

Projet National Mobilisateur

Culture, Société et nouvelles technologies de l'information
(PNM/CSNTI)

Les référents culturels et linguistiques dans la manipulation des interfaces Homme-Machines arabo-latines : étude de cas des jeunes diplômés et des nouveaux professionnels Tunisiens

Mokhtar BEN HENDA

Maître Assistant, Université La Manouba, Institut Supérieur de Documentation

RESUME

L'arabisation des systèmes informatiques constitue désormais un enjeu stratégique au sein de la mondialisation des systèmes d'information et de communication. Le multilinguisme est aussi un créneau de recherche non moins important à travers lequel les langues nationales essaient de garantir leurs places à côté des langues dominantes comme l'anglais ou le français. Les solutions d'Internationalisation entreprises à échelle internationale ont répondu à ce besoin mais leur fiabilité reste aléatoire et limitée. Les Interfaces Homme-Machines bilingues (arabe/latin) constituent notre support d'enquête pour voir le degré de leur adhésion à la langue arabe d'un point de vue technique, culturel et cognitif chez des jeunes diplômés et professionnels tunisiens.

ABSTRACT

Nowadays, the arabisation of computer systems is representing a strategic challenge within the frame of Information and Communication Systems. Multilingualism is also a research framework for national languages to ensure their presence beside dominant languages like English or French. The worldwide solutions for the internationalization of these systems have provides relatively successful results but their reliability is still to be confirmed. Bilingual Human Computer Interfaces (arabic/latin) are our main research topic in this paper aimed at analyzing their adequacy with the Arabic language as apprehended from technical, cultural and cognitive points of view by recent graduates and young professionals in Tunisia.

INTRODUCTION

Deux grands phénomènes socioculturels accompagnent aujourd'hui le processus de la mondialisation : l'internationalisation et la localisation ; deux concepts identifiés dans les milieux des spécialistes par les acronymes I18n et L10n. Bien que diamétralement opposés, ces deux concepts restent intimement inter-reliés dans la conception présumée des précurseurs de la métamorphose de la société globalisante que nous vivons afin d'« ouvrir une ère nouvelle, riche de la singularité des groupes et des individus,

développant au maximum la coopération et l'intelligence collective sociale ... qui ouvrirait le chemin d'une idéologie de l'esprit, d'une écologie » [1].

Conjointement à ce phénomène global de mondialisation, un autre processus de métamorphose sociale est en train d'avoir lieu : la migration progressive de la société industrielle vers un nouveau modèle de société, celui de la société de l'information et des technologies nouvelles qui l'accompagnent. C'était les préférences d'Alvin Toffler et de Tom Surprenant qui annonçaient le non moins fameux concept de la société sans papier.

Or, à ce stade de progression de ces phénomènes sociaux, une série de questionnements s'imposent pour établir des sondages et effectuer des prélèvements sur l'adéquation de ces processus par rapport à leurs environnements et leurs contextes de mutation.

Le questionnement qu'on essaie de soulever dans ce document est d'ordre socioculturel et plus particulièrement linguistique : dans la confluence de la grande multitude de langues qui caractérise la communauté humaine de notre temps, comment cette mondialisation gère ce phénomène multilingue pour asseoir le principe d'In au sein des modèles, des mécanismes et des outils technologiques qu'elle essaie de fournir sur une base d'équilibre et d'équité à toutes les identités socioculturelles et linguistiques présentes ? En d'autres termes, comment l'outil technologique, en tant que maître d'œuvre de cette mondialisation, a-t-il fait face à cette diversité pour satisfaire toutes les aspirations et répondre à toutes les attentes des communautés linguistiques du monde ?

Le travail dans cette étude focalise le concept du multilinguisme arabo-latin pour étudier la profondeur à laquelle la langue arabe a été prise en considération dans les solutions multilingues que la communauté internationale a mises à sa disposition. L'objectif y est de mettre à découvert les niveaux de concordance et les points de divergence entre les solutions technologiques proposées et les soubassements socioculturels du contexte arabe.

La première interrogation posée est relative aux solutions techniques apportées à des strates différentes du processus du traitement de la langue arabe pour surpasser l'obstacle purement scripto-graphique. La langue arabe, comme support d'information scientifique et technique, est appelée à marquer sa présence dans l'univers informationnel et communicationnel mondial qui s'installe depuis les simples pratiques de traitement informatique en monoposte jusqu'aux fonctionnalités plus complexes d'Internet (messagerie électronique, WWW, FTP, Telnet...), de la gestion électronique des documents, de la reconnaissance optique des caractères... comme instrument véhiculaire d'une information écrite, elle est censée rester, comme toute langue de communication, au diapason des évolutions que connaissent les systèmes d'information et de communication en vigueur.

En effet, face à cette expansion et dominance des systèmes ouverts et distribués, comment la partie lourde du multilinguisme (formes typographiques et direction de l'écriture), par analogie à la partie souple (syntaxe et sémantique), trouve-t-elle progressivement sa voie dans les systèmes hybrides actuels d'information et de communication ? S'agit-il d'un fait de dépendance verticale d'une langue source dominante sur une langue cible ? Est-ce plutôt une correspondance horizontale entre deux ou plusieurs langues qui s'alternent ou « cohabitent » au niveau de la même interface ? Dans quelles conditions les solutions techniques déjà apportées au multilinguisme d'interfaçage opèrent-elles ? Associatives (présence simultanée au niveau de l'interface) ou dissociatives (par alternance) ? Pour étudier toutes ces hypothèses, il serait fondamental d'aller en profondeur dans des disciplines de recherche comme les théories linguistiques et la sémiologie, le traitement du signal, le codage binaire et l'algorithmique ; des domaines de recherche qui agissent directement ou indirectement sur le volet fondamental et important de l'adaptation des langues au domaine du traitement informatique (codage de l'information et de sa manipulation physique : saisie, interprétation, transformation, stockage, restitution, échange, reproduction...).

Le présent travail sera par contre limité au domaine particulier des interfaces hommes-machines dans un contexte multilingue arabe où le cognitif de l'homme fait face au procédural de la machine. A chacun ses origines et son référent : l'homme à culture orientaliste face à la machine aux origines occidentales. La question est donc de savoir à quel degré d'adaptation les deux protagonistes

[1] Jacques Robin. Les pièges d'une société de l'information planétaire. Le Monde Diplomatique. Hors-série, M 1909, octobre 1995. pp.67-68

parviennent-ils à créer l'harmonie du système d'information utilisé ; les solutions multilingues arabo-latines sont-elles greffées dans un univers qui apparemment les subit sans pouvoir les rejeter ou sont-elles au contraire introduites sur la base d'une équité objective entre les langues universelles ?

A toutes ces questions correspondront des analyses et des commentaires qui se baseront essentiellement sur un travail d'enquête menée auprès d'un échantillon d'individus particulièrement choisis parmi une tranche de population active, celle des jeunes en phases finales d'études ou en début de carrière professionnelle. L'objectif de ce choix est de cerner les aptitudes et les réactions cognitives d'une catégorie d'utilisateurs qui d'une part vit pleinement la métamorphose technologique de la société tunisienne et qui d'autre part se confronte à une montée en valeur de la langue arabe dans les milieux officiels de l'administration, de l'enseignement et des techniques en Tunisie. L'enquête a ciblé une centaine d'individus auxquels ont été soumises des situations problématiques de manipulation de plates-formes informatiques multilingues (arabo-latines) dans l'objectif d'étudier deux phénomènes essentiels : d'une part l'interaction psycho-cognitive et culturelle de cette population cible avec un système informatique conçu et développé dans une dimension linguistique et culturelle autre, en l'occurrence occidentale ; et d'autre part le taux d'adéquation des solutions multilingues universelles (i.e. Microsoft) avec les caractéristiques phonologiques, morphologiques et syntaxiques propres à la langue arabe.

I – LES ORIGINES ET LES ASPECTS TECHNIQUES DU MULTILINGUISME ARABE-LATIN.

Le fait d'aborder toutes ces questions d'un point de vue ergonomique des interfaces hommes-machines nécessiterait au préalable une présentation générale du concept du multilinguisme et des facteurs qui agissent sur son évolution. L'analyse du système linguistique de la langue arabe et de ses particularités scriptographiques par analogie avec les langues latines seraient un facteur de validation de certains phénomènes qui agissent sur le taux d'appréhension du citoyen arabe, et par conséquent tunisien, aux systèmes informatiques multilingues en cours et conditionnent son adaptation aux deux environnements linguistiques radicalement opposés à travers leurs formes graphiques différentes et leurs directionnalités d'écriture opposées.

1. LA POLEMIQUE MULTILINGUE : UNE HISTOIRE DE CARACTERES.

Toute l'histoire de l'informatique polyglotte remonte aux origines de la codification des caractères traités par l'ordinateur. Partant du principe de la codification binaire des données informatiques (codage par 0 et 1), tout ordinateur nécessite en effet des tables pré-installées de codes et de fontes (ou pages de codes et pages de fontes) pour reconnaître la valeur binaire et la forme graphique (glyphe) des caractères traités. Les Américains ont eu la primauté d'imposer leur propre table de codes ASCII (American Standard Code for Information Interchange) devenue un standard de fait pour le codage des caractères dans les systèmes informatiques du monde entier. Son utilisation limitée à un codage sur 7 bits (éléments binaires) constituait un obstacle à l'universalisation de l'outil informatique. Des extensions de code sur 7 et 8 éléments binaires, sur un ou plusieurs octets (caractères) ont permis la mise en place de tables de représentations multilingues. Ce fut l'ISO 8859 qui normalise une dizaine de jeux de caractères utiles pour un grand nombre de langues à écritures latines, cyrilliques, arabe, grecque et hébraïque, toutes disposant d'une série de caractères diacritiques qui n'avaient pas été pris en considération par l'ASCII. Ce fut ensuite une avalanche de normes nationales et internationales qui ont créé un problème de surabondance de normes donc de manque d'uniformité d'usage. C'est pour résoudre ce problème qu'ont été créés le standard UNICODE et la norme ISO-10646 sur la base d'un codage sur 16 et 32 éléments binaires avec l'objectif de rassembler toutes les langues du monde depuis l'antiquité dans une seule et unique table de codes.

Le problème de la diversité linguistique n'est pas ainsi dire complètement contrôlé. Le domaine du traitement linguistique peut en effet couvrir plusieurs sous-domaines de recherche : d'une part le traitement syntaxique et sémantique des contenus d'information, i.e. la signification des valeurs phonétiques des données, les modalités d'accès aux documents structurés... et d'autre part la représentation graphique et visuelle des contenus. Chaque domaine placera des valeurs différentes dans le traitement de l'information textuelle concernée. La recherche d'information textuelle dans une base de données nécessiterait des techniques complètement différentes de celles requises pour la représentation graphique des mêmes données. Dans le premier cas le traitement concerne plutôt le

contenu des données sans s'occuper de la forme de leur représentation. La définition de la structure d'un document en format SGML, ODA, Hytime... est un exemple d'une procédure de traitement de contenu qui reste indépendante des particularités relatives à la police de caractères. Par contre, le deuxième cas serait plutôt enclin à aborder les effets d'apparence et de présentation visuelle au cours de laquelle les ressources graphiques (fontes, polices, techniques de codage...) seraient très demandées. C'est en quelque sorte le domaine des interfaces graphiques et des méthodes d'affichage synoptique^[2].

De la détermination de ces deux approches dans le traitement des données textuelles, deux domaines de recherche sont actuellement développés en bifurcation d'une discipline générique, en l'occurrence le traitement linguistique de l'information. Le premier est tributaire du domaine de l'analyse structurelle des documents et des méthodes de repérage : interfaces de recherche, interrogation en langage naturel, structure de contenu... Le deuxième domaine dont l'axe central est le format de la représentation du contenu (graphic rendering) couvre l'aspect ergonomique et interactif avec l'utilisateur final à travers la convivialité de la couche interface supérieure des applications : format d'écran, organisation spatiale des données, densité surfacique et informative, multifenêtrage, gestion des couleurs et des contrastes, infographie et codage des données ...

Dans cet enchevêtrement de concepts techniques, notre intérêt porte sur l'un des éléments qui constituent la jonction entre les trois protagonistes rassemblés (information, technologie, communication), en l'occurrence les Interfaces automatisées de dialogue Homme-Machine dans un contexte aussi instable et aléatoire que celui des pays de langue arabe. Notre intérêt sera orienté particulièrement vers des états et des faits relatifs à la société tunisienne. C'est en réalité le point névralgique au niveau duquel se définissent les opérateurs réels de la réussite ou l'échec de l'action d'intégration des forces de la recherche et de la communication scientifique et technique.

Nous nous limitons dans cette étude à l'analyse de la représentation graphique plutôt que de la structuration linguistique du contenu. Nous abordons cette approche du point de vue de l'interopérabilité des systèmes au niveau des différentes phases de traitement technologique, de l'information et de la communication et des solutions fournies et utilisées de part et d'autre (producteurs de ressources, pourvoyeurs de services d'accès, concepteurs de médias, utilisateurs de données) pour répondre aux exigences informatives et communicationnelles d'un contexte bien particulier, celui des environnements multilingues arabo-latins comme c'est le cas de la communauté tunisienne.

2. LA POLEMIQUE DES INTERFACES : UNE HISTOIRE DE BIDIRECTIONNALITE.

Répondre aux interrogations précédentes relatives aux systèmes de codages et de représentation des caractères multilingues n'indique pas pour autant que l'on puisse venir à bout de la polémique des systèmes d'information multilingues. De nos jours, tout un nouveau volet de recherche est en train de prendre place dans la confluence de la polémique linguistique. Il s'agit du domaine non moins important des Interfaces Hommes Machines (IHM), de leurs architectures et de leurs ergonomies qui répondent à des standards de fait en matières d'accès et de diffusion de l'IST. Il y a en effet toute une série de standards, de règles à respecter pour construire une interface utilisateur et tester son degré d'ergonomie et d'acceptabilité par les usagers. Il s'agit là entre autres d'une appréciation des deux critères d'utilité et d'utilisabilité fixés par GRISLIN et KOLSKI comme « *propriétés pertinentes pour le diagnostic de la qualité ergonomique* » de l'IHM^[3].

Si le critère de l'utilité des interfaces est en quelque sorte imposé par les solutions informatiques multilingues des grandes firmes multinationale, l'utilisabilité reste néanmoins tributaire des facultés et des capacités psychocognitives de l'utilisateur arabophone soumis à des inférences socioculturelles bien particulières. A part l'élément graphique (formes des glyphes) qui distinguent les écritures latines de l'écriture arabe, le sens de la bidirectionnalité scripturale, issue de la double culture arabophone et latine de notre population cible, constitue aussi un élément fondamental de ces mêmes inférences. Les deux cultures se démarquent en fait par deux systèmes d'écriture à directionnalité diamétralement

[2] Christophe KOLSKI. Ingénierie des interfaces Homme-Machine : conception et évaluation. Paris : Hermès, 1993.

[3] Martial GRISLIN, Christophe KOLSKI. Évaluation des interfaces Homme-Machine lors du développement des systèmes interactifs. Technique et Science informatiques. Vol.15-n°3/199. p.266

opposée qui certes enrichissent le soubassement culturel de l'utilisateur mais engendrent également des conflits de perception et de logique de suivi (visuel et mental) face à des supports multilingues complexes (bris de lignes et insertion des textes). Si nous abordons cet aspect d'IHM dans notre étude, c'est parce qu'il a toujours été l'un des problèmes les plus épineux pour les utilisateurs finaux autour d'un système informatique bilingue arabo-latin. L'enquête décrite ci-après relève des échantillons de ces difficultés de dialogue homme-machine.

Si nous nous orientons vers cette optique d'évaluation ce n'est nullement pour reprendre l'état de l'art en matière de conception d'interfaces homme-machine, mais plutôt pour étudier l'état de l'art des interfaces (bi)multilingues qui dominent le contexte d'usage de l'information scientifique et technique dans notre contexte régional arabe et national tunisien. Notre champ d'action restera par conséquent rattaché aux solutions logicielles d'une grande envergure d'usage car ; à notre sens, c'est là que se profilent les changements et les évolutions profondes des interfaces Homme-Machine. C'est également à travers cette gamme d'interfaces Homme-Machine que se confirment les valeurs socioculturelles et économiques des actes d'information et de communication entreprises par les individus et les groupes d'individus.

Seulement, puisque la technologie, comme nous l'avons déjà signalée, est un vecteur de suprématie d'une culture d'origine, en l'occurrence celle du monde à dominante linguistique latine, l'information ne saurait échapper à la contrainte qu'exerce la langue du support sur la langue du contenu. Ceci explique les retards observés dans les domaines de la codification électronique des données, de la normalisation des procédés de transcription et d'échange et de l'industrie de la langue orientées vers les langues d'origines non latines.

Notre objectif, dans ce cas particulier, est de démontrer à quel point cette « fatalité » de dépendance à été surmontée (nous dirions plutôt contournée), pour permettre à des langues vernaculaires comme l'arabe, apparemment inadaptées (sur un plan technologique), d'accéder progressivement à l'univers électronique des systèmes d'information et de communication distribués.

Cependant, si le dit contexte relève en apparence d'un paradigme technoculturel évident (Nouvelles Technologies d'une part et Information de l'autre), une large partie du handicap que connaît la langue arabe au niveau de son usage électronique incombe à la réalité des usages qu'on en fait. Comme nous le montrons au courant de notre enquête, des considérations d'ordre culturel et historique agissent de manière très évidente sur la nature et les limites du bilinguisme en vigueur.

II – L'APPROCHE EXPERIMENTALE.

L'enquête menée vise essentiellement d'étudier le concept de l'utilisabilité de la langue arabe au niveau des interfaces multilingues arabe/latin par une tranche de la population active tunisienne. L'objectif de ce questionnement est de soulever les problématiques des différentes formes d'inadaptation entre cette catégorie d'interfaces et les utilisateurs potentiels.

Nous avons été sensibles tout au long de nos travaux de recherche à l'absence d'une littérature abondante et spécialisée axée sur la culture informatique arabe ; sur le questionnement des solutions informatiques proposées actuellement comme consensus implicite et leur confrontation avec les standards de fait de la langue, de ses modes d'usage tels qu'ils sont culturellement établis dans la perception collective et cognitive de ses usagers potentiels.

Généralement conçues par des concepteurs étrangers à la langue ou du moins partiellement connaisseurs des règles linguistiques, calligraphiques et culturelles de la langue arabe, les IHM arabe/latin présentent à notre sens un degré assez important d'incohérence et d'inadéquation entre le système multilingue et l'utilisateur arabophone. Ces incohérences peuvent avoir à notre sens de grandes répercussions sur l'acceptabilité, l'utilisabilité et l'efficacité même de l'interface. Par la soumission de ces solutions ad hoc aux mécanismes linguistiques de fait et d'usage, nous soulèverons beaucoup d'anomalies dont certaines sont justifiées par le référent socioculturel des usagers et d'autres sont induites par l'influence et l'interférence de disciplines connexes (sociolinguistique, transfert de technologies, politiques culturelles...).

1. METHODOLOGIE DE LA RECHERCHE

Faisant partie du domaine de la recherche expérimentale dans le domaine des interfaces Homme-machine, notre étude fait automatiquement recours à l'une des méthodes d'usage pour l'évaluation d'IHM. Nielsen et Molich avancent la théorie des approches dites prédictives et des approches dites expérimentales^[4] alors que Senach établit une distinction entre approches analytiques et approches empiriques^[5] ; Sweeney et Maguire proposent quant à eux trois méthodes basées sur l'utilisateur, la théorie et l'expertise^[6] ; Whitefield et ses collègues organisent leur classification en fonction de la présence ou non de l'utilisateur d'une part, et du système interactif à évaluer d'autre part^[7] ; Coutaz et Balbo^[8] centrent leurs techniques sur plusieurs paramètres tels que le degré d'automatisation, la prise en compte de l'utilisateur, ou encore le savoir utilisé pour l'évaluation.

Notre approche sera quant à elle inspirée essentiellement du premier des trois modèles d'évaluation élaborés par GRISLIN et KOLSKI et qui reposent sur trois catégories d'approches fondamentales à savoir les approches centrées sur les utilisateurs ; les approches centrées sur une expertise humaine ou « papier » et les approches centrées sur une modélisation de l'IHM et/ou de l'interaction Homme-Machine.

Pour accomplir cette évaluation centrée sur les utilisateurs, un questionnaire a été conçu pour être rempli par un enquêteur qui observe un sujet soumis aux opérations de difficultés décrites ci-après. Il s'agit d'une observation et d'un prélèvement d'attitudes et de comportements psycho-cognitifs de l'utilisateur face à une interface bilingue. L'interface choisie est celle de Microsoft Word de la suite Office 97. Ce choix émane de la popularité du produit et de son envergure d'usage quasi unanime parmi la population cible. Des zones de difficultés ont été conçues dans les échantillons d'exercices proposés. L'enquêteur est censé faire l'observation des comportements et des attitudes du sujet conformément aux critères d'évaluation préalablement définis.

L'enquête se présente ainsi sous forme d'une observation des personnes interrogées soumises à des situations de difficultés de traitement de texte automatisé en mode bilingue. Il s'agit en quelque sorte d'un mécanisme de traçage des attitudes et des comportements des uns et des autres pour résoudre ces cas de difficultés préalablement conçues en présence d'un référentiel socioculturel et linguistique particulier et d'une prédisposition psychocognitive singulière. Cet échantillonnage est ensuite croisé avec les variables relatives à l'identification de chacun des participants afin de déterminer les éléments de sortie nécessaires à notre évaluation.

Il est prévu que la conclusion de ce travail mettrait en exergue les problématiques d'usage de toutes les solutions multilingues proposées. La juxtaposition entre l'état de l'art en termes de solutions bilingues concernant l'arabe sera confrontée avec l'état de l'art d'une langue qui paraît parfois se contenter d'accepter la solution informatique développée dans des sphères culturelles qui lui sont étrangères. Ces anomalies seront traduites par des exemples puisés au niveau de certains exemples d'interfaces Homme-Machine arabe/latin d'usage actuel. L'objectif de ce questionnement est avant tout épistémologique : soulever une problématique de recherche rarement exposée au moment de la conception de produits multilingues pour lesquels la langue arabe est prise comme partenaire linguistique.

2. LA PROBLEMATIQUE

-
- [4] J. NIELSEN, R. MOLICH, heuristic evaluation of users interfaces, Actes de colloque CHI'90, Seattle, ACM New York, pp.349-356, 1990
- [5] B. SENACH, P. ALENGRY. Évaluation d'un dispositif d'assistance à la conduite et au dépannage de chaînes de fabrication automatisées. rapport INRIA n°163. Mars 1994.
- [6] M. SWEENEY et al. Evaluating user-computer interaction : a framework. International Journal Man-Machine Studies, 38, 689-711, 1993.
- [7] WHITEFIELD et al. A framework for human factors evaluation. Behavior & Information Technology, vol. 10, n°1, pp.56-79, 1991.
- [8] J. COUTAZ & S. BALBOS. Évaluation des interfaces utilisateur : taxonomie et recommandations. Actes des sixièmes journées sur l'ingénierie des interfaces Homme-Machine, IHM'94, 211-218, Lille, 8-9 décembre 1994.

Le questionnaire est structuré en zones qui retracent les grands axes de l'expérience. Après un prélèvement des caractéristiques du profil du sujet d'un point de vue formation et maîtrise de l'environnement linguistique et informatique, le questionnaire prévoit une série de zones de difficultés qui ciblent successivement le paramétrage de l'interface bilingue et les modalités de saisie et d'organisation des contenus d'information bilingue.

Ces épreuves de difficultés ont été orientées sur quatre thèmes fondamentaux qui constituent la problématique fondamentale de l'interaction de l'utilisateur arabophone avec une interface bilingue comme conçue actuellement par Microsoft :

2.1 CARACTERES NEUTRES ET CHIFFRES CONTEXTUELS

Définitions :

L'une des difficultés majeures toujours observées dans l'interaction avec une interface de saisie bilingue est sans doute le traitement des chiffres et des caractères neutres.

Par caractères neutres il est fait référence aux caractères qui n'ont pas une directionnalité propre, et qui s'alignent par conséquent en fonction de la direction globale du texte. Les caractères neutres prennent les attributs des caractères qui les entourent et s'affiche donc dans une directionnalité relative. Les caractères neutres sont classés en trois catégories distinctes^[9] : les symboles de ponctuation avec le caractère d'espacement et les paires de caractères symétriques, les chiffres et les séparateurs décimaux et enfin les symboles de traçage de lignes ou de boîtes.

Les signes de ponctuation sont généralement des caractères neutres qui prennent la direction du mode d'édition maître actif (Mode Maître Arabe ou Mode Maître Latin=langue et directionnalité active). Une exception peut être observée au niveau des trois caractères contextuels "Virgule", "Point virgule" et "Point d'interrogation" qui, s'ils prennent la direction du mode maître actif, changent aussi en forme du glyphe (typographie) d'un contexte à un autre. La virgule et le point virgule pivotent verticalement ; le point d'interrogation pivote horizontalement. Leur neutralité n'est donc pas pratiquée car ils se relaient en duos avec leurs équivalents linguistiques en fonction du mode maître actif.

Les chiffres constituent également un élément de difficulté majeure dans le processus de saisie et de traitement informatique. Même s'ils prennent la direction du mode maître actif, le mécanisme d'organisation interne des chiffres reste fidèle à la directionnalité gauche-droite définie par les algorithmes internes des machines ; ce qui rend leur traitement ambiguë dans un mode maître arabe à directionnalité Droite-Gauche. L'adjonction des séparateurs décimaux accentue davantage la difficulté d'alignement des blocs de chiffres dans une directionnalité Droite-Gauche car elle engendre, comme il est démontré dans le chapitre suivant, un effet de pivotement (Swapping) complètement déstabilisateur pour un grand nombre de personnes interrogées.

L'effet du pivotement graphique dans un texte Bidi (Swapping)

En effet, l'une des conséquences majeures de la bidirectionnalité est l'effet de pivotement observé au niveau du rendu visuel d'un texte Bidi. Le pivotement s'obtient au niveau de la chaîne des caractères dans un mot. C'est le "swapping" graphique que nous expliquons ce-après. Il s'obtient aussi au niveau d'une suite de mots. C'est le "Swapping" sémantique qui sera démontré dans la section suivante relative aux incises et aux bris.

Le pivotement graphique est le résultat évident de deux facteurs : les caractères neutres et les paires des caractères symétriques qui accompagnent le texte Bidi. Dans l'exemple suivant accompli sous un environnement MS-DOS, il est démontré comment les caractères neutres (chiffres et signes de ponctuation simples ou composés, uniques ou en paires) pivotent de gauche vers la droite de la chaîne de caractères textuels en fonction du mode maître actif (Latin=MML ou Arabe=MMA).

^[9] Microsoft MS DOS User's Guide : Arabic Supplement. Microsoft Corporation, 1988. p.18

MML		MMA
#2302	1	2302 #
/auteur	2	auteur/
قطعة 25	3	قطعة 25
مختارة	4	مختارة
?question	5	question?
[Et.al.]	6	{.Et.al}
[تاريخ]	7	{تاريخ}
{champs}	8	{champs}
{حقل}	9	{حقل}

L'effet de Swapping des caractères neutres sou MS-DOS

- Les lignes 4 et 5 montrent l'alternance entre les duos de caractères contextuels de ponctuation d'un mode maître à l'autre;
- La ligne 5 illustre le cas d'un **pivotement vertical** unique d'un caractère contextuel, en l'occurrence le point d'interrogation. Un pivotement en double est également obtenu par les caractères « virgule » et « point virgule » arabes qui **pivotent verticalement et horizontalement** afin d'apparaître conformément à la contextualité arabe. D'ailleurs, dans la table de code étendue ASMO 708 (équivalente à ISO 8859-6) les caractères spéciaux n'ont pas été reproduits dans la zone étendue excepté les caractères particuliers du signe monétaire international (position 10/04) qui substitue le symbole Dollar (position 02/04), le point d'interrogation (position 11/15), caractère contextuel à pivotement vertical par rapport au point d'interrogation latin (position 03/15), la virgule (10/12) et le point virgule (11/11), deux caractères à pivotement horizontal et vertical par rapport aux caractères latin de même fonction (03/11 & 02/12).
- Les lignes 6 et 7 montrent l'effet de pivotement des paires de caractères symétriques. Le crochet d'ouverture "[" en MML a subi un effet de pivotement en MMA pour garder la même fonction d'ouverture;
- Les deux points de la ligne 7, qui marquent normalement les abréviations, ont subi l'effet de pivotement en MMA afin de garder leur position de fin de mots;
- Dans les exemples 8 et 9, le pivotement n'est pas visible bien qu'il ait eu lieu, car la paire des caractères symétriques ne change pas de visu dans un swapping horizontal ni vertical.

Sous MS Windows par contre, les algorithmes de l'analyse contextuelle et du rendu visuel ou imprimé agissent différemment sur cet aspect de pivotement graphique. L'exemple suivant effectué sur MS Excel illustre cette variabilité dans la direction de l'écriture et l'orientation du texte.

E	D	C	B	A	
Context/GD	Context/DG	Justif.Context	Justif.Gauche	Justif.Droite	
1998	1998	1998	1998	1998	1
#2302	2302#	#2302	#2302	#2302	2
/Auteur	Auteur/	/Auteur	/Auteur	/Auteur	3
/مؤلف	/مؤلف	/مؤلف	/مؤلف	/مؤلف	4
?Question	Question?	?Question	?Question	?Question	5
[Et.al.]	{.Et.al}	[Et.al.]	[Et.al.]	[Et.al.]	6
[تاريخ]	{تاريخ}	[تاريخ]	[تاريخ]	[تاريخ]	7
{Champs}	{Champs}	{Champs}	{Champs}	{Champs}	8
{حقل}	{حقل}	{حقل}	{حقل}	{حقل}	9
حرف	حرف	حرف	حرف	حرف	10
سؤال	سؤال	سؤال	سؤال	سؤال	11
قطعة 25	قطعة 25	قطعة 25	قطعة 25	قطعة 25	12
25pièces	pièces25	25 pièces	25 pièces	25 pièces	13
					14

Effet de Swapping sous MS Excel

- Les colonnes A et B démontrent le déplacement en bloc des chaînes de caractères sans aucun pivotement lorsque la justification est forcée à gauche ou à droite ;

- La colonne C soumet les chaînes de caractères à une justification contextuelle. Seuls les termes commençant par les caractères contextuels arabes "point virgule" et "point d'interrogation" affichent une justification droite. Tous les autres caractères spéciaux, même s'ils précèdent des chaînes de caractères arabes, affichent une justification gauche. Ceci dénote de leur codes binaires puisés dans la partie gauche de la table des codes pour les deux modes arabe et latin.
- La colonne D démontre que lorsque le sens de l'écriture est forcé de droite à gauche, la position des caractères spéciaux passe à droite des termes qu'ils introduisent (i.e. lignes de 8 à 9) ou qu'ils terminent (i.e. ligne 7) ;
- Dans la même colonne D, l'exemple 9/D ne semble pas avoir subi de pivotement car la paire de caractères symétriques n'est pas perceptible dans aucun mouvement de pivot, mais les exemples 7/D et 10/D démontrent le contraire ;
- Entre les cas 2/D et 15 ou 14/D une analogie est également possible : dans le premier cas le chiffre n'est pas traité comme caractère neutre car le dièse qui le précède (#) l'est aussi. Ce qui fait que entre deux caractères neutres, le premier dans la chaîne des caractères pivote autour de l'autre. Les exemples 14/D et 15/D illustrent cette priorité : les chiffres dans ces cas précis pivotent autour des lettres qui sont purement contextuelles.
- La colonne E présente cependant une anomalie à ces règles d'affichage : les caractères spéciaux contextuels arabes ne subissent pas le même effet de l'algorithme Bidi qui les exclut du pivotement bidirectionnel. Les exemples des lignes 11 et 12 sont restés toujours constants.
- Une autre anomalie (par contre justifiée cette fois) est observée également entre les lignes 1, 13 et 14 : la position et la direction des chiffres.
- À part le fait que le chiffre de la ligne 1 soit trié en priorité, le fait qu'il soit aligné à droite malgré le changement du sens de l'écriture (Colonne C) et malgré sa propre direction interne (de gauche à droite), dénote d'un détail très important pour notre étude : les chiffres absolus (non suivis de caractères alphabétiques) sont automatiquement justifiés à droite dans une feuille de calcul ou dans un champ numérique d'une table de bases de données (ce n'est pas le cas de MS Word!).

2.2 INCISES ET BRIS DE TEXTE : GESTION BIDI

Définition des concepts

Les incises sont les textes dans une langue déterminée insérés au milieu d'un texte d'une langue différente. Le phénomène apparaît tout à fait normal entre deux écritures d'un alphabet commun et d'une directionnalité identique (i.e. Français, Anglais), mais il devient beaucoup plus complexe quand il s'agit d'alphabets de formes de caractères (glyphes) différents et de directionnalité opposée (i.e. Français, Arabe). Deux directions opposées, deux modes d'édition opposés, deux modes de justification opposés... sont d'ordre à rendre le traitement multilingue simultané quelque peu difficile.

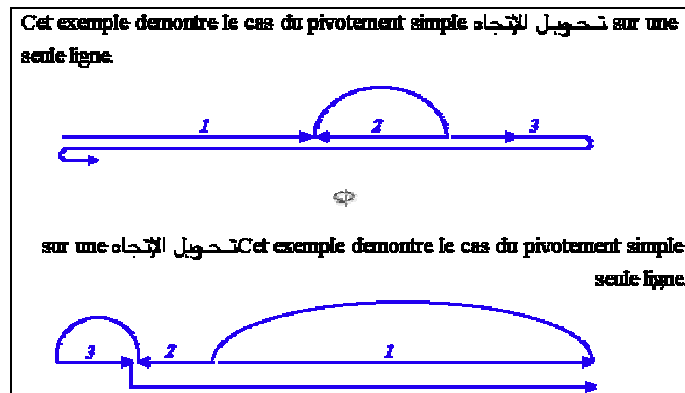
Ce phénomène d'incises devient beaucoup plus complexe dans le cas des bris d'incises. Un bris de texte est la coupure automatique d'un texte en fin de ligne (césure) et le retour à la ligne suivante. La complexité est issue de la directionnalité opposée des deux modes d'écriture Droite-Gauche et Gauche-Droite. La difficulté est adressée dans ces termes : au moment du retour à la ligne, le bris va-t-il s'aligner conformément à la directionnalité de la langue du texte « brisé » ou dans celle du mode maître d'écriture en cours. Des situations pareilles ont toujours marqué la complexité de l'interaction des utilisateurs avec des interfaces bilingues. Un constat est présenté dans la partie d'analyse de l'enquête menée.

L'effet du pivotement sémantique dans un texte Bidi (Swapping)

Le swapping sémantique est l'inversion dans l'ordre des phrases qui résulte d'un changement de direction globale d'un texte Bidi. Cette situation inconfortable est très courante au moment de l'insertion d'incises. Sa complexité est issue de deux phénomènes essentiels : les bris au niveau des incises et l'unicité directionnelle des chiffres. Ces deux aspects constituent deux niveaux de difficultés dans le dialogue avec l'interface machine.

Les bris de texte :

Les deux exemples suivants démontrent le mécanisme de bris (simples et complexe) au niveau d'une incise de texte. Le bris d'un texte uniforme ne constitue aucune difficulté car, selon le mode maître actif, le bris va forcer un retour à la ligne dans le sens global du document. Par contre le bris d'une incise d'une directionnalité opposée posera le problème de l'emplacement et du sens de la continuité du texte coupé sur la ligne suivante.



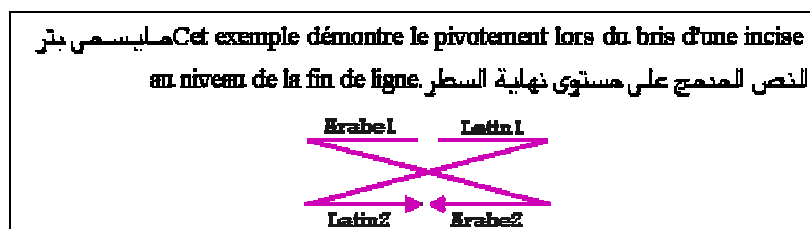
Pivotement d'une incise sur la même ligne

Par contre, dans les deux exemples suivants, le pivotement a lieu sur une incise avec un bris sur deux lignes. Dans le premier cas, l'agencement est correct car le début du texte de chaque langue commence à l'emplacement qui convient à sa lecture normale (début de ligne) et son flux évolue conformément à la directionnalité normale de chaque langue.



Bris normal d'incise arabe en MML

Dans le deuxième exemple, bien que le flux du texte et le retour à la ligne soient conformes à une directionnalité normale, les débuts du texte hôte et de l'incise sont opposés à leurs points de départ normaux. Le texte Bidi ne correspond plus à une juxtaposition faciale des deux chaînes linguistiques car les deux commencements sont au milieu de la ligne. Sur le plan sémantique, le sens de la phrase devient alors moins perceptible.

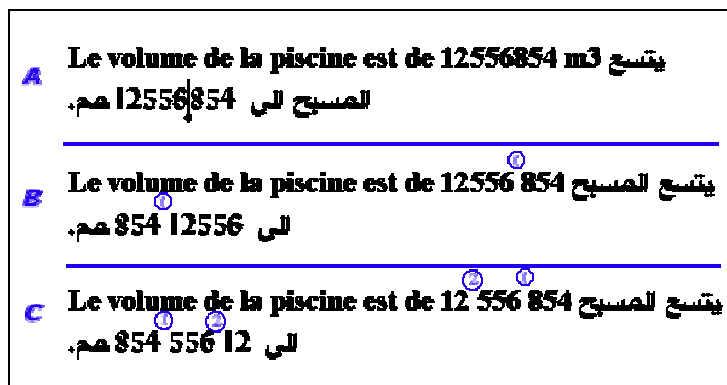


Pivot anormal d'une incise arabe en MML

Les bris de chiffres :

La complexité des deux derniers cas demeure relative devant les bris des chiffres dans une incise arabe. Cette particularité de l'incise arabe chiffrée est issue de la juxtaposition entre la direction droite-gauche de l'incise et la direction unique gauche-droite des chiffres.

Dans l'exemple suivant, le bris au niveau des chiffres dans l'incise arabe a engendré un pivotement successif des blocs de chiffres isolés par rapport à la suite contiguë de départ.



Le pivotement des chiffres dans une suite arabe

L'exemple « A » illustre une incise arabe avec un bloc chiffré dans un mode maître latin qui inclut la même série de chiffres. La forme du curseur qui annonce la direction de la suite au niveau des chiffres relève que le mode d'affichage actif est un mode visuel.

Dans l'exemple « B », un espacement est introduit entre les trois derniers chiffres du nombre en mode latin. Le résultat est un décalage vers la droite des trois chiffres. La directionnalité gauche-droite des chiffres et des lettres reste ainsi intacte.

La même opération est effectuée sur le même nombre en mode arabe. Le résultat est un pivotement des trois chiffres isolés vers la gauche du bloc restant des chiffres.

L'explication de ce phénomène est l'insertion d'un caractère espace entre les chiffres. L'espace est un caractère neutre qui, une fois inséré, divise la chaîne des chiffres en deux blocs non contigus. En mode latin, les deux blocs résultants se suivent dans une orientation gauche-droite normale. En mode arabe, les deux blocs de chiffres séparés par l'espace inséré sont invertis pour respecter la directionnalité droite-gauche de l'incise arabe tout en gardant leur structure interne d'une directionnalité standard de gauche à droite.

L'exemple « C » présente une segmentation en plus du bloc de chiffres. Si la linéarité est sauvegardée en mode latin, le mode arabe force une inversion supplémentaire pour rejeter le nouveau bloc de chiffres à gauche de la suite chiffrée.

Cet aspect de pivotement des chiffres arabes est d'ordre à créer des perturbations dans la structure sémantique des textes multilingues. L'exemple d'une fonction « Rechercher-Remplacer » effectuée pour insérer un espace entre l'indicatif régional et les numéros de téléphone engendrera un revirement dans la structure des numéros téléphoniques. La possibilité d'aérer un chiffre long en espaçant les centaines des milliers et les milliers des millions devient alors impossible car la valeur du chiffre se trouvera ainsi complètement modifiée.

Sans caractères neutres	Avec caractères neutres
Tél : 01234567 الهاتف : 01234567	Tél : (01)234567 الهاتف : 234567(01)
Prix : 98765432 D. السعر : 98765432 د.	Prix : 98 765 432 D. السعر : 432 765 98 د.

2.3 SELECTION DE TEXTE, MODES LOGIQUE ET VISUEL

La sélection de texte dans un document multilingue constitue également une complexité souvent incompréhensible par l'utilisateur. Un texte bilingue dans lequel s'associent les bris de texte, les bris des chiffres, les caractères neutres etc. est parfois un calvaire pour un utilisateur profane qui n'a aucune idée sur le fonctionnement interne du système qu'il est en train d'utiliser. Pour cette raison, nous avons jugé important dans cette partie dans laquelle nous exposons l'environnement général de notre étude et des axes de notre enquête, de donner des éléments explicatifs sur le fonctionnement des systèmes informatiques et tout particulièrement des systèmes multilingues.

Un texte saisi est instantanément stocké dans les registres de la mémoire tampon de l'ordinateur (Backing store) dans un mode séquentiel logique. L'ordre logique de stockage correspond à l'ordre dans lequel les caractères sont saisis. Cette méthode est adoptée d'une part pour des raisons de linéarité dans la gestion de l'espace mémoire et d'autre part pour faciliter l'une des tâches d'édition les plus courantes et non moins complexes comme les combinaisons « Copier/Couper et Coller ». En effet, pour faciliter ces opérations et les rendre transparentes pour le système interne, les caractères gagneraient à être alignés de façon contiguë en un seul bloc. La conversion de l'état de stockage logique et séquentiel des caractères à l'état de rendu visuel ou imprimé des glyphes constitue l'un des grands problèmes du traitement de la langue arabe. Pour optimiser la rentabilité et la transparence des applications multilingues, cette conversion se fait à travers des routines de restitution de texte de bas niveau. Il s'agit des « algorithme bidi ».

Dans l'exemple suivant, les caractères dans les cases de 9 à 15 montrent le terme arabe stocké en mémoire dans une disposition logique séquentielle de gauche à droite (GAD) conformément à la direction du mode maître latin et l'agencement des caractères dans la table des codes en usage, en l'occurrence la table ASMO 708 (ISO/IEC 8859-6).

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
L	'	a	c	c	o	r	d	ا	ل	إ	ت	ف	ا	ق

→ Traitement logique

L'accord الإتفاق

← Traitement visuel

La conversion entre les deux modes logique et visuel en mode maître latin

L'inverse est aussi vrai et le « Backing store » s'effectue de droite à gauche en mode maître arabe. Le rendu visuel ou imprimé est par contre adapté par l'algorithme Bidi qui inverse la disposition des caractères pour leur donner une orientation conforme au mode maître actif. (Droit à Gauche ou Gauche à Droite) et une apparence de glyphes conformément à une analyse contextuelle stricte.

15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
d	r	o	c	c	a	'	L	ق	ا	ف	ت	إ	ل	ا

← Traitement logique

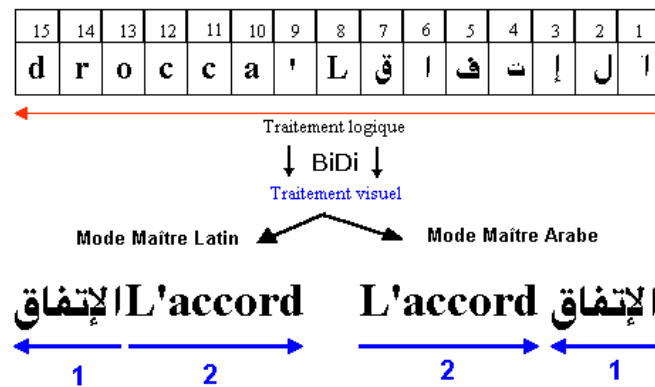
L'accord الإتفاق

← Traitement visuel

La conversion entre modes logique et visuel en mode maître arabe

Si le rendu dans les deux cas est identique c'est suite à l'inversion de saisi des deux concepts opposés. Si la même suite de caractères est saisie dans le même ordre dans les deux modes maîtres d'édition, elle

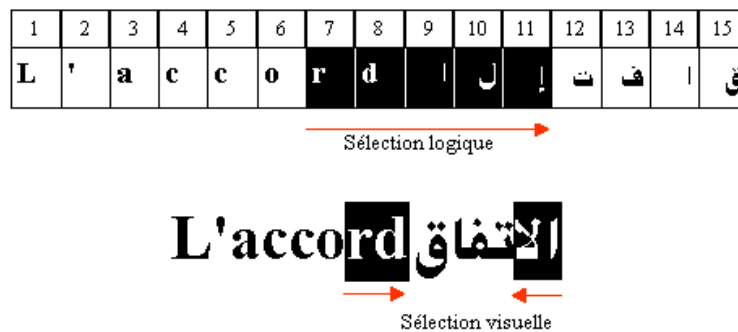
donnera deux rendus opposés car l'algorithme Bidi aurait agit doublement par l'inversion des chaînes de l'incise et l'inversion de sa position par rapport au texte hôte.



Inversion de position des chaînes opposées selon le mode maître actif

La conversion entre traitement logique et rendu physique est également perceptible à travers la fonction de sélection de texte et le mouvement du curseur. Une fois de plus, un mode maître est l'opposé de l'autre. Nous nous contenterons d'œuvrer sur le modèle du mode maître latin essentiellement avec invocation du mode maître arabe quand nécessaire.

Une sélection des chaînes de caractères de l'exemple précédent s'accomplit dans les deux modes à la fois : logique en interne et visuel en externe.



Modes de sélection interne (logique) et externe (visuel) en mode maître latin

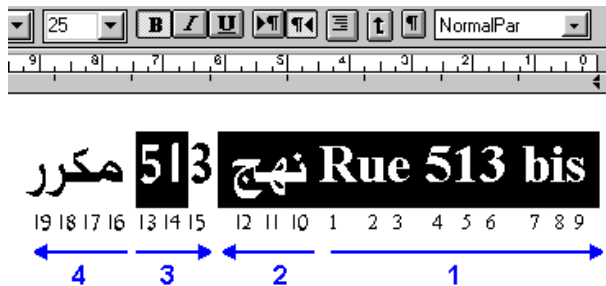
Plus compliqué encore est le cas de la sélection effectuée sur une chaîne de caractères mixtes incluant des chiffres.

L'une des constantes du traitement informatique des données est l'unicité directionnelle des chiffres de gauche à droite (GAD) dans les deux modes maîtres arabe et latin. D'emblée, cette unicité directionnelle des chiffres est en faveur du texte latin qui constituerait ainsi un prolongement directionnel uniforme de l'acte graphique ou de traitement. Ce qui n'est pas le cas pour le texte arabe qui doit dans ce cas de figure traiter les chiffres d'une manière opposée à sa direction normale de droite à gauche (DAG). Ainsi, dans l'exemple suivant édité en mode maître latin, les chiffres gardent leur unicité directionnelle. Une opération de sélection serait plus compliquée du côté du texte arabe que celui du latin.



Unicité directionnelle Gauche à Droite des chiffres en MML

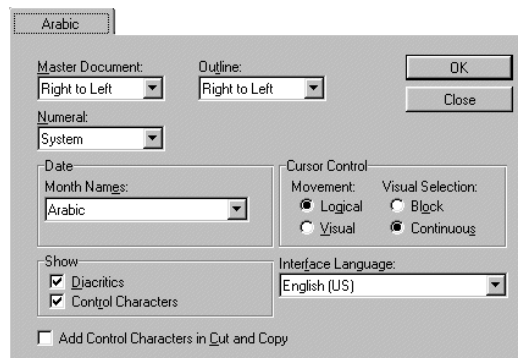
L'inversion vers le mode maître arabe ne fait que compliquer les choses car à part le maintien de la même complexité de la suite des lettres et des chiffres arabes, l'ordre sémantique de la phrase lui-même pivote.



Unicité directionnelle Gauche à Droite des chiffres en MMA et inversion sémantique

Cet exemple est non conforme à l'édition Bidi en mode maître arabe car le texte Bidi de chaque mode maître doit normalement commencer dans la langue qui lui correspond et dans une direction qui lui convient. La forme physique (visuelle) d'un texte Bidi est donc strictement dépendante de sa représentation interne ainsi que du mode maître d'édition dans lequel il est saisi. Le mouvement non linéaire du curseur et de la sélection, la réorganisation sémantique des bris dans les incises et l'unidirectionnalité des chiffres sont des détails très latents et souvent imperceptibles par l'utilisateur profane qui se trouve ainsi facilement perturbé par l'ordre des actions entreprises par les algorithmes qui gèrent les différentes phases d'un texte Bidi.

Le mouvement du curseur qui constitue le repère visuel de l'utilisateur sur le rendu visuel peut cependant être paramétré selon l'un des deux modes en question. Certaines interfaces graphiques courantes sous MS Windows proposent cependant les deux modes à l'utilisateur.



Le paramétrage du curseur en modes logique ou visuel sous MS Word97

2.4 CONVERSION, SAUVEGARDE ET FORMATS DES FICHIERS

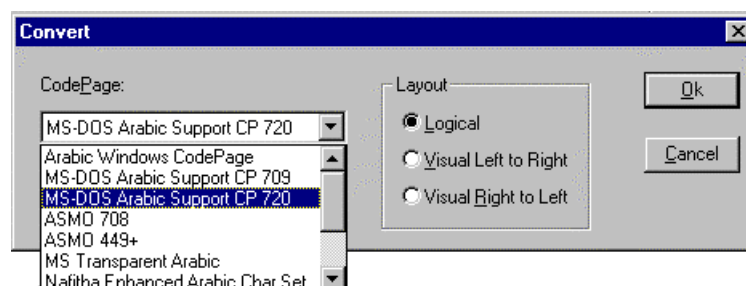
Les codes de formatage des caractères

Face à un utilisateur qui saisit du texte à partir d'un clavier, le système en cours d'utilisation identifie la touche sur laquelle cet utilisateur a appuyé et procède en conséquence à déchiffrer la combinaison binaire du caractère concerné. Une lettre « T » codée en Unicode U+0054 (hexadécimal) sera mémorisée en mémoire pour être ensuite traitée par le programme qui gère l'affichage. Ce dernier, qui pourrait être à titre d'exemple un gestionnaire Windows ou un module particulier de l'application en cours, se chargera d'établir le rapport entre les codes des caractères indexés et l'image des caractères qui leur correspondent pour les afficher sur écran. C'est un rapport systématique entre une page de codes binaires des caractères et une table de fontes qui traduit les codes en formes graphiques.

Les utilisateurs ont généralement tendance à utiliser les outils informatiques dans les formats standards proposés par défaut au moment de leur installation. Il est en effet rare de voir des utilisateurs qui savent distinguer entre un format de fichier et un autre. Les résultats de l'enquête vont le démontrer dans la partie suivante de ce travail.

Or l'univers informatique regorge de normes et de standards pour le codage interne des caractères pour des besoins et des affinités multiples. En effet, il est possible de traiter à la source ou de convertir chaque catégorie de données en plusieurs formes de codage.

L'illustration suivante démontre l'une des complexités de la langue arabe dans le choix du codage approprié. L'origine de cette complexité remonte à l'histoire de la normalisation qui a été marquée par des initiatives disparates et isolées de plusieurs instances nationales et internationales qui ont œuvré chacune de son côté à mettre en place une ou plusieurs normes de reconnaissance et de codification de l'alphabet arabe (i.e. ASMO, ISO, ECMA, UNICODE etc.). Il en résulte que pour une application ou une autre, pour un environnement informatique ou un autre, il est indispensable de savoir quel codage utiliser. Windows est en position d'emporter l'unanimité du format standard. Mais dans des situations particulières, une conversion d'un code à un autre demeure indispensable. Un texte arabe saisi sous Windows nécessiterait une conversion ASCII pour être exploité sous un environnement DOS ou Unix ou tout simplement pour être versé dans une base de données MS-Access.



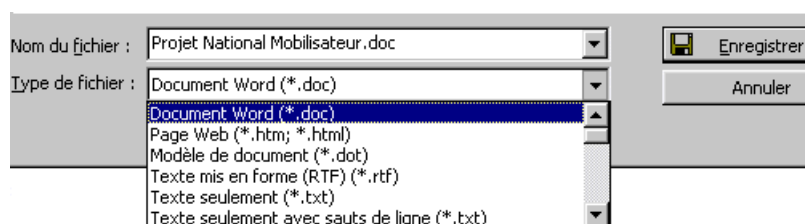
Les différents codes de formatage ou de conversion d'un texte arabe sous Windows

La compatibilité entre les formes de codage en vigueur était l'un des plus importants défis au nouveau contexte informatique. Le passage des applications en cours d'un système de codage à un autre aurait évoqué une grande polémique d'acceptation ou de rejet si un degré de transparence dans la migration d'un code à un autre n'avait pas été garanti. Déjà l'histoire des techniques du codage des caractères ne cesse de connaître des divergences entre l'ASCII et le codage étendu du groupe ISO/IEC 8859. La polémique se situe en réalité au niveau du mécanisme de codage et d'interprétation des codes utilisés au moment de la rédaction ou de la réception d'un corpus textuel.

Les formats des fichiers texte

L'autre aspect de complexité de l'usage des textes arabes est la diversité des formats de sauvegarde des données. Bien que ce soit un phénomène général qui dépend de tout utilisateur sans considération du support linguistique, la langue arabe fait intervenir dans ce contexte une exception particulière due à la diversité des codes de formatage des documents. Les utilisateurs ont toujours tendance à sauvegarder leurs fichiers dans le format standard proposé par le logiciel utilisé ; en l'occurrence le format Word. Or ce format résulte inadéquat si le texte est prévu pour être échangé sur Internet, envoyé dans une base de données tournant sous un environnement DOS, ou tout simplement pour être utilisé par un autre logiciel (i.e. Word Perfect) ou une autre plate-forme (i.e. Macintosh ou Unix). Certes, les filtres de conversion sont opérationnels du côté de la source ou de la destination, mais il est toujours

recommandé de créer à la source des documents adaptés (ou portables) par défaut à toute alternatives de changement de plate-forme ou d'application. Cet aspect reste également énormément inexploré par une grande marge d'utilisateurs.



Les différents formats de sauvegarde d'un texte arabe sous Windows

3. LE DIAGNOSTIC

Suite à cette présentation théorique des axes les plus conflictuels du traitement de la langue arabe dans un environnement informatique, nous exposons en ce qui suit les résultats de l'enquête menée auprès d'un échantillon d'une population cible pour voir le degré de compatibilité entre les outils de l'informatique arabisée et les prédispositions psycho-cognitives de ses utilisateurs.

Notre approche dans ce sens est de reprendre les situations de difficultés exposées à travers des statistiques que nous essayons d'argumenter par des faits socioculturels d'une part et par des éléments de confirmation puisés dans les croisements des questions posées aux sujets interviewés.

3.1 CULTURE INFORMATIQUE, CULTURE LINGUISTIQUE ARABE

De prime abord, l'enquête a révélé que l'échantillonnage a touché une population cible d'une majorité de 55% d'individus entre 25 et 30 d'âge dont 87.2% sont des étudiants de formations DESS et 14,5% en début de carrières professionnelles (secrétaires, Bibliothécaire, chefs de service etc.). 92,7% d'entre eux affirment avoir déjà reçu une formation en NTIC d'une période qui varie entre une et trois ans.

D'autre part, l'enquête a exclusivement couvert un environnement d'études en sciences humaines. Menée dans un campus universitaire des sciences de l'information et de la communication, puis dans des bibliothèques et des centres de documentation professionnels, la population cible est à l'origine littéraire ayant un pré requis linguistique relativement avancé. Ce choix est en réalité décidé pour correspondre à la nature du thème de recherche, en l'occurrence l'approche linguistique des NTICs. Le croisement de l'outil informatique et de la culture linguistique chez l'utilisateur est d'ordre à nous donner plus de visibilité sur l'état de la question posée tout au début de ce travail. Deux axes sont donc à détailler à partir de ces éléments d'information dans les deux tableaux suivants: les pré-requis linguistiques chez la population cible et sa culture informatique de départ.

Tranches d'âge	Connaissance Alphabets			Identification des Chiffres Hindous			Identification de la Kashida	
	Aucun	ABT	ABJD	Confusion avec chiffres arabes	Identification seule	Identification et lecture	Identification sans nom	Identification par nom
18-25	25	8	0	18	5	10	33	0
26-30	26	28	1	30	5	20	55	0
31-40	7	3	1	4	0	6	10	0
Total	58	39	2	52	10	36	95	0

Tab. 1 : Culture linguistique chez la population cible

Parmi les éléments saillants dans ces tableaux, il est à commenter les faits suivants :

A part le fait que 58% de la population cible ne connaisse aucune des deux formes de l'alphabet arabe, le reste se limite à la seule forme enseignée dans les écoles et adoptée dans les écrits officiels. Ceci dénote d'un manque de culture générale dans le domaine linguistique arabe qui définit plusieurs formes d'alphabets parfois utilisés par des érudits ou des écoles classiques dans certains pays arabes.

instinctivement choisi le petit triangle noir du bouton de directionnalité sur la barre des icônes. Personne n'a fait usage des raccourcis claviers.

A notre sens, cette première observation met en évidence l'importance du facteur visuel dans le mécanisme cognitif de l'utilisateur. Le bouton de directionnalité majoritairement choisi, se distingue des autres alternatives par son symbolisme direct : une tête de flèche qui démontre l'orientation Gauche-Droite ; ce qui n'est pas le cas d'un choix de langue textuel (FR, EN, AR...) ou d'un raccourci claviers totalement abstrait. D'ailleurs si 33% ont opté pour le bouton d'alignement gauche pour changer de mode maître actif, bien que les deux fonctions soient totalement indépendantes, ceci dénote encore de la présence d'un penchant implicite chez les interviewés vers les valeurs informationnelles graphiques et visuelles. Le même bouton de directionnalité a également été choisi par 75% des sujets pour passer du clavier arabe vers le clavier latin. Ceci est d'ordre à confirmer la confusion qui règne chez les utilisateurs entre un changement de direction d'écriture et un changement de clavier. Or, il est toujours possible de rédiger en arabe (Droite à Gauche) dans un mode maître latin (orientation d'affichage Gauche-Droite) et vice versa.

Il a été reporté dans l'exemple du texte Bidi précédent qu'au niveau du code du téléphone dans le texte latin (123), 94% des sujets ont tapé les 2 lettres arabes avant les chiffres. Une déduction principale en découle : le phénomène de la bidirectionnalité sémantique n'est pas implicite chez les interviewés sinon ils auraient implicitement déduit que le code est une unité sémantique arabe, donc soumise à une directionnalité Droite-Gauche. Ceci se confirme avec la partie opposée de l'exemple : au niveau du code du téléphone dans le texte arabe (123BT), 62% des sujets ont tapé les 2 lettres latines après les chiffres. L'explication plausible ne peut être que le déclenchement du processus cognitif chez les sujets à partir de l'exemple précédent pour tenir compte de la bidirectionnalité sémantique.

La difficulté majeure dans cet exercice était sans doute le pivotement des chiffres à cause des caractères neutres. Nous avons demandé aux sujets de séparer les blocs du numéro de téléphone par un tiret après le chiffre 9 et un point entre les chiffres 123 et 456 afin de lui donner l'apparence d'un numéro de téléphone avec un indicatif et deux blocs de trois chiffres. Si l'opération s'est déroulée convenablement dans le texte latin, c'est parce que les chiffres ont une directionnalité absolue de Gauche à Droite. L'insertion des caractères neutres comme le tiret ou le point n'a eu aucun effet sur la directionnalité des chiffres qui reste ainsi conforme à celle du texte latin. Par contre, dans le texte arabe, le tiret derrière le chiffre 9 fait pivoter ce dernier à droite du reste des chiffres le traitant ainsi comme un bloc de chiffres indépendant du reste. Le point entre 123 et 456 engendre aussi un effet de pivotement (Swapping) pour les convertir en 456.123. C'est à ce niveau que nous avons voulu voir l'attitude des sujets pour essayer de garder la valeur réelle du numéro de téléphone tout en le séparant en blocs comme prévu. Nous avons observé que pour s'y faire, 48% ont essayé de changer de langue active ; 20% ont essayé vainement la fonction « couper/coller » et 9% seulement ont essayé de réécrire le chiffre formaté en mode maître latin. Ce qui est notoire cependant c'est que 67% d'entre eux n'ont réussi à bien faire la modification qu'à la quatrième tentative. Toutefois 10% seulement ont réussi la transformation dès le premier coup.

Une fois de plus, cet exemple confirme l'ignorance totale des sujets des valeurs directionnelles des chiffres et des caractères neutres dans un environnement bilingue. Ceci dénote d'une interactivité surfacique entre l'utilisateur et la machine qui se limite à un niveau d'interactivité visuelle simpliste.

3.2.2 Incises et bris de textes

L'expérience a continué en deuxième phase avec les difficultés Bidi des incises et des bris de textes bilingues. Les deux situations suivantes leurs ont été proposées : insérer un texte arabe dans un texte latin.

Les textes multilingues présentent des difficultés de direction et d'affichage.

Et insérer un texte latin dans un texte arabe.

La reconnaissance optique des caractères .
constitue encore un problème majeur pour les projets des bibliothèques virtuelles.

La série d'expériences menées pour analyser les réactions de sujets au moment des insertions ou des bris des textes a démontré qu'à ce niveau, il y a une uniformité de comportement. La question était de savoir si dans les deux cas de figures, les sujets allaient d'une part changer de mode maître actif ou rester au mode maître initial pour ajouter l'incise et d'autre part essayer d'aligner le bris selon sa directionnalité d'origine. Les statistiques se rapprochent beaucoup dans ce sens bien qu'elles divisent en deux le nombre des personnes agissant de manières opposées. Ceci met davantage en doute la clarté de la bidirectionnalité graphique dans le mental des sujets.

3.2.3 Sélection de texte : mode logique et mode visuel

La sélection du texte constitue aussi une difficulté majeure pour les utilisateurs des interfaces bilingues car elle met en pratique un traitement bilingue interne dans les registres de la mémoire de la machine (mode logique) opposé à l'effet qu'elle donne au niveau de l'interface utilisateur (mode visuel).

Pour mener cette expérience, nous avons commencé par demander aux sujets la rédaction du texte bilingue suivant composé de chiffres et de lettres dans les deux alphabets arabe et latin sans aucune spécification préalable de l'environnement de travail.

123456 : Le numéro de téléphone est : 123456

L'observation a donné lieu à un chiffre surprenant : dans 78% des cas, les sujets ont commencé la rédaction du texte avec les chiffres de gauche.

Le cas est surprenant car il confirme définitivement que l'analyse sémantique du texte au moment de la rédaction est quasi-absente du mental des utilisateurs. C'est le visuel graphique qui domine le psychocognitif de l'utilisateur au moment de la rédaction. Ce que nous n'avons pas prévu de confirmer c'est si ce phénomène est exclusif à la bidirectionnalité de l'écrit ou s'il est unanime pour toutes les langues.

Or, le deuxième exercice confirme une grande partie de cette constatation. Nous avons demandé aux sujets de faire une sélection à la souris de la partie arabe du texte. 54% ont commencé à partir du premier caractère arabe tout en évoluant vers la gauche ; 34% ont commencé la sélection à partir du chiffre 1 dans la partie arabe. Le premier chiffre confirme le mécanisme de l'apprentissage et de l'éveil des facultés cognitives à partir d'expériences précédentes pour se mettre en position de veille et par conséquent renforcer l'analyse sémantique chez les sujets. Le deuxième chiffre reproduit par contre le modèle de la mémoire visuelle graphique ancrée chez certains dans leur interactivité avec le texte Bidi. Les deux cas ne font que démontrer une fois de plus que les principes de fonctionnement des systèmes bidirectionnels restent très flous dans le psychocognitif des sujets échantillons de notre enquête.

3.2.4 Le formatage et la sauvegarde des fichiers

A la fin de notre enquête, nous avons voulu confirmer les pronostics que nous avons envisagés par une question d'ordre plutôt technique que psychocognitive. Nous avons demandé la sauvegarde du travail réalisé par le sujet en format texte pour savoir à quelle profondeur il dispose d'une culture des systèmes informatiques qu'il utilise. C'était prévu comme complément à la première partie d'évaluation de sa propre maîtrise de l'interface bilingue qu'il utilise. Notre doute était bien fondé car la culture informatique des sujets au niveau système est beaucoup plus pauvre.

22% n'ont pas compris la question du tout. Pour eux sauvegarder un texte sous Word consiste à suivre la commande « Enregistrer Sous » puis à spécifier un nom. Déjà 6% d'entre eux n'ont aucune idée sur l'usage de la zone « Type de fichier » dans la fenêtre de sauvegarde. 44% ont choisi le mode Texte sans se soucier des autres modes texte comme « Texte MS-DOS », « Texte avec saut de ligne » etc.

Quand nous avons demandé de sauvegarder en « Mode Texte » dans l'une des pages de codes appropriées à la langue arabe (i.e. ISO 709, ISO 720, ASMO 449 etc.), 96% nous ont regardé de travers, 4% ont essayé sans pouvoir trouver l'astuce.

CONCLUSION

En réalité un travail d'enquête dans le domaine des interfaces Homme-Machine est beaucoup plus profond que les limites appliquées à ce présent travail de synthèse. Dans le cas particulier de ce travail, il est plutôt question de compléter par un travail de terrain, une étude effectuée dans le cadre d'un travail de thèse réalisée par le même auteur au sein de l'Université de Bordeaux 3 et portant le titre de « Architecture et morphologie des interfaces de communication de l'information scientifique et technique dans un environnement multilingue : le cas du contexte arabo-latin ».

Parmi les éléments que cette enquête a pu fournir pour argumenter les aboutissements de l'étude analytique citée portant sur ce même thème, nous analysons les deux points suivants :

1. Les utilisateurs d'un environnement bilingue arabe-latin sont généralement soumis à un phénomène de perte de repères linguistiques et culturels devant une interface bilingue. La population cible, choisie pour mener cette enquête, constitue en fait un maillon social transitoire entre deux époques et deux générations de dominances culturelles différentes. Après une époque durant laquelle la langue française était une langue de culture et d'enseignement, les nouvelles générations subissent un phénomène d'arabisation massive des enseignements et une hybridation intense des langages de communication. Pris entre deux courants linguistiques mal entretenus, cette tranche de la population a perdu les performances linguistiques des ressortissants des écoles françaises du début du siècle et la notoriété culturelle arabe des érudits et des ulémas zeitouniens. Nous avons pu observer dans cette enquête le phénomène de l'apprentissage « à la carte » que subit un sujet sur une zone de difficulté pour savoir traiter une difficulté similaire suivante. Ceci démontre une absence d'un référentiel culturel ou linguistique qui opère instinctivement pour la résolution des situations bilingues.
2. L'approche des outils proposés pour le traitement bilingue est aussi une partie prenante dans la désorientation des sujets soumis à l'enquête : nous avons pu constater que toutes les solutions bilingues s'accordent sur le traitement Gauche-droite des chiffres sans considération du mode maître actif. Ceci contredit à la source la structure interne des chiffres qui évoluent normalement de Droite à Gauche dans leur traitement arithmétique (unités, dizaines, centaines etc.), leurs modes de prononciation en arabe littéraire (unités, dizaines, centaines etc.) et par conséquent leur mode de transcription qui doit suivre cette directionnalité originelle (unités, dizaines, centaines etc.). Toute l'histoire cette confusion dans le traitement des chiffres est que le système décimal, mis au point par les hindous et repris par les Arabes dans une directionnalité Droite-Gauche à été repris par les occidentaux qui l'ont adapté à leur directionnalité Gauche-Droite et qui nous l'ont retransmis tel-quel. Au lieu de remettre le mécanisme des chiffres dans ses formes d'origine pour correspondre à notre langue arabe, nous avons repris le système doublement transformé ; ce qui engendre automatiquement l'effet de pivotement (Swapping) et la non contiguïté des zones de sélection des textes bilingues qui perturbent le confort des utilisateurs dans leur interaction avec les interfaces bilingues.

La morale de l'histoire dans cette polémique d'interfaces bilingues et du rôle attribué à la langue arabe dans l'accès et la diffusion de l'information locale est beaucoup plus profonde qu'elle n'en a l'air. Nous la situons déjà dans un domaine de recherche fondamentale plutôt que dans la recherche appliquée. Peu s'interrogent en réalité sur cette problématique car déjà les solutions pour la manipulation de la langue arabe existent déjà. Nous étions intrigués dans notre questionnement de ce problème d'interfaces par l'inadéquation et la non-conformité de ces solutions avec les exigences linguistiques et culturelles de la langue arabe. Or par la simple observation de deux facteurs fondamentaux technique et économique, il est possible de s'arrêter sur les origines du problème et de la nécessité d'opter pour une alternative radicale.

Le facteur technique se résume dans le positionnement secondaire de la langue arabe dans toutes les tables de codes des systèmes informatiques du monde : ISO 8859-6, rangée 6 du BMP d'Unicode, etc. face à une domination des caractères latins (et particulièrement anglo-saxons) qui se positionnent en

tête de liste et par conséquent restent omniprésents dans toutes les solutions informatiques mêmes universalistes (Unicode et ISO 10646).

Le facteur commercial est bien évident à travers les gourous de l'informatique mondiale comme Microsoft, qui par un phénomène d'économie d'échelle parviennent à dominer le monde par leurs solutions informatiques qui parfois ne correspondent pas aux réalités et culturelles des zones linguistiques concernées. La littérature spécialisée a toujours parlé de solutions ici et là développées par des PME qui proposent des alternatives d'arabisation. Seulement, même si elles peuvent apporter des solutions plus proches de la réalité linguistique concernée, elles ont beaucoup de difficulté, voire même il leur est impossible, d'entrer dans une logique de concurrence commerciale avec les barons du logiciel et des systèmes informatiques du marché mondial.

La langue arabe et sa présence dans le contexte informatique international dépendront à notre sens des apports de la recherche fondamentale qui orientera la recherche appliquée vers des solutions plus souples. Elle dépendra aussi d'une prédisposition socioculturelle à renforcer le rôle de la langue arabe au niveau de l'enseignement et des usages.

BIBLIOGRAPHIE

1. Jacques Robin. Les pièges d'une société de l'information planétaire. Le Monde Diplomatique. Hors-série, M 1909, octobre 1995. pp.67-68
2. Christophe KOLSKI. Ingénierie des interfaces Homme-Machine : conception et évaluation. Paris : Hermès, 1993.
3. Martial GRISLIN, Christophe KOLSKI. Évaluation des interfaces Homme-Machine lors du développement des systèmes interactifs. Technique et Science informatiques. Vol.15-n°3/199. p.266
4. J. NIELSEN, R. MOLICH, heuristic evaluation of users interfaces, Actes de colloque CHI'90, Seattle, ACM New York, pp.349-356, 1990
5. B. SENACH, P. ALENGRY. Évaluation d'un dispositif d'assistance à la conduite et au dépannage de chaînes de fabrication automatisées. Rapport INRIA n°163. Mars 1994.
6. M. SWEENEY et al. Evaluating user-computer interaction : a framework. International Journal Man-Machine Studies, 38, 689-711, 1993.
7. WHITEFIELD et al. A framework for human factors evaluation. Behavior & Information Technology, vol. 10, n°1, pp.56-79, 1991.
8. J. COUTAZ & S. BALBOS. Évaluation des interfaces utilisateur : taxonomie et recommandations. Actes des sixièmes journées sur l'ingénierie des interfaces Homme-Machine, IHM'94, 211-218, Lille, 8-9 décembre 1994.
9. Microsoft MS DOS User's Guide : Arabic Supplement. Microsoft Corporation, 1988. p.18.

