

Comparaison de trois logiciels utilisables en terminologie : FOXBASE+, MC4, TEXTO

N. Maurice, E. Blanchon, G. Otman et J. Boissy

Volume 36, numéro 1, mars 1991

La terminologie dans le monde : orientations et recherches

URI : <https://id.erudit.org/iderudit/002091ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

[Découvrir la revue](#)

Citer cet article

Maurice, N., Blanchon, E., Otman, G. & Boissy, J. (1991). Comparaison de trois logiciels utilisables en terminologie : FOXBASE+, MC4, TEXTO. *Meta*, 36(1), 207–213.

COMPARAISON DE TROIS LOGICIELS UTILISABLES EN TERMINOLOGIE: FOXBASE+, MC4, TEXTO

N. MAURICE, E. BLANCHON, G. OTMAN ET J. BOISSY
Centre de terminologie et de néologie, Paris, France

Une enquête réalisée par la Société des traducteurs du Québec en mars 1988 révélait que 39 % des services terminologiques étaient alors informatisés, 52 % en voie d'informatisation et que 64 % des terminologues participaient à l'informatisation de leurs activités. Cette phase d'informatisation implique l'acquisition de matériel mais aussi — et surtout — l'adoption d'un logiciel adapté à la fois à la configuration matérielle choisie ou imposée, aux besoins terminologiques spécifiques du service et aux connaissances informatiques des utilisateurs potentiels de la base de données terminologiques, utilisateurs-concepteurs et utilisateurs-consultants.

Cet article présente les caractéristiques essentielles de trois logiciels considérés dans la perspective stricte d'une application terminologique. Ces trois logiciels appartiennent à — et par là même représentent — trois familles distinctes de SGBD :

- FOXBASE+ : (dorénavant FOX+) un logiciel classique de gestion de bases de données de type dBASE, le plus «généraliste» des trois,
- TEXTO : un logiciel à vocation documentaire (parce que les terminologues font aussi de la recherche documentaire dans 58 % des cas selon l'enquête précitée de la STQ), le plus «polyvalent» des trois,
- MC4 : un logiciel à vocation terminologique multilingue (parce que terminologie et traduction sont un couple économiquement inséparable), le plus «spécifique» des trois.

QUELQUES ASPECTS HISTORIQUES ET TECHNIQUES

FOX+, conçu en 1986, produit à vocation grand public, est un clone de dBASE3 auquel il emprunte son langage de programmation et sa conception générale. TEXTO, plus ancien (1975) est né de la rencontre d'ingénieurs chimistes et d'informaticiens qui, confrontés à la mise en place de systèmes d'information, ont créé la société CHEMDATA dont TEXTO est la production-phare. Pour reprendre leur phraséologie commerciale, TEXTO est un *progiciel de gestion documentaire* capable de *gérer la mémoire d'une entreprise* sous ses aspects les plus variés. MC4, quant à lui, est une production universitaire (Université de Clermont-Ferrand II), œuvre d'un homme (J.M. Henning) et d'une association (Terminformatique) travaillant à une base de données terminologique multilingue consacrée à l'industrie du bâtiment et aux travaux publics. MC4 a été conçu en 1985 à partir d'un noyau dBASE3 dont il partage certaines caractéristiques.

Quelles que soient les spécificités et particularités de ces trois produits, ils présentent l'avantage de fonctionner sur la même configuration informatique, configuration très courante de surcroît, soit :

- système d'exploitation MS-DOS sur compatible PC
- disque dur d'au moins 10 Mo
- mémoire centrale de 512 Ko (640 Ko pour TEXTO).

TEXTO peut être utilisé dans sa version standard (on parle alors de mode *ligne à ligne*) ou couplé au logiciel semi-intégré WINDOWS (mode *windows*) qui offre un confort de travail et une efficacité bien supérieure à la version nue.

Si MC4 ne nécessite aucune formation préalable, TEXTO et dans une moindre mesure FOX+ requièrent cet effort qui, le cas échéant, peut se traduire par un investissement financier supplémentaire. MC4 et FOX+ se situent dans la même fourchette de prix alors que TEXTO est un produit haut de gamme dont le prix avoisine, toutes options confondues, les 35 000 F.

USAGES GÉNÉRAUX D'UN SYSTÈME DE GESTION DE BASES DE DONNÉES EN TERMINOLOGIE

Ces usages correspondent en fait aux besoins fondamentaux du terminologue et aux différentes orientations que peut prendre la pratique terminologique : traduction, information scientifique et technique, normalisation et planification. Concrètement, donc, une terminologie se présente sous la forme d'un fichier (hier manuel, aujourd'hui informatisé) dans lequel chaque fiche est organisée autour d'une notion. Ces fichiers constituent la base de produits terminologiques secondaires : vocabulaires systématiques, lexiques bilingues ou multilingues, thesaurus...

La vocation du fichier terminologique est de mettre à la disposition de ses utilisateurs (concepteurs, consultants, traducteurs, normalisateurs...) les informations linguistiques correspondant à leurs besoins. Il est donc primordial que ces informations soient d'une consultation et d'une communication aisées. L'informatique possède intrinsèquement le pouvoir de répondre aux trois exigences fondamentales du terminologue gestionnaire d'une base de données terminologiques : stockage, consultation et communication de données.

Les trois logiciels présentés ici assurent bien évidemment cette triple fonction. L'intérêt de notre étude comparative réside donc dans l'analyse de la façon dont ces fonctions sont assurées par chacun de ces produits. Notre analyse s'articule autour de trois axes principaux qui prennent en compte en priorité les opérations de conception et de gestion d'une base de données sans perdre de vue pour autant les aspects concernant la consultation, l'édition et la mise à jour de fichiers terminologiques.

CONCEPTION DES FICHIERS

STRUCTURE DES FICHIERS

MC4 offre à l'utilisateur la possibilité de constituer simultanément plusieurs fichiers terminologiques, le trait distinctif de chaque fichier étant constitué par le rattachement d'une notion à un domaine ou sous-domaine. Il en résulte que le terminologue doit avoir découpé de manière précise et stable son domaine terminologique en sous-domaines, même si la réaffectation d'une notion à un autre domaine est envisageable. FOX+ laisse à l'utilisateur le soin de structurer sa base en fonction de ses besoins spécifiques; ces besoins doivent être rigoureusement analysés en amont de la réalisation du fichier proprement dit car de la structure même de la base dépend pour une bonne part son efficacité ultérieure. FOX+ et TEXTO autorisent la constitution de plusieurs fichiers parallèles reliables entre eux à condition qu'ils possèdent au moins un champ en commun qui fera office de *pont*. Cette possibilité de lien est particulièrement appréciable dans le cas de fichiers multilingues. Chaque fichier monolingue peut rester autonome, géré par un terminologue spécialiste d'une langue, tous les fichiers étant interconnectés par ailleurs par des numéros de notions communs par exemple. MC4 est conçu pour gérer des bases de données multilingues. Il peut gérer

jusqu'à dix langues de travail, définies au gré de l'utilisateur, les seules contraintes restant le recours à l'alphabet latin et le travail limité à un *couple de langues* (une langue source et une langue cible) défini en début de session.

FORMAT DES FICHIERS ET DES CHAMPS

FOX+ exige l'utilisation de formats fixes tant au niveau de la fiche qu'au niveau des champs. La taille d'une fiche est limitée à 4 000 caractères, celle d'un champ à 254. La définition du nom et du contenu d'un champ implique donc nécessairement la détermination de sa longueur. La limite des 254 octets (à peine quatre lignes de texte) peut être — maladroitement — contournée par l'enchaînement de champs correspondant au même contenu (exemples : définition ou contexte). On peut aussi recourir à un champ *mémo* beaucoup plus large (plusieurs milliers de caractères) mais qui présente l'inconvénient de ne pas être interrogeable (recherche sur des chaînes de caractères par exemple).

MC4 contient des champs prédéfinis impossibles à modifier. Ces champs aux dénominations *transparentes* et *souples* correspondent bien aux besoins généraux de tout service terminologique. Ainsi nous trouvons successivement les champs *terme*, *remarque*, *caractéristique*, *domaine*, *terme équivalent*, *définition*, *liens notionnels*, *bureau émetteur*, *type* et *code de fiabilité*.

Dans TEXTO, il n'y a pas de format fixé mais des limitations globales : la taille d'un enregistrement est limitée à 4 000 caractères, celle d'un champ ne l'est que par rapport à ce plafond, le nombre maximum de champs par enregistrement étant de 99.

Ce paramétrage, pour fastidieux et arbitraire qu'il puisse paraître à première vue, présente cependant un caractère décisif car, mal apprécié, il peut compromettre gravement une gestion rationnelle de l'espace disponible en mémoire.

MODIFICATIONS DE STRUCTURE

C'est indéniablement TEXTO qui offre la plus grande souplesse sur ce plan. MC4 est totalement rigide et FOX+ très résistant. Dans FOX+ la commande «MODIFY STRUCTURE» permet de modifier le nom d'un champ, sa taille ou son type (caractère/numérique). On peut également ajouter ou supprimer un champ (le renommer avec certaines précautions), la suppression entraînant la disparition du contenu. Dans TEXTO, la structure d'un fichier est définie par la création d'un *document de paramètres*. La modification du fichier passe par la modification de ce document. On peut, à loisir, ajouter, supprimer ou déplacer des champs; renommer un champ, par contre, peut entraîner une perte de données.

À nos yeux, ce critère de possibilité de modification est essentiel car aussi soigneusement qu'ait été préparée la structuration d'une base, des ajustements peuvent s'avérer nécessaires après coup. Il faut pouvoir disposer d'une marge de manœuvre, non seulement pour pouvoir rectifier les éventuelles erreurs de conception, mais aussi pour être en mesure de garantir la constante adaptation d'un produit à des besoins ou à des objectifs évolutifs.

GESTION DES OPÉRATIONS COURANTES

Le terminologue gestionnaire d'une base de données doit se soucier à la fois de la cohérence interne de sa base sans laquelle il n'est pas de recherche efficace et des modalités pratiques de saisie; de l'ergonomie et de la convivialité d'un logiciel dépend en effet l'allègement de tâches répétitives et fastidieuses.

GESTION DE LA COHÉRENCE

Le maintien de la cohérence d'une base de données implique des procédures de contrôle et de validation ainsi qu'une mise à jour régulière des index. Sur ce point, MC4 apparaît comme le logiciel le plus vigilant mais aussi le plus contraignant à manipuler. En effet, MC4 gère des procédures de contrôle et de validation à toutes les étapes de la saisie des données. Dès la saisie du terme *vedette*, MC4 vérifie qu'un homographe ne figure pas déjà dans la base. Les autres vérifications portent sur la sélection des langues de travail, sur les champs reliés à des tables de valeurs (domaine, lien ou caractéristique par exemple) ou sur le remplissage obligatoire du champ *domaine*. À noter que le fait d'opérer la validation d'un champ interdit tout retour en arrière. Il en résulte que pour effectuer une correction, il faut achever la création de la fiche puis quitter le menu «CRÉATION» pour passer dans les menus «MODIFICATION» ou «SUPPRESSION». FOX+ et TEXTO n'exercent de contrôle véritable que sur les champs numériques. Ils contrôlent tous deux la longueur des champs et leur éventuel dépassement. TEXTO indique l'atteinte de la limite par un signal sonore qui n'interrompt pas la saisie alors que l'enregistrement des données est interrompu. FOX+ fait retentir un signal sonore en fin de champ, ce signal n'interdisant nullement de poursuivre la saisie puisque le saut vers la tête du champ suivant est automatique. Sur FOX+, la validation d'une fiche est automatique en fin de saisie d'un enregistrement et l'ensemble des ajouts-modifications effectués lors d'une session est enregistré au cours des procédures de fin de session, de même qu'en TEXTO mode *windows*. Par contre, en mode ligne à ligne, TEXTO offre le choix entre *saisie simple* et *saisie validée*. Dans le premier cas, la validation est automatique après chaque fiche. En *saisie validée*, après chaque fiche, un menu apparaît à l'écran, qui permet de valider la fiche, de la corriger, de substituer une information à une autre, de procéder à une modification sur une série de fiches ou de lister la fiche.

MISE À JOUR DES INDEX

Elle est automatique sur FOX+ en fin de session. À noter que dans le cas où l'on élimine des termes vedettes, les index ne sont modifiés que lorsque la suppression logique d'une fiche a été confirmée par sa suppression physique. Sur MC4, les suppressions sont seulement logiques, jamais physiques. Mais, aux yeux de l'utilisateur, la liste des termes qu'il peut visualiser grâce à la fonction «CONSULTATION» est constamment à jour même si elle ne correspond pas à l'index des termes dont elle est extraite. Il en résulte que ces fiches, jamais détruites physiquement, continuent à encombrer la mémoire.

Les fonctions de mise à jour sur TEXTO dépassent largement les possibilités offertes par FOX+ et MC4. TEXTO gère jusqu'à dix index sur le même fichier. De plus, si, au sein d'un champ, certains éléments sont distingués par un *séparateur* spécifique dûment déclaré dans le document de paramètres, ils sont indexés distinctement. Les mises à jour d'index peuvent être immédiates (après chaque fiche), différées (en fin de session) ou par lots (pour un ensemble de fiches définies). Au moment de ces mises à jour, TEXTO peut signaler le fait que certaines données sont déjà présentes dans l'index. TEXTO, grâce au logiciel utilitaire TEXUTIL, permet également de procéder à la réorganisation totale d'un fichier dont l'index a pu se désorganiser progressivement surtout si l'on utilise fréquemment les commandes de correction.

ACCÈS ET CONSULTATION DES DONNÉES

Stocker des données terminologiques est une activité dont l'intérêt n'est justifié que si ces données sont redistribuées à leurs destinataires potentiels. Un gestionnaire de base de données terminologiques doit donc pouvoir fournir ce type d'information à la

demande, quelle que soit la perspective envisagée: traduction, information technique, normalisation... L'accès aux données passe nécessairement par le truchement des dénominations de notions, qu'il s'agisse de termes en langue source ou langue cible, de synonymes ou de standards.

ACCESSIBILITÉ DES DONNÉES

Un logiciel performant doit permettre un accès différencié aux données selon le type d'usage qu'on s'apprête à en faire. En effet, il n'est pas souhaitable qu'un consultant lambda puisse accéder aux fonctions d'enregistrement et de modification et ainsi mettre en péril l'intégrité même de la base. Par ailleurs, toute information contenue dans la base ne doit pas être systématiquement consultable. Un fichier peut être jugé confidentiel et, à ce titre, exiger des procédures d'accès contrôlé. La gestion d'une sorte de fichier d'utilisateurs dotés de droits distincts n'est prévue sur aucun de ces trois logiciels.

CONSULTATION DES DONNÉES: MODES DE QUESTIONNEMENT DE LA BASE

FOX+ demande à l'utilisateur de spécifier tous ses critères de recherche par des commandes plus ou moins complexes. Toutefois, cette diversité de commandes de tri et de recherche n'est accessible que si le terminologue est formé à la syntaxe dBASE ou assisté d'un informaticien. Fort de ces connaissances ou de cette assistance, le terminologue peut transformer sa base-dictionnaire en base-corpus, véritable objet d'investigation terminologique. À titre d'exemple, il devient ainsi facile de mesurer dans quelles proportions sont représentées les différentes catégories grammaticales ou encore d'évaluer la diversité — ou l'uniformité — des sources utilisées.

TEXTO offre une convivialité plus importante. Son intérêt majeur sur ce plan réside dans la possibilité d'utiliser très simplement les opérateurs booléens ET-OU-SAUF regroupables sur plusieurs niveaux de parenthèses pour éviter les ambiguïtés. TEXTO permet également de procéder à des recherches débouchant sur des résultats de type statistique.

MC4, quant à lui, est nettement conçu en vue d'une utilisation de type dictionnaire à partir d'un couple de langues et d'un domaine constituant la base de la consultation.

Les fonctions de tri par chaînes de caractères, accessibles sur les trois logiciels, permettent en néologie de dégager, par exemple, les procédés de formation par préfixes gréco-latins.

ÉDITION DES DONNÉES

La consultation est la forme minimale de la communication de l'information terminologique. L'édition permet de faire de ces données une exploitation dégagée de toute contrainte spatio-temporelle. La souplesse de l'outil informatique doit permettre à la fois l'impression de données en *édition de travail*, suffisante pour une communication ponctuelle d'informations et l'impression de produits élaborés (lexiques, vocabulaires, dictionnaires...) nécessitant un effort de mise en page et de présentation typographique à la fois agréable et fonctionnelle.

Les concepteurs de MC4 ont fait un réel effort pour tenir compte de ces exigences d'édition de données. Ainsi, sans même recourir à des procédures de traitement de texte, on peut obtenir une impression de fichiers de qualité correcte. Le module d'édition de MC4 — bien que contraignant et quelque peu fastidieux — permet une édition de lexiques alphabétiques multilingues d'aspect honorable pour un usage interne. TEXTO propose la création de *fichiers d'édition* (modèles d'édition prédéfinis et réutilisables) qui

peuvent porter sur une fiche isolée, une séquence de fiches ou l'ensemble du fichier dans un environnement convenablement ergonomique.

Ni TEXTO, ni MC4 n'offrent d'options typographiques (italiques, gras...) pourtant tout à fait nécessaires pour un produit dictionnairique où la qualité d'impression conditionne la vitesse et l'efficacité de lecture. FOX+ laisse l'utilisateur très démuni : procédures d'impression complexes et instructions sibyllines. Si un informaticien compétent peut extraire de FOX+ des merveilles d'édition (mises en page, options typographiques...), l'utilisateur béotien doit se satisfaire de texte imprimé au kilomètre à la limite de la lisibilité.

GESTION DU SGBD

Les performances d'un SGBD dans la conduite des opérations courantes sont sans doute celles auxquelles l'utilisateur du système est le plus directement sensible. Elles interviennent généralement pour une large part dans le choix du logiciel. Mais le terminologue doit être conscient qu'une gestion satisfaisante des données suppose à l'arrière-plan une bonne gestion du système tout entier. Ces opérations de gestion des données ont pour objectif principal de garantir la cohérence de la base par une réorganisation régulière des fichiers. Elles assurent également la sécurité des données grâce à des procédures de sauvegarde et de copie ou même à des procédures de restauration de données en cas de détérioration grave.

RÉORGANISATION DES FICHIERS

Dans FOX+, la réorganisation de la base se fait automatiquement au fur et à mesure des modifications apportées. MC4 et TEXTO sont pourvus de programmes utilitaires (UTILMC4 et TEXUTIL respectivement). Ces programmes permettent de réorganiser les fichiers pour en maintenir la cohérence, garantir la rapidité des temps de réponse et permettre une occupation rationnelle de l'espace disque. Cette réorganisation reste limitée dans MC4 à une simple réindexation des termes, des définitions et des liens notionnels. TEXTO permet de réorganiser totalement le fichier avec renumérotation des termes si nécessaire.

SÉCURITÉ DES DONNÉES

Ce souci de sécurité doit être obsessionnel dans l'esprit du gestionnaire s'il ne veut être exposé à des pertes irrécupérables. La première mesure de prudence est la recopie régulière sur supports externes. Il est donc très appréciable que le SGBD prenne en charge les procédures de sauvegarde. Et plus elles seront faciles, rapides et conviviales, moins le terminologue-gestionnaire les négligera et mieux la sécurité sera assurée. Cette évidence mérite d'être soulignée car son intérêt n'est hélas perceptible qu'*après l'accident*.

COMMUNICATION DES DONNÉES

Les possibilités de transfert de données, que ce soit en mode émission ou réception, sont un autre paramètre important pour l'intégration du SGBD dans un environnement informatisé. Ni FOX+, ni MC4 ne possèdent de procédures de téléchargement ou téléversement ou d'exportation de fichiers. Les concepteurs de TEXTO ont mis au point le logiciel LOGOTEL, complémentaire de TEXTO, destiné spécifiquement aux procédures d'échanges de données entre sites TEXTO.

Les fichiers de MC4, bien que non directement transférables, ont été conçus pour pouvoir être intégrés à la base EURODICAUTOM, atout non négligeable.

CONCLUSION

Nous ne concluons pas en attribuant à ces logiciels des étoiles dignes d'un guide du consommateur éclairé. Notre objectif n'est pas de chercher à établir un palmarès en décrétant tel logiciel supérieur à tel autre. L'activité terminologique étant par essence protéiforme, la question que chaque utilisateur doit résoudre pour son propre environnement est la suivante : quel logiciel pour quels besoins ?

Ainsi MC4 offre la possibilité de travailler dans une perspective multilingue sans que ce soit pour autant une obligation. FOX+ présente l'intérêt de laisser une réelle marge de manœuvre au terminologue puisque celui-ci configure librement sa base de données en fonction de ses intentions. TEXTO, enfin, permet de faire de la terminologie avec des procédures de recherche documentaire riches et exploitables.

Nous nous sommes volontairement limités ici à trois logiciels, laissant ainsi dans l'ombre nombre de produits — connus et inconnus de nous — dont certains sont peut-être et probablement plus performants que ceux présentés ici. Si nous avons pu simplement éclairer la lanterne du terminologue en quête de logiciels, ne serait-ce qu'en jetant un faible halo sur la forêt des critères de choix, nous aurons modestement atteint notre objectif.

BIBLIOGRAPHIE

MAURICE, N. (à paraître) : *Évaluation des logiciels MC4, AQUILA et FoxBASE+ dans une perspective terminologique*, Collection Recherches en sciences de l'information, Association des documentalistes et bibliothécaires spécialisés.

INFORMATIONS TECHNIQUES**MC4**

Concepteur/diffuseur : TERMINFORMATIQUE
 Complexe scientifique des Cézeaux
 B.P. 45
 63170 AUBIERE (France)

Année de création : 1985
 Dernière version : 1990
 Prix : 6 000 TTC (juillet 1989)

FOX+

Concepteur : FOX SOFTWARE
 Diffuseur pour la France : A. B. SOFT
 27, rue de Montevideo
 75016 PARIS (France)

Année de création : 1986
 Dernière version : 1990
 Prix : 8 950 F HT pour la version anglaise

TEXTO

Concepteur/Diffuseur : CHEMDATA S. A.
 Tour Saône-Croix Rousse
 17, quai Joseph Gillet
 69316 LYON CEDEX 04

Dernière version : 1990
 Année de création : 1975
 Prix : TEXTO : 20 000 F HT
 LOGOTEL : 10 000 F HT