

Un modèle de mini-banque de terminologie bilingue

Jean Baudot, André Clas et Maurice Gross

Volume 26, numéro 4, décembre 1981

URI : <https://id.erudit.org/iderudit/002006ar>

DOI : <https://doi.org/10.7202/002006ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Baudot, J., Clas, A. & Gross, M. (1981). Un modèle de mini-banque de terminologie bilingue. *Meta*, 26(4), 315–331. <https://doi.org/10.7202/002006ar>

Un modèle de mini-banque de terminologie bilingue

JEAN BAUDOT
ANDRÉ CLAS
MAURICE GROSS

INTRODUCTION

On sait que les découvertes scientifiques et les inventions techniques entraînent aussi des créations ou des modifications langagières. Les choses nouvelles reçoivent une appellation nouvelle. Ainsi, on peut voir dans un catalogue d'un fabricant de pièces électroniques qu'il y a, sur une période d'environ cinq ans, remplacement de 50% des objets et de leur appellation. Si les nouvelles inventions apportent de nouveaux mots, elles peuvent parfois aussi créer de nouvelles possibilités de « traitement » de ces mots. Ainsi l'invention de l'ordinateur a donné naissance aux recherches en traduction automatique qui, à leur tour, ont favorisé les travaux sur les dictionnaires automatiques. La *1961 International Conference on Machine Translation of Language and Applied Language Analysis* de Teddington, qui provoque la rencontre de J. A. Bachrach et de L. Hirschberg, donne naissance aux premières recherches sur la création d'un dictionnaire automatique multilingue, le Dicautom, pour les besoins de la Haute Autorité de la Communauté Européenne du Charbon et de l'Acier (Luxembourg). Le Dicautom visait à rassembler en mémoire électronique un grand corpus consultable automatiquement qui enregistrerait toutes les possibilités de traduction des termes dans leurs contextes et tous les renseignements utiles sur l'origine des textes. Pour faciliter la consultation, puisqu'au fond il s'agissait d'un dictionnaire multilingue phraséologique, on avait constitué une liste des mots-clé les plus importants. Les recherches ont ensuite été orientées vers la constitution de dictionnaires de locutions (LOCFRA, LOCNE, LOCDE, LOCIT) en français, néerlandais, allemand et italien pour permettre aux traducteurs d'interroger l'ordinateur sur des phrases à traduire en les comparant aux « phrases-exemples » consignées dans le dictionnaire. À défaut de la même phrase, l'ordinateur devait donner les phrases les plus proches possibles ou la traduction des unités lexicales isolées. On cherchait donc à réaliser un système de consultation terminologique semi-automatique en même temps qu'une banque de données documentaires.

Le projet Dicautom devint en 1973, après une période d'interruption, l'Eurodicautom, c'est-à-dire une banque de terminologie qui contient les termes avec ou sans contexte et leurs équivalents en six langues. L'Eurodicautom est conçu

comme une banque de terminologie spécifiquement au service des traducteurs de la Commission économique européenne.

En 1968, lors d'un colloque organisé par l'Office de la langue française, fut lancée l'idée d'une banque de terminologie au Canada par l'un des soussignés. Cette banque, qui était une entreprise tripartite — Office de la langue française du Québec, Bureau des traductions du Secrétariat d'État du Canada et Université de Montréal —, fut officiellement lancée en 1970. Les recherches se poursuivirent à l'Université de Montréal jusqu'en 1976 et, le stade d'exploitation étant atteint, la banque (appelée TERMIUM) fut mise à la disposition des gouvernements. Le Bureau des traductions du Secrétariat d'État du Canada continua l'exploitation et l'enrichissement de cette banque et poursuivit les recherches pour améliorer le logiciel. La banque devint la Banque de terminologie du Canada. Là encore, la banque est conçue principalement pour répondre d'abord aux besoins des traducteurs.

L'Office de la langue française renonça à l'entente avec l'Université de Montréal et entreprit, en 1973, la mise sur pied de la Banque de terminologie du Québec. Cette banque vise avant tout à traiter et à diffuser les renseignements terminologiques nécessaires à la francisation du Québec.

La maison Siemens a créé, en 1967, son système TEAM, une banque multilingue au service des traducteurs et rédacteurs de l'entreprise.

En 1966, le Bundesprachenamt de la République fédérale d'Allemagne créa aussi une banque de terminologie, appelée par la suite LEXIS, pour faciliter la traduction de textes techniques vers l'allemand. Il faut signaler encore la création en 1972, par l'Association française de normalisation (AFNOR), de NORMATERM, banque de terminologie conçue pour l'exploitation de la terminologie technique des normes françaises.

Il n'est pas question ici de se livrer à un inventaire exhaustif, ni à une description détaillée des banques de terminologie qui existent dans le monde, mais simplement à rappeler à grands traits quelques faits qui permettent de situer un ensemble de recherches et de besoins. Toutes ces banques fonctionnent sur de gros ordinateurs et, très souvent, les systèmes d'interrogation sont assez complexes. Il est vrai que leur objectif est de recenser toutes les données terminologiques et de satisfaire les besoins de la clientèle dans tous les domaines de l'activité humaine. En fait, comme les données sont subdivisées en domaines spécialisés, on peut dire que ces banques de terminologie sont en réalité des banques de terminologies ou, si l'on veut, un ensemble de petites banques spécialisées.

Les nouvelles inventions techniques, notamment la miniaturisation, qui ont donné lieu à la fabrication de petits ordinateurs, permettent d'envisager des créations de mini-banques sur des mini-ordinateurs à la portée de tous.

1. LES BESOINS DU TRADUCTEUR

Le traducteur est appelé à exercer son métier dans les divers secteurs de l'activité humaine, mais une très grande part de son travail se situe dans les domaines de ce qu'on appelle la traduction technico-scientifique. Le traducteur

est alors appelé à se spécialiser dans un ou plusieurs domaines très spécifiques où sa compétence est d'autant plus évidente qu'il possède une bonne compréhension de sa matière et connaît la terminologie fondamentale. Il n'a donc recours à des aides à la traduction que s'il hésite sur des notions complexes ou si la terminologie est peu courante, neuve ou même inexistante.

Les nombreux glossaires et lexiques qu'il est appelé à consulter se révèlent souvent insuffisants. En effet, c'est un fait connu qu'un listage alphabétique linéaire alourdit la consultation dès que les entrées sont constituées d'expressions formées de plusieurs mots. La mise à jour de ce listage est aussi rendue complexe, car elle nécessite une refonte totale qui est toujours consommatrice de temps et onéreuse. Si le traducteur se trouve, par exemple, devant une expression comme

(1) declining balance method of depreciation,

il risque de feuilleter fort longtemps ses dictionnaires pour y consulter, parfois sans succès, les articles de chacun des quatre mots principaux!

Les éditeurs de lexiques ont parfois tenté de remédier à cette difficulté de consultation soit en incluant des renvois, soit en adoptant une méthode systématique pour déterminer le rang alphabétique d'une expression. Dans ce dernier cas, on a fait souvent apparaître l'expression au rang de son mot principal, c'est-à-dire en donnant l'entrée au pivot de l'expression ou à la «tête» de son agencement syntaxique. Cette méthode veut pallier ainsi les difficultés dues aux différences de structure syntaxique des expressions de langues différentes. On sait, par exemple, qu'en français les syntagmes nominaux commencent plutôt par le mot principal, alors qu'en anglais, on le rencontre le plus souvent à la fin.

Cette approche n'est cependant rentable que si l'expression recherchée se trouve effectivement dans le dictionnaire. Souvent, malheureusement, les expressions à traduire n'apparaissent pas, comme telles, dans les lexiques consultés.

C'est précisément dans ces cas que le traducteur a besoin d'aide. Lorsque l'expression à traduire n'est pas relevée par les lexiques consultés, le traducteur souhaite pouvoir consulter des expressions similaires pour lesquelles des équivalents ont été répertoriés. Souvent, il s'agit d'expressions dont certains constituants sont identiques. Or, il est pratiquement impossible de repérer dans un lexique une expression ayant un même constituant, si celui-ci n'est pas la «tête de l'expression». Par exemple, s'il y a une entrée pour

(2) declining balance approach,

rangée alphabétiquement à «approach», cette adresse ne pourrait être dépistée bien qu'elle puisse être utile au traducteur confronté avec l'expression (1).

De même pour traduire

(3) vitesse de recombinaison,

si cette expression n'est pas lexicalisée, il est avantageux de pouvoir repérer le terme apparenté

(4) taux de recombinaison.

On a souvent constaté que la terminologie d'un domaine technique est constituée, dans sa plus grande partie, d'expressions. C'est ainsi que, dans un vocabulaire spécialisé en énergie solaire, on trouve la répartition suivante : substantifs, 15% ; expressions nominales, 82% ; codes ou abréviations, 3%.

Si l'on analyse un lexique portant sur le financement des entreprises, on obtient une répartition analogue : adjectifs, 1% ; substantifs, 16% ; expressions nominales, 83%. Il est clair que le syntagme nominal est le type d'expression le plus fréquent.

Pour traduire ces « empilages » de mots, il est très souvent intéressant de rechercher les autres expressions du même domaine, car celles-ci fourniront des indices utiles, les mêmes constituants ayant la plus forte probabilité de définir les mêmes notions. Ainsi, pour traduire

(5) inflatable occupant restraint system released by radar sensor,
il y a intérêt à chercher, dans le même domaine (nouvelles techniques de transport), si

- inflatable system,
- occupant,
- restraint system,

apparaissent comme constituants d'autres expressions. Ce type de repérage peut, bien sûr, se faire « à la main », mais risque d'être long et décourageant, alors qu'avec une banque de terminologie informatisée, cette forme de repérage devient un jeu facile et automatique.

2. UNE MINI-BANQUE DE TERMINOLOGIE BILINGUE

Qu'est-ce qu'une mini-banque de terminologie ? C'est bien sûr une petite banque par comparaison aux grandes banques de terminologie dont chaque entrée est une fiche constituée de plusieurs champs. C'est aussi une banque qui peut être informatisée sur un petit système informatique, mini ou micro-ordinateur, dont les coûts sont du même ordre de grandeur que d'autres machines de bureau. Essentiellement, cependant, une mini-banque a un contenu informationnel limité sous deux aspects : a) *elle ne donne accès qu'à la terminologie d'un seul domaine spécialisé restreint*, b) *elle ne contient que les termes, leurs synonymes et leurs équivalents dans l'autre langue*. Ces limites doivent être imposées si l'on veut la rendre exploitable sur un mini-système informatique peu coûteux. Une mini-banque n'en est pas moins utile pour autant.

Un seul domaine

Même si, dans la traduction d'un texte technique ou scientifique, on a affaire à plusieurs domaines en même temps, en général, les difficultés terminologiques d'un texte se situent dans un domaine bien particulier et c'est dans ce domaine que le traducteur a besoin d'informations. Si une mini-banque met à la disposition du traducteur la terminologie d'un domaine, cela ne signifie pas que le système qu'il utilise est limité à ce seul domaine. Le système informatique est un système d'interrogation de banque de données, et la banque

est un support d'informations interchangeable. On peut donc disposer d'une bibliothèque de mini-banques dans divers domaines et sélectionner la mini-banque dont on a besoin lors de la traduction d'un texte spécifique. Ceci n'est guère différent du traducteur qui choisit tel ou tel lexique selon le domaine dans lequel il travaille.

Des fiches limitées

Dans la mini-banque que nous décrivons ici, les fiches ne contiennent que les informations essentielles. Une fiche correspond à une notion; elle contient les termes et les synonymes qui délimitent cette notion dans les deux langues et des renseignements grammaticaux pertinents. Des index contiennent tous les constituants de chacun des termes, synonymes ou équivalents (la sélection des constituants et leurs rôles seront décrits plus loin). À la différence des grandes banques de terminologie, les fiches ne contiennent donc ni définition, ni contexte, ni source. Cette restriction s'impose pour limiter la taille du fichier qui ne doit pas dépasser les capacités de stockage des petits systèmes informatiques. Il est possible cependant d'inclure tous les renseignements susceptibles d'être codés (marques d'usage, pondération, etc.). Le traducteur, qui dispose d'une certaine connaissance du domaine dans lequel il travaille, ne devrait pas être gêné par l'absence de renseignements explicatifs et confirmatifs dans son activité courante. Lorsqu'il fait face à un problème terminologique, il connaît bien souvent la notion en cause, mais désire savoir s'il existe un équivalent attesté. S'il n'en existe pas, il a besoin d'indices pour lui permettre de fabriquer un équivalent approprié. Par exemple, pour traduire

(6) under linked signal conditions

le traducteur sera heureux de pouvoir repérer la fiche :

(7e) linked traffic signal control

(7f) signalisation coordonnée — coordination des signaux.

Comme pour tout système d'interrogation de banque de données, la véritable utilité d'une mini-banque de terminologie dépend de la structure du contenant informatique et de la nature du contenu terminologique.

3. LE CONTENANT INFORMATIQUE

Au niveau du matériel, un petit système informatique (mini ou micro-ordinateur) équipé d'une ou plusieurs consoles d'interrogation et d'une mémoire auxiliaire interchangeable (disquette) est suffisant pour implanter une mini-banque. Les disquettes que l'on trouve présentement sur le marché devraient permettre le stockage de deux à trois mille fiches et donc satisfaire certains besoins.

L'intelligence dans la structure

On sait que l'intelligence (artificielle) d'un système de consultation de banque de données est répartie entre le logiciel d'interrogation et la structure des données. Or, dans le passé, on a toujours eu tendance à incorporer la plus grande partie de l'intelligence dans le logiciel, avec d'importantes conséquen-

ces : programmes de grande taille ne pouvant fonctionner que sur de gros systèmes informatiques, temps d'interrogation souvent longs, logiciels difficiles à modifier, complexité du langage et des méthodes d'interrogation. Il est possible de remédier à plusieurs de ces désavantages en attachant plus d'importance à la structure des données. Ainsi, en incorporant le plus d'intelligence possible dans la structure des données, on rentabilise la consultation et on déplace le poids du traitement informatique dans la bonne direction. En effet, dans un système de banque de données il y a essentiellement deux types de logiciel important : d'une part le logiciel de création et de mise à jour de la banque, d'autre part le logiciel d'interrogation. La complexité de ces logiciels est inversée selon que la banque de données est peu ou très structurée. Le tableau ci-dessous illustre schématiquement cette relation :

Structuration (intelligence) des données	Logiciel d'interrogation	Logiciel de création et de mise à jour
peu	lourd	léger
beaucoup	léger	lourd

On interroge une banque de terminologie bien plus souvent qu'on ne la crée ou qu'on ne la met à jour. Il est donc normal d'alléger l'activité informatique la plus fréquente. Le but d'une mini-banque est d'être exploitable sur un petit système. Sa création et sa mise à jour peuvent être exécutées sur un système de plus grande taille. On peut ainsi concevoir l'existence d'un organisme producteur de mini-banques que de petits services de traduction peuvent acquérir et exploiter dans leurs bureaux.

Une complexité transparente

Un aspect important de l'utilisation d'une banque de terminologie informatisée réside dans sa facilité d'interrogation. Dans une exploitation (interrogation) en mode dialogué, l'utilisateur communique avec le système à l'aide d'un langage de commande pour formuler ses requêtes. Ces langages sont souvent complexes et l'utilisateur non-informaticien doit suivre une période de formation pour en maîtriser le maniement. Plus le logiciel d'interrogation est sophistiqué, plus le langage de commande est étendu et plus grands les risques d'erreur lorsque l'utilisateur formule des requêtes. Un système de banque de terminologie informatisée est complexe, mais cette complexité doit être transparente pour l'utilisateur si l'on veut que le système soit une ressource facilement accessible et un outil utile et efficace. Deux moyens permettent de se rapprocher de cet objectif.

D'une part, en concentrant l'intelligence du système dans la structure des données, le logiciel d'interrogation est allégé et son langage de commande associé simplifié. Le nombre de commandes différentes est restreint (une dizaine au plus) et facile à mémoriser.

D'autre part, en rappelant à l'utilisateur, à chaque étape du dialogue interrogatif, quelles sont les commandes qu'il peut soumettre pour poursuivre sa requête, on évite ainsi les erreurs les plus courantes. En effet, dans le déroulement dialogué d'une requête, les commandes ne sont pas toutes valides à chaque étape et il est difficile de se rappeler quel est le sous-ensemble de commandes acceptables à un moment donné.

Nous croyons que les commandes ci-dessous suffiront à la consultation d'une mini-banque :

- ouverture de la séance d'interrogation
- formulation d'une (nouvelle) requête
- spécification de la langue de départ (si elle est différente de la dernière spécification soumise)
- soumission d'un terme et demande d'équivalents
- demande de consultation de synonymes
- soumission d'un constituant et demande de consultation des termes contenant ce constituant
- demande d'équivalents de certains termes choisis parmi ceux qui contiennent le constituant soumis
- retour à la consultation de la liste des termes contenant le constituant soumis
- demande d'assistance
- terminaison de la séance d'interrogation.

Structure des données

Pour faciliter le repérage des termes, toute la structure des données est basée sur la notion de constituants. Dans la préparation d'une fiche, chaque vedette (entrée principale, synonyme ou équivalent) est accompagnée de la liste de ses constituants. Vedettes et constituants forment les entrées d'un index qui permet l'accès aux fiches. Dans l'index, chaque entrée est transcrite en graphie normalisée (l'expression est comprimée, en ne conservant que lettres et chiffres, en éliminant les accents et la distinction majuscule-minuscule) et est accompagnée de trois séries de pointeurs, c'est-à-dire d'un système de renvois. On a donc

- des pointeurs sur les équivalents dans l'autre langue,
- des pointeurs sur les synonymes dans la même langue,
- des pointeurs sur les expressions plus larges, dans la même langue, et dont l'entrée est un constituant.

Cet ensemble de pointeurs constitue un réseau structurel reliant constituants et constitués. Certaines liaisons sont unidirectionnelles, d'autres bidirectionnelles selon les nécessités de parcours. La structure est entièrement symétrique par rapport à chacune des langues. Lorsque l'utilisateur soumet un terme ou un constituant le logiciel n'a qu'une recherche à effectuer : localiser cette expression dans l'index. Si l'expression existe dans l'index, elle constitue un point d'entrée dans la structure des données et toute la suite du traitement de la requête consistera dès lors en parcours le long des branches du réseau. Cette

approche est simple (logiciel réduit exigeant peu d'encombrement en mémoire d'ordinateur) et rapide (faible temps de réponse).

Réponses du système

Lorsque l'utilisateur soumet une expression en début de requête, le système peut y répondre de quatre façons différentes :

- l'expression n'est pas dans la banque ;
- l'expression apparaît comme un terme dans la banque, avec ou sans synonymes et un ou plusieurs équivalents ;
- l'expression apparaît comme constituant dans la banque et donne accès à une liste de termes (expressions plus larges ayant l'expression soumise comme constituant) ayant chacun un ou plusieurs équivalents ;
- l'expression apparaît à la fois comme terme et comme constituant dans la banque (avec les propriétés des deux cas précédents).

Dans chacun des trois derniers cas, la suite des réponses du système dépendra des commandes de l'utilisateur.

Par exemple, l'utilisateur qui soumettrait :

(8) peak hour

risque d'être intéressé de savoir que, bien que cette entrée constitue un terme pour lequel il existe un équivalent, plusieurs expressions plus larges sont également des termes de la banque :

- (9) peak hour factor
- (10) peak hour traffic
- (11) peak hour volume
- (12) vehicular flow at the peak hour.

Le logiciel d'interrogation est écrit et décrit dans un langage informatique de haut niveau pour qu'il soit transportable et exploitable sur différents petits systèmes informatiques. L'utilisateur qui dispose déjà d'une terminologie dans un domaine donné peut demander à l'organisme producteur de la lui constituer en mini-banque structurée.

4. LE CONTENU TERMINOLOGIQUE

Une mini-banque est une aide informatique à l'activité de traduction. À la différence des grandes banques de terminologie qui cataloguent, de façon continue, les résultats de recherches terminologiques, une mini-banque est bâtie à partir d'une terminologie établie. La fiche est une équivalence entre deux ensembles de termes synonymiques :

$S_1 \leftrightarrow S_2$

par exemple :

(13) $\left\{ \begin{array}{l} \text{densité des atomes dopants} \\ \text{concentration du dopage} \end{array} \right\} \leftrightarrow \left\{ \begin{array}{l} \text{doping level} \\ \text{dopant density} \\ \text{doping density} \\ \text{dopant concentration} \end{array} \right\}$

on considère comme synonymes d'expressions entières, les expressions réduites :

- (14) { dendritic web growth }
 { dendritic growth }

et les abréviations :

- (15) { chemical vapour deposition }
 CVD }

En plus des renseignements grammaticaux, chaque terme peut, le cas échéant, être accompagné de quelques paramètres codés (marques d'homographie, marques d'usage, etc.), si ces renseignements supplémentaires sont nécessaires pour éviter des ambiguïtés ou des interprétations fautives.

Dans sa représentation interne, la fiche est limitée à ces renseignements. Par contre, lors de la constitution des index, tous les constituants de chaque terme sont incorporés aux index, et tous les pointeurs nécessaires sont associés à chaque entrée.

À titre d'exemple, si la mini-banque était limitée à la fiche (13) ci-dessus et la fiche (16) suivante :

- (16) { dopage } ↔ { doping }

les index auraient une structure interne analogue au tableau ci-dessous :

Index français

1. atome [6]
2. atome dopant [6]
3. concentration [4]
4. concentration du dopage
 (D) (E) (G) (H) (A)
5. densité [6]
6. densité des atomes dopants
 (D) (E) (G) (H) (A)
7. dopage (F) [4]
8. dopant [6]

Index anglais

- A concentration (D)
- B density (E) (G)
- C dopant (D) (E)
- D dopant concentration
 (4) (6) (A) (G) (H)
- E dopant density
 (4) (6) (D) (G) (H)
- F doping
 (7) (A) (H)
- G doping density
 (4) (6) (D) (E) (H)
- H doping level
 (4) (6) (D) (E) (G)
- I level (H)

Convention :

1, 2, ..., A, B, ... : adresses des entrées dans les index ;

○ : pointeur (bidirectionnel) sur équivalent dans l'autre langue ;

△ : pointeur (bidirectionnel) sur synonyme dans la même langue ;

□ : pointeur (unidirectionnel) sur entrée plus large ayant au moins un équivalent dans l'autre langue.

Au fur et à mesure de la constitution de la mini-banque, les entrées correspondantes sont ajoutées (en graphie normalisée) aux index si elles n'y apparaissent pas déjà et des pointeurs sont ajoutés aux entrées existantes. Ce sont les

pointeurs de type □, les plus utiles à la recherche de patrons terminologiques, qui sont les plus susceptibles d'être ajoutés aux entrées existantes.

Avec une telle structure un terme peut être repéré de plusieurs façons :

- par la soumission du terme intégral
- par la soumission d'une variante graphique du terme (qui n'affecte pas la graphie normalisée). Par exemple :

(17) jonction p-n

peut être repéré par :

jonction p/n
jonction P-N
jonction (p, n)
etc.

et

(18) time clock

par

timeclock
time-clock

- par une abréviation du terme
- par l'un quelconque de ses constituants.

6. LA PRÉPARATION DES FICHES

Dans sa forme externe, la fiche contient, pour chacune des langues, chaque terme de l'ensemble synonymique (= ES) accompagné de ses paramètres pertinents et, s'il s'agit d'une expression de plusieurs mots, de ceux de ses constituants qui correspondent à des sous-expressions significatives (= SXS). Nous nous limiterons ici à l'illustration du choix des constituants. La fiche (13), par exemple, aurait la forme suivante :

LANGUE 1	LANGUE 2
1. densité des atomes dopants	A doping level
1.1 densité	A.1 doping
1.2 atome	A.2 level
1.3 dopant	
1.4 atome dopant	
2. concentration du dopage	B dopant density
2.1 concentration	B.1 dopant
2.2 dopage	B.2 density
	C doping density
	C.1 doping
	C.2 density
	D dopant concentration
	D.1 dopant
	D.2 concentration

Légende :

1, 2, ... termes synonymes en langue 1
A, B, ... termes synonymes en langue 2
1.1, 1.2, ... SXS pointant sur 1, etc...
A.1, A.2, ... SXS pointant sur A, etc...

L'aspect le plus important dans la préparation des fiches réside dans le choix (et la forme) des constituants. La méthodologie que nous proposons ici est basée sur deux critères : la simplicité et l'analyse syntaxique.

La simplicité

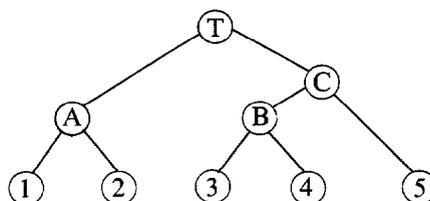
Comme on l'a mentionné précédemment, chaque mini-banque est bâtie à partir d'une terminologie établie pour en faciliter la consultation par des moyens informatiques. La préparation des fiches ne sera vraisemblablement pas assurée par des terminologues et devra suivre des règles simples pour garantir l'uniformité du contenu terminologique. Il est essentiel, lorsque des données doivent être préparées par des personnes différentes et souvent à des moments et des endroits différents, d'éviter dans la méthodologie tout critère susceptible d'être interprété. Il vaut parfois mieux sacrifier certaines informations si l'on n'a pas l'assurance qu'elles seront systématiquement choisies de manière uniforme. C'est ainsi que les SXS sont retenues en appliquant des critères de nature, et non de rôle, et dans une forme morphologique simple. La catégorie grammaticale des entrées est incluse dans le seul souci de fournir une indication utile et de permettre la désambiguïsation d'éventuelles homographies. Quelques statistiques et consultations préliminaires indiquent que cinq catégories grammaticales sont suffisantes ; ce sont :

- V : verbes et groupes verbaux
- N : substantifs et groupes nominaux
- ADJ : adjectifs, participes, groupes adjectivaux
- ADV : adverbes, groupes adverbiaux
- X : autres (noms propres, codes, abréviations, etc.)

Il n'apparaît pas nécessaire de pousser beaucoup plus loin la description grammaticale quand l'observation révèle que la plus grande partie des termes est constituée de groupes nominaux.

L'analyse syntaxique

Le choix des SXS d'un terme de plus d'un mot est basé sur le résultat de son analyse en constituants immédiats, limitée aux mots. Une telle analyse peut être représentée sous forme arborescente :



où :

- 1, 2, ... sont des constituants de premier niveau
- A, B, ... sont des constituants de niveaux supérieurs
- T... est le terme analysé

La sélection et la transcription des SXS se réduisent à l'application de trois règles élémentaires :

Règle 1. Au premier niveau d'analyse, seuls sont retenus comme SXS les constituants qui sont :

- des substantifs
- des verbes
- des adjectifs (ou participes en position d'adjectifs)
- des adverbes
- des noms propres
- des codes ou abréviations

Règle 2. Aux niveaux supérieurs, ne sont considérés comme SXS que les constituants qui sont :

- des groupes verbaux
- des groupes nominaux (*sans* déterminant)
- des groupes adjectivaux
- des groupes adverbiaux
- des juxtapositions de codes

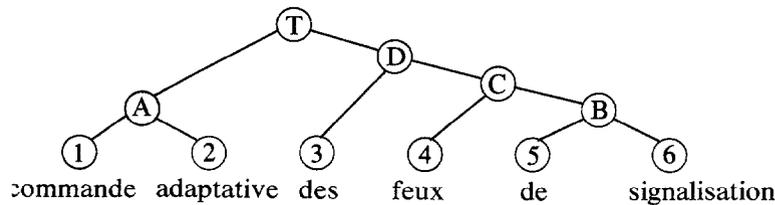
Règle 3. La transcription des SXS sur la fiche externe se fait en leur donnant une forme morphologique fixe (singulier pour substantifs et groupes nominaux, masculin singulier pour adjectifs et groupes adjectivaux, infinitif pour verbes et groupes verbaux).

Cette dernière règle présente le double avantage de standardiser la préparation des fiches et d'alléger le logiciel de repérage en éliminant les problèmes d'accord.

À titre d'exemple, voici comment on préparerait la fiche externe correspondant à la notion définie par l'équivalence entre les ES (19) et (20) suivants :

(19)	{ commande adaptative des feux de signalisation }	(19-1)
	{ commande des signaux par le trafic }	(19-2)
(20)	{ vehicle actuated traffic signals }	(20-1)
	{ vehicle actuated signalization }	(20-2)

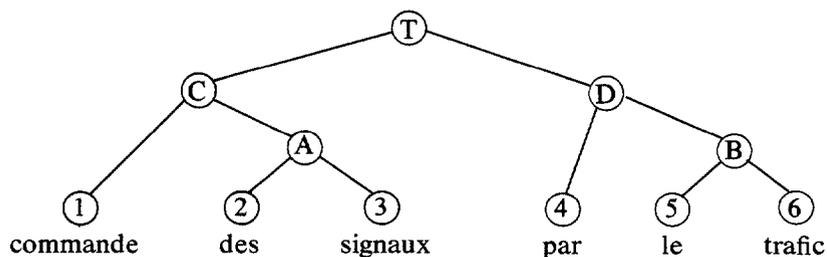
L'analyse (19-1) donne :



et le choix des SXS retient les constituants 1, 2, 4, 6, A et C :

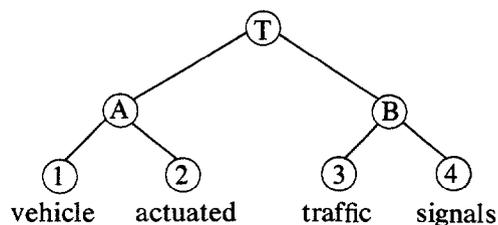
- commande (règle 1)
- adaptatif (règles 1 et 3)
- feu (règles 1 et 3)
- signalisation (règle 1)
- commande adaptative (règle 2)
- feu de signalisation (règles 2 et 3)

De la même façon, dans (19-2) :



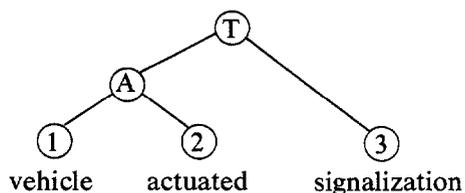
on retient les SXS 1, 3, 6 et C.

Dans (20-1) :



on retient 1, 2, 3, 4, A et B ;

et dans (20-2) :



on retient 1, 2, 3 et A.

6. LE RÔLE DES CONSTITUANTS

En se limitant à l'application de quelques règles simples, la préparation des fiches devient ainsi une entreprise relativement aisée. Il est même possible d'envisager le cas où chaque demi-fiche serait préparée par des personnes unilingues.

De plus, en limitant les moyens de repérage des termes à leurs seuls constituants significatifs on évite de surcharger inutilement soit les index (en y incluant des parties d'expressions non fonctionnelles), soit le logiciel (en lui imposant de balayer les expressions avec des fenêtres de taille variable). Par exemple, le terme :

(21) diode à barrière de Schottky

pourra être repéré par :

— barrière de Schottky (SXS)

et non par

— diode à barrière (pas un constituant)

— de Schottky (constituant non SXS : règle 2).

De même, dans :

(22) chemical vapour deposition

on pourra repérer ce terme avec le SXS :

— vapour deposition

alors que :

— chemical vapour

n'est pas un constituant pour ce terme.

Pour le traducteur qui recherche de l'assistance pour la traduction d'expressions techniques, ce que nous appelons ici un SXS est un élément logique dans la constitution des expressions, il correspond assez bien à la notion intuitive de constituant.

Dans la structure informatique des données, l'inclusion des SXS dans les index représente un compromis raisonnable. Dans le vocabulaire spécialisé de l'énergie solaire (mentionné précédemment) la longueur des termes, mesurée en mots, se répartit ainsi :

1 mot	: 18%	5 mots	: 5%
2 mots	: 38%	6 mots	: 2%
3 mots	: 26%	7 mots	: 1%
4 mots	: 9%	plus de 7 mots	: 1%

longueur moyenne : 2,6 mots.

La nature des SXS se répartit comme suit :

N	: 80%
ADJ	: 15%
X	: 5%

La préparation des fiches engendre en moyenne 2,5 SXS par terme. Toutefois l'inclusion des SXS ne multiplie pas la taille des index par 2,5, et ceci pour les deux raisons suivantes :

— du point de vue du nombre des entrées, comme les SXS sont ramenées à une forme morphologique fixe, nombreuses sont celles qui apparaissent plusieurs fois dans les fiches externes mais qui n'apparaissent qu'une fois dans les index internes avec autant de pointeurs qu'il y a de termes auxquels elles se réfèrent ;
— du point de vue de l'encombrement, la longueur des SXS est toujours inférieure à celle des termes dont elles sont des constituants.

En première approximation, on peut estimer que l'inclusion des SXS dans la structure des données devrait doubler la taille des index. L'impact sur la quantité totale de stockage nécessaire au système devrait se solder par un accroissement de 20 à 25% par rapport à une mini-banque sans SXS. Il en résulte, par contre, un allègement considérable du logiciel de repérage.

7. LES DIFFICULTÉS

La conception et la constitution d'une mini-banque, telle qu'elle a été décrite ici, soulèvent néanmoins certaines difficultés qui, dans la pratique, doivent être résolues ou contournées.

Les mots fréquents

Une conséquence de l'application stricte de la **Règle 1** dans la préparation des fiches est que certaines SXS de premier niveau risquent d'atteindre un taux d'occurrences tel que leur utilité s'en trouve diminuée. Ainsi, dans le vocabulaire de l'énergie solaire, l'adjectif «solaire» est trop fréquent et l'utilisateur de la mini-banque qui demanderait tous les termes qui contiennent ce mot comme constituant se verrait inondé d'une liste trop longue à consulter. Pour remédier à ce type de référence excessive, on pourrait nuancer la **Règle 1** en spécifiant qu'on ne retient au premier niveau d'analyse que les constituants pertinents, c'est-à-dire ceux qui semblent présenter un bon rendement de repérage.

Dans une méthodologie de préparation de données, une clause de ce genre est fragile car elle est hautement sujette à interprétations et peut par là compromettre l'uniformité du contenu. Le critère de simplicité exige que les règles soient faciles à suivre sans faire appel à des évaluations subjectives. Une solution possible consiste à faire appel à une disposition informatique automatique qui appliquerait une guillotine dès que le nombre de pointeurs d'une SXS dépasse un certain seuil prédéterminé (ceci revient à limiter à un maximum donné le nombre de pointeurs unidirectionnels dont peut être dotée toute entrée dans l'index). Cette façon mécanique de déterminer la non-pertinence d'un constituant (et d'en limiter les conséquences) est probablement plus sûre et maintient la simplicité méthodologique.

Les homographies

Si deux termes (ou SXS) homographes ont des catégories grammaticales distinctes, ils seront automatiquement traités comme des entrées différentes. Par exemple :

(23) photovoltaïque
apparaîtra deux fois dans l'index (catégories N et ADJ).

Il peut cependant arriver que deux termes aient une même catégorie grammaticale tout en appartenant à deux ES distincts (la probabilité de ce phénomène augmente lorsqu'il y a beaucoup d'abréviations). Par exemple :

(24) {délestage} ↔ {advisory diversion}

et

(25) {déviation recommandée} ↔ {advisory diversion}
{alternative route}

Une solution simple consiste à sous-catégoriser, comme dans les dictionnaires, le terme

— advisory diversion

en deux termes :

— advisory diversion 1

— advisory diversion 2

et à signaler à l'utilisateur, qui soumettrait ce terme dans une requête, qu'il existe deux fiches (notions) consultables pour ce terme.

jeure mémoriser des fiches plus complètes (avec définition, contexte, source, etc.) soit au même endroit que les fiches internes, soit par un système de renvois à d'autres mémoires secondaires.

Note. La plupart des exemples ont été extraits de deux vocabulaires du Bureau de terminologie de la Commission des communautés européennes :

— Vocabulaire spécialisé : Nouvelles sources d'énergie.

III Énergie solaire. C.C.E. 1980.

— Glossaire — Nouvelles techniques de transport — C.C.E. 1975.

Ces ouvrages ont été aimablement mis à notre disposition par M. R. Goffin du Bureau de terminologie de la C.C.E.