

**LE FRANÇAIS
DANS LES TECHNOLOGIES DE L'INFORMATION**

Conventions et pratiques en usage au Québec

SOMMAIRE

AVANT-PROPOS

INTRODUCTION

I LA TERMINOLOGIE DES TECHNOLOGIES DE L'INFORMATION

II LES CARACTÈRES

1. Les systèmes d'écriture
2. La translittération
3. Les jeux de caractères
4. Les signes de ponctuation
5. Les caractères spéciaux

III LES MOTS : TRI, CLASSEMENT, COMPARAISON, RECHERCHE, FUSION

1. Un système à quatre niveaux
2. Les outils FRANCIS
3. Les particularités de l'ordre du classement du français

IV LES ÉLÉMENTS TEXTUELS FRANÇAIS UTILISÉS COMME DONNÉES STRUCTURÉES

1. Les noms de personnes
2. Les noms de sociétés et d'organismes publics ou privés
3. Les adresses postales
4. Les adresses électroniques
5. Les numéros de téléphone et de télécopie
6. Les données numériques
7. Les abréviations
8. Les sigles et les acronymes

9. Les coupures de mots en fin de ligne

V LA PRÉSENTATION DES TEXTES

1. La pagination
2. Les marges
3. Les interlignes
4. L'espacement
5. La division d'un texte
6. Les formats de papier

VI LE BALISAGE DES TEXTES

VII LES INTERFACES PERSONNE-MACHINE

1. Le poste de travail
2. Le dialogue personne-machine
3. Les besoins des personnes qui ont des limitations fonctionnelles

VIII EXIGENCES LÉGALES ET RÉGLEMENTAIRES

1. La propriété intellectuelle et les droits d'auteur
2. La protection des renseignements personnels

CONCLUSION

Coordination

Michèle Drouin-Rochette
Secrétariat du Conseil du trésor

Rédaction

Danielle Stanton

Révision linguistique

Liliane Bourbeau

Office de la langue française

Le contenu de cette publication est diffusé dans Internet à l'adresse suivante:

<http://www.tresor.gouv.qc.ca/frantech.htm>

Pour plus d'information, veuillez contacter

Alain La Bonté

Secrétariat du Conseil du trésor

Téléphone: + 1 418 528-6144

Télécopieur: + 1 418 646-3571

Courrier électronique:

alb@sct.gouv.qc.ca

M. Jian Yang
Office de la langue française

Téléphone: + 1 514 873-3840
Télécopieur: + 1 514 864-3948
Courrier électronique :
Jian@olf.gouv.qc.ca

© **Gouvernement du Québec 1997**

AVANT-PROPOS

Depuis déjà quelques années, le Secrétariat du Conseil du trésor et l'Office de la langue française unissent leurs efforts pour promouvoir une meilleure utilisation de la langue et de la culture françaises dans l'Administration publique. Ainsi, en 1992, le gouvernement adoptait une politique d'utilisation du français dans les technologies de l'information qui vient reconnaître ainsi toute la portée économique, sociale et culturelle du français dans les technologies.

Nous nous associons de nouveau en offrant modestement un premier guide unique sur la façon dont se traduit la francisation des technologies en termes concrets. Ce premier guide, malgré les imperfections d'une première version, présente les conventions et pratiques en usage au Québec et la façon dont ces conventions et pratiques peuvent être implantées en standard ou de manière normalisée dans les logiciels.

Ce guide dresse la liste des conventions de saisie, de traitement et de présentation propres à la langue française. Ces trois volets, pour un utilisateur de technologies de l'information, représentent la base essentielle d'une communication adéquate.

À l'heure du déploiement des inforoutes, ce guide encourage la mise en place de moyens de communications qui contribuent à l'épanouissement de la langue française à travers le monde et à faire connaître les particularités propres à la société québécoise. Il y a là aussi matière à conception de meilleurs logiciels, mieux adaptés aux langues et aux cultures du monde. C'est un domaine où le Québec excelle déjà et où il devrait continuer à exceller.

Ce guide sera périodiquement remis à jour en fonction de l'évolution des technologies et des usages. Nous vous encourageons à cet effet à bien vouloir nous faire part de vos critiques constructives.

Nicole René
Présidente
Office de la langue française

Pierre Roy
Secrétaire
Conseil du trésor

INTRODUCTION

L'évolution des technologies de l'information est en train de modifier en profondeur l'activité humaine. Les nouvelles technologies abolissent graduellement les frontières traditionnelles, créent de nouveaux espaces économiques et culturels, entraînent le développement de nouveaux secteurs d'activité. Cette évolution ne peut cependant masquer une obligation : la nécessité de respecter les langues, les cultures et les usages propres à chaque société. Cet impératif devient encore plus aigu avec l'augmentation rapide du nombre d'utilisateurs des technologies et l'émergence des cyberroutes.

La situation géographique du Québec, son appartenance à l'ensemble nord-américain, lui impose un double défi : il doit développer des moyens de communication qui contribuent à l'épanouissement de la langue et de la culture de la majorité, dans le respect des langues et des cultures des minorités; il doit participer activement à l'activité du continent de façon à profiter des échanges avec ses partenaires. De telles contraintes resteront toujours difficile à concilier. Un des moyens privilégiés d'y arriver, c'est d'adapter les outils technologiques, afin qu'ils répondent à deux ordres de besoins : ceux relevant des règles du français québécois et des conventions d'écriture adoptées par la majorité, ceux imposés par la traduction, c'est-à-dire le passage d'une langue à une autre.

Le Québec forme la seule société française en Amérique du Nord. En vertu de la Charte de la langue française, adoptée par le gouvernement en 1977, le français est la langue officielle du Québec. Par ailleurs, environ le cinquième de la population québécoise a gardé ou adopté une autre langue d'usage, l'anglais principalement, mais aussi les langues autochtones.

Selon un principe reconnu internationalement, tout utilisateur de technologies de l'information doit avoir accès à une information qui respecte les conventions en vigueur dans son pays ou son territoire. En d'autres mots, les technologies ne doivent pas représenter de nouvelles contraintes pour l'utilisateur; elles doivent plutôt offrir de nouvelles possibilités qui permettent à chaque individu de développer ses habiletés et

d'exploiter son plein potentiel. Les outils informatiques doivent continuellement être adaptés pour les rendre conformes aux systèmes d'écriture et aux conventions culturelles en usage dans chaque société.

Il faut également tenir compte des besoins particuliers de certains groupes de la population. À cet égard, les personnes qui vivent avec un handicap physique ou mental présentent un grand défi. Ces personnes forment un groupe très hétérogène : leurs limitations peuvent être d'origines diverses; elles prennent différentes formes (motrices, sensorielles, cognitives); elles peuvent être de courte ou de longue durée. Au Québec, la notion de handicap est définie, non pas par rapport à l'individu seul mais en fonction de ses relations avec la société, des barrières à l'intégration sociale qu'il rencontre. En d'autres mots, le degré d'accessibilité aux activités normales de la vie en société détermine la nature et le degré du handicap. Or, si les technologies sont adaptées, elles constituent un puissant instrument susceptible de réduire les limitations fonctionnelles et de favoriser ainsi l'intégration sociale de tous les individus.

Depuis plusieurs années, des organismes internationaux et nationaux s'emploient à définir des normes et standards pour orienter le développement technologique. À l'échelle internationale, il faut notamment relever les travaux de l'ISO (Organisation internationale de normalisation) et de ses membres, dont le Conseil canadien des normes, ainsi que ceux de la CEI (Commission électrotechnique internationale). Par ailleurs, des groupes d'utilisateurs, comme GSE (Guide SHARE Europe) revendiquent -- en publiant des manifestes -- le respect des langues et des cultures dans les technologies de l'information. Plus récemment, la Société Internet a aussi adopté des standards qui prennent en compte plusieurs facteurs culturels et linguistiques.

Le gouvernement du Québec participe activement à ces travaux. Il a de plus élaboré des outils pratiques pour la francisation des données informatiques, outils regroupés sous l'acronyme FRANCIS.

Ce recueil constitue un pas de plus. Il dresse la liste des conventions d'écriture en usage au Québec applicables aux technologies de l'information. Le recueil comprend huit parties, qui portent respectivement sur :

- . la terminologie des technologies de l'information;
- . la représentation des caractères;
- . le tri, la comparaison, la recherche et le classement des mots;
- . les éléments textuels français utilisés comme données structurées;
- . la présentation des textes français;
- . le balisage des textes;
- . les interfaces personne-machine;
- . les exigences légales et réglementaires.

Le recueil a été conçu à l'intention des concepteurs, des constructeurs et des distributeurs de matériel et de logiciels qui mènent des activités sur le marché québécois ainsi que de toute personne qui prépare et diffuse des documents sur support informatique au Québec.

I LA TERMINOLOGIE DES TECHNOLOGIES DE L'INFORMATION

Quand un secteur se développe, de nouveaux concepts naissent, de nouvelles réalités doivent être nommées dans toutes les langues nationales. La terminologie des technologies de l'information évolue rapidement. À l'échelle internationale, elle fait l'objet de nombreux travaux multilingues dont ceux, très importants, de l'ISO et d'un comité technique formé conjointement par l'ISO et la Commission électrotechnique internationale (voir ISO/CEI JTC1/SC1)

Au Québec, l'Office de la langue française mène également plusieurs travaux en collaboration avec d'autres organismes, notamment le Conseil du trésor et le CAFI (Comité d'action pour le français dans l'informatique). L'Office a publié plusieurs lexiques de termes techniques et des guides généraux pour soutenir la francisation. Il mène également des activités de normalisation sur les nouveaux termes techniques utilisés dans les technologies de l'information, en collaboration avec l'industrie.

L'Office participe en outre à des échanges internationaux avec d'autres pays francophones et collabore aux travaux d'organismes de la francophonie internationale. Enfin, il gère la Banque de terminologie du Québec, un dictionnaire informatisé de la terminologie technique en français. L'Office de la langue française et la Banque de terminologie du Québec sont accessibles dans le réseau Internet ([HTTP://WWW.olf.gouv.qc.ca](http://www.olf.gouv.qc.ca)).

II LES CARACTÈRES

Les ordinateurs ont été d'abord conçus pour traiter des nombres. Ce n'est que par la suite qu'ils ont été adaptés pour traiter les signes typographiques, c'est-à-dire les lettres, les signes de ponctuation et les signes diacritiques. On imagine tout le chemin parcouru quand on sait qu'aujourd'hui on dispose de logiciels de reconnaissance optique des caractères et d'outils de traitement de la voix.

Toutefois, les ordinateurs actuels ont gardé la caractéristique d'origine de leurs ancêtres. Encore aujourd'hui, tout signe typographique doit en effet être traduit en une valeur numérique, plus exactement en une suite unique de nombres binaires (bits), qui est interprétée par l'ordinateur. Selon les normes en usage, chaque signe typographique est représenté par huit bits, ou un octet. Des normes d'avant-garde vont au-delà de huit bits, mais le principe reste le même.

Les produits informatiques actuels permettent de représenter de mieux en mieux le code particulier de chaque langue nationale, dont le français. Il s'agit là d'un développement important, qui facilitera les communications dans le respect des traits culturels de chaque société.

1. Les systèmes d'écriture

Les systèmes d'écriture d'aujourd'hui sont le fruit de plusieurs siècles d'histoire qui témoignent d'autant de manières de penser, d'appréhender le monde. L'écriture a suivi plusieurs voies au cours des âges : les hiéroglyphes, les idéogrammes, les pictogrammes, les syllabaires et les alphabets.

Chaque code a évolué différemment et développé sa propre logique. Ainsi, selon les langues, les lettres de l'alphabet latin changent de prononciation :

en turc, la lettre *c* se prononce *dj*; en italien, selon la voyelle qui suit *k* ou *tch*; en français dans les mêmes circonstances, *k* ou *s*; l'Inuttitut utilise des caractères très anciens, pictographiques, ou encore un syllabaire inventé par les missionnaires pour enseigner la Bible aux Inuits.

Dans plusieurs langues, dont le français, les caractères accentués, la ponctuation ou les signes diacritiques constituent des sous-systèmes lentement élaborés, dont il faut absolument tenir compte dans les opérations de tri, de comparaison, de recherche, de classement, si l'on veut produire une information exacte et qui corresponde à l'usage.

Par ailleurs, pour des raisons culturelles, le développement technologique ne peut, en aucun cas, conduire à la radiation de plusieurs siècles d'histoire, comme ce fut le cas, notamment, pendant un quart de siècle alors que les premiers ordinateurs étaient incapables de reproduire les majuscules accentuées. Les logiciels conçus pour une langue donnée ou pour un territoire donné doivent donc comprendre le répertoire complet des caractères en usage.

2. La translittération

L'explosion de l'information et la multiplication des échanges entre les sociétés ont multiplié les besoins de traduction. Les technologies de l'information doivent donc être en mesure de passer d'un code d'écriture à un autre, c'est-à-dire de représenter les caractères d'un système d'écriture donné (alphabétique, syllabique...) par les caractères d'un autre système d'écriture. Ce processus s'appelle la translittération.

L'ISO a depuis longtemps édicté des règles de correspondance d'unité graphique à unité graphique : le même signe doit être rendu toujours de la même façon, et en principe un signe d'un alphabet ne doit correspondre qu'à un seul signe de l'autre. La norme ISO 3602 énonce que l'on ne peut rendre un signe par deux signes que lorsque l'alphabet latin n'offre pas de possibilité raisonnable de faire autrement. Toutefois, ces règles sont en

révision en Europe parce qu'elles sont loin de faire l'unanimité, la translittération étant un processus hautement dépendant des langues impliquées. Cet aspect délicat doit être considéré au Québec, en particulier pour la retranscription des noms propres étrangers dans l'écriture latine. Ceux qui s'intéressent à la question doivent se tenir informés de l'évolution des travaux européens et des travaux du Comité technique 46 (SC2) de l'ISO.

3. Les jeux de caractères

Un jeu de caractères est défini comme l'ensemble des combinaisons reconnues par une unité informatique pour représenter des caractères en vue de leur traitement ou de leur transmission.

La représentation du système d'écriture du français exige à lui seul un jeu de plus de 134 caractères :

- . les 26 lettres de l'alphabet, minuscules et majuscules (52 caractères);
- . les 14 lettres accentuées (28 caractères);
- . les dix chiffres (10 caractères);
- . les signes de ponctuation, communs à toutes les langues européennes (38 caractères);
- . les ligatures «ae» et «oe» (4 caractères);
- . les guillemets français, ouverts et fermés (2 caractères);
- . quelques lettres d'origine étrangère pour certains mots du français.

Dans un contexte multilingue, il faut toutefois prévoir idéalement l'utilisation d'autres caractères, soit à tout le moins l'ensemble des caractères communs aux langues qui utilisent l'alphabet latin affectés de

leurs signes diacritiques, auxquels il faut ajouter les caractères d'usage national et international. Cela représente au moins 180 caractères de plus que les caractères de l'alphabet latin non accentués. Au total, la norme Télétex de base, adoptée en 1980 par le Comité consultatif international télégraphique et téléphonique (CCITT) prévoit 330 caractères différents pour l'ensemble des langues qui utilisent l'alphabet latin. Les limites de la technologie actuelle obligent cependant à faire des compromis.

Au Québec, le compromis minimal qui puisse être fait c'est de garantir l'accès au répertoire complet de l'alphabet latin n^o 1 (ISO/CEI 8859-1). Dans les faits, cela conduit à recommander les jeux de caractères suivants, par ordre de priorité :

- . le codage ISO/CEI 8859-1, lorsque cela est possible;
- . le codage IBM 850, pour les applications DOS;
- . le codage IBM EBCDIC 037, sur ordinateurs de grande puissance;
- . tout autre codage qui comprenne les 191 caractères de l'alphabet latin n^o 1.

Ces recommandations valent pour les jeux de caractères à huit bits.

Récemment, l'ISO a adopté le premier jeu de caractères normalisé destiné à couvrir l'ensemble des langues : le code JUC, pour *Jeu universel de caractères codés sur plusieurs octets* (ISO/CEI 10646). Avec 33 000 caractères, le JUC couvre d'ores et déjà la plupart des systèmes d'écriture en usage d'aujourd'hui; il pourra théoriquement être augmenté jusqu'à deux milliards de caractères. Le nouveau code oriente l'évolution des technologies. Pour l'heure, il y a encore peu de produits sur le marché qui s'y conforment mais la situation pourrait évoluer rapidement, la plupart des manufacturiers ayant reconnu qu'il s'agit de la solution de l'avenir.

4. Les signes de ponctuation

Les signes de ponctuation usuels, soit la virgule, le point, le point-virgule, le point d'interrogation, le point d'exclamation, les points de suspension, les

parenthèses, les guillemets et les tirets, sont utilisés sensiblement pour les même fins en français que dans les autres langues latines et en anglais. Les règles d'espacement entre les signes de ponctuation et les mots sont toutefois différentes. Pour les règles usuelles en français, se référer à la section 4 de la partie V.

Sur l'utilisation des signes de ponctuation, deux particularités doivent être soulignées :

- . les guillemets français (en forme de chevrons « ») sont toujours utilisés pour encadrer une citation, mettre en valeur un mot, un groupe de mots ou une locution étrangère, ou encore pour signaler l'emploi douteux d'un mot ou d'une expression. Les guillemets anglais (*quotes*) ne sont généralement pas utilisés en français.

Note : Si la citation entre guillemets forme une phrase complète, le point doit précéder les guillemets finals. Par ailleurs, un espace doit suivre les guillemets ouvrants et un autre doit précéder les guillemets fermants.

- . Selon l'usage québécois, le tiret indique la nullité dans les tableaux, les rapports comptables, les factures, etc. alors que les guillemets indiquent la répétition. Cette convention est particulière au Québec; en France, c'est exactement l'inverse, le tiret marquant la répétition et les guillemets signifiant la nullité.

5. Les caractères spéciaux

- . *La barre oblique (/)*

- Elle sert à exprimer une fraction ou un rapport (3/4; km/h).

- Elle est utilisée dans la correspondance :
 - . pour faire état des références (N/Réf.);
 - . pour séparer les initiales d'identification (AL/mp);
 - . pour abréger l'expression «aux soins de» (a/s de).
- Elle sert à marquer une opposition (le cycle ouvert/fermé; les rapports dominant/dominé).

Note : On ne peut pas utiliser la barre oblique pour indiquer le féminin d'un mot, dans les désignations de métiers et professions. Dans ce cas, on écrira et la forme masculine et la forme féminine. La barre oblique n'est pas non plus indiquée pour séparer les éléments d'une date.

. *La perluète (&)*

La perluète est utilisée en français seulement dans les désignations de sociétés (Pouliot & Associés; Dupuis & Fils). Dans tous les autres cas, on lui préfère la conjonction « et ».

. *Le « a commercial » (@)*

Jusqu'à récemment, ce caractère n'était pas utilisé en français. Il est maintenant employé dans les adresses de courrier électronique du réseau Internet.

Dans l'adresse Internet, le @ agit comme aiguilleur : il délimite le nom de la personne et l'adresse du domaine. Ce symbole, qui remplit en quelque sorte la fonction de « timbre de la poste électronique », est reconnu et accepté par tous les serveurs Internet. D'autres symboles sont par ailleurs utilisées dans d'autres protocoles, par exemple le point virgule et le point d'exclamation.

III LES MOTS : TRI, CLASSEMENT, COMPARAISON, RECHERCHE, FUSION

Chaque langue a développé au cours de son histoire des modes de classement des mots qui lui sont propres. Ces méthodes de tri naturelles doivent être respectées, c'est-à-dire reproduites de façon automatique par les ordinateurs. Mais les ordinateurs et leurs programmes se sont longtemps révélés peu efficaces dans les opérations de tri et de classement alphabétique, et ce non seulement pour le français mais également pour plusieurs langues importantes comme l'allemand, l'espagnol, l'arabe, le chinois, le thaï et même l'anglais.

Afin de régler ce problème, le Canada a adopté une norme nationale en 1996 (CAN-CSA Z243.4.1 - 1996). Cette norme fait la synthèse des méthodes de classement suivies par les dictionnaires français et anglais et normalise un ordre de classement lexical prévisible à 100 %. La norme canadienne est également applicable intégralement pour l'allemand, le portugais, l'italien et le néerlandais, et elle nécessite une adaptation mineure pour l'espagnol.¹

1. Un système à quatre niveaux

Pour bien comprendre la solution adoptée par le Canada et le Québec, il faut garder à l'esprit que, dans l'univers informatique, les données alphabétiques doivent être traitées à la manière des nombres ou de façon analogue. Or, on sait que la représentation externe des nombres ne correspond pas aux structures de données internes propres aux ordinateurs. En informatique, les nombres réels sont plutôt codés selon des registres binaires à virgule flottante, qui comportent trois parties : un bit pour le

¹Au moment de mettre sous presse, la norme internationale ISO/CEI 14651 était en préparation; il s'agit d'une extension de la norme canadienne.

signe (pour déterminer s'il s'agit d'un nombre négatif et positif); un certain nombre de bits pour l'ordre de grandeur (exposant du nombre); des bits supplémentaires selon le niveau de précision désiré.

Par exemple, le nombre +2,5 peut être représenté sous la forme binaire suivante :

<i>Signe</i>	<i>Ordre de grandeur</i>	<i>Mantisse</i>
0	0000001	1100000 0000000
+	2^1	X 0,1100000...0000000 (nombre en base 2)

On peut évidemment tronquer l'une ou l'autre de ces informations; tout dépend du niveau de précision que l'on désire obtenir. Ainsi, dans cet exemple, si l'on ne veut se préoccuper que du signe du nombre, un seul bit est requis; si l'on a besoin de connaître son ordre de grandeur, quelques bits supplémentaires sont nécessaires, et ainsi de suite.

La même logique s'applique pour la représentation des caractères alphabétiques. Dans un système d'écriture, les lettres constituent l'information de base. Les signes diacritiques (accents, trémas, cédille...) apportent des précisions supplémentaires. D'autres éléments ajoutent encore d'autres précisions, par exemple l'emploi de majuscules et de minuscules, les signes de ponctuation, les caractères spéciaux et les espaces.

Il est possible de restructurer les données alphabétiques selon une méthode de représentation analogue à celle qui est utilisée pour les nombres. Pour le français, on a recours à un système à quatre niveaux, tel qu'il est illustré dans l'exemple qui suit.

Exemple

Le patronyme « La Bonté » sera traduit de la façon suivante :

<i>Information de base :</i>	<i>[code indiquant les caractères alphabétiques et les chiffres de la chaîne de caractères, le cas échéant] : labonte</i>
<i>1^{er} niveau de précision :</i>	<i>[code indiquant un accent sur le « e »]</i>
<i>2^e niveau de précision :</i>	<i>[majuscule] [minuscule] [majuscule] [...]</i>
<i>3^e niveau de précision :</i>	<i>[code indiquant qu'il y a un espace requis en troisième position]</i>

Ce système à quatre niveaux, codé en binaire, permet une foule d'applications, soit le tri, la comparaison, la recherche de l'information désirée et sa représentation dans la forme voulue : avec ou sans accents, en majuscules ou en minuscules, avec les espaces et les caractères spéciaux prévus. Le résultat des manipulations est toujours totalement prévisible. Enfin, le système est applicable à tout mécanisme de classement actuellement sur le marché, même ceux de conception très ancienne qui peuvent trier de l'information binaire directement.

1.1 La fusion de données accentuées et non accentuées

Pour fusionner des données accentuées et non accentuées provenant de fichiers différents, on s'en remettra à l'information de base, sans autre niveau de précision. Ainsi, dans l'exemple précité, « LABONTE », « LA BONTE », « LA BONTÉ » et « La Bonté » seront repérés dans le voisinage immédiat de la forme « labonte », qu'ils ont en commun.

2. Les outils FRANCIS

Le gouvernement du Québec a élaboré des outils logiciels qui permettent de franciser les données nominatives dans bon nombre d'environnements technologiques, selon la méthode qui vient d'être décrite sommairement. Ces outils correspondent en fait à des fonctions de traitement de chaînes de caractères (mots ou syntagmes) qui respectent les règles de classement des dictionnaires français.

Les outils FRANCIS sont de trois ordres : des fonctions de base, des fonctions d'exploitation et des fonctions de soutien.

2.1 Les fonctions de base

Les fonctions de base sont la « traduction de clés » et la « traduction de clés inverse ».

La fonction « traduction de clés » permet de réinterpréter une chaîne de caractères, c'est-à-dire de la convertir en une nouvelle séquence de nombres codés qui peut être utilisée directement pour la comparaison ou le tri. Cette nouvelle séquence, ou clé étendue, est formée de quatre sous-clés, selon le système à quatre niveaux décrit plus haut.

Sommairement, la conversion d'une chaîne de caractères se déroulera comme suit :

- réception de la chaîne d'entrée;
- sélection de la matrice de caractères correspondant à une page de code (spécifiée dans les paramètres d'appel);
- traduction de chaque caractère de la chaîne en utilisant l'équivalent trouvé dans la matrice;

- réduction des sous-clés « indices relatifs aux accents » et « indices relatifs aux majuscules et minuscules » (1^{er} et 2^e niveaux de précision) selon des règles strictes;
- concaténation de chacune des composantes de la clé;
- retour à la clé étendue.

La fonction « traduction de clés inverse » permet de reconstituer la chaîne de caractères originale à partir d'une séquence de nombres obtenue avec la fonction précédente.

Les fonctions de base sont conformes à la norme canadienne CAN/CSA Z243.4.1 - 1996.

2.2 Les fonctions d'exploitation des répertoires des données nominatives

Le gouvernement du Québec a constitué quatre répertoires pour l'utilisation des majuscules et des minuscules accentuées dans les grands systèmes, selon l'usage français. Ce sont :

- un répertoire des noms de personnes;
- un répertoire des prénoms;
- un répertoire des noms de voies publiques;
- un répertoire des noms de villes.

Des fonctions d'exploitation sont associées à ces répertoires afin de convertir des données alphabétiques non accentuées en leur équivalent accentué. Par exemple, « QUEBEC » deviendra « QUÉBEC » ou

« Québec », « LASALLE » deviendra « LaSalle », etc.

Seul le format des caractères (accent et hauteur de casse) est touché; les fonctions ne modifient pas l'orthographe des mots traités.

2.3 *Les fonctions de soutien*

Ces fonctions ont été élaborées pour soutenir le traitement de listes ou de chaînes de caractères accentuées. Elles sont au nombre de six :

- la fonction de comparaison (de deux chaînes de caractères);
- la fonction de conversion majuscule/minuscule des chaînes de caractères ;
- la fonction de fusion (à partir de deux listes triées au préalable);
- la fonction de repérage (qui permet de repérer une sous-chaîne dans une chaîne de caractère);
- la fonction de tri interne (qui permet de trier une liste en mémoire);
- la fonction de tri externe (qui permet de trier une liste sur support externe).

2.4 *L'environnement technologique*

- Les outils logiciels FRANCIS s'appliquent sur une plate-forme maxi-ordinateur dans un environnement MVS ou sur plate-forme UNIX, DOS, Windows et Macintosh.

3. Les particularités de l'ordre de classement du français

Vu l'ampleur des données qu'il faut souvent traiter en informatique, le classement alphabétique de mots français doit être très rigoureux et

dépasser les évidences intuitives, tout en les respectant. Les règles générales suivantes s'appliquent :

- . Pour les mots qui contiennent des lettres accentuées, on ne tient pas compte des accents, sauf pour les homographes absolus (mots qui s'écrivent avec les mêmes lettres), par ailleurs très nombreux en français.
- . L'ordre de classement des homographes est déterminé selon les règles suivantes : le caractère non accentué précède le caractère marqué d'un accent aigu, qui vient, dans l'ordre, avant l'accent grave, l'accent circonflexe et le tréma. Il est important de noter que c'est la dernière différence dans le mot -- quand il y en a plus d'une -- qui a préséance.

Exemples : « du » précédera « dû », « pêcher viendra avant « pêcher », etc.; par ailleurs, « pêche » précédera « péché ».

Pour traiter des documents anglais et français, selon l'usage canadien et français, on se référera à la norme CAN-CSA Z243.4.1 - 1996. Les règles édictées dans cette norme permettent un tri conforme aux conventions culturelles et une prévisibilité absolue des résultats.

IV LES ÉLÉMENTS TEXTUELS FRANÇAIS UTILISÉS COMME DONNÉES STRUCTURÉES

Les traditions d'usage ont donné lieu à des règles précises de présentation des divers éléments textuels utilisés comme données informatiques. D'autres règles d'écriture relèvent de la logique interne, syntaxique et textuelle, du français.

Les travaux de l'ISO et de ses partenaires ont permis d'adopter des normes de présentation pour les éléments linguistiques et numériques les plus courants dans les échanges internationaux : code de pays, code d'unités monétaires, etc. Quant aux éléments qui appartiennent en propre à une société, elles font l'objet de « locales ». De façon précise, *une locale est un sous-ensemble définissant l'environnement informatique de l'utilisateur lié aux conventions culturelles et linguistiques*. Ce sous-ensemble comprend une ou plusieurs catégories :

- . la classification des caractères et la conversion en majuscules-minuscules (LC-CTYPE);
- . l'ordre de classement (LC-COLLATE);
- . les formats pour la date et l'heure (LC-TIME);
- . les formats des données numériques et non monétaires (LC-MONETARY);
- . les formats des messages de diagnostic et d'information et des réponses interactives (LC-MESSAGES).

Les informations qui suivent sont conformes, le cas échéant, à la *locale* de langue française pour le Canada (qui est partie intégrante de la norme CAN/CSA Z243.230-1996), pour les aspects particuliers au Québec, et aux normes de l'ISO pour les aspects généraux.

1. Les noms de personnes

Les noms et prénoms s'écrivent en minuscules, sauf pour la première lettre qui est en majuscules. Quand un nom est précédé d'un mot-outil (article ou préposition), ce dernier commence généralement par une majuscule, quoique l'usage soit encore flottant pour certains patronymes. En fait, l'utilisation d'un mot-outil en minuscule devrait indiquer en principe une origine nobiliaire, mais cette règle n'est plus tellement respectée.

Exemples : Des Champs, De Margerie, De Foy, De Fontenay (et de Fontenay), De Granpré (et de Granpré), etc.

2. Les noms de sociétés et d'organismes publics ou privés

De façon générale, il est d'usage d'employer la majuscule seulement pour le premier terme des désignations de sociétés ou d'organismes.

Exemples : la Société générale de financement; l'Union des producteurs agricoles; la Société des alcools du Québec; l'Assemblée nationale.

Toutefois, dans le cas des ministères du gouvernement, l'usage est de mettre une majuscule au domaine qu'ils gèrent, ministère s'écrivant avec une minuscule.

Exemples : le ministère de la Santé et des Services sociaux; le ministère de l'Emploi.

Dans le cas des raisons sociales composées d'un générique et d'un spécifique, l'usage est de mettre une majuscule aux deux termes.

Exemples: les Ameublements Dion inc.; le Magasin d'alimentation Bernard enr.

Enfin, il en est de même pour les désignations officielles d'établissements scolaires.

Exemples : l'Université de Montréal; le Collège de Sherbrooke; le Collège Saint-Charles-Garnier.

3. Les adresses postales

Une adresse postale complète comprend le nom de la personne à qui l'envoi est adressé, le nom de la société ou de l'organisme auquel elle appartient le cas échéant, le numéro du domicile et le nom de la rue, le nom de la ville et de la province, le code postal, le nom du pays dans le cas d'un envoi à l'étranger. La présentation de l'adresse doit respecter les conventions suivantes :

- . chacun de ces éléments est placé sur une ligne (ils ne sont pas séparés par des virgules);
- . une virgule sépare le numéro du domicile du nom de la voie de communication;
- . le nom de la voie de communication doit toujours être précédé des mots *RUE*, *BOULEVARD (BOUL.)*, *AVENUE (AV.)*, *CÔTE*, *CHEMIN*, selon le cas;
- . lorsque les mentions *EST*, *OUEST*, *NORD*, *SUD* sont nécessaires, elles s'écrivent avec une majuscule et sont placées immédiatement après le nom de la rue;
- . le nom de la province s'écrit entre parenthèses, après le nom de la ville ou du village;

- . le code postal (alphanumérique) est composé de deux séries de trois caractères, séparées par un espace.

Exemple : *MONSIEUR PIERRE UNTEL
DIRECTION DES COMMUNICATIONS
ASSOCIATION DES MÉDECINS DU QUÉBEC
1554, BOUL. DE MAISONNEUVE OUEST
MONTRÉAL (QUÉBEC)
H2L 3R9*

Note : En vertu des nouvelles règles de Postes Canada, l'adresse devrait être écrite entièrement en majuscules, accentuées le cas échéant.

4. Les adresses électroniques

4.1 Les adresses de courrier électronique de l'Internet

La popularité de l'Internet a entraîné la généralisation des adresses de courrier électronique. En effet, pour accéder au réseau, chaque utilisateur doit avoir sa propre adresse.

Toutes les adresses de l'Internet sont compilées et gérées à l'échelle mondiale par le Network Information Centre (NIC). Le NIC délègue cependant une partie du travail à chaque pays. Au Canada, la responsabilité des adresses de l'Internet relève du Network Corporation Center. Au Canada, chaque service relié au réseau doit donc faire appel au NCC pour obtenir une classe d'adresses, qu'il pourra par la suite attribuer à ses clients, et pour s'enregistrer. Sur le territoire canadien, le NCC délègue à son tour la gestion des nomenclatures d'adresses à divers services. Au gouvernement du Québec par exemple, cette responsabilité est assumée par les Services gouvernementaux du Conseil du trésor.

Au Québec, les adresses électroniques de l'Internet incluent souvent les éléments suivants, dans l'ordre :

- . le nom propre de l'utilisateur en minuscules (prénom et nom au complet sans espace, nom abrégé ou encore pseudonyme)²;
- . le symbole @;
- . l'endroit (le nom, complet ou abrégé, de l'organisme ou du serveur), en minuscules ou en majuscules;
- . un nom de code à deux lettres pour indiquer la province (le « sous-domaine » dans le langage de l'Internet); *qc*, dans le cas du Québec, en vertu d'une norme adoptée par Postes Canada qui attribue à chaque province un code à deux lettres selon le modèle retenu par l'ISO pour les pays (ISO 3166)³;
- . un nom de code à deux lettres pour indiquer le pays (le « domaine » dans le langage de l'Internet) : *ca*, pour Canada, en vertu de la norme ISO 3166 (voir note 3).

Note : Un point sépare généralement ces éléments l'un de l'autre (sauf évidemment le @, qui marque lui-même une séparation entre deux éléments).

²Comme les adresses de l'Internet ne permettent pas l'usage des caractères accentués, certains préfèrent utiliser leurs initiales ou encore un pseudonyme.

³Il est à noter que la norme ISO 3166 précise que le code de pays doit être en majuscules, ce qui n'est pas discriminé dans les adresses électroniques de l'Internet.

Recommandations relatives au courrier électronique

Le serveur

. Tout serveur de courrier électronique branché à l'Internet devrait être mis à niveau pour soutenir le protocole ESMTP (*Extended Simple Mail Transfer Protocol*).

. Dans le choix d'un serveur, il est recommandé d'opter pour celui qui offre un soutien complet des communications en mode 8 bits. La majorité des problèmes de transmission dans les autres langues que l'anglais proviennent de communications en mode 7 bits.

Les logiciels

. D'ici à ce que le jeu universel de caractères sur plusieurs octets soit généralisé, tout logiciel de courrier électronique devrait être configuré pour transmettre selon le code ISO/CEI 8859-1. Ce jeu de caractères s'impose graduellement sur toutes les plates-formes reliées à l'Internet, autant en Europe de l'Ouest que dans les deux Amériques. Par ailleurs, il y a correspondance bi-univoque entre ce dernier et le code IBM 850, utilisé par certains logiciels de courrier sous DOS ou OS/2; il existe des polices qui permettent de visualiser les caractères codés selon IBM 850 sous Windows, même si ce système utilise normalement le code ISO/CEI 8859-1. Quant aux logiciels de courrier Macintosh, ils transmettent selon le code ISO/CEI 8859-1, même s'ils disposent d'un codage différent. La correspondance est donc assurée.

. Les logiciels de choix sont ceux qui offrent le soutien du standard de balisage MIME (*Multipurpose Internet Mail Extension*). Le MIME identifie les jeux de caractères transmis, les modes de codage ainsi que le genre de données (voix, image, sons, texte).

La transmission

. On évitera de transmettre en mode QP (*quoted printable*). Cette option de configuration des logiciels conformes à MIME transmet toutefois en mode 7 bits. Le résultat, c'est que les balises MIME

4.2 Les autres adresses de l'Internet

Plusieurs utilisateurs sont déjà familiers avec des adresses qui donnent accès à divers fichiers électroniques comme les Telnet, Gopher, FTP (File Transfer Protocol), Usenet et Archie et W3 (*World Wide Web*). Les adresses qui permettent d'accéder aux sites du W3 méritent qu'on s'y arrête parce qu'elles sont de plus en plus répandues.

Le premier élément d'une adresse Web est formé par le sigle HTTP. Ce sigle désigne le protocole qui permet l'emballage et le transfert des fichiers entre les serveurs du réseau. Le deuxième élément est constitué généralement par l'identification du Web, le sigle WWW. Les autres éléments donnent les précisions nécessaires selon une syntaxe analogue à celle des adresses du courrier électronique de l'Internet.

Exemples : L'adresse de la vitrine du gouvernement du Québec est :
<http://www.gouv.qc.ca>

*L'adresse du site de ministère de l'Agriculture,
des Pêcheries et de l'Alimentation du Québec est :*
<http://www.agr.gouv.qc.ca/mapaq/mapaq.html>

5. Les numéros de téléphone et de télécopie

Les numéros de téléphone sont composés en Amérique du Nord d'un indicatif régional, à trois chiffres, d'un code local, à trois chiffres également, et d'un code personnel, à quatre chiffres. Il est d'usage d'écrire l'indicatif régional entre parenthèses, suivi du code local et du code personnel, ces deux derniers éléments étant séparés par un trait d'union.

Exemple : (416) 524-8745

Entre usagers d'une même région téléphonique, on omet généralement d'indiquer le code régional. Par ailleurs, en contexte international, on s'en remet à la convention suivante : le code du pays est précédé du signe + et suivi du numéro de téléphone; le code 1 s'applique exceptionnellement à toute l'Amérique du Nord plutôt qu'à un seul pays.

Exemple : +1 416 524-8745

6. Les données numériques

6.1. L'écriture des nombres

Le système international d'unités (SI) est en vigueur au Québec depuis 1970. Il comporte les particularités suivantes :

- Les abréviations des unités de mesures ne sont pas suivies d'un point abrégatif et ne prennent jamais la marque du pluriel.

Exemples : km (kilomètre), cm (centimètre), mm (millimètres).

- Le signe décimal est la virgule (sur la ligne).

Exemples : 27,2 km; 4,35 l

- Les nombres de cinq chiffres et plus sont séparés par tranches de trois chiffres et ce, tant pour la partie des unités que pour celle des fractions. Il est recommandé d'utiliser soit un demi-cadratin ou un quart de cadratin typographiques comme séparateurs de tranches, soit un espace insécable (caractère 192 dans l'alphabet latin n^o 1) pour les applications informatiques.

Exemples : 1 807 915; 0,014 75 mais 4232; 0,0147.

6.2. Les dates

De façon générale, la présentation numérique d'une date doit être constituée de la manière suivante : quatre chiffres pour indiquer l'année, deux chiffres pour le mois et deux autres pour la journée. Si l'on veut utiliser des séparateurs, on peut mettre un trait d'union ou encore un espace entre ces trois éléments.

Exemple : 19960322 ou 1996 03 22 ou 1996-03-22.

Note : Avec l'arrivée de l'an 2000, l'usage de deux chiffres pour indiquer l'année est à proscrire, afin d'éviter toute confusion. Les systèmes informatiques qui ne respecteraient pas cette injonction devront être adaptés.

6.3 Les heures

La période de 24 heures constitue la base de référence pour la présentation numérique de l'heure. L'indication de l'heure est suivie de celle des minutes, les deux éléments étant séparés par les deux points (sans espace).

Exemple : 17:30 (et non 5 h 30 P.M., qui correspond à l'usage américain).

Lorsqu'on veut indiquer la date et l'heure dans une communication, on peut donc l'écrire comme suit : 1996032217:30 ou 1996 03 22 17:30 ou 1996-03-22-17:30.

Dans un texte suivi, on préférera cependant la présentation alphanumérique; dans ce cas, on utilise l'abréviation *h*, séparée du chiffre qui précède par un espace.

Exemple : Le 22 mars 1996, à 17 h 30.

Lorsqu'il s'agit d'une heure précise, on l'indique comme suit : 17 h (et non 17 h 00).

Quand il est nécessaire d'indiquer la durée exacte, on recourra alors aux abréviations du système international d'unités (*h, min, s*)

Exemple : L'athlète a couru le marathon en 2 h 13 min 23 s.

Pour plus de précision et quand le contexte le requiert, on peut indiquer les mentions HNE (heure normale de l'Est) ou HAE (heure avancée de l'Est). Le changement d'heure s'effectue au Québec à 2 h. le dimanche suivant le premier samedi d'avril et le dernier samedi d'octobre (CAN/CSA Z243.230-1996). L'heure normale de l'Est est en usage durant la période d'hiver.

Le fuseau horaire : Pour des raisons pratiques, tout le Québec continental est sur le même fuseau horaire, à l'heure de l'Est. Seules les Îles-de-la-Madeleine sont à l'heure atlantique (une heure plus tard). Le changement d'heure s'effectue au Québec à 2 h les dimanches suivant le premier samedi d'avril (heure avancée de l'Est) et le dernier samedi d'octobre (heure normale de l'Est). Le changement d'heure est régi au Canada par des lois provinciales (pour plus de détails, se référer à la norme CAN/CSA Z234.4-1989).

6.4 Les unités monétaires

Le symbole d'unité monétaire est placé après la valeur numérique (et ses décimales le cas échéant), et séparé de cette dernière par un espace.

Exemples : 250 \$; 24,95 \$; 1288,01 \$

Le symbole d'unité monétaire peut être précédé du symbole *M*, pour « million » ou encore *MM*, pour « milliard ». Les deux symboles sont alors accolés.

Exemples : 98 M\$; 12 MM\$.

Note : Ces usages sont réservés aux tableaux et aux textes de nature statistique ou financière.

Dans les échanges internationaux et lorsqu'on veut éviter toute ambiguïté, on s'en remettra à la norme ISO 4217, qui précise les codes des devises. Le code attribué au dollar canadien est CAD; le code USD désigne le dollar américain, le code FRF s'applique au franc français, et ainsi de suite. Le code précède immédiatement le chiffre; les deux éléments sont séparés par un espace (sur la même ligne).

Exemples : CAD 48; USD 789

7. Les abréviations

L'abréviation d'un mot se fait toujours avant une voyelle, lorsqu'il n'existe pas d'abréviation conventionnelle; le terme ainsi abrégé se termine par un point.

Exemples : div. (divers); compl. (complément); part. (partie).

Le français comporte par ailleurs plusieurs abréviations d'usage courant, notamment :

- . dans les formules de politesse : M^{lle} (mademoiselle); M^{me} (madame); M^{mes} (Mesdames); M. (monsieur); MM. (Messieurs); S.V.P. (s'il vous plaît);
- . pour les titres ou fonctions D^r (docteur); P^r (professeur);
- . dans les toponymes : S^t (saint); S^{te} (Sainte);
- . pour indiquer un ordre de classement : 1^{er} (premier); 2^e (deuxième); n^o ou n^{os} (numéro et numéros).

Note : On remarquera l'emploi du point abrégatif dans tous les cas où la dernière lettre du mot ne fait pas partie de l'abréviation. Sauf pour les exemples précités, les abréviations, tout comme les symboles d'unités de mesure, ne prennent pas la marque du pluriel.

Les majuscules accentuées des mots abrégés doivent également apparaître, le cas échéant, dans les abréviations.

Exemple : N.-É. (Nouvelle-Écosse).

8. Les sigles et les acronymes

Les sigles et les acronymes s'écrivent en majuscules non accentuées.

Exemples : REER (régime enregistré d'épargne retraite); ENAP (École nationale d'administration publique).

Note : Dans la langue parlée, les sigles se prononcent toutefois selon les règles de la phonétique française.

Quelques acronymes se sont avec le temps transformés en véritables mots; c'est le cas en particulier de *laser* et de *cégep* (collège d'enseignement général et professionnel). Ils sont alors traités comme tels.

9. Les coupures de mots en fin de ligne

De façon générale, les coupures de mots en fin de ligne doivent obéir à la division syllabique. Les règles que l'on peut utiliser dans les applications informatiques pour reconnaître les syllabes sont les suivantes :

- . De façon générale, une consonne placée entre deux voyelles introduit une nouvelle syllabe : *pa/ri/té*, *fé/bri/le*, *La/bra/dor*.
- . Lorsqu'il y a deux consonnes placées entre deux syllabes, la première appartient à la syllabe précédente, la seconde, à la syllabe suivante.

Toutefois *l* ou *r*, précédés d'une autre consonne (différente de *l* ou *r*), forment avec cette consonne un groupe inséparable (*bl*, *cl*, *fl*, *gl*, *pl*, *br*, *cr*, *dr*, *fr*, *gr*, *pr*, *tr*, *vr*).

Les groupes *ch*, *ph*, *gn* *th* sont également inséparables.

- . Lorsqu'il y a trois consonnes consécutives à l'intérieur d'un mot, ordinairement les deux premières terminent une syllabe et l'autre commence une nouvelle syllabe. Toutefois, les groupes *bl*, *cl*..., *br*, *cr*..., cités plus haut, commencent généralement un syllabe.
- . Dans le cas des mots composés, la division se fait entre les éléments (un élément ne pouvant être scindé, pour des raisons de compréhension) : *fer/à/cheval*; *court-/circuit*.

Les divisions à éviter :

- . les mots en fin de page;
- . les nombres, les pourcentages, les sigles, les dates;
- . les mots de moins de quatre lettres;
- . les coupures qui isolent une seule lettre sur une ligne (par exemple, é/cole);
- . les coupures après une apostrophe (par exemple, aujourd'hui);
- . les coupures qui isolent le titre de civilité ou le titre honorifique du nom propre (la coupure est toutefois permise entre le prénom et le nom;
- . les coupures entre deux voyelles; on peut toutefois, dans tous les cas, séparer un préfixe du radical du mot (par exemple, *pré/avis*, *anti/allergique*).

Malgré ces quelques règles, le français est moins complexe que l'anglais pour traiter les coupures de mots en fin de ligne. Aussi, au lieu de recourir aux dictionnaires comme on doit le faire avec les logiciels de langue anglaise, il est possible en français de prévoir un algorithme simple. Cette solution requiert moins d'espace de mémoire et améliore la rapidité du traitement.

V LA PRÉSENTATION DES TEXTES

Dans sa forme la plus simple, un texte est composé de trois parties : l'introduction, le développement et la conclusion. Dans sa forme la plus complexe, il comprend : un titre, des parties préliminaires, le corps du texte et des pages annexes. Les parties préliminaires qu'il est possible d'introduire sont, dans l'ordre : l'avant-propos, le sommaire, la table des matières, la table des illustrations, la liste des tableaux. Les annexes peuvent être, dans l'ordre : un ou des appendices, un ou des index, une bibliographie.

1. La pagination

La page de titre, les pages préliminaires ainsi que les pages d'intertitres (lorsque les intertitres apparaissent sur une page seule) ne sont pas paginées; toutes ces pages sont toutefois comptées dans la pagination. La pagination commence à partir du début de l'introduction et se termine à la fin des pages annexes, en incluant les illustrations et les tableaux, le cas échéant.

2. Les marges

Il n'y a pas de norme qui régit l'emploi des marges. L'usage québécois veut toutefois que la marge du haut soit d'environ 5 cm pour la première page de chacune des parties du corps du texte (introduction, développement ou même chacune des parties du développement --sections, chapitres --, conclusion) et de 2,5 cm pour les autres pages. La marge du bas ainsi que les marges de gauche et de droite varient généralement entre 2,5 et 4 cm.

Enfin, l'usage veut que la marge de gauche soit de dimension au moins égale, sinon supérieure, à la marge de droite.

Par ailleurs, afin de faciliter les échanges internationaux, on s'assurera que les textes puissent être imprimés autant sur du papier de format 17 x 22 cm que sur du papier de format A4 (standard international), ce dernier étant légèrement moins large. Des marges de 4 cm à droite et à gauche, et de 2,5 cm en haut et en bas apparaissent indiquées.

3. Les interlignes

Il est d'usage d'employer un interligne différent pour distinguer le corps du texte des divers ajouts. On doit recourir à l'interligne simple pour les éléments suivants : titre de plus d'une ligne, notes en bas de page, tableaux, citations détachées du texte, bibliographie. Pour les autres éléments du texte, on aura recours à un interligne supérieur. Enfin, l'interligne doit être au moins doublé entre chaque paragraphe du texte.

4. L'espacement

Les règles suivantes s'appliquent selon les divers signes de ponctuation :

- . Point : pas d'espace avant, un espace après.
- . Virgule, point-virgule, point abrégatif, point de suspension (à l'intérieur d'une phrase) : pas d'espace avant, un espace après.
- . Deux points, tirets : un espace avant, un espace après.
- . Point d'interrogation, point d'exclamation : pas d'espace avant, un espace après.
- . Trait d'union, barre oblique : pas d'espace avant ni après.

5. La division d'un texte

Un texte complexe peut être divisé de plusieurs manières. Une première méthode consiste à utiliser les diverses formes typographiques pour bien marquer les chapitres, sections, sous-sections : majuscules, minuscules en caractères gras, soulignement, italique, etc. On peut également avoir recours au système numérique international (1, 1.1, 1.1.1, etc.) ou encore au système de classement traditionnel utilisant lettres et chiffres, dans l'ordre suivant : les chiffres romains (I, II, III, IV...; les lettres majuscules (A, B, C, D...); les chiffres arabes (1, 2, 3, 4...); les lettres minuscules (a, b, c, d...). Ces éléments sont séparés de l'intertitre par un point, un tiret, un gros point, un losange ou même une parenthèse fermante dans le cas des deux derniers.

6. Les formats de papier

Dans l'usage courant, les formats les plus usuels au Québec sont :

- . le format 17 x 22 cm (8,5 x 11 po), pour les lettres, les rapports, les communiqués, les documents reprographiés, etc.
- . le format 17 x 28 cm (8,5 x 14 po), pour les documents juridiques et administratifs principalement.

Note : Le gouvernement du Québec a adopté des normes en matière d'impression, à l'intention de ses fournisseurs. Ces normes précisent trois niveaux de qualité : édition prestige, édition soignée, édition informative. De plus, le gouvernement exige des produits qui satisfont aux exigences des ISO 9000. Pour plus de renseignements, on consultera : *Le Guide de l'imprimerie et du papier, Normes gouvernementales en matière d'impression*, Gouvernement du Québec, 1995, 180 p.

VI LE BALISAGE DES TEXTES

Le balisage (ou marquage) des textes n'est pas une idée nouvelle. Traditionnellement, nous utilisons plusieurs marqueurs dans la communication écrite : les signes de ponctuation, les titres et intertitres, les paragraphes, la pagination, le soulignement, l'italique, etc. Avec l'arrivée de l'ordinateur, ces codes ont été reproduits dans l'environnement informatique.

Mais l'édition électronique a entraîné de nouveaux besoins. Des systèmes de balisage inédits sont devenus nécessaires afin de pouvoir échanger plus commodément des documents et de les publier à meilleur compte, sur divers supports : électronique, papier, disque compact. La norme SGML (*Standard Generalized Markup Language*) aussi désignée ISO/CEI 8879, offre plusieurs avantages. Elle a été élaborée au début des années 1980 et sa pertinence s'est imposée avec le développement de l'Internet, depuis 1989.

Le succès du W3 fournit en effet une preuve irréfutable de l'intérêt qui existe à adopter un système de marquage normalisé pour les documents. La réussite du W3 tient en effet à trois éléments : une méthode normalisée d'adressage (les URL), un protocole de communication client-serveur (le HTTP), et enfin une convention de balisage (le HTML). Or, on sait que le HTML (*Hypertext Markup Language*) est une application particulière du SGML, application destinée essentiellement aux serveurs du WWW.

Précisément, le SGML est un système de balisage descriptif, c'est-à-dire qu'il fournit des informations sur la structure du texte : titre, paragraphes, chapitres, références, etc. Diverses utilisations du document deviennent ainsi possibles : on peut, par exemple, concevoir des systèmes de repérage où la requête préciserait que les termes recherchés doivent se trouver dans le titre, les intertitres, etc.; on peut aussi passer plus facilement d'un mode

de publication à un autre, puisque le document n'est pas encombré de marques propres à un producteur particulier de logiciels (comme c'est le cas avec les logiciels de traitement de texte) ou encore de caractéristiques typographiques qui permettent une seule utilisation possible.

Le SGML présente beaucoup d'intérêt, en particulier pour l'échange de données sur les inforoutes et pour la gestion documentaire.

Le gouvernement du Québec soutient le développement du SGML. Plusieurs organismes, dont le Conseil du trésor, la Société de l'assurance automobile du Québec, la Régie des rentes du Québec, le Centre francophone d'étude et de recherche en information des organisations (CEFRIO) mènent divers travaux afin d'intégrer le SGML à leurs technologies et leurs activités.

Note : Pour le codage des caractères accentués dans les documents dirigés dans l'Internet, il est recommandé de coder comme un seul caractère les lettres accentuées (selon le jeu de caractères ISO/CEI 8859-1, qui constitue la norme minimale explicite) plutôt que comme des entités SGML; par exemple, on écrira « é » plutôt que « é ».

VII LES INTERFACES PERSONNE-MACHINE

Les utilisateurs des technologies de l'information doivent avoir accès à un poste de travail qui respecte les caractéristiques de leur langue d'usage. L'application de ce principe est devenue impérative avec la généralisation de la micro-informatique.

Depuis plusieurs années, le gouvernement du Québec encourage le développement et l'achat de dispositifs et d'équipements offrant un soutien adéquat du français. Il a notamment élaboré des règles à l'intention de ses fournisseurs afin de doter les ministères et les agences gouvernementales de postes de travail francisés.

1. Le poste de travail

Un poste de travail doit être francisé de manière cohérente, c'est-à-dire qu'il doit permettre d'harmoniser à la fois le codage, les échanges d'information et le traitement de l'information. Pour remplir ces fonctions, il doit comprendre les éléments suivants :

- . un clavier conforme à la norme CAN CSA Z243.200-1992 (niveau de conformité 2a); **[AJOUTER ICI UN HYPERLIEN: «À peu près tout ce qu'il faut savoir sur le clavier canadien: <http://www.olf.gouv.qc.ca/techno/pages/f06.html> et <http://www.olf.gouv.qc.ca/techno/images/clavier4.gif>]**
- . un pilote de clavier approprié, qui permet de reproduire le jeu de caractères complet du clavier, quels que soient la plate-forme et le codage utilisés;

- . des pilotes vidéo qui peuvent reproduire le jeu de caractères complet du clavier;
- . une imprimante dotée des mêmes attributs;
- . les versions françaises des logiciels, qui offrent un dialogue interactif entièrement francisé et qui permettent l'utilisation des caractères français à l'écran, dans les fichiers ainsi que dans les opérations de tri, de classement, de comparaison et de recherche⁴;
- . l'affichage des fonctions en français (claviers, écrans, imprimantes et autres éléments); on peut aussi avoir recours à des pictogrammes (se référer à la norme ISO 9995-7);
- . la documentation en français sur tous ces éléments.

⁴La recherche devrait idéalement permettre de repérer des lettres accentuées, même si ces lettres ne sont pas incluses dans la requête, et vice-versa. Ce besoin peut être satisfait si les fonctions sont conformes à la norme CAN/CSA 2243.200 - 1996 et à la norme ISO/CEI 14651 (en préparation au moment où ces lignes sont écrites).

Illustration du clavier canadien

VOIR PLUS HAUT, page 54 : [Hyperlien](#)

Poste de travail

VOIR Danielle Gagnon et Michèle Drouin-Rochette (8-6141)

2. Le dialogue personne-machine

Les systèmes d'exploitation doivent offrir à l'utilisateur un environnement convivial. Les menus, les commandes, les messages d'information et les outils d'aide doivent apparaître en français à l'écran. Le choix des icônes ne doit pas être marqué culturellement, mais plutôt s'aligner sur l'usage international. Enfin, l'emploi de couleurs doit être contrasté afin d'être facilement lisible par tous, notamment ceux qui sont affectés d'une des formes les plus fréquentes du daltonisme.

3. Les besoins des personnes qui ont des limitations fonctionnelles

L'utilisation des technologies de l'information est en train de se généraliser. Bientôt, la majorité des foyers québécois auront leur micro-ordinateur. Au travail, l'informatique a maintenant intégré plusieurs activités et fonctions. Dans la vie courante, les transactions bancaires, les activités commerciales -- même les plus simples comme l'achat d'un article dans un magasin de détail ou le paiement de l'addition au restaurant -- passent de plus en plus souvent par les technologies de l'information.

Les consommateurs, les travailleurs, et dans un avenir rapproché les citoyens, doivent être en mesure d'utiliser les technologies de l'information d'usage courant. Les technologies doivent donc être adaptées afin qu'elles soient faciles d'utilisation et, surtout, accessibles à tout le monde. Or, certains groupes de la population vivent avec des limitations fonctionnelles, pour des raisons d'âge ou encore parce qu'elles ont un handicap physique : troubles de la vue, de la motricité, de l'audition, etc.

Des interfaces sont maintenant disponibles pour les gens qui ont des troubles de la vue : interfaces graphiques (grossissement des caractères à l'écran), interfaces combinant le son et le toucher. Mais de grands besoins restent à combler à l'heure actuelle, notamment la représentation des caractères et la conception d'un clavier en braille.

D'autres besoins sont également pressants, comme le soutien à la téléphonie textuelle (pour les gens qui ont des troubles du langage), la standardisation internationale des caractères employés pour le télétexte (sous-titrage des émissions de télévision), la standardisation de l'utilisation des couleurs et autres effets spéciaux, le soutien au développement des langages symboliques utilisés par les personnes handicapées (Bliss, pictogrammes).

Plusieurs travaux sur ces questions sont menés par l'Union européenne, le cadre du programme TIDE (Telematics for Disabled and Elderly People). Ces travaux portent sur :

- . le braille : standardisation des caractères et conversion des textes en braille;
- . le Bliss : élaboration d'un jeu de caractères graphiques pour les symboles Bliss;
- . la représentation des langages par signes;
- . la téléphonie textuelle;
- . le télétexte (en vue de favoriser l'échange d'émissions de télévision accessibles aux malentendants);
- . les symboles pour la signature et l'adressage (utilisés, par exemple, pour indiquer qu'une application destinée aux personnes handicapées est disponible).

Le Québec suit l'évolution de ces travaux afin d'adapter et d'intégrer les résultats le plus rapidement possible à son environnement technologique.

VIII LES EXIGENCES LÉGALES ET RÉGLEMENTAIRES

Les lois ont pour but ultime de protéger les droits de tous les citoyens. La législation a dû être adaptée avec la généralisation des technologies de l'information, pour des raisons évidentes. Deux aspects cruciaux doivent être traités avec beaucoup d'attention : la propriété intellectuelle et les droits d'auteur; la protection des renseignements personnels.

1. La propriété intellectuelle et les droits d'auteur

Les brevets d'invention pour de nouvelles technologies sont délivrés par l'Office de la propriété intellectuelle, qui relève d'Industrie Canada (Ottawa). L'Office administre également le Bureau des droits d'auteur.

Outre les brevets d'invention, l'Office de la propriété intellectuelle émet également des titres pour :

- . les dessins industriels, pour la forme, le modèle ou l'ornementation d'un objet utilitaire manufacturé;
- . les marques de commerce : les mots, symboles ou dessins, ou toute combinaison de ces éléments utilisés pour distinguer les biens et services produits par une personne de ceux produits par une autre personne;
- . les topographies de circuits intégrés : la configuration tridimensionnelle des circuits électroniques utilisés dans des produits de circuit intégré, également appelés micropuces

Par ailleurs, le Bureau des droits d'auteur reçoit les demandes d'enregistrement des droits pour des oeuvres originales :

- . les oeuvres littéraires, incluant les programmes d'ordinateur;
- . les oeuvres dramatiques (films, vidéos...);
- . les oeuvres musicales;
- . les oeuvres artistiques (peintures, dessins, cartes, sculptures, photographies et oeuvres architecturales).

L'enregistrement officiel n'est pas nécessaire, le droit d'auteur étant automatiquement protégé au Canada. Mais il est préférable de demander un certificat d'enregistrement afin d'indiquer quelle œuvre visée est protégée.

Le Canada est signataire de la Convention de Berne et de la Convention universelle sur le droit d'auteur. Le droit d'auteur canadien est donc valable dans tous les pays qui ont adhéré à ces conventions; c'est le cas de la plupart des pays.

2. La protection des renseignements personnels

L'accès à l'information gouvernementale et la protection des renseignements personnels sont encadrés, au Québec, par deux lois.

La Loi sur l'accès aux documents des organismes publics et sur la protection des renseignements personnels cherche à établir un nécessaire équilibre entre la nécessaire transparence administrative et la préservation de la vie privée. Afin de protéger les citoyens, la loi établit le principe selon lequel tous les renseignements personnels sont confidentiels et ne doivent pas être divulgués sans le consentement de la personne concernée. Les organismes publics sont, en quelque sorte, « fiduciaires » de l'information qui leur est confiée. Selon ce principe, toute personne sur qui on détient des renseignements a le droit :

- . d'y avoir accès;
- . de pouvoir faire rectifier ces renseignements s'ils sont inexacts;
- . de donner son consentement avant que ces renseignements ne soient diffusés à des tiers.

Le même principe et les mêmes règles sont formulés dans la *Loi sur la protection des renseignements personnels dans le secteur privé*. La loi stipule par ailleurs des règles strictes relativement à la collecte, au stockage et à l'utilisation des renseignements personnels. Ainsi, elle oblige notamment les entreprises à :

- . prendre et appliquer des mesures de sécurité propres à assurer le caractère confidentiel des renseignements qu'elle détient;
- . veiller à ce que les renseignements qu'elle utilise au moment de prendre une décision touchant une personne soient à jour et exacts;
- . ne pas communiquer des renseignements personnels à des tiers sans le consentement explicite de la personne concernée.

Les objectifs poursuivis par le gouvernement du Québec quant à la transparence de l'Administration publique et la protection des renseignements personnels ont des implications importantes sur les technologies de l'information.

Les organismes responsables

- . L'Office de la propriété intellectuelle relève du ministère de l'Industrie du Canada. Ses bureaux sont à Ottawa.
- . La Commission d'accès à l'information relève du gouvernement du Québec. Elle a des bureaux à Québec et à Montréal.

CONCLUSION

Le domaine des technologies de l'information connaît un développement accéléré. Les choses bougent très vite. Les concepteurs, les distributeurs et les diffuseurs doivent se tenir continuellement au fait des dernières avancées, afin d'adapter continuellement les outils qu'ils mettent à la disposition du public. Plusieurs outils de recherche et, en particulier, le réseau Internet offrent maintenant l'accès instantané à l'information et permettent de suivre, presque au jour le jour, l'évolution des travaux nationaux et internationaux de normalisation.

Ces progrès constituent autant de pas vers la mise en place de moyens de communication qui respectent les codes et les pratiques en usage dans chaque société, et qui contribuent à leur épanouissement. Tel est le seul but qui est visé par ce recueil.

BIBLIOGRAPHIE

Normes internationales

ISO/CEI 2382	Technologies de l'information -- Vocabulaire
ISO 639	Codes de langues
ISO 3166:1993	Codes pour la représentation des noms de pays
ISO/DIS 3166-1:1995	Codes pour la représentation des pays et de leurs subdivisions -- Partie 1 : Codes pays (Révision de l'ISO 3166:1993)
ISO 4217:1995	Codes pour la représentation de monnaies et types de fonds
ISO/CEI 8859-1:1997	Traitement de l'information -- Jeux de caractères graphiques sur un seul octet -- Partie 1 : Alphabet latin n ^o 1
ISO/CEI 8879	Langage normalisé de balisage généralisé (SGML)
ISO/CEI 9995-7:1994	Technologies de l'information -- Disposition des claviers conçus pour la bureautique -- Partie 7 : Symboles employés pour la représentation des fonctions.
ISO/CEI 10646-1:1993	Jeu universel de caractères codés à plusieurs octets -- Partie 1 : Architecture et table multilingue

ISO/CEI 14755	Méthode de saisie de caractères du jeu universel de caractères à l'aide d'un clavier ou d'autres unités d'entrée
ISO/CEI CD 14651	Classement international et comparaison de chaînes de caractères utilisant le jeu universel de caractères codés pour plusieurs octets

Normes canadiennes

CAN/CSA Z243.230-1996	Conventions canadiennes minimales de localisation des logiciels
CAN/CSA Z243.4.1-1996	Méthode canadienne de classement alphabétique
CAN/CSA Z243.200-1992	Claviers canadiens pour le français et l'anglais

Ouvrages techniques

LaBonté, A. (Comité Canada-Québec sur le développement du logiciel d'expression française), *Guide conceptuel pour la réalisation de logiciels de langue française à portée multilingue*, 1987, 23 p.

LaBonté, A., *Les séquences de classement : du chinois à l'anglais en passant par le français* (exposé présenté à la rencontre anniversaire de SHARE European Association, à Aalborg), Québec, ministère des Communications, 1988, 12 p. et fig.

LaBonté, A., *Règles de classement alphabétique en langue française et procédure informatisée pour le tri*, Québec, ministère des Communications, 1988, 6 p.

LaBonté, A., *Fonction de systèmes Soutien des langues nationales*, Québec, ministère des Communications, 1988. 4 p.

LaBonté, A., *Technique de réduction Tris informatiques à quatre clés*, Québec, ministère des Communications, 1989, 6 p.

LaBonté, A., *Quand « Z » vient-il avant « a » Algorithme de tri respectant langues et cultures* (exposé présenté lors de la 75^e conférence de SHARE, à la Nouvelle-Orléans), Québec, ministère des Communications, 1990, 9 p. et fig.

Huard, G, et al., *Le SGML en documentation juridique et gouvernementale : potentiel et mise en oeuvre*, Les Publications du Québec, 1995, 86 p.

Québec (Gouvernement), *Le Guide de l'imprimerie et du papier*, Normes gouvernementales en matière d'impression, Québec, 1995, 180 p.

Québec (Ministère des Communications), *Francis -- Outils de francisation de donr*

Ouvrages généraux

Grevisse, M., *Le bon usage*, Éditions J. Duculot, Gembloux (Belgique), 1975, 1322 p.

Cajolet-Laganière, H. ET N. Guilloton, *Le français au bureau* (4^e éd. mise à jour et augm.), Cahiers de l'Office de la langue française, Les Publications du Québec, 1996.

Vocabulaires et lexiques sur les technologies de l'information, publiés par l'Office de la langue française*

Vocabulaire du micro-ordinateur

Vocabulaire de la bureautique

Lexique de la micro-électronique

Vocabulaire d'Internet

* *Ces ouvrages sont vendus dans les points de service des Publications du Québec.*