

Une troisième version du " Dicautom " .

J.A. Bachrach et L. Hirschberg.

Notre dictionnaire automatique fournit la traduction des termes d'un texte, commentée par des exemples, à des équipes de traducteurs et d'interprètes qui travaillent sur les quatre langues de la Communauté Européenne et sur l'anglais.

Il se différencie d'un dictionnaire, fait de papier et de carton, essentiellement par les moyens automatiques dont il dispose pour le tri, le stockage, la mise à jour, la consultation ou l'impression. Mais s'il est plus riche en exemples et manipulé en ordinateur, il a été réuni par des lexicologues, des terminologues et des traducteurs, comme un dictionnaire ordinaire.

L'équipe responsable des versions successives se composait des membres du Bureau de Terminologie de la Haute Autorité de la Ceca, J.A. Bachrach, R. Denti, P. Fleurbaay, J. Goetschaelckx, H. Lench, C. Noss, R. Weisgerber, J. Wilson, et du Centre de Linguistique Automatique Appliquée de l'Université libre de Bruxelles, J. Blois, Fr. Decresy, L. Delhaye, R. Dorrepaal-Van Beek, L. Hirschberg, N. Hollaar, J. Mommens, E. Morlet.

La description présentée ici concerne une troisième version du système, fonctionnant sur une IBM 360/40, achevée en décembre 1966 et mise au point au départ de la langue allemande. La première version sur IBM 1401 et au départ du français, date de 1962 (" La Traduction Automatique " - IV-3-1963). La deuxième a été présentée à Venise par J.A. Bachrach au " Séminaire International de l'OTAN sur la Documentation Automatique " (juillet 1963) et publiée depuis en divers endroits. Elle fonctionne sur IBM 1410, au départ du français et du néerlandais. Ainsi qu'il ressort du rapport " Language and Machine ", publié fin 1966 par la National Academy of Sciences, Washington, il n'existe qu'une seule formule qui se rapproche de la nôtre, à savoir celle adoptée par le service de traduction de l'armée allemande à Mannheim.

Au cours de cette nouvelle expérience nous avons précisé les définitions des concepts linguistiques, simplifié et systématisé les codages manuels. Nous avons réanalysé et transformé le processus et réorganisé les fichiers. Enfin, en tenant compte des possibilités nouvelles de la machine, d'une mémoire de 256K, et en fixant les modes d'exploitation des programmes, nous en avons considérablement amélioré le rendement.

Avant d'entrer dans le détail de la description du système, il est bon de donner un aperçu de son utilisation et des motivations qui ont guidé sa conception.

Notre dictionnaire comporte un mot ou une locution de la langue source, sa ou ses traductions dans les langues cibles, et pour chaque unité des exemples de contextes et leurs traductions. Ce qui est valable pour la langue générale, l'est également pour la langue technique ; la traduction d'un terme ne dépend pas seulement du domaine, de la matière traitée, de l'époque où il a été employé, mais aussi des contextes dans lesquels il se place.

Toutes ces données sont réunies dans le dictionnaire.

Quelle sera la source de nos exemples ? Les terminologues puisent de préférence leurs informations dans des documents originaux : manuels, rapports, catalogues, brochures scientifiques ou publicitaires et surtout dans les revues spécialisées. L'emploi de textes traduits en plusieurs langues serait plus simple, mais risquerait d'entraîner la confection d'un langage artificiel et de perpétuer des erreurs. Il importe de se servir dans chaque langue de la terminologie propre au secteur en question dans le pays d'origine de cette langue.

Le terminologue reçoit du traducteur des questions, sous forme de phrases-demandes ou de fragments de phrases dans lesquelles le traducteur a senti des pièges de traduction. Du stock de phrases-exemples que nous avons accumulées dans notre fichier - qui s'élève à quelque six mille unités aujourd'hui, mais peut arriver à cent mille - l'ordinateur doit extraire les exemples de traduction les "plus proches possibles" de ceux de la demande. La présence des contextes dans les réponses permet au traducteur de choisir lui-même la traduction des mots, en contrôlant le bien-fondé des traductions proposées par les exemples, relativement aux cas qui l'occupent. S'il y a doute, plusieurs réponses sortiront, parmi lesquelles le traducteur éliminera lui-même les réponses inadéquates. (Illustration I)

Une autre condition importante pour l'utilisateur est que les questions puissent être formulées librement. En aucun cas, la formulation de la question ne nécessite une connaissance même partielle de la réponse : tout l'effort de construction du système incombe au réalisateur. L'utilisateur souligne simplement un fragment de quelques mots. Pour que ce fragment de texte à traduire puisse être découpé sans précautions particulières, les terminologues doivent veiller à opérer un découpage judicieux pour obtenir les phrases-exemples.

L'absence de tout codage au moment de la demande, implique un codage effectué lors de la construction du dictionnaire, qui donne des lois d'identification des formes du texte courant avec les unités du lexique. Une telle structure se justifie par le fait que le programme est conçu pour consulter, fréquemment et en vue d'une très grande quantité de textes à traduire, un dictionnaire de phrases relativement limité, de sorte que le codage à l'entrée est le plus économique. (Illustration II)

Illustration I

demande 000000012

eine § traenke dieser art besteht aus einem robusten verzinkten § stahlblechgehaeuse

réponses

traenke als automatische traenken haben sich tropftraenken und rinnentraenken bewaehrt
fra gli abbeveratoi automatici si sono affermati quelli a goccia e quelli a livello costante

bestehen aus die verteilerkoepe der melkzeuge bestehen aus gedruckten nichtrostenden stahlblechteilen
blech i distributori delle unita di mungitura sono composti di pezzi a rilievo in lamiera di acciaio inossidabile

bestehen aus luftheizungsoefen bestehen aus einem inneren ofenmantel aus stahl und einem aeusseren stahlblechmantel
blech le stufe ad aria calda sono costituite di un rivestimento del forno in acciaio e di un rivestimento metallico esterno

stahl selbsttragende pfannen aus verzinktem stahlblech
blech elemente autoportanti in lamiera d acciaio zincata
verzinken

blech ventilatoren fuer unterdachtroeknung werden zum
gehaeuse ortsfesten einbau in stahlblechgehaeuse montiert geliefert
i ventilatori per l essiccamento dell erba in fienile sono disponibili con involucri in lamiera d acciaio per montaggio stabile nel fienile

demande 000000013

hochfoerderrinne mit schubstangen und § sperrechen

pas de réponse.

Trois modes de consultation paraissent adéquats selon le rythme et le volume des questions.

" En direct " tout d'abord, le programme se trouve en permanence tout entier dans la machine, munie de tous les fichiers, et simultanément d'ailleurs avec des programmes appelés pour d'autres travaux. Le demandeur possède sur sa table un clavier d'appel qu'il actionne à tout moment et même pour une seule question. La réponse apparaît sur un écran de télévision ou se dactylographie sur une machine à écrire. Nous arriverons à ce mode de fonctionnement lorsque notre ordinateur sera organisé en multiprogrammation et lorsque nous posséderons des unités de stockage permanent pour nos dictionnaires.

L'organisation actuelle permet de travailler " en différé ", c'est-à-dire de réunir jusqu'à mille phrases demandées en une journée ; les faire perforer par une dactylo sur des cartes IBM et de distribuer les réponses quelques heures plus tard.

Mais une organisation bien plus efficace, lorsqu'il s'agit de traduire systématiquement de grands volumes de textes, pourra être mise en oeuvre bientôt. Il existe déjà en Europe des imprimeries qui font corriger et justifier automatiquement en ordinateur les textes à imprimer. Au moment où l'imprimeur passe les textes de la bande linotype en ordinateur nous pouvons les introduire et les conserver sur bandes magnétiques sans avoir à les perforer. Dès lors si ces textes sont destinés à être traduits, nous pouvons les passer sans aucun travail humain supplémentaire sur notre programme de consultation de dictionnaire. A la sortie, nous obtiendrons, en plus et indépendamment du texte imprimé, une préparation du volume entier, qui nous informera, phrase par phrase, sur les difficultés de traduction vers une ou plusieurs langues et la manière de les surmonter.

Mais avant que nous ne soyons en possession de cet immense fichier et d'une réponse instantanée à toutes nos questions sur un écran de télévision les terminologues doivent répondre aux besoins journaliers par des moyens plus classiques. En aucun cas leurs obligations courantes ne peuvent être diminuées ni ralenties sous prétexte de construction pour le futur. Par ailleurs une réalisation à longue échéance ne peut être effectuée sérieusement en marge de la connaissance des besoins pratiques actuels. C'est pourquoi nous nous servons de leur propre fichier, pour notre dictionnaire automatique, mais aussi pour l'édition de glossaires phraséologiques à l'aide d'un programme qui évite toutes les manipulations ordinaires de la typographie.

Ce que le terminologue introduit dans la machine, pour construire le dictionnaire c'est le contenu d'une fiche, c'est-à-dire un fragment de texte plus ou moins long (jusqu'à vingt mots environ) ainsi qu'un code-matière et une référence bibliographique.

Illustration II

T1PV	SOLIVE		SOLIVES-HOURDIS
T1PV	SOLLICITATION		SOLLICITATION
T1PV	SOLLICITATION		SOLLICITATIONS
T3PF	SOLLICITE		SOLLICITES
T1PM	SONDAGE		SONDAGE
T1PV	SORTE		SORTE
T1PV	SOUDABILITE		SOUDABILITE
T3PI	SOUDABLE		SOUDABLE
T1PM	SOUDAGE		SOUDAGE
T9	SOUDER		SOUDE
T9	SOUDER		SOUDEE
T9	SOUDER	_____	SOUDER
T9	SOUDER		SOUES
T3PE	SOUDEUR		SOUDEUR
T3PE	SOUDEUR	_____	SOUDEUSE
T1PV	SOUDURE		SOUDURE
T1PV	SOUDURE	_____	SOUDURES
T3PF	SOUFFLE		SOUFFLEE
T1PV	SOUFFLERIE		SOUFFLERIE
T1PM	SOUFRE		SOUFRE
T1PV	SOURCE		SOURCE
T8	SOUS		SOUS-FACE
T8	SOUS		SOUS-GROUPES
T8	SOUS		SOUS-SOL
T3PF	SOUTERRAIN		SOUTERRAINES
T1PM	SOUTIEN		SOUTIEN
T3XF	SPATIAL		SPATIALE
T3XF	SPECIAL		SPECIAL
T3XF	SPECIAL		SPECIALE
T3XF	SPECIAL		SPECIAUX
T1PV	SPECIFICATION		SPECIFICATION

Ce fragment contient une difficulté de traduction et le contexte qui conditionne l'emploi de cette tournure particulière. Si nous avons en français, " PONT A POUTRES EN TREILLIS " , le même énoncé est repris dans cinq ou six langues sur la même fiche. Il servira de "phrase-exemple" (au singulier) qui ira dans le fichier.

F PONT A POUTRES EN TREILLIS

D FACHWERKBALKENBRUECKE

I PONTE A TRAVI A TRALICCIO (Illustration III)

N VAKWERKLIJGERBRUG

E TRUSSED GIRDER BRIDGE

Le premier travail de la machine est effectué par un programme du genre "KWIC" qui attache cette fiche toute entière, avec les énoncés, dans toutes les langues, en autant d'endroits différents qu'elle contient de mots: pont, poutres, treillis, vakwerkliggerbrug, bridge etc. En chaque endroit, la fiche va rejoindre d'autres fiches conçues comme elle. (Illustration IV)

En somme, le même stock de phrases-exemples, existant actuellement en cinq langues, est utilisé pour cinq dictionnaires, un dictionnaire par langue de départ. Toutefois, le programme est prévu pour six langues. Il s'ensuit que cinq réponses peuvent être données à partir d'une langue, vers cinq langues différentes. Cela équivaut finalement à trente dictionnaires bilingues.

Insistons maintenant sur le fait que les listes-dictionnaires finales contiennent des "unités lexicales" et non pas des mots, car ces unités sont parfois plus longues, comme "plaque de façade" ou "sauerstoffgeblasenes Eisen", tenus pour des unités. Parfois elles sont plus courtes. Ainsi il faut trois unités lexicales "Rost" "Stab" et "Eisen" pour reconstituer automatiquement c'est-à-dire reconnaître le mot "Roststabeisen" dans le texte d'une demande.

C'est le constructeur qui choisit l'unité lexicale et marque pour chacune d'elles les phrases du stock qui la concernent. Il emploie des critères d'observation basés sur les phrases-exemples et leurs traductions.

Lorsqu'un mot se traduit de plusieurs manières différentes, notre système donne au lexicologue le choix entre deux solutions. Ou bien il crée pour le mot plusieurs rubriques, chaque rubrique contenant d'autres phrases-exemples, et nous disons qu'il s'agit d'autant d' "homonymes" différents de ce mot. Ainsi en allemand "Raum" se traduira dans de nombreux contextes par "espace", dans d'autres par "chambre" ou "local" et enfin nous lui avons même trouvé la traduction "surface". Nous créons des "homonymes", lorsque le nombre de traductions n'est pas trop grand, et les

contextes où ces traductions s'imposent sont nombreux.

Au contraire, nous avons trouvé pour "Lager" plusieurs traductions très particulières qui ne convenaient chacune qu'à un contexte bien défini. Pour définir les cas où il fallait choisir ces traductions, le mieux était de faire des "groupes locutionnels" de ces contextes, c'est-à-dire des unités indépendantes composées de plusieurs autres unités. Notre lexique allemand contient de cette manière "Kugellager" (roulement à bille), "Gleitlager" (appui glissant), "Obstlagerhaus" (dépôt ou entrepôt de fruits), "Lager-behaelter" (cellule de stockage).

Illustration IV

- PONT PARAPETS DE ROUTES ET DE PONTS
 D STRASSEN- UND BRUECKENGELAENDER
 I PARAPETTI STRADALI E DI PONTI
 N BORSTWERINGEN VOOR WEGEN EN BRUGGEN
 E ROAD AND BRIDGE RAILINGS
- PONT PONT A POUTRES EN TREILLIS ET ARTICULATIONS INTERMEDIAIRES
 D FACHWERKBALKENBRUECKE MIT ZWISCHENGELLENKEN
 I PONTE A TRAVI A TRALICCIO ARTICOLAZIONI INTERMEDIE
 N VAKWERKLIJGERBRUGGEN MET SCHARNIERENDE DIAGONALEN
 E LATTICE GIRDER BRIDGES WITH INTERMEDIATE JOINTS
- PONT PONT A POUTRES EN TREILLIS
 D FACHWERKBALKENBRUECKE
 I PONTE A TRAVI A TRALICCIO
 N VAKWERKLIJGERBRUG
 E TRUSSED GIRDER BRIDGE
- PONT PONT BASCULANT
 D KLAPPBRUECKE
 I PONTE LEVATOIO., PONTE RIBALTABILE
 N Klapbrug., Ophaalbrug., Basculebrug
 E BASCULE BRIDGE
- PONT PONT CANTILEVER EN TREILLIS
 D FACHWERK-GERBERTRAEGER-BRUECKE
 I PONTE A SBALZO A TRALICCIO
 N VAKWERK-GERBERLIJGERBRUG
 E CANTILEVER TRUSS BRIDGE

Les "groupes locutionnels", comme les autres unités lexicales, possèdent dans notre dictionnaire des phrases-exemples qui leur sont attachées. Mais il s'agit de reconnaître de telles unités dans le texte d'une demande.

C'est donc par une analyse morphologique au sens le plus large de ce terme, allant de la reconnaissance d'une forme et de sa découpe à la découverte d'un groupement de plusieurs formes dans une phrase, que commencent les programmes de consultation automatique de nos dictionnaires. Ils comportent deux phases importantes en dehors de routines de tri et de mises en page, la deuxième grande étape étant le choix des meilleures phrases-exemples à donner en réponse à une certaine demande. Ce choix est fait parmi les phrases attachées aux rubriques du dictionnaire qui viennent d'être reconnues par l'analyse morphologique.

Les phrases-réponses les meilleures sont choisies par programme. Nous avons élaboré dans ce but des règles précises et un calcul assez compliqué. Mais en un mot il est permis de dire que ce sont les phrases parmi celles du stock qui ont en elles chacune le plus grand nombre possible d'unités lexicales reconnues dans une demande par la machine.

Le dictionnaire se compose de quatre fichiers indépendants, ce qui accélère la consultation automatique et facilite la correction éventuelle de ces fichiers.

Le premier est un dictionnaire alphabétique de mots ou de radicaux, accompagnés de leurs codes morphologiques et de deux numéros de renvoi, le numéro de la liste des groupes locutionnels auxquels ils appartiennent et le numéro de la liste des phrases-exemples dans lesquelles ils apparaissent comme mots simples. L'une des deux listes peut d'ailleurs ne pas exister.

Ce premier fichier est suffisamment dépouillé pour qu'il puisse être chargé en mémoire rapide au moment de la consultation du dictionnaire. Une table d'adresses constituée par les deux premières lettres des mots se forme au même moment dans cette mémoire. Cela permet une analyse morphologique très rapide, sans même trier alphabétiquement les mots d'une demande. (Illustration V)

Toutes les hypothèses de découpes sont faites pour chaque mot, de gauche à droite. Le programme tient compte aussi de l'existence possible, de désinences finales à la fin d'une séquence de lettres.

De plus il suit des règles d'acceptation et de rejet des diverses découpes, dont nous ne donnerons qu'un exemple d'application. Ainsi on voit dans l'illustration V, que "SERIEN|BAUTEN" a été découpée en deux mots, tandis que la découpe "BAU|KASTEN|SYSTEM" a été rejetée (un signe "moins" se trouve devant le sigle DE) au profit de la découpe "BAUKASTEN|SYSTEM". Il est évident qu'il refuse aussi toute hypothèse qui ne recouvre pas convenablement une séquence de lettres sans blancs.

Illustration V

MCT		00000000	0000C5D8	00000
MCT	VORGEFERTIGTE	0000C5A0	0000C6DC	0000C
DE		00000000	00000000	00000
F	VOR	00000000	0000C658	0000C
F	GEFERTIGTE	0000C618	00000000	0000C
MCT	STANDARDBAUTEN	0000C5D8	0000C874	0000C
DE		00000000	0000C7B4	00000
F	STANDARD	00000000	0000C744	0000C
F	BAUTEN	0000C71C	00000000	0000C
DE		0000C708	0000C814	00000
F	STANDARD	00000000	0000C7F0	0000C
F	BAUTEN	0000C7C8	00000000	0000C
DE		0000C784	00000000	00000
F	STANDARD	00000000	0000C850	0000C
F	BAUTEN	0000C828	00000000	0000C
MCT	UND	0000C6DC	0000C894	0000C
MCT	SERIENBAUTEN	0000C874	0000C990	0000C
DE		00000000	00000000	00000
F	SERIEN	00000000	0000C918	0000C
F	BAUTEN	0000C8D4	00000000	0000C
MCT	IM	0000C894	0000C9E4	0000C
DE		00000000	00000000	00000
F	IM	00000000	00000000	0000C
MCT	BAUKASTENSYSTEM	0000C990	0000CBE4	0000C
-DE		00000000	0000CA8C	00000
F	BAU	00000000	0000CA44	0000C
F	KASTEN	0000CA24	0000CA68	0000C
F	SYSTEM	0000CA44	00000000	0000C
-DE		0000CA10	0000CB08	00000
F	BAU	00000000	0000CAC0	0000C
F	KASTEN	0000CAA0	0000CAE4	0000C
F	SYSTEM	0000CAC0	00000000	0000C
-DE		0000CA8C	0000CB84	00000
F	BAU	00000000	0000CB3C	0000C
F	KASTEN	0000CB1C	0000CB60	0000C
F	SYSTEM	0000CB3C	00000000	0000C
DE		0000CB08	00000000	00000
F	BAUKASTEN	00000000	0000CBC0	0000C
F	SYSTEM	0000CB98	00000000	0000C
MCT	.	0000C9E4	0000C5BC	0000C
MCT		0000CBE4	00000000	00000
0001	VOR			
0002	FERTIG			
0003	STANDARD			
0004	BAU			
0005	BAUTEN			
0006	SERIE			
0007	IM			
0008	BAUKASTEN			
0009	SYSTEM			
TEMPS	FN SEC=	000003.6		

Tout le processus n'a duré que 3,6 secondes pour une demande de six mots dont quatre ont été découpés en deux.

Le numéro de liste des groupes locutionnels est le premier argument de tri du deuxième fichier, qui donne ainsi pour chaque mot les numéros des groupes locutionnels auxquels il appartient dans le dictionnaire. Les mots eux-mêmes ne réapparaissent pas dans cette deuxième liste, mais seulement leurs caractéristiques relativement à chaque groupe, qui permettent de vérifier si les conditions d'existence du groupe sont remplies dans un texte que l'on analyse. (Illustration VI)

Après la consultation des deux premiers fichiers, les mots d'une demande ont été découpés en fragments ou (et) réunis en groupes locutionnels. Chaque fragment peut avoir aussi été identifié avec plusieurs homonymes.

A chaque homonyme libre, comme à chaque fragment d'un groupe locutionnel, correspond une liste des numéros des phrases-réponses variables pour lui. C'est le troisième fichier qui contient ces listes de numéros de phrases, sans contenir les phrases elles-mêmes. Il mentionne aussi les langues-cibles pour lesquelles ces phrases existent. On remarquera le gain de temps obtenu en effectuant l'analyse jusqu'ici sans se soucier de la langue cible. Ce n'est qu'au moment de la consultation du troisième fichier pour la recherche de la meilleure phrase-réponse que se fait la séparation du processus par langues demandées.

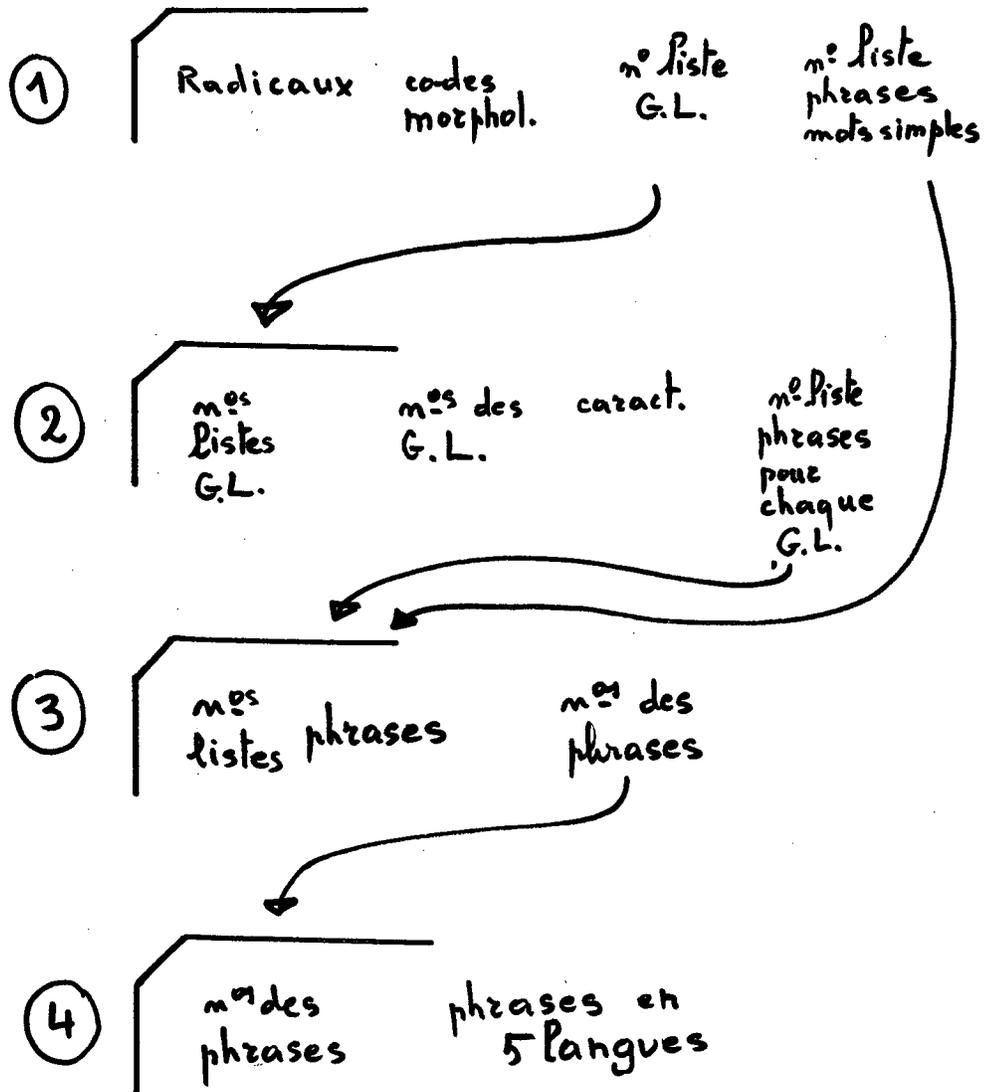
Après consultation de ce troisième fichier, le programme attache à chaque radical découpé dans une demande plusieurs paquets de numéros de phrases-réponses, chaque paquet correspondant à un homonyme ou à une hypothèse de groupe locutionnel. On obtient ainsi un tableau de numéros de phrases où chaque phrase est citée un certain nombre de fois. Nous attribuons à chaque phrase une cote le plus souvent égale à ce nombre, moyennant certaines précautions dans les cas de découpes multiples.

On choisit ensuite pour chaque fragment de la demande la phrase-réponse qui le concerne et qui possède la cote la plus élevée. Cela constitue en même temps un choix parmi ses homonymes. Cependant plusieurs homonymes peuvent être acceptés, avec la meilleure phrase-réponse de chacun d'eux, si elles ont toutes la même cote.

L'appartenance à un groupe locutionnel n'élimine pour un mot toutes ses autres hypothèses que si ce groupe a, dans le texte envisagé, une valeur de certitude qui est déterminée par programme. On se base sur des critères de certitude que nous ne pouvons énumérer ici. Disons par exemple qu'un verbe séparable n'est reconnu avec certitude que dans le cas où la particule est collée au radical verbal. Dans les autres cas on permet à plusieurs hypothèses d'entrer en compétition.

Ce n'est qu'au moment d'éditer les résultats que l'on retire du dernier fichier les phrases-exemples elles-mêmes, qui viennent d'être choisies.

Illustration VI



Ce quatrième fichier, trié par numéros de phrases, peut être beaucoup plus volumineux que les autres et tenir, s'il le faut, dans une mémoire-mars.

Il nous reste à justifier quelques détails de l'analyse morphologique adoptée ici. Elle se caractérise tout d'abord par la comparaison des unités lexicales à partir de la gauche. Les avantages de reconnaissance des formes à partir de la gauche sont liés au fait que la machine consulte d'abord le lexique des formes, et non pas des listes fermées de désinences, en commençant par la droite. Le premier avantage, sauf pour des unités lexicales de moins de quatre lettres et des préfixes, réside dans le fait que la machine va plus vite, parce qu'elle crée beaucoup moins d'hypothèses inutiles, ainsi que nous l'avons vérifié expérimentalement, en simulant les deux méthodes.

Mais il y a un deuxième avantage de loin le plus important surtout pour des langues comme l'allemand et le néerlandais qui agglomèrent en un mot des unités lexicales que l'on peut, -et que l'on doit, - séparer dans le lexique. En effet le programme peut effectuer en une seule fois les découpes des mots et des désinences, jusqu'à épuisement et sans retour en arrière.

Enfin, nous avons groupé les innombrables déclinaisons des substantifs allemands en admettant une seule liste de désinences pour tous, définie par le code signifiant substantif. (Illustration VII)
La simplification des déclinaisons allemandes au point de les ramener toutes à une seule, facilite énormément le codage du dictionnaire, si nous le comparons au codage nécessaire pour l'analyse morphologique élaborée par le groupe de Grenoble, qui tient compte de vingt-trois listes distinctes de désinences pour les substantifs allemands.

Que sacrifions-nous à cette simplification ?

Il n'y a aucun danger, dans le cas présent, à admettre pour un substantif des désinences qu'il ne prendra en réalité jamais. En effet si cela pouvait mener à confondre deux substantifs, il suffirait de donner toutes leurs formes dans le dictionnaire et de les traiter comme des mots invariables. Nous n'avons pas rencontré cette situation sur plusieurs milliers de mots de textes techniques traités ainsi. De toutes manières, dans le but d'obtenir une découpe rapide des composés, faite de gauche à droite et en un seul passage, nous donnons deux formes dans le dictionnaire pour les substantifs qui prennent un "umlaut" ou pour tout autre changement de radical.

Notre analyse perd une seule information, celle de savoir à quel cas et à quel nombre appartient un substantif donné dans un texte. Elle convient donc pour les applications pour lesquelles ces renseignements seraient superflus, telles que les consultations automatiques de dictionnaires qui ne seraient pas suivies d'analyses syntaxiques.

Illustration VII

mots qui les acceptent	noms		adjectifs		verbes			
	F	I	FetI	F	forts	faibles	participes	I
blanc	o	o	o	o	o	o		o
CHEN	o							
CHENS	o							
E	o	o		o	o	o		o
EM				o				
EN	o	o		o	o	o		
END							o	
ER			o	o				
ES	o	o		o				
EST			o		o	o		
ET					o	o	o	
ETE						o		
ETEN						o		
ETEST						o		
ETET						o		
LEIN	o							
LEINS	o							
N	o	o			o	o		
ND							o	
NEN	o							
R				o				
S	o	o						
SE	o							
SEN	o							
SES	o							
ST					o	o		
T					o	o	o	
TE						o		
TEN						o		
TEST						o		
TET						o		

F = finales I = intermédiaires.

Au moment de conclure, il est bon de passer en revue quelques points que soulève la lexicographie classique et de voir dans quelle mesure ils nous concernent ici.

Il s'agit principalement du choix des termes à introduire dans un dictionnaire, c'est-à-dire de la délimitation d'un domaine sémantique et du relevé du plus grand nombre possible de termes que s'y rapportent, ensuite des relations sémantiques entre les termes c'est-à-dire de la structuration sémantique à l'intérieur du lexique choisi; et enfin des problèmes d'adressage, c'est-à-dire de l'organisation matérielle en vue de la consultation, emplacement des termes, répétition éventuelle, orthographes multiples etc...

Si nous n'avons pas centré notre exposé sur ces questions, c'est que le traitement en ordinateur les résoud assez naturellement à condition d'introduire en machine les données complètes et primitives.

L'adressage tout d'abord est à ce point aisé que l'on peut multiplier à souhait les orthographes et que les problèmes d'emplacement et de répétition ne se posent plus.

En introduisant en machine des fragments de textes authentiques, nous notons en même temps leur référence bibliographique et un ou plusieurs codes matières définissant le contexte dont ils ont été extraits. Ces informations sont par là-même et sans autre travail attachées aux termes contenus dans ces fragments. Comme la mémoire de la machine est pratiquement illimitée, on ne doit refuser aucun terme à l'entrée, ni séparer explicitement des domaines, tels la sidérurgie, l'architecture etc... Tout au plus devra-t-on attacher à certaines unités lexicales un signe spécifiant qu'elles n'ont aucun intérêt par elles-mêmes dans un dictionnaire technique (par exemple pour des mots comme "faire", "grand", "froid" etc.).

Une liste parfois longue de codes-matières se trouve attachée à chaque unité par le truchement des diverses phrases-exemples. Il s'établit ainsi spontanément une structuration sémantique complexe à l'intérieur du système que l'on pourrait découvrir par des techniques taxonomiques. Elle est plus riche que la structure en arborescence (analogue à la classification décimale universelle) la seule qui soit réalisable dans un dictionnaire habituel, à côté du classement alphabétique.

Au moment de formuler une demande, sous forme de "phrase difficile à traduire", le champ sémantique est délimité implicitement par le fait que la machine ne cherche que des phrases réponses aussi proches que possible de la demande. Le traducteur reçoit un glossaire qui ne se rapporte qu'à son texte et dans l'ordre d'apparition des termes.

Une mise à jour régulière est prévue qui supprime le problème d'une certaine complétude au départ, et permet d'enregistrer rapidement ces néologismes qui font défaut dans les dictionnaires ordinaires, souvent dépassés dès leur sortie de presse ! Les pertes de temps considérables, occasionnées par des recherches fastidieuses et souvent vaines, seront réduites au strict minimum.

L'établissement d'une véritable collaboration entre les compilateurs du dictionnaire et les utilisateurs mettra à la disposition de tous les données dont disposent quelques-uns et qui risquent, autrement d'être enfouis à tout jamais dans quelque obscur fichier personnel.

Cet instrument souple, puisqu'adaptable à toutes sortes de demandes, peut non seulement faciliter considérablement le travail de son utilisateur, mais lui offre aussi, sans effort supplémentaire de sa part, la possibilité d'un travail plus rapide et de meilleure qualité, ainsi qu'une spécialisation sans peine.

Systeme maniable, facile à étendre et à corriger, l'outil de travail tel que nous le proposons offre de multiples avantages, et de loin de favoriser l'élimination du traducteur humain, il pourra bien au contraire constituer une aide appréciable dans son travail.

Notes bibliographiques.

I.A. MELCUK . Analyse morphologique en traduction automatique.
"Problèmes de Cybermétique" Vol. 6 - 1961 , pp.207-276.

L.E. DOSTERT , R.R. MACDONALD . Georgetown University Machine Translation Research project . General report 1952 - 1963 (n°30) june 1963.

G. LAURENT , A. BERNARD-GEORGES , D. LEVENBACH . Etude Morphologique du substantif allemand -Grenoble, avril 1962 - EUR 525f. 1964.

J. BLOIS , Fr. DECRESY , J. MOMMENS . Analyse morphologique automatique du français . Université libre de Bruxelles . Mars 1963.

DICAUTOM . - La " Traduction Automatique " - 1963 , Vol.IV n°3 (Mouton)

L. HIRSCHBERG . Dictionnaires automatiques pour traducteurs humains.
" Journal des Traducteurs " , Montréal , Vol.10, n°3 (1965)pp. 78-86.

J.A. BACHRACH . Une méthode en terminologie . Comptes rendus du
" Colloque de Linguistique Appliquée " -Ceca Luxembourg- avril 1966.

L. HIRSCHBERG . Conception, construction et utilisation de dictionnaires automatiques multilingues. Comptes rendus du " Colloque de Linguistique Appliquée " - Ceca, Luxembourg - avril 1966.