



Guide de configuration
de quelques logiciels et environnements Internet
pour l'usage du français

Version 1.0

Viagénie

<http://www.viagenie.qc.ca>

Mars 2001

Table des matières

1. INTRODUCTION	4
2. SYSTÈMES D'EXPLOITATION	6
2.1 CLAVIERS ET PARAMÈTRES RÉGIONAUX	6
2.1.1 <i>Macintosh</i>	7
2.1.2 <i>Windows</i>	10
2.1.3 <i>Unix</i>	11
2.1.4 <i>Linux</i>	12
2.1.5 <i>Sun Solaris</i>	14
2.1.6 <i>Solaris 7 x86</i>	15
2.1.7 <i>Solaris sparc</i>	15
2.1.8 <i>x86 ou Sparc</i>	16
2.1.9 <i>Applications X-Windows - XFree86</i>	16
3. COURRIER ÉLECTRONIQUE	18
3.1 NORMES RELATIVES AU COURRIER ÉLECTRONIQUE	18
3.1.1 <i>Politique de l'IETF pour les tables de caractères et les langues</i>	18
3.1.2 <i>Les normes pour la composante client du courrier électronique</i>	19
3.1.3 <i>Le support du français dans les logiciels de courrier électronique Internet</i>	20
3.2 ENVOI D'UN MESSAGE ENCODÉ UTF-8 AVEC MESSENGER DE NETSCAPE COMMUNICATOR	20
3.3 SERVEUR DE COURRIER ÉLECTRONIQUE.....	21
4. PUBLICATION WEB	23
4.1 SERVEURS.....	23
4.1.1 <i>Exemple de la configuration d'un serveur Apache multilingue</i>	24
4.2 CLIENTS.....	25
4.2.1 <i>Netscape</i>	25
4.2.2 <i>Explorer</i>	26
5. TELNET ET FTP	27
5.1 CLIENT TELNET	27
5.1.1 <i>Unix</i>	27
5.1.2 <i>Windows</i>	27
5.1.3 <i>SecureCRT</i>	28
5.2 SERVEUR TELNETD.....	28
5.2.1 <i>Unix</i>	29
5.2.2 <i>Solaris (Sun)</i>	29
5.3 CLIENT FTP.....	30
5.4 SERVEURS FTP.....	30
5.4.1 <i>Unix</i>	30
5.4.2 <i>WU-FTPD</i>	31
5.4.3 <i>La prochaine version du protocole FTP</i>	32
6. GESTIONNAIRES DE LISTES	33
6.1 MAJORDOMO	33

7. COMMERCE ÉLECTRONIQUE.....	35
7.1 INTERNET OPEN TRADE PROTOCOL - IOTP.....	35
7.2 SECURE ELECTRONIC TRANSACTION - SET.....	36
7.3 EXTENSIBLE MARKUP LANGUAGE - XML.....	37
7.3.1 <i>Les profils pour les différentes applications.....</i>	<i>37</i>
7.3.2 <i>Propriétés linguistiques de XML.....</i>	<i>38</i>
8. CONCLUSION	39

1. Introduction

Les technologies de l'information ont traditionnellement utilisé l'*American Standard Code for Information Interchange* (ASCII) pour la représentation des caractères. Ce jeu de caractères ne permettant pas de représenter les langues autres que l'anglais, il a été étendu sous la série des jeux ISO-8859 afin de permettre à de nombreuses autres langues d'être représentées. Ainsi, le français est représenté par les jeux ISO-8859-1 et ISO-8859-15. Par contre, la série des jeux ISO-8859 ne permettent pas la représentation des langues asiatiques. Celles-ci ont d'ailleurs leurs propres jeux de caractères (Shift-JIS, JIS, ISO-2022).

De nos jours, un jeu de caractères commun pour toutes les langues devient une nécessité, d'autant plus que la mondialisation des marchés et l'internationalisation d'Internet s'intensifient. Ce jeu, appelé jeu universel de caractères a été proposé par le consortium Unicode et par ISO sous la norme ISO-10646. Il permet de représenter presque toutes les langues utilisées dans le monde, dont le français. C'est pourquoi ce jeu est de plus en plus utilisé ; il est en fait appelé à devenir le jeu de caractères standard de tous les environnements informatiques. Cependant, dans sa forme originale, ce jeu est incompatible avec les protocoles de communications et plusieurs logiciels. Pour corriger ce problème, plusieurs modes de transformation ont été conçus. Le plus couramment utilisé est le format UTF-8 qui est compatible avec l'ASCII. D'ailleurs, Unicode est souvent mal référencé sous le nom UTF-8.

Même si le jeu universel de caractères gagne en popularité, il n'en demeure pas moins que l'internaute se voit souvent contraint de configurer lui-même son ordinateur pour obtenir un environnement informatique en français. De plus, que ce soit pour les systèmes d'exploitations ou pour les applications courantes, tels que celles de courrier électronique et de publication Web, une même norme n'est pas utilisée avec constance. Ainsi, le jeu de caractères généralement rencontré dans Unix est ISO-8859-1, quoique certains fournisseurs de serveurs Unix supportent UTF-8, tel que Sun dans Solaris 8. Les administrateurs de serveurs, les gestionnaires de listes ainsi que tous ceux qui conçoivent des sites de commerce électronique rencontrent la même

difficulté à trouver un jeu adapté au français commun à leur système d'exploitation et à leurs logiciels.

Le but de ce guide est donc de décrire les configurations requises pour permettre l'utilisation du français dans les environnements couramment utilisés aujourd'hui.

2. Systèmes d'exploitation

À la base même de tout système informatique, le système d'exploitation est le cœur de l'ordinateur. La configuration de ce système adapte l'ordinateur à son utilisateur. Que ce soit par un clavier qui donne un accès facile aux lettres accentuées ou par la sélection de paramètres prédéterminés relatifs à notre culture, la francisation commence avec la configuration du système d'exploitation.

2.1 Claviers et paramètres régionaux

Claviers

Le clavier est le dispositif qui permet à l'humain de communiquer avec la machine. Périphérique de base, le choix d'un bon clavier est primordial au français intégral dans Internet. Le Canada a émis une norme pour obtenir un clavier qui répond autant au besoin des francophones qu'à celui des anglophones ; c'est la norme CAN/CSA Z243.200-92 approuvée par l'ACNOR (ou CSA). Grâce à cette norme, les internautes canadiens utilisent tous le même type de clavier, peu importe le système d'exploitation (Windows, Dos, MacIntosh, UNIX, etc.). De plus, un tel clavier donne un accès direct aux principales lettres accentuées. Cependant, c'est le pilote de clavier, logiciel inclus dans le système d'exploitation, qui permet à l'ordinateur de reproduire le jeu de caractères voulu par l'utilisateur, et cela indépendamment du clavier et de la plate-forme.

Paramètres régionaux

Intégrés au système d'exploitation, les paramètres régionaux définissent les particularités propres à une culture ou à un pays. Ainsi, l'utilisateur d'un ordinateur peut choisir les critères qui permettront d'afficher la date, la monnaie, les nombres, etc. selon ses préférences. Ces paramètres sont définis dans la localisation d'un environnement informatique.

2.1.1 Macintosh

Clavier

Depuis plus d'une dizaine d'années, Apple suit la norme CAN/CSA Z243.200-92 pour ses claviers et le pilote correspondant, peu importe la version du système d'exploitation. Le pilote de clavier pour cette norme est identifié : ISO-Canadien.

Pour que la version canadienne-française du clavier soit utilisée par l'ordinateur, il suffit de suivre les étapes suivantes :

1. Aller dans la *Pomme*
2. Choisir *Tableau de bord*
3. Sélectionner *Frappe Clavier*
4. Choisir *Canada-ISO* (ou *Canada-CSA* selon la version de la norme de votre clavier)

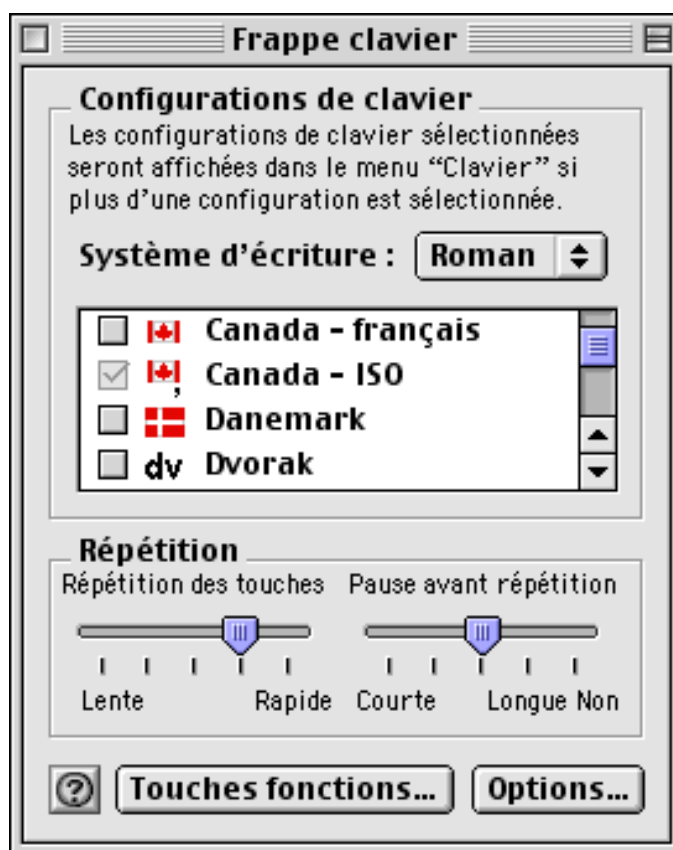


Figure 2-1 Choix du clavier sous Macintosh

Paramètres régionaux

Lors de l'installation du système d'exploitation Mac OS 9, si vous choisissez Canada comme pays de résidence, toutes les paramètres régionaux (incluant le clavier) seront configurés selon le standard canadien français.

Pour que le format canadien-français des dates et des heures soit utilisé par l'ordinateur, il suffit de suivre les étapes suivantes :

1. Aller dans la *Pomme*
2. Choisir *Tableau de bord*
3. Sélectionner *Date & Heure*
4. Choisir *Formats de date* et sélectionner *Canada - Français*

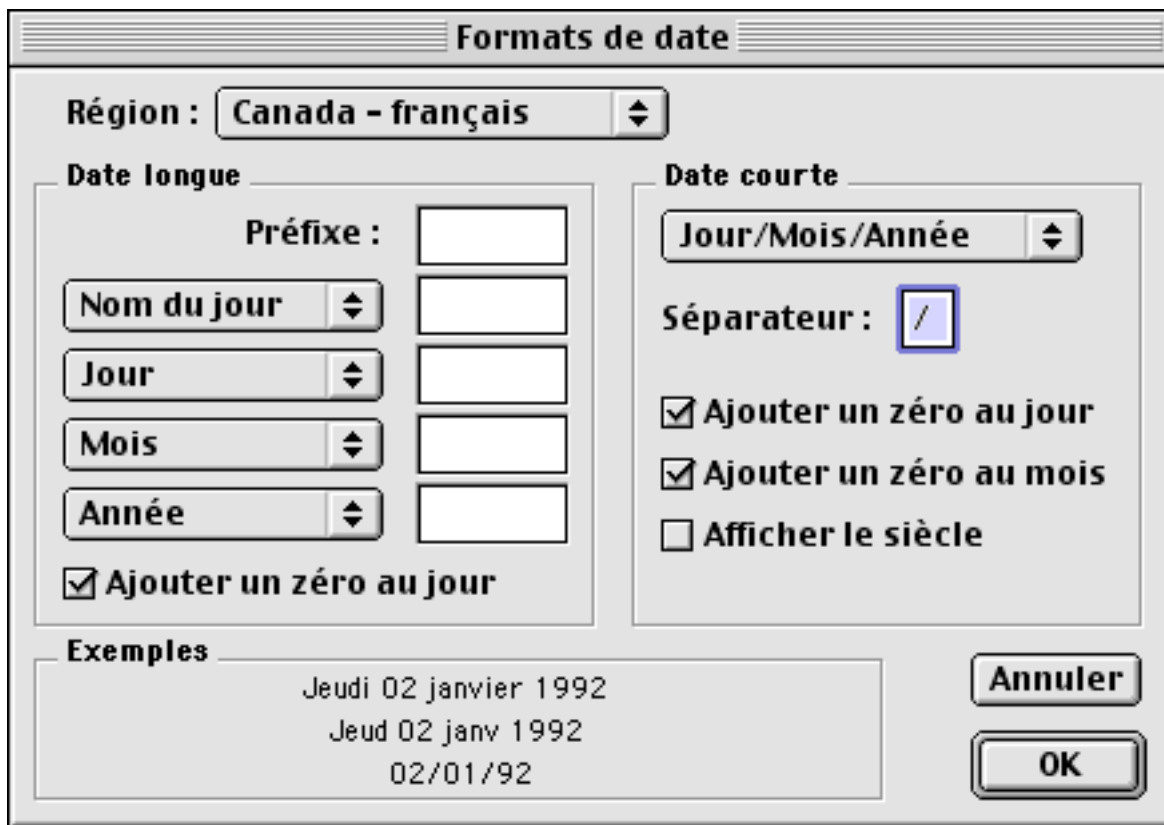


Figure 2-2 Choix des paramètres régionaux sous Macintosh

5. Toujours dans *Date & Heure*, choisir *Formats d'heure* et sélectionner *Canada - Français*.

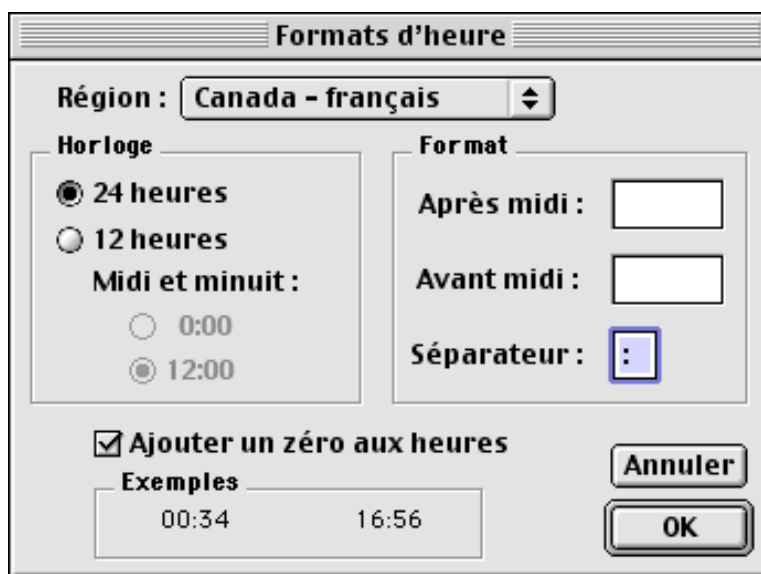


Figure 2-3 Choix de la date et de l'heure sous Macintosh

Pour les nombres :

1. Aller dans la *Pomme*
2. Choisir *Tableau de bord*
3. Sélectionner *Nombres*
4. Sélectionner *Canada - français*.

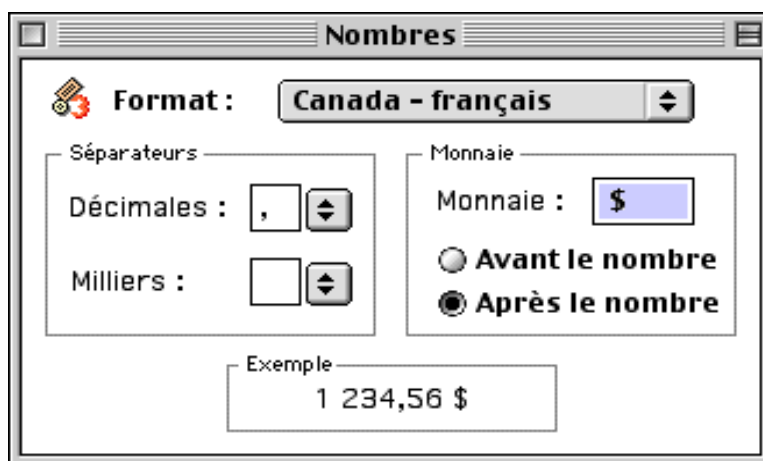


Figure 2-4 Choix des nombres sous Macintosh

2.1.2 Windows

Claviers

Le système d'exploitation Windows offre la possibilité d'avoir différents types de claviers, dont plusieurs en français. Il existe plusieurs pilotes de clavier conformes à la norme CAN/CSA Z243.200-92 pour Windows (à partir de 3.1) et pour DOS.

Pour que la version canadienne-française devienne celle utilisée par l'ordinateur, il suffit de suivre les étapes suivantes :

1. Aller dans *Paramètres*
2. Choisir *Panneau de configuration*
3. Sélectionner *Clavier*
4. Choisir *Paramètres régionaux d'entrée*
5. Sélectionner *Français-Canada*

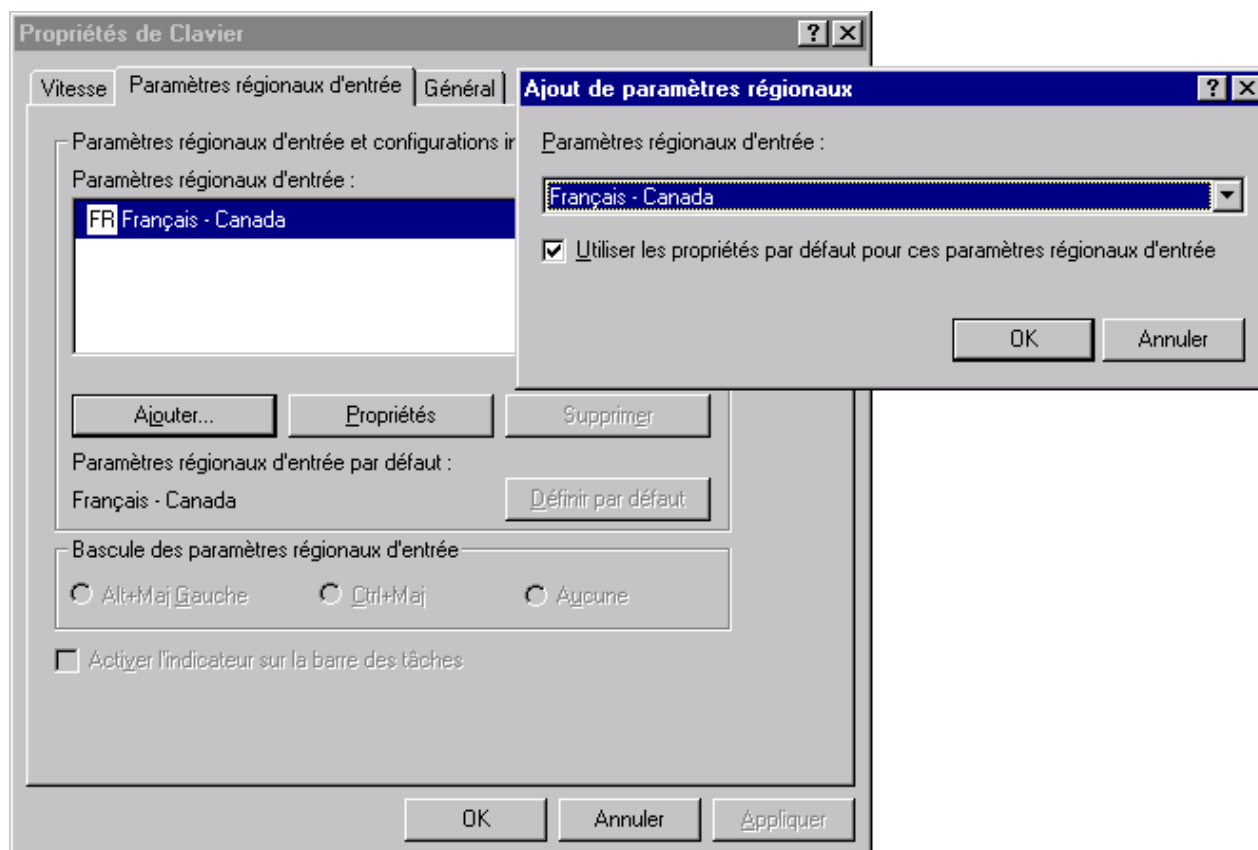


Figure 2-5 Choix du clavier sous Windows

À noter que, pour la version Windows NT 4.0, le pilote de clavier ACNOR peut être difficile à installer. Dans ce cas, le *Service Pack 6* pour Windows NT 4.0 et le *Service Pack 5a* pour le serveur Windows NT 4.0 doivent être téléchargés à partir du site Web de Microsoft à l'URL suivante : <http://www.microsoft.com/downloads>

Pour Windows2000, le pilote pour le clavier ACNOR est intégré.

Paramètres régionaux

Dans Windows, la sélection des paramètres régionaux se fait à partir du panneau de configuration. Ainsi, même si les paramètres par défaut ne sont pas conformes aux préférences de l'utilisateur, ceux-ci peuvent être facilement modifiés. Pour installer les paramètres régionaux du Canada-français il suffit donc d'aller dans l'option *Paramètres régionaux* et de sélectionner Canada-français. Ces options incluent le format d'affichage des devises, des nombres, de l'heure et de la date.

2.1.3 Unix

Claviers

Il existe plusieurs versions d'Unix. Seule la version Linux de Red Hat sera traitée dans ce guide. Il faut préciser que les autres versions peuvent être francisées.

Unix est utilisé en mode commande (*shell*) ou en mode graphique (*X-Windows*). La francisation du mode commande consiste à permettre l'utilisation des caractères codés sur 8 bits par le gestionnaire de terminal et à ce que le gestionnaire de commande permette à son tour l'utilisation de ceux-ci. Le mode graphique est couvert dans la section X-Windows.

2.1.4 Linux

Normalement, lors de l'installation de Linux, le système d'exploitation demande quelle langue et quel clavier l'utilisateur veut installer. C'est pourquoi un clavier canadien-français est nécessaire dès l'installation. Cependant, si la langue et le clavier n'ont pas été spécifiés lors de l'installation, il est possible de spécifier un nouveau choix à l'aide des commandes suivantes :

Commande Explication

`loadkeys cf` Charge en mémoire la table de configuration du clavier canadien-français.

Puisque `loadkeys` est exécuté par des scripts au moment du démarrage, le fichier qui lance `loadkeys` doit être édité afin que les modifications soient prises en compte à chaque nouveau démarrage. Pour ceci :

1. Trouver le script du démarrage `/etc/rc.d/init.d/keytable` , `/etc/rc.d/rc.local`, `init.d/keytable` ou `rc.keymap.`, selon le cas.
2. Trouver le nom exact du fichier lançant `loadkey` avec la commande suivante : `find /etc/rc.d -type f | xargs grep loadkeys`
3. Éditer ce fichier
4. Dans `.inputrc` de votre répertoire « HOME »


```
set meta-flag on      Permet la saisie de caractères 8-bits
set convert-meta off   Empêche la conversion des caractères 8-bits en
représentation 7-bits. Ex : é = \351
set output-meta on    Permet l'affichage des caractères 8-bits
```
5. Pour empêcher le gestionnaire de terminal de couper le 8^{ième} bit, vous devez insérer la commande suivante dans le fichier d'initialisation de votre interpréteur de commande de votre répertoire « HOME » :

`stty pass8` ou **`stty -parenb -istrip cs8`**

Interpréteur de commande	Fichier d'initialisation
Tcsh	<code>.tcshrc</code>
csh	<code>.cshrc</code>

bash

.bash_profile

Attention, RedHat et Debian constituent des exceptions puisque leur table de clavier par défaut se trouve dans `/etc/sysconfig/keyboard` pour RedHat et dans `/etc/kbd/config` pour Debian.

Paramètres régionaux

Dans la majorité des versions d'Unix, les paramètres régionaux sont appelés *Locale*. Chacun des paramètres est configuré par une variable d'environnement qui sera interrogée par les applications. Les applications utilisant les fonctions standardisées pour le traitement des paramètres régionaux traiteront et afficheront les données selon ce qui est défini dans les tables de paramètres correspondant à la langue spécifiée par ces variables :

LC_COLLATE définit les équivalences de caractères pour les comparaisons, pour les ligatures (æ peut être équivalent à ae) et pour les césures.

LC_CTYPE définit les caractères affichables

LC_MONETARY définit le format et le symbole de la monnaie utilisée

LC_NUMERIC définit le format numérique : regroupement, marqueur décimal...

LC_MESSAGES définit la langue des messages

LC_TIME définit le format de la date, les noms des jours et des mois

LC_ALL définit une valeur par défaut des variables précédentes : si la variable LC_ recherchée n'est pas définie, le contenu de la variable LC_ALL est prise en compte, sinon les fonctions standardisées prendront le contenu de la variable LANG.

MM_CHARSET définit la table de caractères pour les applications supportant MIME.

Le format standardisé pour la valeur de ces variables est :

langue[_PAYS[.CHARSET]][@variante]

Les éléments à l'intérieur des crochets sont optionnels, par exemple : 'fr', 'fr_CA', 'fr_CA.ISO-8859-15'. Les fonctions standardisées chercheront du plus spécifique au moins spécifique lors du traitement des paramètres locaux. Par exemple si on spécifie 'fr_CA.ISO-8859-15', l'ordre de recherche sera : « fr_CA.ISO-8859-15 », « fr_CA » puis « fr ». Pour le Québec, les valeurs sont « fr_CA », ou plus précisément « fr_CA.ISO-8859-1 ».

Autres variables d'environnement :

LANG valeur différentes des précédentes, car elle contient le code langue au format ISO : *fr, en, de*

LANGUAGE liste les *Locales*, par ordre de préférence, séparées par deux points (fr:es:de:en), c'est une particularité GNU, fort utile si certaines pages de manuel ou les messages d'un programme n'existent que dans une langue.

Exemple d'utilisation de la variable LANGUAGE :

```
bash# export LANGUAGE=es_ES
bash# ls fichier_n_existant_pas
ls: fichier_n_existant_pas: No existe el fichero o el directorio
bash# export LANGUAGE=de_DE
bash# ls fichier_n_existant_pas
ls: fichier_n_existant_pas: Datei oder Verzeichnis nicht gefunden
bash# export LANGUAGE=en_US
bash# ls fichier_n_existant_pas
ls: fichier_n_existant_pas: No such file or directory
bash# export LANGUAGE=fr_FR
bash# ls fichier_n_existant_pas
ls: fichier_n_existant_pas: Aucun fichier ou répertoire de ce type
```

2.1.5 Sun Solaris

Claviers

La compagnie Sun Microsystems supporte plusieurs langues dans Solaris et SunOs. Les claviers multilingues et les paramètres régionaux sont bien intégrés au système d'exploitation.

L'installation de ce système d'exploitation permet de spécifier les paramètres et initialise l'environnement de façon automatique.

2.1.6 Solaris 7 x86

Lors de l'installation, après le balayage des périphériques, il faut appuyer sur F4 « Devices Tasks ». Choisissez *Set keyboard configuration*, puis choisir le clavier *French-Canadian*. Selon le type de clavier, il faut finalement sélectionner *Standard 101 keys* ou *Windows 105 keys*.

Le système demandera alors de choisir la langue, l'utilisateur francophone optera alors pour *French*. Il est ensuite possible de choisir la variante linguistique ; *Canada ISO8859-1* sera ainsi le choix du francophone vivant au Canada. Cette étape terminée, l'installation se fait de façon identique à tous les Solaris.

2.1.7 Solaris sparc

Normalement, le clavier français de Sun est automatiquement détecté par Solaris. Si vous n'avez pas un clavier français, vous pouvez toujours forcer le clavier avec les options LANG et LC_TYPE dans la fenêtre *shell*. Il faut ajouter ces variables dans les fichiers de configuration des interpréteurs de commandes, localisés dans /etc sous les noms suivants :

profile pour sh ou le shell de type bourne

cshrc pour csh ou le shell de type c-shell

2.1.8 x86 ou Sparc

En mode graphique, lors de l'ouverture d'une session (*login*), cliquer sur le bouton *options* et choisir la langue *fr_CA*, si elle est disponible, ou *fr*. Ceci configurera l'environnement pour recevoir les messages en français et permettra l'utilisation du clavier avec les accents. Si vous choisissez *fr*, vous aurez le paramètres régionaux de la France et non ceux du Canada.

2.1.9 Applications X-Windows - XFree86

X-Windows est un environnement graphique sous Unix. Le gestionnaire X-Windows le plus répandu est XFree86. Disponible sur la majorité des saveurs d'Unix, la configuration de X-Windows est sensiblement la même d'une version d'Unix à l'autre. Seuls les répertoires peuvent être différents. XFree86 peut être utilisé avec des gestionnaires de fenêtres plus spécialisés comme Gnome et KDE. Le paramétrage de l'environnement sera plus convivial sous Gnome ou KDE. Ce guide ne montre que la façon de configurer KDE.

X-Windows utilise une table de correspondance pour lier les touches du clavier aux caractères affichés. Pour avoir un clavier canadien-français dans la majorité des applications X-Windows, il faut utiliser le fichier *Xmodmap* (URL : <http://www.linux-quebec.org/download/Xmodmap.cf>). Ce pilote de clavier remplace le clavier US afin d'utiliser le français dans toutes les applications de Gnome, KDE, Netscape 4.5, etc. Pour activer ce pilote, il suffit de placer le fichier dans le répertoire */etc/X11/xinit* et de le renommer *.Xmodmap* (« *.Xmodmap* » étant le nouveau nom du fichier)

Dans les versions récentes de KDE, le clavier canadien-français peut être activé à partir du *Control center*. Pour ce faire, il suffit de :

1. Aller à la section *Internationalized keyboard*
2. Choisir *Add*
3. Sélectionner *Language : Canadian French, charset : latin*

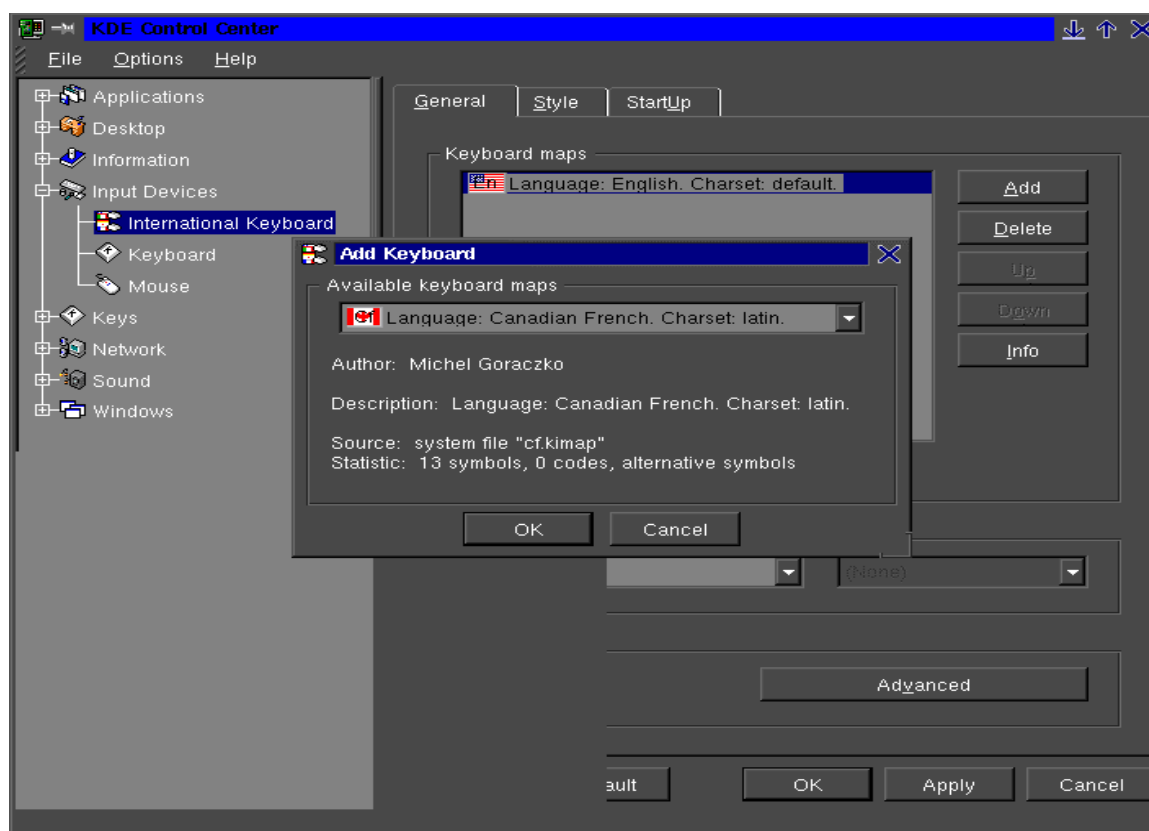


Figure 2-6 Configuration d'un clavier canadien français sous KDE

Suite à l'installation du nouveau clavier, l'icône du gestionnaire de clavier apparaîtra au bas et à droite de l'écran. Il sera possible de basculer du clavier anglais au clavier français en cliquant sur l'icône.



3. Courrier électronique

En quelques années, le courrier électronique est devenu essentiel au travail de bureau et même à la correspondance personnelle. Malgré cette omniprésence, certains internautes ont encore de la difficulté à envoyer et recevoir des messages contenant des lettres accentuées et des caractères non latins. Cependant, quoique méconnues par plusieurs, il existe des normes qui régissent l'internationalisation du courrier électronique. Les producteurs de logiciels ont emboîté le pas, et maintenant il est possible d'avoir des courriers écrits dans notre langue.

3.1 Normes relatives au courrier électronique

L'internationalisation du courrier électronique est normalisée par l'IETF. Elle se situe principalement au niveau des logiciels de courrier électronique et du protocole SMTP (*Simple Mail Transfert Protocol*).

Puisqu'il existe de nombreux produits de courrier électronique, il est difficile de vérifier la conformité d'un produit face aux normes d'Internet. En effet, très peu de fabricants référencent ces normes dans leurs fiches techniques. Cependant, il est possible de demander aux fabricants de préciser les normes auxquelles le produit se conforme en demandant les numéros de RFC sur lesquels sont basés leurs produits.

3.1.1 Politique de l'IETF pour les tables de caractères et les langues

Au début d'Internet, les protocoles utilisaient l'ASCII codé sur 7 bits. Cependant le transport des caractères internationalisés devenant nécessaire avec l'usage grandissant d'Internet, le RFC 2277 fut établi pour définir les spécifications nécessaires aux chaînes de texte utilisant l'UTF-8. Ce RFC spécifie aussi qu'une étiquette de langue peut être associée au texte. Ces étiquettes de langue sont définies dans le RFC 1766.

3.1.2 Les normes pour la composante *client* du courrier électronique

En-tête des messages :

Dans le RFC 2047, norme relative à MIME, il est défini que le texte internationalisé peut être supporté dans les en-têtes de courrier électronique. Ceux-ci étant nécessaires pour l'envoi des messages, c'est dans l'en-tête que l'on trouve l'adresse Internet du destinataire ainsi qu'un champ pour décrire l'objet du message.

Le RFC 2231 normalise des paramètres et des spécifications pour MIME afin de supporter le *charset* et les étiquettes de langue pour l'utilisation de caractères internationalisés dans les en-têtes de courrier électronique. Ainsi, par exemple, l'en-tête pourrait contenir la spécification du *charset* UTF-8 et l'étiquette de langue FR afin que les logiciels de courrier électronique puissent comprendre et afficher des caractères internationalisés dans l'en-tête du message reçu.

Corps des messages :

Le RFC 2045, qui décrit la norme MIME, définit les spécifications du *charset* en usage pour le texte dans le corps du message. Ainsi, pour les francophones, le *charset* ISO-8859 ou ISO 10646 avec encodage UTF-8 est nécessaire afin que le logiciel de courrier électronique puisse accepter et afficher tous les caractères accentués présents dans le corps du message. Le RFC 2046, autre document relatif à MIME, décrète que le *charset* par défaut est l'ASCII lorsque aucun n'est spécifié.

Affichage de l'en-tête et du corps de messages :

L'affichage d'un courrier électronique avec du texte internationalisé implique que le logiciel de courrier électronique client puisse comprendre le *charset* et qu'il soit en mesure de l'afficher. L'affichage a une relation directe avec la capacité qu'a le système d'exploitation d'afficher les caractères internationalisés tel que défini dans les *charset*s. Un logiciel de courrier électronique donné pourrait supporter seulement la compréhension du *charset* dans le message sans être en mesure d'afficher les caractères internationalisés.

Ainsi, un message de courrier électronique échangé entre deux internautes francophones pourraient être rédigé par le premier sous l'environnement Windows 2000 avec spécification du charset UTF-8 et de l'étiquette de langue FR et serait bien reçu par le destinataire avec le même logiciel de courrier fonctionnant sous un système d'exploitation ne supportant pas UTF-8. Par conséquent, le logiciel de courrier reconnaîtrait le charset en usage, mais il ne serait pas en mesure d'afficher tous les caractères internationalisés.

3.1.3 Le support du français dans les logiciels de courrier électronique Internet

Dans un contexte d'usage de la langue française avec les logiciels de courrier électronique client, il est nécessaire que les logiciels de courrier électronique supportent un des charsets contenant les caractères français dans la rédaction des messages et dans leur affichage. L'affichage implique que le système d'exploitation supporte UTF-8.

De plus, il est important de s'assurer que les logiciels de courrier électronique sont conformes aux RFC 2045 à 2049 qui définissent l'usage de MIME dans les échanges de courrier électronique, ce qui permet l'usage de caractères sur 8 bits.

Dans un contexte d'internationalisation, il importe de s'assurer que le produit analysé est conforme aux RFC 2024 et RFC 2045 pour l'utilisation d'UTF-8 et des étiquettes de langues dans les en-têtes et les corps des messages.

3.2 Envoi d'un message encodé UTF-8 avec Messenger de Netscape Communicator

1. Ouvrir *Messenger* de Netscape
2. Avant la rédaction d'un message, se rendre dans le menu *Affichage* puis *Encodage*
3. Sélectionner l'encodage Unicode UTF-8
4. Écrire le message puis l'envoyer

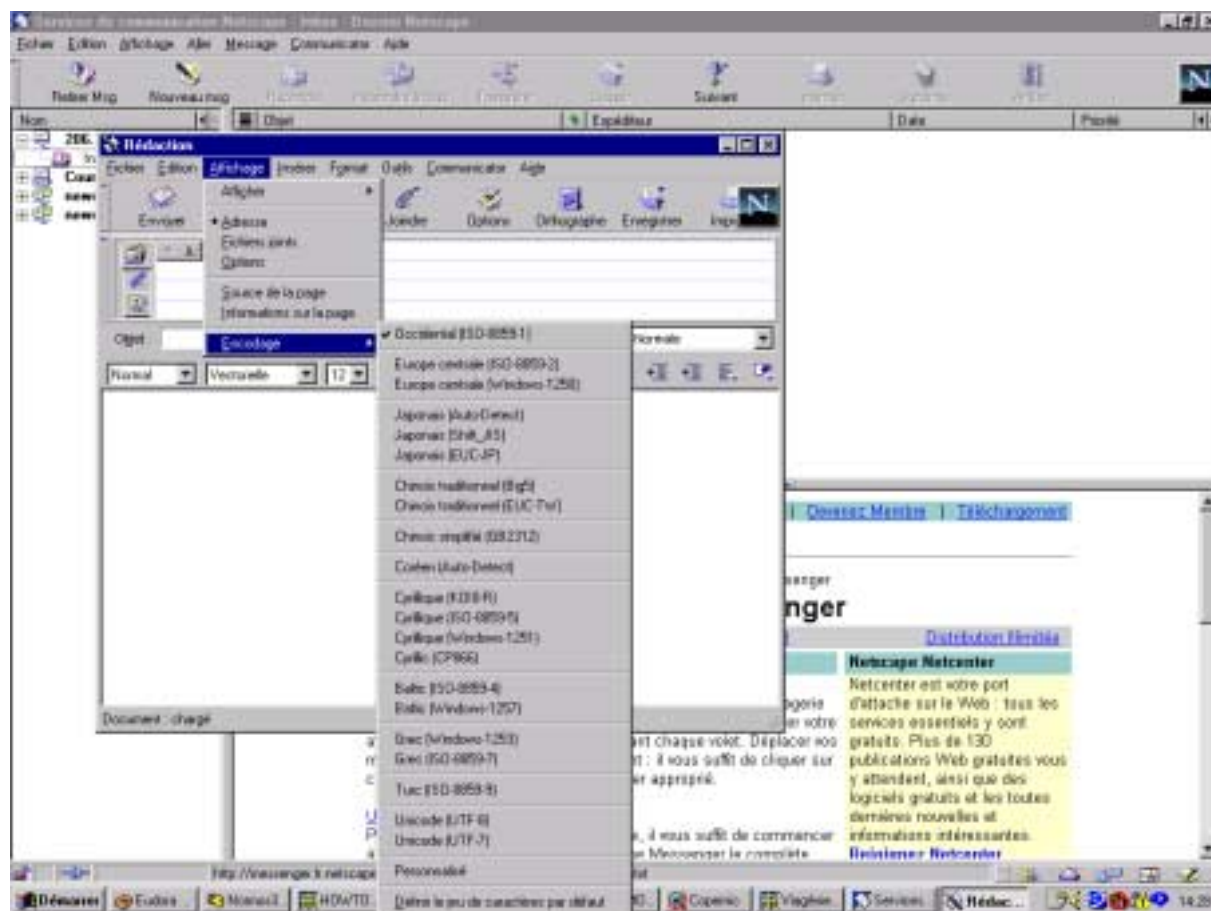


Figure 3-1 Sélection d'UTF-8 dans Netscape Messenger

La réception d'un message encodé UTF-8 nécessite que l'utilisateur sélectionne manuellement l'affichage Unicode UTF-8 tel que montré ci-dessus.

3.3 Serveur de courrier électronique

Le RFC 1652 définit une extension au protocole SMTP afin qu'il puisse supporter nativement des messages 8 bits-MIME par rapport à un usage conventionnel sur 7-bits. Le protocole SMTP, tel que présenté dans le document *Problèmes reliés à l'internationalisation d'Internet*, dispose maintenant de cette propriété qui rend ses échanges plus souples (support 8-bits et support 7-bits) avec les autres serveurs de courrier électronique dans Internet.

Ainsi, tel que mentionné par l'Internet Mail Consortium (<http://www.imc.org>), il est recommandé que tout nouveau serveur de courrier électronique déployé dans Internet soit conforme au RFC 1652, c'est-à-dire qu'il supporte les échanges 8 bits-MIME sous SMTP.

4. Publication Web

De l'individu qui a un site Web pour montrer des photos de sa famille à la multinationale qui vend des produits par le commerce électronique, la publication Web est primordiale à l'échange d'informations diverses entre les individus, les compagnies et les différents paliers de gouvernements. Cette information qui circule dans le Web doit donc être accessible à tous, peu importe leur langue. Pour cela, les serveurs doivent être configurés pour que le lecteur ait automatiquement accès aux pages écrites dans sa langue, de même que le lecteur doit avoir un navigateur qui supporte l'affichage des caractères internationaux.

4.1 Serveurs

Comment exploiter la fonctionnalité de négociation de la préférence de langue du protocole HTTP pour la mise en place d'un serveur Web multilingue ? Le protocole HTTP version 1.1 permet de déployer des serveurs Web pouvant supporter la publication dans plusieurs langues. Cette négociation se fait de façon transparente pour l'utilisateur. Le navigateur envoie une liste de langues au serveur (cette liste est définie par l'utilisateur et est présentée par ordre de préférence). Si le serveur possède le document demandé dans une des langues suggérées, il retournera le document dans la langue se rapprochant le plus de la préférence de l'utilisateur.

Par exemple, un navigateur possède la liste de langues suivante : « fr, es, en », ces lettres représentent respectivement le français, l'espagnol et l'anglais. Le navigateur demande le document « fichier.html » au serveur Web. Le serveur possède ce document en anglais et en espagnol. Le serveur retournera la version espagnole au navigateur puisque cette version se rapproche le plus de la préférence de l'utilisateur.

Cette fonctionnalité nécessite l'activation de la préférence de langue dans le navigateur Web des utilisateurs, tel que présenté dans section suivante.

4.1.1 Exemple de la configuration d'un serveur Apache multilingue

Description du scénario présenté :

Serveur Apache HTTP version 1.3

Support de HTTP/1.1

Langues supportées par le serveur Web : français, anglais

Ordre de préférences des langues : 1. français ; 2. anglais

Arborescence type des fichiers pour Apache HTTP server

/usr/local/www/data Répertoire de base du site Web multilingue

Exemple du fichier de configuration Apache (extrait des lignes pertinentes seulement) httpd.conf

```
# Répertoire racine des documents
```

```
DocumentRoot "/usr/local/www/data"
```

```
# Fichier de d'accueil HTML par défaut
```

```
DirectoryIndex index
```

```
# Spécifications des langages et de l'extension des fichiers associés selon les étiquettes de langue supportées dans le RFC 1766
```

```
AddLanguage en .en
```

```
AddLanguage fr .fr
```

```
# Ordre de priorité de langages en cas d'égalité durant la négociation
```

```
LanguagePriority fr en
```

La directive *AddLanguage* permet de définir le suffixe utilisé pour chacune des langues supportées. La directive *AddLanguage fr .fr* spécifie que le fichier composé d'un suffixe « .fr »

doit être utilisé pour le code de langue « fr ». Par exemple, pour la page d'accueil il faut trois copies du fichier *index.html* :

index.html Page utilisée si le navigateur ne supporte pas la préférence de langue. Peut être un lien symbolique vers la page d'accueil française *index.fr.html*.
index.fr.html Page d'accueil française.
index.en.html Page d'accueil anglaise.

4.2 Clients

Tel que présenté dans la section précédente, le serveur Web Apache supporte différentes langues par l'entremise des paramètres du *Content Negotiation*. L'effet réel de la modification du navigateur est qu'il enverra systématiquement la préférence de langue de l'utilisateur *Accept-Language* à chaque connexion vers le site Web.

Voici maintenant l'exemple de la configuration des deux principaux navigateurs Internet, soit Netscape et Explorer, pour que ceux-ci supportent la langue française par défaut lorsqu'ils négocient avec des serveurs Web supportant le français.

4.2.1 Netscape

Menu Edition -> Préférences

Choix de la langue de préférence des pages Web

Navigateur -> Langues -> Ajouter / Supprimer

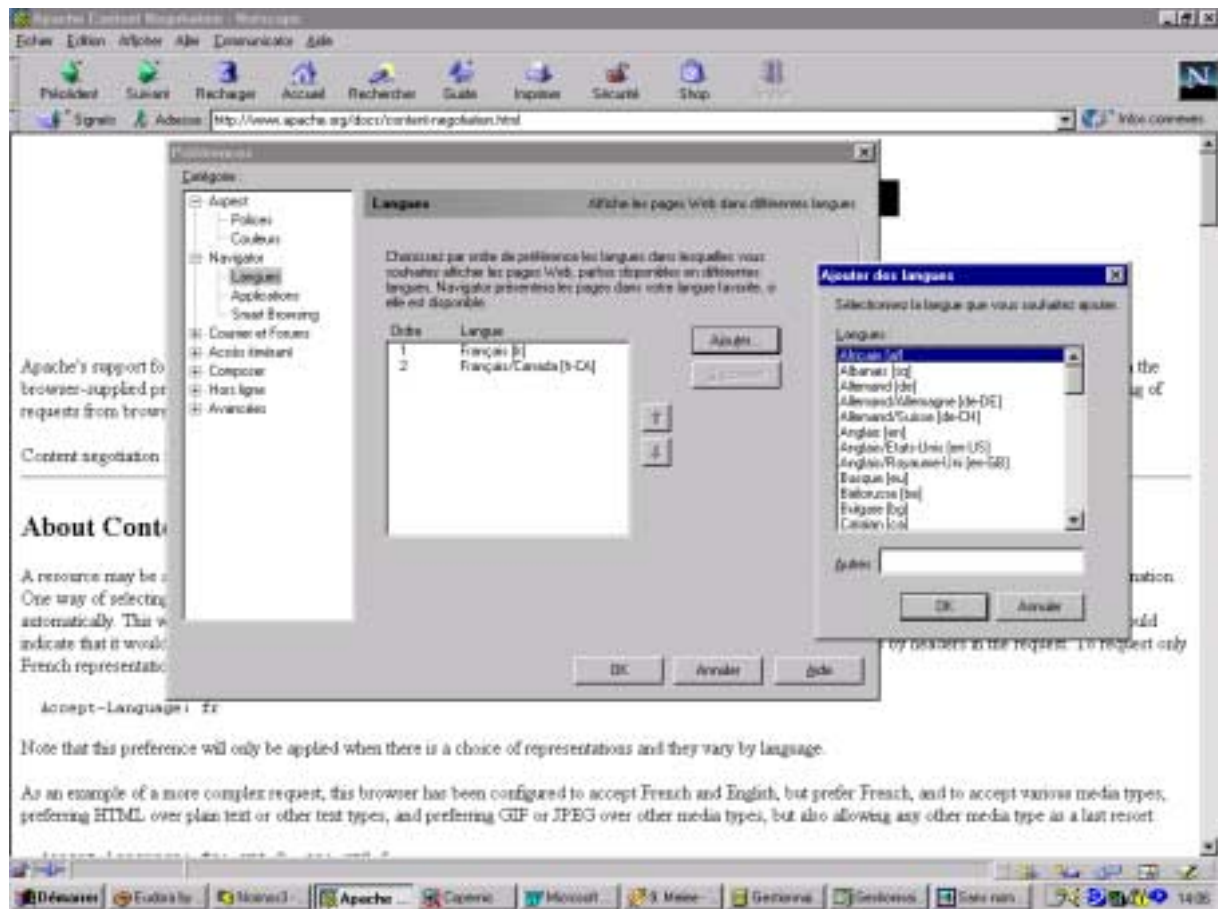


Figure 4-1 Configuration de la préférence de langue dans Netscape

4.2.2 Explorer

Menu Outils -> Options Internet

Options Internet -> (onglet Général) Langues... -> Ajout...

5. Telnet et FTP

Telnet et FTP sont des services réseaux. Telnet permet d'accéder à un ordinateur en mode terminal via le réseau, et FTP est un service de transfert de fichiers. Telnet et FTP fonctionnent sous le modèle client/serveur, donc un logiciel client contacte le logiciel serveur pour qu'un utilisateur accède au service.

5.1 Client Telnet

Le protocole Telnet permet l'accès à un ordinateur à partir d'un terminal distant. Que ce soit avec Unix, Windows ou SecureCRT, l'utilisation du format 8 bits est possible dans le client Telnet.

5.1.1 Unix

Pour la majorité des clients Telnet d'Unix, l'option -8 permet de forcer l'utilisation des 8 bits. De cette façon les accents pourront être utilisés lors de la session.

Exemple :

```
telnet -8 serveur-a-rejoindre
```

5.1.2 Windows

Le client Telnet de Microsoft Windows supporte les caractères 8 bits dans sa configuration de base. Il n'y a donc pas de programmation supplémentaire à faire.

5.1.3 SecureCRT

Le client Telnet de SecureCRT (www.vandyke.com) est un programme qui offre un client Telnet et un ssh. Tout comme Windows, SecureCRT supporte les caractères 8 bits dans sa configuration de base. Cependant, il arrive que les caractères 8 bits ne sont pas affichés correctement. Il faut alors vérifier que la case *strip 8th bit* ne soit pas cochée dans les paramètres de session.

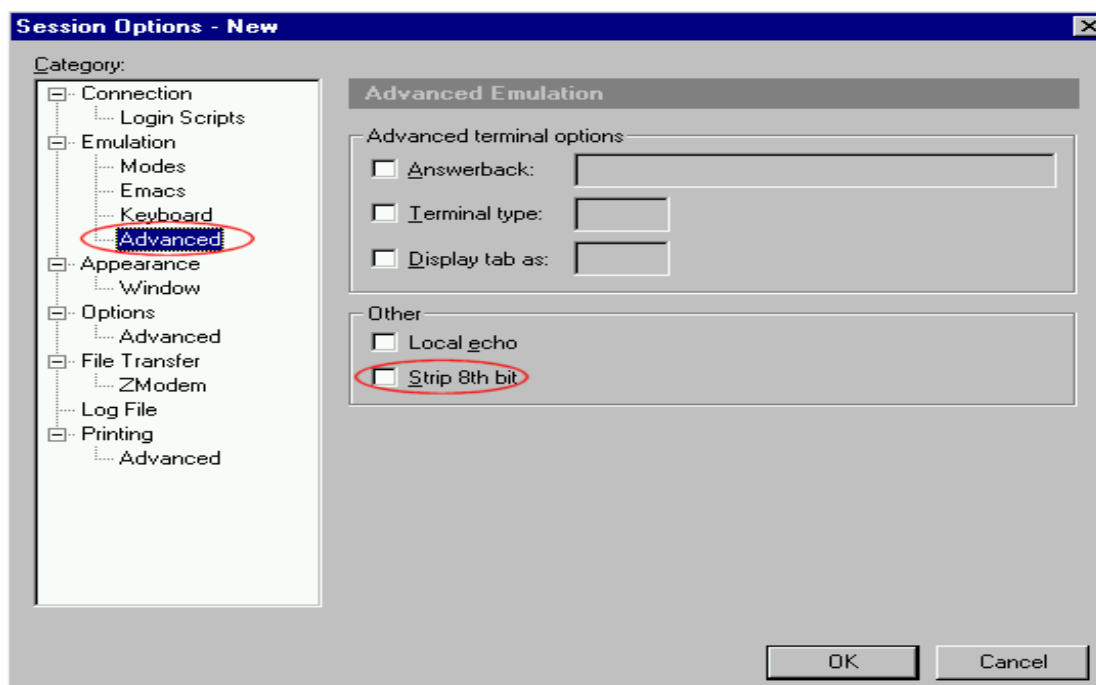


Figure 5-1 Configuration de SecureCRT

5.2 Serveur Telnetd

Telnetd est le logiciel serveur qui permet aux logiciels clients d'accéder à un ordinateur. Il est très peu utilisé sous Windows et fréquemment utilisé sous Unix.

5.2.1 Unix

Le serveur Telnetd (ou in.telnetd) possède différents fichiers que l'on peut utiliser pour franciser ce serveur. Voici une description de ces fichiers.

/etc/motd	contient du texte qui sera affiché juste avant d'invoquer l'interpréteur de commande.
/etc/login.conf	contient les différents paramètres à attribuer lors d'une session Telnet (ou tout autre programme utilisant le programme <i>login</i>). Les paramètres pertinents à la francisation dans <i>login.conf</i> sont : <pre> default:\ :copyright=/etc/COPYRIGHT:\ :welcome=/etc/motd:\ :charset=ISO_8859-1:\ :lang=fr_CA.ISO_8859-1:\ </pre>
copyright	pointe sur un fichier qui contient des informations supplémentaires de copyright.
welcome	pointe sur un fichier à imprimer avant de retourner à l'invite du shell.
charset	permet d'assigner la variable MM_CHARSET.
lang	permet d'assigner la variable LANG.

5.2.2 Solaris (Sun)

La bannière Telnetd de Solaris est gérée différemment. Vous devez la changer dans le fichier */etc/default/telnetd*. Pour ce faire, il faut insérer la ligne suivante :

```
BANNER=""Nouvelle bannière de `uname -n`\r\n"
```

Ce qui donnera comme résultat la bannière suivante :

Nouvelle bannière de ordinateur.nomdomaine.qc.ca

login :

5.3 Client FTP

Le transfert de fichiers en français fait par FTP ne pose généralement pas de problèmes. En effet, les clients FTP récents permettent d'utiliser les caractères encodés sur 8 bits dans leur configuration initiale.

Cependant, les accents sont codés selon des tables de caractères qui peuvent être incompatibles entre plates-formes hétérogènes. Windows utilise des tables de caractères propriétaires (appelées *Code Page*) tandis qu'Unix et Mac utilisent la norme ISO-8859. Ainsi le caractère « é » n'a pas la même valeur dans la table CP1251 que dans la norme ISO-8859-1. Ceci rend difficile la francisation des messages et des noms de fichier, car le protocole n'a pas été conçu pour permettre les liens entre les différentes tables de caractères.

5.4 Serveurs FTP

Les serveurs FTP permettent l'échange de fichiers entre ordinateurs. La version actuelle de ce protocole ne supporte pas le multilinguisme. La plupart des serveurs FTP envoient des messages de bienvenue ou spécifiques à un sous-répertoire seulement en anglais. Ces messages sont stockés dans des fichiers et peuvent être modifiés pour transmettre de l'information en français.

5.4.1 Unix

Le serveur FTP de base possède 2 fichiers pouvant être modifiés pour accepter les caractères internationaux. Ces fichiers sont :

`/etc/ftpwelcome` le contenu sera envoyé avant de demander le nom de l'utilisateur.

`/etc/ftpmotd` le contenu sera envoyé après l'authentification de l'utilisateur. Attention, celui-ci est relatif au répertoire de départ du processus du serveur FTP. Si `ftpd` est lancé avec comme répertoire de base `/usr/ftp`, le fichier devra être placé dans `/usr/ftp/etc/ftpmotd` pour le transfert de fichier avec usagers anonymes. Pour les usagers du système, une copie devra être placée dans `/etc/ftpmotd`.

5.4.2 WU-FTPD

Ce serveur FTP est un des plus employés dans Internet. Son fichier de configuration `/etc/ftpaccess` permet de configurer les bannières qui seront envoyées au client. Attention, certains fichiers sont relatifs au répertoire de départ du processus serveur FTP, surtout dans le cas des usagers anonymes.

Exemple du contenu de `ftpaccess` :

```
limit    local    20    Any        /etc/msgs/msg.toomany

readme   README*          cwd=*

message  /welcome.msg        login
message  .message       cwd=*
```

Le paramètre *limit* permet de spécifier le message qui sera envoyé lorsque le nombre maximal d'utilisateurs est atteint. Ainsi, pour franciser ce message, il suffit de modifier le fichier `/etc/msgs/msg.toomany`.

La commande *readme* permet de spécifier le message à envoyer après l'exécution d'une commande, dans le cas de « `cwd=*` » (qui permet de changer le répertoire courant). Tous les

fichiers dont le début est *README* et qui se trouvent à l'intérieur du nouveau répertoire seront envoyés au client. Donc, pour envoyer un message en français il suffit d'éditer ces fichiers.

5.4.3 La prochaine version du protocole FTP

Une nouvelle version du protocole FTP est en cours de développement. Le RFC 2640 modifie le protocole FTP afin de permettre de choisir la langue des messages envoyés au logiciel client FTP. Ce RFC définit les commandes *lang* et *feat* qui permettent au client de spécifier la langue de son choix. Ainsi, la commande *feat* permettra de connaître les langues supportées par le serveur et la commande *lang* permettra de choisir la langue parmi celles listées dans *feat*.

La phase finale de normalisation de ce RFC est en cours. Lorsque ce RFC deviendra un standard officiel, les serveurs FTP supporteront cette nouvelle fonctionnalité.

6. Gestionnaires de listes

Le service de gestionnaires de listes donne l'occasion aux internautes d'échanger des messages électroniques sur un sujet précis à toutes les personnes inscrites sur une même liste. Un francophone peut écrire ses messages en utilisant les lettres accentuées, cependant les requêtes faites auprès du gestionnaire de listes ainsi que ses réponses ne se font qu'en anglais.

Majordomo est le principal gestionnaire de listes utilisé, c'est pourquoi seul celui-ci sera abordé dans ce chapitre.

6.1 Majordomo

La version 1.94 (la plus récente au moment de la conception de ce guide) ne supporte pas l'internationalisation des messages. Elle est unilingue anglaise et accepte partiellement le danois. Cependant, la version 2 de Majordomo offrira le support de langue. Cette version est actuellement en version alpha. Un groupe de français ont entrepris la traduction des messages de Majordomo. Ils ont traduit une bonne partie des messages, mais il en reste encore beaucoup à faire. Cette version offrira aussi le support de langue en fonction de la préférence de l'abonné.

Voici donc comment configurer Majordomo 2 pour l'utilisation des messages français :

Dans le fichier `mj_cf_defs.pl` :

il faut remplacer la valeur « en » de l'option `default_language` pour « fr ».
 « `'default_language' => 'en',` » devient « `'default_language' => 'fr',` »

Si la version utilisée ne possède pas le répertoire `files/fr` :

Créer le répertoire `files/fr`

Télécharger les fichiers de message traduits du site (<http://savage.iut-blagnac.fr/projets/developpement/majordomo2.shtml>) et les placer dans le répertoire requis dans *files/fr*

7. Commerce électronique

Polyvalent, flexible, accessible aux clients peu importe où ils habitent : le commerce électronique reflète ce que sont les affaires de la nouvelle génération. Une des applications d'Internet fort prometteuse, à condition que celle-ci supporte les transactions multilingues afin que le commerçant puisse offrir un service adéquat à ses clients. Pour l'échange d'information multilingue, le commerce électronique est principalement régi par trois protocoles. Le premier protocole est l'IOTP qui définit les paramètres de la structure du commerce, le second est le SET qui régit le paiement alors que le troisième, XML, donne une grande flexibilité à l'entrée et à la publication de données.

7.1 *Internet open trade protocol - IOTP*

Le protocole IOTP définit une structure interopérable pour le commerce électronique dans Internet. Cette structure est indépendante du système de paiement utilisé et intègre les systèmes de paiement comme SET, Mondex, Cybercash, etc. IOTP gère les cas où le site de commerce en tant que tel, le gestionnaire du paiement, le gestionnaire de livraison des biens ou de services et le fournisseur du support à la clientèle sont effectués par différents intervenants ou par un seul. Le langage XML est utilisé pour structurer les données échangées (voir section 7.3 sur le XML).

Les transactions se font sous forme de messages IOTP échangés entre les divers intervenants. L'utilisateur spécifie les langues de préférences et les encodages de caractères à utiliser lors des échanges. Les intervenants pourront alors répondre à l'utilisateur en utilisant une des langues et un des encodages préférés. Afin d'indiquer ses préférences, l'utilisateur configurera l'en-tête de message « Message ID component » avec les éléments *xml:lang*, *LangPrefList* et *CharSetPrefList*. Ainsi, l'option *xml:lang* spécifie la langue et l'encodage utilisé dans le message alors que *LangPrefList* et *CharSetPrefList* spécifient la liste des langues et celle des encodages suggérés pour la réponse à ce message. Cette liste est fournie uniquement à titre informationnel et l'intervenant n'est pas tenu d'y répondre.

L'exemple suivant montre l'utilisation du français dans un échange IOTP :

```

<OtpMessage>
  <TransRefBlk ID='M1.1'>
    <TransId
      ID='M1.2'
      Version='1.0'
      OtpTransId='1234567@somewhere.net'
      TransTimeStamp='2000-06-01 11:45:00-20' >
    </TransId>
  <MsgId
    ID='M1.2'
    xml:lang='fr-ca'
    LangPrefList='fr,en'
    CharSetPrefList='iso-8859-1'
    SoftwareId='Logiciel xyz v1.0'
    TimeStamp='2000-06-01 11:45:01-34' >
  </MsgId>
  <TransRefBlock>
</OtpMessage>

```

7.2 Secure Electronic Transaction - SET

SET est un protocole de paiement électronique développé conjointement par Mastercard et VISA. Il facilite le paiement en ligne par carte de crédit et assure la sécurité pour protéger le numéro de carte de crédit de l'acheteur. Ce protocole accepte le multilinguisme de sorte que le détenteur de la carte de crédit peut préciser la langue de communication qu'il préfère. Ainsi, les messages envoyés au client seront envoyés dans la langue qu'il a spécifiée.

De plus, plusieurs fonctionnalités de SET offrent le multilinguisme. Par exemple, l'information donnée par le marchand peut être envoyée dans la langue du détenteur, de même que le formulaire électronique peut être affiché et rempli dans cette même langue. La langue est identifiée selon la norme décrite dans le RFC 1766.

7.3 Extensible Markup Language - XML

XML est un langage de balisage extensible. C'est un sous-ensemble du *Standard Generalized Markup Language* (SGML). SGML est le standard de fait dans le milieu de la Gestion Électronique Documentaire (GED). Il a été conçu pour publier, dans Internet et les intranets, des documents comportant une structure formelle exploitable. Son avantage principal réside dans sa capacité à pouvoir décrire n'importe quel domaine de données grâce à son extensibilité. C'est-à-dire qu'il permet de structurer un document en définissant le vocabulaire et la syntaxe des données que le document va contenir.

Les échanges de données du commerce électronique se font de plus en plus par XML, car ce langage s'utilise facilement lors d'échanges de données ou de documents entre sites distants ou entre applications différentes. Le XML peut donc être vu comme un format pivot garant des données échangées malgré l'hétérogénéité des bases de données et des logiciels. IOTP est un bon exemple de l'utilisation de XML dans un environnement de commerce électronique. IOTP définit les balises dans un cadre XML pour l'échange d'informations pertinentes à une transaction de commerce électronique.

7.3.1 Les profils pour les différentes applications

Un autre avantage est que XML permet la création de données qui ne sont pas dépendantes d'un environnement. Ces données peuvent donc être interprétées par un grand nombre de dispositifs : ordinateurs habituels, téléphones cellulaires, terminaux braille, WebTV, etc.

7.3.2 Propriétés linguistiques de XML

Les spécifications de XML indiquent explicitement que XML utilise la norme internationale de codage des caractères sur 31 bits ISO 10646 (encodée avec UTF-8), qui couvre la plupart des langues connues. Cette norme est de plus en plus adoptée par les nouveaux logiciels et protocoles, mais d'autres tables de caractères sont encore en utilisation. Les spécifications XML expliquent comment indiquer dans un fichier XML le codage du jeu de caractères utilisé.

XML définit l'option *xml:lang* pour préciser la langue et *encoding* pour préciser la table de caractères si elle est différente de celle proposée comme valeur par défaut. Par exemple, pour spécifier la langue canadienne-française et l'utilisation de la table iso-8859-1 dans un ensemble XML :

```
<?xml encoding='iso-8859-1'?>  
<Donnees-quelquonque xml:lang='fr-ca'>  
  Du texte  
</Donnees-quelquonque>
```

8. Conclusion

Si le français n'a pas été une priorité au début d'Internet, l'utilisation grandissante de ce moyen de communication oblige les manufacturiers à fournir des produits supportant le multilinguisme. La difficulté est que la configuration pour franciser un environnement diffère non seulement d'un système d'exploitation à l'autre, mais aussi parmi les différentes versions d'un même système d'exploitation (Unix). Par contre, les manufacturiers tendent vers une uniformisation de l'utilisation des tables de caractères en adoptant de plus en plus ISO-10646 et UTF-8. Ceci facilitera les échanges entre les différents systèmes d'exploitation. Par ailleurs, les organismes de normalisation tels que l'IETF et le W3C travaillent à internationaliser Internet par l'adaptation au multilinguisme de leurs principaux protocoles.

Ce guide donne des exemples de configuration, mais ne couvre pas toutes les plates-formes. Il donne une piste pour que l'utilisateur ait une idée où aller chercher de l'information plus spécifique et quoi faire lorsqu'il veut un environnement adapté à la langue française.

A. Glossaire

RFC	<i>Request For Comments</i> , document technique normalisé par l'IETF pour les protocoles utilisés dans Internet
MIME	<i>Multipurpose Internet Mail Extensions</i> ou capacité d'utiliser les caractères codés sur 8-bits dans les protocoles Internet ne supportant que 7-bits
UTF-7	<i>Unicode Transformation Format</i> avec encodage sur 7-bits, tel que spécifié par la RFC-2152
UTF-8	<i>Unicode Transformation Format</i> avec encodage sur 8-bits
CHARSET	Méthode de conversion d'une séquence d'octets vers une séquence de caractères. UTF-7, UTF-8 et ASCII sont des exemples de charset. Les charsets sont en usage dans le monde Internet et définis par l'Internet Assigned Numbers Authority (IANA). Le RFC 2278 donne de l'information sur les charsets tels que définis dans Internet.

B. Références

Apple (2000-04-11) Apple [en ligne] Adresse URL : <http://apple.com>

Azim Manjee (1995-02) Le Clavier de votre ordinateur est-il normalisé? [en ligne] Adresses URL : <http://www.cimbcc.qc.ca/chip/pages/chip9502.htm#clavier>, transcrit par Michel Paul dans la revue CHIP, vol.11 no 7

eFA Group (2000-04-30) *Portail sur XML*, Adresse URL : <http://www.efagroup.com/>

Guyllhem Aznar (1999) Le Francophones-HOWTO : Linux & la langue française, version 3.2, Adresse URL : <http://www.oeil.qc.ca/~guyllhem/french/Francophones-HOWTO.html>

Leygnac Marc (1998-12-14) Clavier multilingue [en ligne] Adresses URL : <http://www.ac-poitiers.fr/pedago/outil/astuces/trucs/clavier/clavier.htm>, Académie de Poitiers

Linux-France (2000-02-12) *Réponses aux questions fréquemment posées dans les forums Usenet fr.comp.os.linux* [en ligne] Adresse URL : <http://www.linux-france.org/article/fcol-faq/fcol-faq.html>

Linux-Québec (2000-03-12) *La FAQ de Linux-Québec* [en ligne] Adresse URL : <http://www.linux-quebec.org/faq.html#7>, site consulté le 4 avril 2000

Office de la langue française (1999-10-22) *Le Signet, la référence branchée en terminologie* [en ligne] Adresse URL : <http://www.olf.gouv.qc.ca>

RFC 1652 : Klensin, J. et al (1994) *SMTP Service Extension for 8bit-MIMEtransport*, Adresse URL : <http://www.normos.org/ietf/rfc/rfc1652.txt>

RFC 1766 : Alvestrand, H. (1995) *Tags for the Identification of Languages*, Adresse URL : <http://www.normos.org/ietf/rfc/rfc1766.txt>

RFC 2024 : Chen, D., Gayek, P., Nix, S. (1996) *Definitions of Managed Objects for Data Link Switching using SMIPv2*, Adresse URL : <http://www.normos.org/ietf/rfc/rfc2024.txt>

RFC 2801 : Burdett, D. (2000) *Internet Open Trading Protocol - IOTP*, Adresse URL :

<http://www.normos.org/ietf/rfc/rfc2801.txt>

Sun (2000-04-10) *SunSolve Home* [en ligne] Adresse URL : <http://sunsolve.sun.com/>

Viagénie (2000-04-01) *Problèmes reliés à l'internationalisation d'Internet* [en ligne] Adresse

URL : <http://www.ceveil.qc.ca/viag1.html>

Viagénie (2000-04-01) *Rapport de mission : les noms de domaine internationalisés* [en ligne]

Adresse URL : <http://www.ceveil.qc.ca/ietf47.html>

W3C (1998) Extensible Markup Language (XML) 1.0, URL : <http://www.w3.org/TR/REC-xml>

W3C (1998) Langage de balisage extensible (XML) 1.0, traduction de Patrick Andries, Samira Cuny, François Yergeau, URL : http://babel.alis.com/web_ml/xml/REC-xml.fr.html

Secure Electronic Transaction

<http://www.setco.org/>

Extensible Markup Language (XML)

<http://www.gutenberg.eu.org/pub/GUTenberg/publications/HTML/FAQXML/faqxml-fr.html>

<http://www.abf.asso.fr/enrichi/xmldoc.htm>

<http://www.stars.com/Software/XML/>

<http://www.bvamyfra.fr/xmldef.htm>

<http://www.multimania.com/pensarguet/technique.htm#XML>