVRE,éCUS
V I V R E > -IVRE,éCUE
V I V R E > -IVRE,éCUES

```

flag *z:

```

# V du 3e gp sur le modèle de 'acquérir'
é R I R > -éRIR,IS
é R I R > -éRIR,ISE
é R I R > -éRIR,ISES

```

```

# V du 3e gp sur le modèle de 'couvrir'
V R I R > -RIR,ERT
V R I R > -RIR,ERTS
V R I R > -RIR,ERTE
V R I R > -RIR,ERTES

```

```

# certains V du 3e gp en -ir
[˘O] I R > -IR,U
[˘O] I R > -IR,US
[˘O] I R > -IR,UE
[˘O] I R > -IR,UES

```

```

# V du 3e gp sur le modèle de 'valoir'
# et 'vouloir'
L O I R > -OIR,U
L O I R > -OIR,US
L O I R > -OIR,UE
L O I R > -OIR,UES

```

```

# V du 3e gp sur le modèle de 'peindre',
# 'joindre', 'craindre'
I N D R E > -DRE,T
I N D R E > -DRE,TS
I N D R E > -DRE,TE
I N D R E > -DRE,TES

```

```

# V du 3e gp sur le modèle de 'prendre',
E N D R E > -ENDRE,IS
E N D R E > -ENDRE,ISE
E N D R E > -ENDRE,ISES

```

```

# V du 3e gp sur le modèle de 'coudre'
C O U D R E > -DRE,SU
C O U D R E > -DRE,SUS
C O U D R E > -DRE,SUE
C O U D R E > -DRE,SUES

```

```

# V du 3e gp sur le modèle de 'moudre'
M O U D R E > -DRE,LU

```

```

M O U D R E > -DRE,LUS
M O U D R E > -DRE,LUE
M O U D R E > -DRE,LUES

# V 'naître' et 'renaître'
N A î T R E > -AîTRE,é
N A î T R E > -AîTRE,és
N A î T R E > -AîTRE,éE
N A î T R E > -AîTRE,éES

# écrire/décrire et V en -SCRIRE
[és] C R I R E > -RE,TS
[és] C R I R E > -RE,TE
[és] C R I R E > -RE,TES

# V du 3e gp : inclure, occlure
C L U R E > -RE,SE
C L U R E > -RE,SES

# V de la famille de 'croître', sauf 'croître'
# et 'recroître'
C R O î T R E > -OîTRE,u
C R O î T R E > -OîTRE,us
C R O î T R E > -OîTRE,ue
C R O î T R E > -OîTRE,ues

# Lemmes supplémentaires pour les emplois réflexifs de certains verbes
# des premier et deuxième groupes (concerne uniquement l'élision du
# pronom 'se'
#
# Verbes du 1er groupe : lemmes = s' + 3e pers. sing. IND PR avec drapeau p
# = s' + 3e pers. sing. IND IMP avec drapeau p
#
# Verbes du 2e groupe : lemme = s' + 3e pers. sing. IND IMP avec drapeau q

flag p:
E > NT
E > -E,ÈRENT
E > RA
E > RONT
E > RAIT
E > RAIENT

A I T > -T,ENT
A I T > -AIT,A # = -IT
A I T > -IT,NT
A I T > -AIT,ÂT
A I T > -IT,SSENT

```

```

flag q:
  Ì S S A I T > -ÏSSAIT,IT
  [˘Ï] S S A I T > -SSAIT,T
  S S A I T > -AIT,E
  S S A I T > -AIT,ENT
  S S A I T > -T,ENT
  S S A I T > -IT,NT
  Ì S S A I T > -SSAIT,T
  S S A I T > -SSAIT,RENT
  Ì S S A I T > -ÏSSAIT,ÎT
  I S S A I T > -ISSAIT,ÎT
  S S A I T > -SSAIT,RA
  S S A I T > -SSAIT,RONT
  S S A I T > -SSAIT,RAIT
  S S A I T > -SSAIT,RAIENT

```

Chapitre 7

Ajout de nouveaux dictionnaires

Ce chapitre expose les étapes à suivre pour créer un dictionnaire spécialisé contenant des noms, des adjectifs et des verbes. Celui-ci peut bien entendu aussi contenir des adverbes et d'autres mots invariables. Pour ces derniers, aucune précaution particulière n'est requise, il suffit de les faire figurer dans le dictionnaire tels quels.

En cas de besoin, vous pouvez vous reporter au chapitre 6, mais la lecture de ce dernier n'est pas nécessaire pour comprendre ce qui suit.

7.1 Ajouter des noms et des adjectifs

7.1.1 Le cas trivial

Le cas le plus simple est celui d'un nom ou d'un adjectif épïcène ne commençant pas par une voyelle ou un « h » muet. Il suffit de l'entrer dans le dictionnaire et de lui adjoindre le drapeau **S**¹ s'il prend un simple « s » au pluriel, ou le drapeau **X** dans tous les autres cas. Par exemple :

(7.1) *célébration/S*

(7.2) *bateau/X*

(7.1) permettra de vérifier *célébration* et *célébrations*. (7.2) sera la caution de *bateau* et *bateaux*.

Si le pluriel est identique au singulier, le drapeau n'est pas nécessaire.

7.1.2 Féminin et masculin

Certains noms et adjectifs possèdent des formes féminines et masculines. Dans le cas d'un mot qui ne commence pas par une voyelle ou un « h » muet, il suffit d'entrer la forme au féminin singulier, accompagnée du drapeau **F**, sauf si on a affaire à un

1. La casse du drapeau est importante et donc **S** n'est pas équivalent à **s**).

mot qui finit en *-eur* au masculin (cf. 5.2), auquel cas il faut utiliser le drapeau **G**. Quelques exemples :

(7.3) *nominative*/F

(7.4) *malheureuse*/F

(7.5) *boudeuse*/G

(7.3) représente *nominative*, *nominatives*, *nominatif* et *nominatifs*. (7.4) permet d'obtenir *malheureuse*, *malheureuses* et *malheureux*. (7.5) est équivalent à *boudeuse*, *boudeuses*, *boudeur* et *boudeurs*.

Il faut également utiliser le drapeau **G** pour les mots en *-al* qui font leur pluriel en *-als*. Par contre, on utilisera le drapeau **F** pour ceux qui font leur pluriel en *-aux*. Par exemple (on part toujours du féminin singulier) :

(7.6) *fatale*/G

(7.7) *dentale*/F

(7.6) engendrera *fatale*, *fatales*, *fatal* et *fatals*. (7.7) permet d'obtenir *dentale*, *dentales*, *dental*, *dentaux*.

Les drapeaux **F** et **G** couvrent la quasi-totalité des correspondances entre masculin et féminin du français : vous pouvez en principe vous y fier et, dans le doute, vérifier ce qui est produit par le drapeau (cf. 7.4 pour ce faire). Seuls les cas les plus irréguliers ne sont pas traités : *andalou/andalouse*, *beau/belle*, *favori/favorite*, *fou/folle*, *tiers/tierce*. Il y a peu de chance qu'un dictionnaire spécialisé doive contenir des mots présentant de telles caractéristiques.

7.1.3 Le problème de l'apostrophe

Comment gérer l'élision dans le cas d'un mot commençant par une voyelle ou un « h » muet ? Il faut distinguer deux cas :

- le mot est strictement masculin ou féminin, ou possède une forme identique au masculin et au féminin ;
- le mot possède des formes différentes au masculin et au féminin.

Dans le premier cas, on entre le mot au singulier dans le dictionnaire, accompagné des drapeaux **LMS** ou **LMX** (selon le pluriel du mot, cf. 7.1.1). Par exemple :

(7.8) *éclaircissement*/LMS

qui représente *éclaircissement*, *éclaircissements*, *l'éclaircissement*, *d'éclaircissement* et *d'éclaircissements*.

Dans le second cas, l'opération est un peu plus délicate : il faut entrer la forme du féminin singulier du mot, accompagnée des drapeaux **LMF** ou **LMG** (cf. 7.1.2), et le masculin singulier précédé de « l' ». Par exemple :

(7.9) *ancienne*/LMF

(7.10) l'*ancien*

où (7.9) engendre *ancienne*, *anciennes*, *ancien*, *anciens*, *l'ancienne*, *d'ancienne*, *d'anciennes*, *d'ancien* et *d'anciens*, mais pas *l'ancien* qui doit être séparément spécifié. Pourquoi cela? Regardons le « code » des drapeaux L et M:

flag L:
[aââeèèéêiîoôuh] > L'

flag *M:
[aââeèèéêiîoôuh] > D'

Ces deux règles indiquent que tout mot commençant par une voyelle (accentuée ou non) ou par un « h » muet peut être précédé de *l'* ou de *d'*. La différence entre elles se situe au niveau de l'astérisque (*) qui précède M mais pas L. Cet astérisque signifie que le préfixe est libre de se combiner avec n'importe quel suffixe également doté de l'astérisque (et réciproquement). Ceci a pour conséquence que (7.11) est autorisé mais pas (7.12):

(7.11) *d' + ancienne + s*

(7.12) *l' + ancienne + s*

Cette restriction interdit au système d'accepter ou de produire les formes erronées **l'anciens* et **l'anciennes*, mais nécessite alors que la forme correcte *l'ancien* soit spécifiée séparément. Le problème ne se pose pas pour le drapeau M.

Vous avez sans doute remarqué que je n'opère pas de distinction entre noms et adjectifs pour l'attribution des drapeaux. On pourrait cependant se demander si les drapeaux L et M, représentant respectivement l'élision des déterminants *le* et *de*, doivent vraiment être attribués aux adjectifs. Il ne faut pas oublier qu'en français la plupart des adjectifs peuvent se trouver antéposés au nom qu'ils qualifient, comme dans (7.13).

(7.13) L'ignoble individu.

Certains adjectifs, appartenant à des catégories bien précises, ne sont toutefois jamais antéposés. Pour ceux-ci, et seulement pour eux, on peut se passer des drapeaux L et M. Les catégories d'adjectifs concernées sont les suivantes:

- les adjectifs de couleur;
- les adjectifs d'appartenance (nationalité, origine géographique, religion, etc.);
- les adjectifs de forme (*oval* par exemple);
- les adjectifs verbaux formés sur le participe passé (cf. en fin de section 7.2.3, page 71).

7.1.4 Utilisation des autres préfixes

Le fichier `français.aff` de Francais-GUTenberg permet de gérer les préfixes latins suivants: *a(d)-*, *con-*, *dé-*, *dis-*, *ex-*, *in-*, *pré-*, *re-* et *sub-*. Leur utilisation peut

faciliter et accélérer la composition d'un dictionnaire.

7.2 Ajouter des verbes

Dans les pages qui suivent, l'adjonction de verbes est examinée. Pour simplifier les explications, les développements qui suivent partent du principe que les verbes à ajouter appartiennent aux premier et deuxième groupes de conjugaisons². Comme *Français-GUTenberg* contient en principe tous les verbes du troisième groupe, il est peu probable que vous ayez à vous soucier des verbes du troisième groupe (sinon, reportez-vous à la section 7.2.7).

7.2.1 Éléments de base

Plaçons-nous d'abord dans le meilleur cas de figure, où nous avons affaire à un verbe qui ne commence ni par une voyelle, ni par un « h » muet, pour éviter le problème de l'élision (cf. 7.2.3). Il faudra entrer les premières personnes des temps de l'indicatif suivants : présent, imparfait, passé simple et futur, accompagnés respectivement des drapeaux u, v, w et x. Prenons l'exemple du verbe *charmer* :

(7.14) charme/u

(7.15) charmais/v

(7.16) charmai/w

(7.17) charmerai/x

Ces quatre lemmes permettent d'engendrer toutes les formes du verbe *charmer* à l'exception de l'infinitif et du participe passé. Voici les règles, extraites de *français.aff*, qui leur sont appliquées (ne figurent que les lignes pertinentes pour le verbe *charmer*) :

flag *u:

E	>	S
E	>	NT

flag *v:

A I S	>	-S,T
[^çE] A I S	>	-AIS, IONS
[^çE] A I S	>	-AIS, IEZ

2. Les verbes du premier groupe sont les verbes en *-er* à l'exception de *aller*. Les verbes du deuxième groupe sont les verbes en *-ir* qui font leur première personne du présent en *-is* et leur participe présent en *-issant*. Dans tous les autres cas, le verbe appartient au troisième groupe. (Pour un linguiste, cette classification est peu satisfaisante du point de vue scientifique, mais ce n'est pas ce qui nous intéresse ici.)

A I S	>	-S,ENT
A I S	>	-AIS,ONS
[^çE] A I S	>	-AIS,EZ
A I S	>	-IS,NT
flag *w:		
A I	>	-I,S
A I	>	-AI,A # = -I
A I	>	-AI,âMES
A I	>	-AI,âTES
[^çE] A I	>	-AI,èRENT
A I	>	-I,SSE
A I	>	-I,SSES
A I	>	-AI,âT
A I	>	-I,SSIONS
A I	>	-I,SSIEZ
A I	>	-I,SSENT
flag *x:		
A I	>	-I,S
A I	>	-AI,A # = -I
A I	>	-AI,ONS
A I	>	-AI,EZ
A I	>	-AI,ONT
A I	>	S
A I	>	T
A I	>	-AI,IONS
A I	>	-AI,IEZ
A I	>	ENT

Leur interprétation est triviale et je la laisse comme exercice au lecteur.

7.2.2 Infinitif, participe passé et adjectif verbal formé sur le participe présent

Concernant le participe passé, il faut distinguer deux cas :

- il ne s'accorde jamais (ce qui concerne principalement les verbes intransitifs se conjuguant avec l'auxiliaire *avoir*), par exemple *bifurquer* ou *frissonner* ;
- il est susceptible de s'accorder en genre et nombre (ce qui est le cas de la majorité des verbes français).

Dans le premier cas, on entre simplement l'infinitif et le participe passé dans le dictionnaire. Par exemple :

(7.18) *bifurquer*

(7.19) *bifurqué*

Dans le second cas, on entre l'infinitif suivi du drapeau *y*. Par exemple :

(7.20) *charmer/y*

Le drapeau *y* se compose, entre autres, des règles suivantes :

flag **y*:

E R	>	-ER, é
E R	>	-ER, éS
E R	>	-ER, éE
E R	>	-ER, éES

Une question du même type se pose pour le participe présent qui est engendré par le drapeau *v* (*charmant* est créé à partir de (7.15) dans notre exemple). En effet, bien souvent, au participe présent correspond un adjectif verbal (et parfois un substantif) susceptible de s'accorder en genre et en nombre (*charmants, charmante, charmantes*). Quand c'est le cas, on peut ajouter le drapeau *c* au lemme à l'imparfait. Dans notre exemple, (7.15) deviendrait (7.21).

(7.21) *charmais/vc*

Les règles du drapeau *c* sont :

flag * <i>c</i> :		# adjectif verbal accordé
A I S	>	-IS, NTS
A I S	>	-IS, NTE
A I S	>	-IS, NTES

Ce dernier mécanisme est très pratique, mais il faut savoir que l'orthographe française recèle quelques pièges dans ce domaine : il faut dans certains cas s'abstenir d'utiliser le drapeau *c*. En effet, les graphies du participe présent et de l'adjectif verbal peuvent parfois être différentes. Quelques exemples :

(7.22) a. Pierre, *négligeant* toute prudence, appuya sur le champignon.

b. Pierre se montre toujours très *négligent* au volant.

(7.23) a. Paul asséna son dernier argument à Jean, le *convainquant* ainsi du bien-fondé de son point de vue.

b. Paul a été très *convaincant* face à Jean.

(7.24) a. Le journaliste continua de poser ses questions, *fatiguant* les coureurs.

b. Les questions du journaliste étaient *fatigantes*.

Au vu de ce qui précède, il faut tenir compte des points suivants :

– à certains participes présents en *-quant* correspondent des adjectifs en *-cant* ;

- à certains participes présents en *-quant* correspondent des adjectifs en *-gant* ;
- à certains autres participes présents (en *-ant*) correspondent des adjectifs verbaux en *-ent*.

Grevisse (1988) donne au §1887 une liste d'adjectifs « verbaux » qui ne se forment pas sur le participe présent.

7.2.3 Le problème de l'élision

Lorsqu'on est confronté à un verbe commençant par une voyelle ou un « h » muet, la situation est un peu plus compliquée et il faut être extrêmement prudent par rapport à deux choses :

- les types d'élision possibles devant le verbe considéré (« j' », « n' », « m' », « t' », « s' » et « l' ») ;
- le traitement d'un éventuel adjectif verbal et des participes passés (leur ajouter des drapeaux sans réfléchir entraînerait des effets de bord catastrophiques pour la validité du dictionnaire).

Les pronoms susceptibles de s'élider relèvent de différentes catégories :

- le pronom *je* est toujours susceptible de s'élider en « j' » ;
- il en va de même pour la négation : *ne* est toujours susceptible de s'élider en « n' » ;
- certains pronoms qui font fonction d'objet direct peuvent s'élider (*me*, *te*, *le*) ;
- certains pronoms qui font fonction d'objet indirect également (*me*, *te*) ;
- de même que certains pronoms réfléchis (*me*, *te*, *se*).

<i>Sujet</i>	<i>Obj. dir.</i>	<i>Obj. indir.</i>	<i>Réflex.</i>
je → j'	me → m'	me → m'	me → m'
tu → ∅	te → t'	te → t'	te → t'
il/elle/on → ∅	le → l'	lui → ∅	se → s'
nous → ∅	nous → ∅	nous → ∅	nous → ∅
vous → ∅	vous → ∅	vous → ∅	vous → ∅
ils/elles → ∅	les → ∅	leur → ∅	se → s'
Négation : ne → n'			

TABLE 7.1 – *Les pronoms du français et l'élision*

La table 7.1 synthétise les points ci-dessus. On observe un grand syncrétisme entre les différentes formes, ce qui n'est pas de nature à nous faciliter la tâche. Aux différentes formes d'élision correspondent les drapeaux j, n, m, t, l, s :

flag j: # pour accompagner le système à 5 lemmes
[aââèèèiîoûh] > J'

flag *n:	# élision de la négation
[aàâeèèéiîoδuh]	> N'
flag *m:	# élision du pronom 'me'
[aàâeèèéiîoδuh]	> M' # tu m'aides
flag *t:	# élision du pronom 'te'
[aàâeèèéiîoδuh]	> T' # je t'énerve
flag *l:	
[aàâeèèéiîoδuh]	> L' # je l'attends
flag s:	
[aàâeèèéiîoδuh]	> S' # s'attendre

Il ne faut *jamaï*s combiner ces drapeaux avec *c* et *y* (voir plus bas pour les détails). L'élision de *se* n'est pas prise en charge par le système (sauf pour l'infinitif, par le biais du drapeau *s*): voir les sections 7.2.5 et 7.2.6 pour ce qui concerne les emplois réflexifs et les verbes essentiellement pronominaux. Nous ne nous en occuperons plus dans cette section.

Quels drapeaux employer et dans quels cas? Le drapeau *j* doit, dans tous les cas, précéder les quatre premiers lemmes du verbe, mais pas l'infinitif. Tout comme le drapeau *L* (cf. 7.1.3), il n'est pas précédé de l'astérisque (*), ce qui l'empêche de se combiner avec un suffixe pour donner des résultats aberrants (cependant, comme pour le drapeau *L*, cela présente quelques inconvénients — cf. 7.2.4).

Le drapeau *n* concerne tous les lemmes du verbe.

Si le verbe accepte un objet direct, chaque lemme devrait être en principe accompagné des drapeaux *m*, *t* et *l*. En fait, il faut bien réfléchir au sens du verbe: si l'objet direct ne peut pas être une personne les drapeaux *m* et *t* ne devraient en principe pas être spécifiés. On peut prendre comme exemples les verbes *accélérer* ou *ébruiter*.

Si le verbe n'accepte pas d'objet direct, mais seulement un objet indirect, seuls les drapeaux *m* et *t* entrent en ligne de compte (cf. table 7.1).

Le critère sémantique invoqué pour les compléments d'objets directs s'applique également aux compléments indirects. Mais il ne faut cependant pas surestimer son importance: le français dispose d'une construction appelée datif éthique, dont les grammairiens parlent peu mais qui est très usitée. Rentrent dans cette catégorie toutes les constructions du type:

(7.25) Je peux *t'*aménager de la place dans mon laboratoire.

(7.26) Il faudra *m*'améliorer la conclusion de cet article.

qui indiquent, la plupart du temps, le bénéficiaire de l'action accomplie. Je ne crois pas qu'on puisse sérieusement considérer toutes ces formes comme incorrectes et il faut donc en tenir compte dans l'attribution des drapeaux.

Si le verbe est intransitif, les choses deviennent fort simples : on ignore les trois drapeaux *m*, *t* et *l*.

Avant d'aller plus loin, une remarque complémentaire sur les drapeaux *m* et *t* s'impose : ils autorisent des formes erronées comme **m'absolvons* et **t'énergiez*. Malheureusement, la seule solution pour éviter cela consisterait à supprimer ces drapeaux et à entrer « à la main » toutes les élisions des pronoms *me* et *te* dans le dictionnaire : vu leur nombre, ce n'est pas envisageable.

Concernant le cinquième lemme, l'infinitif, on le fera figurer accompagné des drapeaux *nM*³, dans tous les cas, et des drapeaux *m*, *t* et *l*, quand il y a lieu (cf. ci-dessus), ainsi que le drapeau *s* dans certains cas (cf. 7.2.5).

Que faire si le participe passé peut s'accorder ? On fait figurer la forme féminine du participe passé accompagnée du drapeau *F* (cf. 7.1.2). Si ce participe passé est à l'origine d'un substantif, on utilise la forme féminine accompagnée des drapeaux *LMF*, suivie de la forme masculine précédée de « *l'* » (cf. 7.1.3). Mais on ne fera pas de même pour un adjectif : contrairement aux adjectifs formés sur le participe présent, les adjectifs formés sur le participe passé ne peuvent jamais être antéposés en français, *i.e.*, utilisés comme épithètes⁴. Ils ne seront donc jamais précédés d'un déterminant. On peut s'en rendre compte empiriquement, en comparant les quatre exemples ci-dessous :

(7.27) L'homme assailli.

(7.28) *L'assailli homme.

(7.29) La question assaillante.

(7.30) L'assaillante question.

Que faire si le participe présent peut être mis en rapport avec un adjectif verbal ? Au vu de ce qui précède, il convient d'utiliser la forme féminine accompagnée des drapeaux *LMF*. Il n'est pas nécessaire d'ajouter la forme masculine précédée de « *l'* », sauf si le verbe est intransitif (le drapeau *l* n'accompagne pas le lemme de l'imparfait) ou si l'adjectif verbal ne se forme pas sur le participe présent (voir les explications concernant le drapeau *c* en fin de section 7.2.2).

La table 7.2 présente quelques exemples pour clarifier les idées.

3. Je vous rappelle que le drapeau *M* gère le préfixe *d'*.

4. Ils partagent d'ailleurs cette intéressante propriété, comme déjà signalé à la fin de la section 7.1.3 (page 65), avec les adjectifs de couleur, d'appartenance et de forme.

TABLE 7.2 – *Verbes et élision : quelques exemples*

charme/u	alunis/jnu j'alunisse ^a	ébruite/jnlu
charmais/vc	alunissais/jnv	ébruitais/jnlv
charmai/w	alunis/jnw ^b	ébruitai/jnlw j'ébruitasse
charmerai/x	alunirai/jnx j'alunirais	ébruiterai/jnlx j'ébruiterais
charmer/y	alunir/nM aluni	ébruiter/snLM ébruitee/F s'ébruite/p ^c s'ébruitait/p
archaise/jnu	outrepasse/jnlu	assiège/jnmtlu
archaisais/jnv	outrepassais/jnlv	assiégeais/jnmtlv
archaisante/LMF		assiégeante/LMF
l'archaisant		
archaisai/jnw	outrepassai/jnlw	assiégeai/jnmtlw
j'archaisasse	j'outrepassasse	j'assiégeasse
archaïserai/jnx	outrepasserai/jnlx	assiégerai/jnmtlx
j'archaïserais	j'outrepasserais	j'assiégerais
archaïser/nM	outrepasser/nLM	assiéger/nmtLM
archaïsé	outrepassée/F	assiégée/LMF l'assiégé

^a Pour les formes avec élision du pronom *je*, voir 7.2.4.^b En principe, on peut utiliser une seule entrée, *alunis/jnuw*, en lieu et place de *alunis/jnu* et *alunis/jnw*.^c Pour les formes avec élision du pronom *se*, voir 7.2.5.

7.2.4 Premières personnes du conditionnel et des subjonctifs présent et imparfait avec élision de *je*

Lorsqu'un verbe commence par une voyelle, il faut généralement ajouter « à la main » les premières personnes des subjonctifs présent et imparfait et du conditionnel précédées du pronom *je* élide (« j' »). (Pour les verbes du premier groupe, la première personne du subjonctif présent est identique à celle du présent de l'indicatif, il n'est donc pas nécessaire de l'ajouter. Pour les verbes du second groupe, la première personne du présent du subjonctif est identique à celle de l'imparfait du subjonctif ; il est donc inutile de la faire figurer deux fois.) Voyez les exemples de la table 7.2.

7.2.5 Élision et emplois réflexifs

Les drapeaux présentés jusqu'ici ne permettent pas de régler le problème de l'élision du pronom *se* (dans *il s'aime* ou *ils s'aiment*). Ce problème présente deux caractéristiques :

- il ne concerne que la troisième personne, singulier ou pluriel, des différents temps du verbe ;
- il ne concerne pas tous les verbes (certains ne sont pas, ou rarement, utilisés de manière réflexive).

On procédera de la manière suivante :

- pour les verbes du premier groupe, on entrera les deux lemmes suivants :
 - « s' » + la troisième personne du singulier de l'indicatif présent, accompagnée du drapeau p (par exemple : s'abonne/p) ;
 - « s' » + la troisième personne du singulier de l'indicatif imparfait, accompagnée du même drapeau p (par exemple : s'abonnait/p) ;
- pour les verbes du deuxième groupe, on entrera un seul lemme : « s' » + la troisième personne du singulier de l'indicatif imparfait, accompagnée du drapeau q (par exemple : s'abêtissait/q) ;
- pour les verbes du troisième groupe (cf. 7.2.7), il faudra malheureusement entrer toutes les formes « à la main » (mais il y a très peu de chance que vous ayez affaire à ce genre de verbes) ;
- quel que soit le groupe auquel appartienne le verbe considéré, on ajoutera au lemme de l'infinitif, le drapeau s.

Pour un exemple, voyez la table 7.2.

7.2.6 Élision et verbes *essentiellement pronominaux*

Un problème de même nature, mais de plus grande ampleur, se pose avec les verbes dits « essentiellement pronominaux » comme *s'absenter*. La meilleure solution

avec de tels verbes est d'en entrer manuellement toutes les formes dans le dictionnaire, à l'exception de celles qui peuvent être engendrées par les drapeaux **p** ou **q** (cf. 7.2.5). Ils sont heureusement peu nombreux (et seuls les verbes commençant par une voyelle ou un « h » muet posent problème) et il est peu probable que de tels cas apparaissent dans la composition d'un dictionnaire spécialisé. Français-GUTenberg en contient un certain nombre.

L'infinif et le participe présent de ces verbes doivent être entrés tels quels, accompagnés des drapeaux **s**, **m** et **t**.

Voici un exemple (le verbe *s'absenter*) :

m'absente
t'absentes
s'absente/p
absentons
absentez
m'absentais
t'absentais
s'absentait/p
absentions
absentiez
m'absentai
t'absentas
absentâmes
absentâtes
m'absenterai
t'absenteras
absenterons
absenterez
m'absentasse
t'absentasses
absentassions
absentassiez
m'absenterais
t'absenterais
absenterions
absenteriez
absentant/smt
absentée/F
absenter/smt

7.2.7 Les verbes du troisième groupe

Les verbes du troisième groupe doivent être traités comme les autres, selon les points qui précèdent. La seule difficulté supplémentaire est que les drapeaux gérant la conjugaison proprement dite ne sont pas toujours les mêmes que pour les deux autres groupes. D'autre part, il n'existe rien d'équivalent aux drapeaux p et q (cf. 7.2.5) pour les verbes à emplois réflexifs. En principe, Français-GUTenberg contient tous les verbes du troisième groupe, vous ne devriez donc pas avoir à vous en soucier.

Si vous disposez d'un répertoire *Bescherelle*, vous pouvez vous reporter à la table 7.3 pour savoir quels drapeaux employer.

TABLE 7.3 – Drapeaux et conjugaisons du *Bescherelle*

Conjugaison	IND PR	IND IMP	PS	FUT	INF
1–2	(<i>e. m.</i>)	(<i>e. m.</i>)	w	x	(<i>e. m.</i>)
3–5	—	—	—	—	—
6–19	u	v(c) ⁵	w	x	(y) ⁶
20	g	v(c)	w	x	(y)
21	—	—	—	—	—
22	(<i>e. m.</i>)	v(c)	w	x	(y)
23	u	v(c)	w	x	(z)
24	u	v(c)	w	x	(y)
25	u	v(c)	w	x	(y)
26	u	v(c)	w	x	(z)
27–29	u	v(c)	w	x	(y)
30	u	v	w	x	(<i>e. m.</i>)
31–32	g	v(c)	w	x	(y)
33	u	v(c)	w	x	(z)
34	u	v(c)	w	x	(<i>e. m.</i>)
35	g	v(c)	w	x	(y)
36	(<i>e. m.</i>)	v(c)	w	x	(<i>e. m.</i>)
37	(<i>e. m.</i>)	(<i>e. m.</i>)	(<i>e. m.</i>)	(<i>e. m.</i>)	(<i>e. m.</i>)
38	u	v(c)	w	x	(y)
39–40	g	v(c)	w	x	(<i>e. m.</i>)
41	g	v ⁷ c ⁸	w	x	(<i>e. m.</i>)
42	u	v(c)	w	x	(<i>e. m.</i>)
43	u	d	w	x	(<i>e. m.</i>)

5. Il ne faut appliquer le drapeau c qu'aux verbes qui ne commencent pas par une voyelle ou un « h » muet (cf. début de la section 7.2.3).

6. Même remarque que pour le drapeau c (et également valable pour le drapeau z).

7. Implique de compléter le dictionnaire avec les entrées suivantes : *sachions*, *sachiez* et *sachant*.

8. Le drapeau c engendre *savant(e)(s)*.

TABLE 7.3 – Drapeaux et conjugaisons du *Bescherelle*

Conjugaison	IND PR	IND IMP	PS	FUT	INF
44	g	v(c)	w	x	(e. m.)
45–46	(e. m.)	(e. m.)	(e. m.)	(e. m.)	(e. m.)
47	u ou g ⁹	v(c)	w	x	(z)
48	u	v(c)	w	x	(z)
49	g	v	w	x	y
50	(e. m.)	(e. m.)	(e. m.)	(e. m.)	(e. m.)
51	g	v	w	x	y
52	(e. m.)	(e. m.)	(e. m.)	(e. m.)	(e. m.)
53	u	v(c)	w	x	(y)
54	g	v(c)	w	x	(z)
55–56	g	v(c)	w	x	(y)
57–59	u	v(c)	w	x	(z)
60	g	v(c)	w	x	y
61	g	v(c)	(p. l.)	x	(y)
62	g	d(c)	w	x	(y)
63	g	v(c)	w	x	(e. m.)
64	g	v(c)	w	x	(y)
65	g	v(c)	w	x	(z)
66	g	v	w ¹⁰	x	(e. m.)
67	u ou (e. m.) ¹¹	v(c)	w	x	y ou z ¹²
68	g	v(c)	w	x	(y)
69	u	v(c)	w	x	(y)
70	(e. m.)	—	—	x	(y)
71	uw	v(c)	(p. l.)	x	(y ou z ¹³)
72	g	vc	w ¹⁴	x	(e. m.)
73–74	g	v	w	x	z
75–76	(e. m.)	v(c)	w	x	(y)
77	g	v(c)	w	x	(y)
78	gw	d(c)	(p. l.)	x	(y)

9. Pour *prévaloir* : g. Pour tous les autres : u.

10. Le verbe *paître* n'a ni passé simple, ni subjonctif imparfait.

11. Le suffixe ne fonctionne que pour *croître*, il faut entrer manuellement les formes des verbes de la même famille, à cause du circonflexe présent uniquement sur le « i » de la troisième personne du singulier (présent de l'indicatif). (Et parce que je ne veux pas introduire un nouveau drapeau juste pour trois verbes.)

12. Pour *croître* et *recroître* : y. Pour les autres : z.

13. Pour *conclure*, *exclure* et *reclure* : y. Pour *inclure* et *occlure* : z.

14. Seul *résoudre* a un passé simple.

TABLE 7.3 – Drapeaux et conjugaisons du *Bescherelle*

Conjugaison	IND PR	IND IMP	PS	FUT	INF
79	(<i>e. m.</i>)	v(c)	w	x	(<i>e. m.</i>)
80	g	v(c)	w	x	(z)
81	gw	v(c)	(<i>p. l.</i>)	x	(y)
82	u	v(c)	w	x	(y)

7.3 Entrer des abréviations

Votre dictionnaire peut également contenir des abréviations. Si une abréviation contient un point final, celui-ci ne doit pas être tapé. Par contre vous pouvez utiliser des points dans le corps de l'abréviation, de même que des tirets et des apostrophes. Par exemple, pour l'abréviation « c.-à-d. » de « c'est-à-dire », vous entrerez `c.-à-d` (sans point final).

7.4 Vérification de la validité des entrées

Comment vérifier la validité d'une entrée de votre dictionnaire spécialisé? Il suffit d'invoquer `ISPELL` de la manière suivante :

```
echo lemme/drapeau(x) | ispell -d francais -e
```

et de vérifier que les mots engendrés ne comportent pas de monstruosité.

Si vous désirez valider une liste de mots, taper :

```
cat fichier.dico | ispell -d francais -e
```

7.5 Dernières étapes

Une fois que votre dictionnaire est au point, copiez le répertoire `dicos` et lancez `makehash`. Une fois que vous aurez vérifié que votre travail vous donne satisfaction, vous pouvez m'envoyer votre dictionnaire accompagné d'un fichier `.desc` pour qu'il soit inclus dans la prochaine distribution de `Francais-GUTenberg`. Et si vous avez besoin d'aide pour composer votre dictionnaire, n'hésitez pas à me contacter.

Bibliographie

CATACH Nina 1995: *L'orthographe*, Nathan Université, Paris, 3^e édition.

GREVISSE Maurice 1988: *Le bon usage*, Duculot, Paris/Gembloux, 12^e édition.

MONTAGUE Richard 1974, « The proper treatment of quantification in ordinary English », in *Formal Philosophy: Selected Papers of Richard Montague*, éd. par R. H. Thomason, New Haven, Yale University Press.

PYTHOUD Christophe 1996: *Problèmes de la correction automatique de l'orthographe lexicale du français à travers une étude de cas: le correcteur orthographique ISPELL et le dictionnaire Francais-IREQ*, Université de Lausanne, mémoire de licence.

— 1998: « Francais-GUTenberg: un nouveau dictionnaire pour ISPELL. problèmes résolus et intégration de contributions extérieures », *Cahiers GUTenberg*, n° 28–29, pp. 252–275.