Vilmos Bárdosi — László Csink
(Université Eötvös Loránd de Budapest)

Le traitement des locutions idiomatiques par micro-ordinateur

1. Remarques préliminaires

Depuis quelques années, un projet de recherches est en cours à la Chaire de Français de l'Université Eötvös Loránd de Budapest sur les locutions idiomatiques du français. Les recherches portent sur des problèmes théoriques de la phraséologie /problèmes de définition, possibilités de classification des unités idiomatiques du français, recherches d'équivalences idiomatiques dans une seconde langue, en l'occurrence le hongrois, etc./ ainsi que sur la mise en pratique lexicographique des résultats de ces recherches sous la forme d'un nouveau dictionnaire bilingue /français-hongrois/ de locutions.

La première phase du travail, qui vient de se terminer, a permis de préciser certaines notions de base, de dégager quelques méthodes lexicographiques peu ou pas encore utilisées en phraséologie et de rédiger un petit dictionnaire français-hongrois des locutions.

1.1. En ce qui concerne les notions de base, c'est sans doute celle de locution qui est la plus difficile à définir. Qu'est-ce qu'on entend donc par locution au sens phraséologique du terme?

Parmi les éléments de la langue qu'il faut acquérir pour s'exprimer, on trouve non seulement les mots au sens traditionnel du terme, mais aussi des ensembles de mots plus ou moins imprévisibles. Les étrangers qui apprennent le français font quotidiennement la fâcheuse constatation que connaître le sens des mots simples comme vers et nez, ainsi que les règles de syntaxe qui permettent de les assembler, ne suffit pas pour comprendre et à fortiori pour bien employer par exemple: tirer les vers du nez à qn 'faire parler habilement qn'. Ainsi, un lexique ne se définit pas seulement par des éléments minimaux, ni par des mots simples et complexes, mais aussi par des suites de mots convenues, fixées, dont la signification n'est guère prévisible et qu'on appelle en général les l o c u t i o n s ou expressions idiomatiques. Les linguistes regroupent tous ces éléments sous le terme communément accepté d'u n i t é s p h r a s é o l o g i q u e s . Malgré les travaux fondamentaux sur ce sujet de Ch. Bally [1] — surtout son Précis de stylistique où le chapitre IV. s'intitule "La phraséologie" —, ce terme est tombé un peu en désuétude en français, à tel point qu'aujourd'hui la phraséologie est souvent

ignorée en France ou identifiée avec la stylistique. Nous employons ce terme par la suite avec le sens que définit ainsi H. Burger:

> La phraséologie est un domaine de la description des langues où entrent des unités normalement plus grandes que les mots — quelquefois des syntagmes — qui peuvent constituer à eux seuls — mais pas forcément — des phrases et dont la signification globale /la signification idiomatique/ est en voie d'intégration ou ne peut être interprétée de façon régulière. [2]

Cette définition est assez générale pour accueillir une grande diversité de combinaisons des unités du lexique. En allant des groupements les plus libres vers les groupements les plus figés, on pourrait avoir dans une classification possible — car il en existe d'autres également — huit catégories:

1/ Les périphrases verbales très répandues actuellement dans le style des mass media: reményt táplál vmi iránt = 'remél' — caresser un espoir = 'espérer'.

2/ Les clichés, les lieux communs: a teljesség igénye nélkül — sans prétendre à l'exhaustivité.

3/ Les expressions imagées, employées au sens abstrait: töri a fejét — se casser la tête.

4/ Les termes géminés: se füle, se farka — n'avoir ni queue ni tête.

5/ La figure étymologique: éli az életét — vivre sa vie.

6/ Les locutions idiomatiques proprement dites: sutba dob — jeter le manche après la cognée.

7/ Les comparaisons idiomatiques: úgy áll rajta, mint tehénen a gatya — cela lui va comme un tablier à une vache.

8/ Les proverbes, les dictons, les adages, les citations: Ajándék lónak ne nézd a fogát — A cheval donné on ne regarde pas la bride; a kocka el van vetve — les dés sont jetés.

Qu'est-ce qui caractérise ces combinaisons énumérées, qui entrent toutes dans la phraséologie?

Elles se composent de deux ou plusieurs mots ayant une relation syntagmatique entre eux. Cette relation est — souvent depuis longtemps — plus ou moins liée et fermée. Ce ne sont pas nous qui générons ces unités au cours de la communication, mais — tels des éléments préfabriqués — elles existent dès le départ dans notre entendement. Leur rôle n'est pas de marquer des relations grammaticales, mais d'évoquer des images, de rendre le message plus expressif. Comme leur signification globale n'est pas la somme des significations concrètes des éléments constitutifs, elles ont la valeur d'un mot. Leur synonyme est aussi souvent un mot particulier. Leur emploi est général dans la langue. Comme ces unités

PHRASÉOLOGIE ⟷ PARÉMIOLOGIE

Unités phraséologiques non idiomatiques	Unités phraséologiques idiomatiques = IDIOTISMES	
1/ reményt táplál caresser un espoir	3/ töri a fejét se casser la tête	8/ a kocka el van vetve les dés sont jetés
2/ a teljesség igénye nélkül sans prétendre à l'exhaustivité	4/ se füle, se farka n'avoir ni queue ni tête	Ajándék lónak ne nézd a fogát A cheval donné on ne regarde pas la bride.
	5/ éli az életét vivre sa vie	
	6/ sutba dob jeter le manche après la cognée	
	7/ úgy áll rajta, mint tehénen a gatya cela lui va comme un tablier à une vache	

Tableau 1.

ont des valeurs métaphoriques, leur rôle est avant tout stylistique et consiste à renforcer l'expressivité du style.

En systématisant un peu les groupements qui entrent dans la phraséologie, voir le tableau 1., on exclura à priori — comme l'avaient fait des spécialistes de la question tel Gábor O. Nagy — la catégorie des proverbes en les renvoyant dans la p a r é m i o l o g i e. Pour les catégories qui restent ainsi, on pourrait établir deux groupes selon le degré de l'intégration de la signification. Dans le premier groupe entreraient les périphrases verbales et les clichés, les lieux communs, c'est-à-dire les ensembles dont les éléments disposent d'une certaine possibilité de commutation /Bally parle de "groupements usuels", Coseriu de "solidarités lexicales"/. On les appellera avec József Juhász u n i t é s p h r a s é o l o g i q u e s n o n i d i o m a t i q u e s. On ne s'en occupera pas non plus par la suite. Mais on s'attachera plus particulièrement aux autres groupements /les catégories 3 à 7/, qui seront appelés u n i t é s p h r a s é o l o g i q u e s i d i o m a t i q u e s /le "discours répété" de Coseriu, le "modismo" de Casares/ ou avec un terme technique que l'on voudrait introduire et employer par la suite: i d i o t i s m e s. Dans la définition de l'idiotisme, les caractéristiques énumérées tout à l'heure sont particulièrement dominantes.

Rappelons-les:

a/ une signification idiomatique globale qui ne s'explique pas par l'adjonction des significations concrètes, analytiques des éléments constituants;

b/ l'inaltérabilité des éléments constituants /dans tirer les vers du nez à qn il est par exemple impossible de remplacer 'tirer' par 'traîner' ou 'ver' par 'larve' sans perdre cette signification idiomatique/;

c/ une structure grammaticale et lexicale qui s'écarte souvent de la norme ou est archaïque;

d/ des valeurs métaphoriques particulières qui représentent en général un surplus informationnel par rapport aux équivalents périphrastiques simples /dans tirer les vers du nez a qn il s'agit d'un secret qu'on arrache à quelqu'un adroitement, en le faisant parler habilement/.

Ce sont précisément ces idiotismes et en particulier leur approche lexicographique qui forment l'objet de nos investigations.

1.2. C'est conformément aux caractéristiques énumérées ci-dessus des locutions que, dans la pratique lexicographique, nous avons rédigé un premier recueil de locutions françaises qui est actuellement sous presse [3]. Sa structure, peut-être inhabituelle — signalée et explicitée déjà ailleurs [4] — se présente comme suit.

1.2.1. Le choix des locutions

Le fonds phraséologique d'une langue, au sens large du terme, est immense. Ceci est aussi valable pour le français. Il fallait donc adopter certains critères qui protègeraient de l'arbitraire, dans toute la mesure du possible, la sélection de notre corpus. Nous n'avons retenu d'abord que les locutions verbales ou adverbiales /prendre la clé des champs; à brûle-pourpoint/ et les comparaisons idiomatiques /se ressembler comme deux gouttes d'eau/, qui sont les plus productives en français. Nous en avons recueilli, dans un premier temps, un millier. On ne trouvera pas de locutions dites substantivales /une grosse légume/, d'ailleurs très fréquentes dans la langue, qui sont en voie de lexicalisation ou déjà lexicalisées, parce qu'elles auraient par trop étoffé le recueil. Par ailleurs, il n'est pas toujours facile de décider si une locution est verbale ou substantivale. Dans ces cas-là, nous avons été guidés par des considérations pragmatiques. Ont été classées substantivales, donc éliminées, les unités qui ne peuvent être complétées que par être, devenir ou des verbes similaires /être ou devenir une grosse légume/. Dans les cas limites, nous avons opté pour le maintien de la locution quand elle peut être complétée par toute une série de verbes /être un remède de cheval ⟶ prendre ou demander ou donner ou prescrire un remède de cheval/. Font exception à ce prin-

cipe quelque vingt locutions — essentiellement dans le premier et le dernier chapitre, de nature descriptive —, locutions fréquentes qui, pour des raisons didactiques, doivent figurer dans un recueil de ce genre.

1.2.2. Le regroupement thématique des locutions

La majorité des dictionnaires utilise le classement par ordre alphabétique de la matière. Il est bien évident que ce n'est pas la seule possibilité de classement. Il arrive souvent qu'on ne trouve pas, dans une situation de communication donnée, le mot ou l'expression adéquat à notre message. Et on ne trouvera pas la locution cherchée dans le dictionnaire parce qu'on ne connaît pas son ou ses éléments sous le/s/quel/s/ elle figure. Cependant nous sommes parfaitement en mesure de définir le concept, la notion qui, à un niveau plus abstrait en général, recouvre notre locution particulière. Autrement dit, nous pouvons définir par voie déductive, et en général à l'aide de substantifs, le concept-clé pour chaque locution. Ce concept-clé ne doit pas être confondu avec le mot-clé de la locution, qui est l'élément comportant le plus d'information et sous lequel les dictionnaires alphabétiques rangent traditionnellement les locutions. Par exemple le mot-clé de la locution prendre la mouche est 'mouche', alors que son concept-clé est 'COLÈRE'.

Aussi, contrairement à la pratique lexicographique, notre dictionnaire prend-il comme point de départ du classement les concepts-clés /p.ex.: 'TRAVAIL', 'AMOUR', 'FAIM', 'RICHESSE', etc./, pour lesquels il propose à chaque fois des séries synonymiques de locutions. Une locution qui, en raison de sa polysémie, peut s'intégrer dans plusieurs séries synonymiques, figurera dans chacune de celles-ci. Théoriquement nous aurions pu entreprendre la rédaction d'un vaste réseau de concepts-clés ordonnés alphabétiquement. Que nous n'ayons pas choisi cette solution tient essentiellement à la raison suivante: il nous a semblé plus utile et plus efficace pour l'apprentissage du français langue étrangère d'organiser les concepts-clés non pas dans un ordre alphabétique formel, mais dans des chapitres dont le principe ordonnateur est l'homme: sa réalité physique, ses actes et son comportement, son intellect, ses états d'âme, sa vie sociale et ses rapports avec l'univers.

A l'intérieur des chapitres, ce n'est pas l'ordre alphabétique, mais la synonymie des concepts-clés ainsi que le fait que ceux-ci peuvent s'appeler par voie associative qui seront le principe ordonnateur. Ainsi on aura par exemple SYMPATHIE - AMITIÉ - AMOUR - GALANTERIE - MARIAGE - FAMILLE. Dans les sous-chapitres de ce type, l'ordre des locutions est déterminé soit par les valeurs stylistiques /on aura une

suite de locutions littéraires, neutres, familières, populaires, vulgaires/, soit par une logique interne facilement identifiable /dans le premier chapitre, viennent par exemple d'abord les locutions caractérisant la taille, le corps pour arriver ensuite à celles qui sont relatives aux cheveux, aux yeux, aux oreilles/. On a eu également parfois recours à la combinaison de ces deux principes. On comprendra qu'il n'aurait pas été logique, par exemple, de donner dans le sous-chapitre MARIAGE - FAMILLE d'abord la locution porter la culotte 'exercer l'autorité dans un ménage' avant la locution monter en graine 'avancer en âge et n'être pas encore mariée' uniquement parce que culotte précède graine dans l'alphabet. Par ailleurs, le regroupement thématique éclaire mieux les rapports synonymiques et associatifs des locutions, en permettant d'énumérer les unes après les autres celles qui expriment un même concept. Par exemple: dévisser son billard, casser sa pipe, fermer son parapluie pour le concept-clé 'MOURIR'. Donc l'ordre alphabétique n'a été conservé que dans les index qui figurent dans la dernière partie de l'ouvrage.

1.2.3. La structure des entrées

Conformément à la pratique lexicographique française, les locutions sont données à l'infinitif. Les adjectifs apparaissent seulement au masculin. Les lo-

cutions jamais utilisées à l'infinitif ou dont l'emploi à l'infinitif est inhabituel, figurent sous leur forme la plus fréquente /p.ex.: on le ferait rentrer dans un trou de souris/.

Les parenthèses ( ) désignent un élément facultatif de la locution: se faire des cheveux (blancs).

La barre oblique / désigne les variantes dans une locution; elles sont toujours données par ordre de fréquence, la première variante étant la plus fréquente.

La prononciation correcte de certains mots peu fréquents a été donnée entre crochets [ ] avec les signes de transcription de l'APhI /Association Phonétique Internationale/.

Les compléments obligatoires des locutions ont toujours été signalés /faire la courte échelle à qn, tirer dans les pattes de qn, crier haro sur qn, etc./.

Dans la définition des locutions, nous avons essayé de donner tous les renseignements possibles sur leurs emplois. Les renseignements grammaticaux comme les renseignements situationnels figurent entre < >. Par exemple: se mettre martel en tête 'se faire du souci <utilisé surtout au négatif impératif>. Ou par exemple, à propos de la locution tomber à l'eau il a fallu remarquer que le sujet de la locution ne peut pas être une personne, mais seulement un objet. C'est ce que nous avons représenté sous la forme suivante: un projet ou cela < Snc. = sujet nom de chose > tombe à l'eau.

1.2.4. La valeur stylistique des locutions

La majorité des locutions figurant dans ce recueil appartiennent au niveau de langue neutre. Un certain nombre de locutions littéraires, familières voire vulgaires ont également été retenues en raison de leur fréquence d'emploi. Il est évidemment très important pour les usagers étrangers de tenir rigoureusement compte des indications stylistiques, qui leur permettront, à l'occasion, d'éviter de faux emplois, des malentendus ou des situations pénibles. Pour les indications stylistiques, nous avons essentiellement adopté celles du Petit Robert qui définit les niveaux de langue comme suit:

Argotique: emploi limité à un milieu particulier, surtout professionnel /p.ex. argot scolaire/, mais inconnu du grand public. Pour les mots d'argot passés dans le langage courant, voir: populaire.

Familier: usage parlé et même écrit de la langue quotidienne /conversation courante, etc./; mais ne s'emploierait pas dans les circonstances officielles, solennelles, dans les ouvrages qui se veulent sérieux.

Littéraire: mot ou locution qui n'est pas d'usage familier, qui s'emploie surtout dans la langue écrite élégante.

Populaire: qualifie un mot ou un sens courant dans la langue parlée des milieux populaires /souvent argot ancien répandu/, qui ne s'emploierait

pas dans un milieu social élevé, cultivé.

Vieilli: mot ou sens encore compréhensible de nos jours, mais qui ne s'emploie plus naturellement dans la langue parlée courante.

Vieux: mot, sens ou emploi de l'ancienne langue, incompréhensible ou peu compréhensible de nos jours et jamais employé, sauf par effet de style /archaïsme/.

Vulgaire: mot, sens ou emploi choquant qu'on ne peut utiliser entre personnes bien élevées, quelle que soit leur classe sociale.

Occasionnellement nous avons encore utilisé les qualifications stylistiques suivantes /cf. Hors-texte 2./:

Euphémique /euph/: expression atténuée d'une notion dont l'expression directe aurait quelque chose de déplaisant.

Ironique /iron/: pour se moquer /souvent par antiphrase/.

Moderne /mod/: insiste sur le fait qu'un sens, un emploi est d'usage actuel.

Péjoratif /péj/: avec mépris, en mauvaise part.

Plaisant /plais/: emploi qui vise à être drôle, à amuser.

Poétique /poét/: mot de la langue littéraire, utilisé seulement en poésie.

Régional /rég/: mot ou emploi particulier au français parlé dans une ou plusieurs régions, mais qui n'est pas d'usage général ou qui est senti comme propre à une région.

1.2.5. Les notes en bas de page

Tous les mots qui ne figurent pas dans Francia-Magyar Kéziszótár /Budapest, Akadémiai Kiadó, 1966/ ou qui sont pris dans un sens spécial, par exemple argotique /le mot portugaise 'oreille' dans la locution avoir les portugaises ensablées/, ont été donnés en notes. Les chiffres renvoient dans ce cas-là au mot en question. Nous avons également éclairci en notes l'étymologie, les circonstances socio-culturelles de la formation de telle ou telle locution dont l'origine est obscure, discutée ou utile à connaître dans l'apprentissage du français langue étrangère. Ici les chiffres renvoient non pas à un mot, mais à la locution entière. Nous n'avons pas donné d'explications pour les locutions dites transparentes, utilisées uniquement dans un sens métaphorique /par exemple: avoir le bras long/.

1.2.6. Les index

La partie dictionnaire est suivie de trois index. Le premier énumère alphabétiquement les concepts-clés français /en majuscules/ utilisées dans le dictionnaire ainsi que leurs synonymes les plus

fréquents /en minuscules/. Les chiffres romains renvoient aux chapitres du dictionnaire.

Le deuxième index donne, dans l'ordre alphabétique de leurs mots-clés, toutes les locutions figurant dans le dictionnaire. Le chiffre romain renvoie au chapitre correspondant du dictionnaire et le chiffre arabe au numéro sous lequel figure la locution en question dans ce chapitre. Par exemple: Anglaise: filer à l'~, XX-31 = chapitre XX, locution 31. Le mot hongrois donné entre parenthèses est le mot-clé de la locution hongroise équivalente sous lequel celle-ci peut être retrouvée dans le troisième index, celui des équivalents idiomatiques hongrois.

Ce troisième index ne contient que les équivalents hongrois de type phraséologique et ignore les simples équivalences lexicales. Les locutions hongroises sont données dans l'ordre alphabétique de leurs mots-clés. Nous les avons toujours fait suivre, entre parenthèses, des mots-clés des locutions françaises. Les chiffres romains et arabes localisent, une fois de plus, la locution cherchée: Par exemple: Angolosan: ~ távozik /anglaise, XX-31/.

En ce qui concerne les comparaisons idiomatiques, elles figurent dans l'index français sous le substantif /beau comme un astre → astre; boire comme un trou → trou/. Dans l'index hongrois elles sont données sous leur premier élément, adjectif ou verbe,

sauf si ce premier élément a plusieurs variantes. Par exemple: gyáva, mint a nyúl ⟶ gyáva, mais fekszik/áll/ül, mint egy darab fa ⟶ fa.

2. Le micro-ordinateur dans le traitement des locutions

Il y a à peu près un an, nous sommes arrivées à une nouvelle phase de travail avec la possibilité du traitement par ordinateur de la base de données toujours croissante. En effet, il est apparu rapidement que le traitement régulier, rapide et permettant une analyse parallèle selon plusieurs points de vue d'une base de données même relativement petite /1000 à 2000 idiotimes/ est manuellement impossible. C'est alors que nous avons mis en route un projet dont le but était de réaliser un logiciel, conçu pour la configuration du micro-ordinateur la plus répandue en Hongrie /le Commodore 64 avec un lecteur de disquette 1541 et une imprimante du type MPS 801/ et capable de traiter selon plusieurs points de vue 2000 idiotismes au maximum de n'importe quelle langue à caractères latins.

Pour ce faire, il fallait écrire deux logiciels de base. Un premier, que nous avons baptisé ENREGISTROMAT, pour l'enregistrement des idiotimes sur micro-ordinateur et un deuxième, appelé IDIOMAT, à l'aide duquel les données /les locutions/ enregistrées peuvent être traitées, recherchées et analysées se-

lon les points de vue donnés ci-dessous. Il est important de souligner que notre ENREGISTROMAT est capable de réaliser la base de données /l'ensemble des idiotismes à traiter par IDIOMAT/ de n'importe quelle langue à caractères latins.

Nous avons fait les essais d'analyses sur un matériel français. Ayant voulu réduire au maximum l'arbitraire dans le choix du corpus, nous avons enregistré pour les essais faits avec IDIOMAT les unitées phraséologiques les plus faciles à délimiter, à savoir les comparaisons idiomatiques. Ces quelques 70 comparaisons nous ont servi à essayer et à faire évoluer notre IDIOMAT. Comme nous considérons qu'il satisfait à présent à toutes les exigences que nous avons posées, il sera possible maintenant de réaliser, exploitant le corpus du petit dictionnaire décrit sous 1.2., une base de données plus importante.

Dans ce qui suit, nous passerons rapidement en revue les caractéristiques de l'enregistrement des données et le choix des informations d'après lesquelles s'effectue la recherche des idiotismes.

2.1. La structure de la base de données

2.1.1. Enregistrement des données — base de codes

L'enregistrement des données se fait selon l'ordre choisi par l'utilisateur. Aucune organisation préalable — par exemple une mise en ordre alphabétique

– de la matière n'est demandée. Dans le cas de chaque idiotisme il faut choisir un mot sous lequel l'idiotisme pourra être retrouvé. Nous avons appelé ce mot mot-clé. Est considéré comme mot-clé de l'idiotisme son constituant portant le plus d'information et sous lequel il est expliqué dans les dictionnaires. Ce mot-clé est en général, dans la pratique lexicographique, le premier substantif /cf. 1.3.2./.

Notre logiciel classe et traite les idiotismes sous deux mot-clés /si la structure de la locution le permet, bien entendu/. Ainsi par exemple dans les idiotismes tirer le diable par la queue ou attendre qn comme le Messie, les premiers mots-clés sont respectivement, 'diable' et 'attendre', les deuxièmes mot-clés sont 'queue' et 'Messie'. Quel est le principe directeur dans la désignation des mots-clés?

a/ Pour les comparaisons le premier mot-clé est toujours le verbe ou l'adjectif précédant "comme", le deuxième mot-clé est le mot qui suit "comme".

aa/ si celui est un mot composé – par exemple: agile comme des DOIGTS DE FÉE –, c'est le deuxième élément de la composition qui sera le deuxième mot-clé /fée/.

aaa/ si après "comme" il y a nom + complément circonstanciel – par exemple: s'agiter comme UN DIABLE DANS UN BÉNITIER –, c'est le nom /diable/ qui sera toujours le deuxième mot-clé.

b/ Pour les idiotismes qui ne sont pas des comparaisons, le premier mot-clé sera le premier substantif de l'idiotisme, et le deuxième mot-clé le deuxième substantif de la locution.

bb/ dans le cas de trois substantifs, ce qui est d'ailleurs très rare, il faut choisir les deux qui portent le plus d'information.

bbb/ s'il n'y a qu'un substantif dans l'idiotisme - par exemple: avoir du chien -, le premier mot-clé sera le verbe, le deuxième le substantif.

bbbb/ dans les idiotismes à structure attributive - être fou à lier - c'est l'attribut /fou/ qui sera le premier mot-clé et le verbe non-copule /lier/ le deuxième.

bbbbb/ l'absence d'un des deux mots-clés est possible. Par exemple: à l'OEIL.

bbbbbb/ les variantes sont considérées et enregistrées comme deux idiotismes à part, ainsi par exemple dans COUPER/TRANCHER le noeud gordien, "couper" et "trancher" seront à chaque fois des premiers mots-clés.

La rigueur dans la désignation des mots-clés est très importante car une éventuelle recherche d'après ceux-ci n'est possible que si nous donnons le même mot-clé, et avec la même orthographe, que celui qui a été enregistré par ENREGISTROMAT. Voilà pourquoi nous donnons dans notre logiguide une de-

scription beaucoup plus détaillée et systématique du choix des mots-clés.

Outre l'enregistrement d'un mot-clé au minimum, ENREGISTROMAT exige d'autres entrées obligatoires. Ce sont des informations en nombre fini et encodables que nous avons préalablement organisées dans une base de codes. Il s'agit des neuf dictionnaires utilisés commes sources pour le dépouillement du corpus, des quinze qualifications stylistiques selon le niveau de langue, des deux degrés de motivation des idiotismes, des soixante concept-clés étymologiques marquant les domaines d'origine des locutions ainsi que des cinquante structures grammaticales les plus fréquentes /cf. hors-textes 1-4/. Pour ces dernières il va falloir attendre le moment où tout le corpus sera enregistré pour pouvoir les soumettre à une analyse statistique de fréquence. ENREGISTROMAT effectue un contrôle automatique de ces informations encodées et n'accepte pas des entrées n'existant pas dans la base de codes. Dans le cas d'une faute de frappe, d'une faute d'orthographe ou si on tape un chiffre, il fait répéter l'enregistrement de l'entrée. Si une nouvelle information encodable vient à apparaître – par exemple un soixante-et-unième concept-clé étymologique –, ENREGISTROMAT dispose d'un menu capable de l'ajouter à la base de codes déjà existante; ainsi la recherche d'après cette nouvelle information sera-t-elle égale-

ment possible.

Si deux signes codés de la même classe d'informations peuvent également se rapporter à un même idiotisme — par exemple un idiotisme peut être à la fois familier et ironique —, ENREGISTROMAT les considère comme deux idiotismes différents qu'il faut donc enregistrer deux fois.

Nous avons aussi dans le processus de l'enregistrement trois types de données variables imprévisibles qui étaient impossibles à organiser dans la base de codes. Celles-ci sont:

a/ les concepts-clés condensant sémantiquement la signification des locutions /cf. 1.2.2. et 2.2.3./; vu l'importance des rapports synonymiques de ceux-ci, ENREGISTROMAT en accepte deux, dont un obligatoire, pour chaque locution. Par exemple, les concepts-clés possibles de l'idiotisme aller comme le vent sont: RAPIDITÉ et VITESSE.

b/ les périphrases ou définitions développées des idiotismes; elles peuvent ne pas figurer.

c/ les équivalents dans une seconde langue — leur présence n'est pas obligatoire non plus — dont l'ensemble constituera par la suite une nouvelle base de données nécessaire pour les analyses contrastives.

Dans ces trois cas, il est donc possible d'enregistrer des données libres. Le contrôle automatique d'ENREGISTROMAT fonctionne toujours — il ne permet

pas par exemple de mettre des chiffres dans la donnée —, mais ne peut plus signaler les fautes d'orthographe. Pour y remédier, nous aurons pour l'enregistrement de ces trois types de données, à chaque fois une question de contrôle du type "est-ce correct?". S'il y a une faute dans le texte, l'enregistrement peut être répété, si non, on continue.

Ainsi donc l'enregistrement fautif des données peut être évité dans une large mesure, ce qui est indispensable pour une recherche rapide et efficace.

HORS-TEXTE 1.

LISTE DES DICTIONNAIRES DÉPOUILLÉS

L = Lexis. Larousse de la langue française. 1983.

R = Le Petit Robert 1. 1983.

UR = A. Rey — S. Chantreau: Dictionnaire des expressions et locutions figurées. Les Usuels du Robert. 1979.

Q = Dictionnaire Quillet de la langue française. 1948.

GP = M. Lis — M. Barbier: Dictionnaire du Gai Parler. 1980.

MG = E. Rogivue: Le Musée des gallicismes. 1963.

DT = M. Thérond: Du tac au tac. 1953.

SDVV = H. Schick: Synchron-diachrone Untersuchungen zu volkstümlichen Vergleiche des Deutschen, Französischen und Spanischen. 1982.

VVF = W. Widmer: Volkstümliche Vergleiche im Französischen. 1929.

Les dictionnaires n'ont été dépouillés tous les neuf que pour les comparaisons idiomatiques. En effet, Q, MG, DT sont des dictionnaires ayant des chapitres à part sur les comparaisons, alors que SDVV et VVF constituent exclusivement des recueils de comparaisons. Pour les idiotismes qui ne sont pas des comparaisons, nous n'avons dépouillé que les trois premiers /L, R, UR/ qui, à eux trois déjà, proposent un corpus impressionnant.

HORS-TEXTE 2.

LISTE DES QUALIFICATIONS STYLISTIQUES /cf. aussi 1.2.4./

arg = argotique
euph = euphémique, par euphémisme
fam = familier
iron = ironique/ment/
litt = littéraire
mod = moderne
neutr = neutre
péj = péjoratif
plais = plaisant, par plaisanterie
poét = poétique
pop = populaire
rég = régional
vulg = vulgaire
vieilli = vieilli
vx = vieux

HORS-TEXTE 3.

LISTE DES DEGRÉS DE MOTIVATION

ID = idiomatique
PHRAZ = phraséologique

On parlera de motivation au degré phraséologique si les constituants de l'idiotisme sont sémantiquement bien présents, c'est-à-dire que l'image de l'idiotisme a son origine dans une observation quotidienne bien claire ou dans un acte synchroniquement encore réalisable. Par exemple: BLEU COMME LE CIEL; AVOIR LE BRAS LONG; DONNER UN COUP DE POING SUR LA TABLE.

Par contre, le degré idiomatique de la motivation signifiera que, synchroniquement, la signification de l'idiotisme n'est plus transparente — on ne sait plus très bien par exemple pourquoi on dit TIRER LE DIABLE PAR LA QUEUE — ou que l'acte-même exprimé dans et par l'idiotisme n'est pas réalisable /par exemple: SE METTRE MARTEL EN TÊTE; SE CASSER LA TÊTE/.

Il est bien évident qu'entre ces deux degrés il peut y avoir des transitions, des cas problématiques dont l'analyse mériterait un article à part. Dans la pratique du traitement des locutions il nous a toujours semblé qu'il était assez facile de trancher entre ces deux catégories.

HORS-TEXTE 4.

LISTE DES CONCEPTS-CLÉS ÉTYMOLOGIQUES

agriculture
alimentation
aliments
argent
armée
arts
Bible
boire
chasse
cheval
chiffres
circulation
commerce
couleurs
coutumes populaires
cuisine
culture
distraction
équitation
état
faune
fêtes
flore
géminées
guerre
habitation
histoire
inculture
industrie
jeux
justice
lettres
littérature
maison
maladies
matières
médecine
métiers
mort
mots archaïques
mythologie
nom géographique
nom propre de personne
parenté
parties du corps
pêche
phénomènes naturels
religion
sciences
sport
techniques
travail
toilette
transport
unités de temps
univers
ustensiles
vêtement
vie sexuelle
voyage

2.2. Idiomat. La recherche des idiotismes

2.2.1. Logiguide de la recherche

Nous avons essayé de réaliser notre logiciel IDIOMAT de façon que la recherche des idiotismes se fasse le plus rapidement possible. Toutes les informations importantes pour la recherche ont été codées /cf. hors-textes 1-4/. Comme nous l'avons déjà dit plus haut, aucun classement préalable — par exemple alphabétique — des données n'est nécessaire. Le logiciel ne le fera pas par la suite non plus. Ceci est possible parce qu'au cours de l'enregistrement des idiotismes, ENREGISTROMAT établit la liste des caractères de code identifiant l'idiotisme en question et les stocke dans une base de codes à part. Pendant la recherche, toute cette base de codes se trouve dans la mémoire centrale du matériel. La recherche est ainsi très rapide car le nombre des opérations entrée/sortie est relativement petit. La recherche selon les mots-clés se fait d'après des tables de caractères, ce qui assure une fois de plus une recherche extrêmement rapide et efficace.

Après l'entrée d'une information de recherche, IDIOMAT examine la base de code se trouvant dans la mémoire centrale. Indépendamment de la complexité de l'information de recherche, cet examen ne doit être effectué qu'une fois, ce qui fait que la vitesse de la recherche n'est pas fonction de la complexité de

l'information de recherche en question. La seule exception est quand la recherche se fait /aussi/ selon les sources de l'idiotisme /les dictionnaires/. C'est que pour les signes codés des sources, la machine doit effectuer une transformation en système binaire /entre autres pour économiser de la place/, ce qui nécessite un peu de temps.

Les idiotismes retrouvés apparaissent continuellement sur l'écran. A ce moment-là, l'utilisatuer pourra choisir entre plusieurs menus /informations détaillées sur l'idiotisme en question, poursuite de la recherche selon le même point de vue ou fin de la recherche/. Pendant qu'on hésite à choisir un menu, IDIOMAT attend. Pour éviter ces temps d'attente, on aura la possibilité, tout à fait au début, de choisir un menu "recherche rapide". Les numéros d'ordre des locutions recensées seront écrits sur l'écran sans que le programme s'arrête. Ce n'est qu'à la fin de la recherche /après l'examen de toute la base de code/ qu'on pourra demander de lister les numéros d'ordres sur l'écran ou sur l'imprimante. La recherche rapide se fera d'après notre mesurage en 30 secondes pour 1000 idiotismes.

L'impression des informations peut se faire à deux occasions.

a/ si l'on demande au début de la recherche les résultats directement sur imprimante, on aura une liste comme celle qui est donnée comme exemple ci-dessous:

```
PREMIER MOT-CLE      ALLER (CELA LUI VA)
DEUXIEME MOT-CLE
SOURCE
STRUCTURE
CONCEPT-CLE
MOTIVATION
NIVEAU DE LANGUE
ETYMOLOGIE

 25

 ALLER (CELA LUI VA) COMME UN GANT

 26

 ALLER (CELA LUI VA) COMME UNE BAGUE AU DOIGT

 27

 ALLER (CELA LUI VA) COMME UN TABLIER A UNE VACHE

 28

 ALLER (CELA LUI VA) COMME UN TABLIER A UNE VACHE

 29

 ALLER (CELA LUI VA) COMME UN TABLIER A UNE VACHE

 30

 ALLER (CELA LUI VA) COMME UN TABLIER A UNE VACHE
```

b/ Si l'on demande les idiotismes d'abord sur écran et seulement après la fin de la recherche sur imprimante, on aura:

```
PREMIER MOT-CLE      ALLER (CELA LUI VA)
DEUXIEME MOT-CLE
SOURCE
STRUCTURE
CONCEPT-CLE
MOTIVATION
NIVEAU DE LANGUE
ETYMOLOGIE
 25   26   27   28   29   30
```

Une description détaillée de la stratégie de la recherche est donnée dans notre logiquide.

2.2.2. Le choix et l'utilisation des informations de recherche

Comme nous l'avons déjà signalé dans la description du fonctionnement de l'enregistrement des données, IDIOMAT est capable de rechercher les idiotismes selon huit points de vue. Ceux-ci apparaissent au début de chaque recherche dans la grille suivante:

PREMIER MOT-CLÉ = ?

DEUXIÈME MOT-CLÉ = ?

SOURCES /cf. hors-texte 1./ = ?

STRUCTURE = ?

CONCEPT-CLÉ = ?

NIVEAU DE LANGUE /cf. hors-texte 2./ = ?

MOTIVATION /cf. hors-texte 3./ = ?

ÉTYMOLOGIE /cf. hors-texte 1./ = ?

REMARQUE: Actuellement la recherche selon la structure grammaticale n'est pas encore possible. Une fois que toute la base de données sera enregistrée, nous entreprendrons une analyse statistique de celle-ci pour déterminer les 50 structures les plus fréquentes. Elles seront codées puis enregistrées dans la base de codes. A partir de ce moment, la recherche des idiotismes pourra se faire également selon la structure.

En fonction du point de vue de la recherche, il faut mettre à la place du point d'interrogation l'information voulue. Tout comme dans ENREGISTROMAT, il faut faire attention à taper les informations très exactement /en particulier les données codées et les mots-clés/. Si l'on tape par exemple, au lieu de ARGOT, "ergot", IDIOMAT nous signale que, dans la base de codes, il n'y a pas de donnée correspondant à l'information entrée et nous demande de répéter l'entrée de l'information ou de contrôler la base de codes. Pour les informations libres /les concepts-clés par exemple/, IDIOMAT ne peut pas contrôler les fautes de frappe ou de langue de l'utilisatuer. Ainsi si l'on met "impatiance" à la place de IMPATIENCE, il

ne trouvera pas les idiotismes correspondant à ce concept-clé.

IDIOMAT exige au moins une information de recherche, mais la combinaison de plusieurs informations est tout à fait possible aussi. Ainsi l'utilisateur pourra-t-il entreprendre des analyses fort intéressantes sur le rapport entre les concepts-clés et les structures grammaticales ou sur celui entre les degrés de motivation et les structures, etc. Il est également possible de faire un examen statistique exhaustif et comparatif du fonds phraséologique des dictionnaires dépouillés. Il sera très instructif de voir quelles sont les différences entre les dictionnaires, par exemple du point de vue de la qualification stylistique des idiotismes. La comparaison des relations entre concepts-clés sémantiques et concepts-clés étymologiques ne serait pas sans intérêt non plus. Ces considérations, pour le moment à l'état d'hypothèses de travail, mériteront sûrement une analyse plus approfondie. Comme nous l'avons déjà signalé sous 1.2., c'est certainement la recherche selon les concepts-clés qui peut présenter le plus d'avantages pratiques directs. Que signifie ceci exactement?

2.2.3. Le concept-clé en position-clé

Le classement traditionnellement alphabétique des dictionnaires de locutions existants les réduit

pratiquement à n'être utilisables que pour le compréhension d'énoncés lus ou entendus. Dans nombre de cas /traduction, composition, correspondance, discussion, etc./ nous avons à juste titre l'impression que nous n'arrivons pas à exprimer nos pensées avec la même efficacité que dans notre langue maternelle, où — pour les rendre de façon expressive et imagée — nous disposons des expressions idiomatiques les plus diverses. A première vue, c'est une situation inévitable, à laquelle il ne faut cependant pas se résigner. Seulement, pendant le long apprentissage d'une langue étrangère, tel le français dans notre cas, les sources — surtout les dictionnaires d'idiotismes — accessibles pour l'instant ne nous sont que d'un secours très limité. Et ceci surtout dans une utilisation de ces dictionnaires à fin d'expression.

> On distinguera le niveau de l'<u>expression</u> /production/, c'est-à-dire la génération de locutions et de phrases en langue étrangère et le niveau de la <u>compréhension</u> /réception/, c'est-à-dire le décodage d'énoncés en langue étrangère.

Alors que le classement alphabétique est très efficace pour retrouver les simples équivalences lexicales /à l'aide de l'alphabet connu et employé universellement, on aura vite l'équivalent en français de n'importe quel mot hongrois/, dans le cas des idiotismes, ce classement sémasiologique ne

peut en aucun cas être suffisant et efficace. Les dictionnaires d'idiotismes sont en général composés de façon que les idiotismes puissent être retrouvées d'après leurs mots-clés dont on pense qu'ils portent l'information, et qui apparaissent dans l'ordre alphabétique. Les idiotismes se composant en général de plusieurs éléments, il est difficile de décider dans la majorité des cas quel est le mot-clé sous lequel le dictionnaire en question permet de retrouver un idiotisme voulu. Ce mot-clé peut évidemment, à l'occasion, ne pas être le vrai. Ainsi par exemple un idiotisme comme il n'a pas invité le fil à couper le beurre, peut être retrouvé dans trois dictionnaires sous trois entrées différentes. Et encore faut-il connaître ces mots-clés! Le choix des mots-clés est donc ainsi constamment exposé à l'arbitraire des lexicographes et ne sera jamais aussi stable et universel que l'emploi de l'alphabet. De plus, le point de départ probable d'un Hongrois pour trouver un idiotisme dans une langue étrangère, c'est-à-dire l'un des éléments de l'idiotisme hongrois /feltalál ou spanyolviasz/, ne coïncide absolument pas ou seulement en partie avec l'élément apparaisant au même niveau de l'idiotisme dans la langue d'arrivée. /D'ailleurs dans le cas d'équivalences totales ou partielles entre les idiotismes français et hongrois, le piège des faux amis idiomatiques guette encore souvent l'usager de la langue./

Or,par la force des choses, le simple usager de la langue ne connaissant, le plus souvent, pas exactement ou pas du tout les mots-clés ou les autres éléments constitutifs des idiotismes dans la langue étrangère, ne trouvera un idiotisme correspondant parfaitement à son message qu'après avoir feuilleté tout le dictionnaire — ce qui est absurde — ou par une chance extraordinaire en ouvrant le dictionnaire juste à la page où se trouve l'expression cherchée. Pour identifier, comprendre, apprendre, traduire une locution lue ou entendue, ces dictionnaires alphabétiques sont bien sûr d'une valeur indiscutable.

Un autre problème est que les index — si index il y a — ne sont pas en général exhaustifs et, qui plus est, sont souvent très défaillants et ne facilitent aucunement le maniement des dictionnaires. Ces imperfections rapidement évoquées des dictionnaires donnent lieu à formuler une hypothèse de travail qui peut se résumer ainsi. Les constituants /les mots-clés/ de l'idiotisme ainsi que l'idiotisme lui-même peuvent être rangés dans la catégorie du p a r t i c u l i e r à laquelle correspond — indépendamment des langues — la catégorie du g é n é ra l qui recouvre un concept plus abstrait, qu'on pourrait appeler aussi la paraphrase ou le concept-clé de l'idiotisme. Dans le cas de notre exemple /il n'a pas inventé le fil à couper la beurre/, ce général peut se traduire par des mots particuliers

comme 'bêtise', 'sottise'. L'emploi d'un dictionnaire qui part dans son classement du particulier pour retrouver le particulier, ne peut être efficace. Logiquement et méthodologiquement il serait plus fondé de partir du général, de ce qui est ou peut être connu de tout le monde, c'est-à-dire la paraphrase de l'idiotisme — pour arriver jusqu'au particulier.

Quels seraient donc les avantages d'une recherche partant des concepts-clés?

a/ Tout le monde peut plus facilement trouver, même dans une langue étrangère, — au terme d'une activité mentale condensatrice — le concept-clé d'éléments linguistiques /en l'occurrence les idiotismes/ qui lui sont encore inconnus. En donnant cette information à IDIOMAT, notre utilisateur aura tout de suite une liste plus ou moins longue de locutions particulières, parmi lesquelles il pourra choisir celle qui lui convient le mieux pour une situation de communication donnée.

b/ Pour l'apprentissage de la langue, le principe de champ onomasiologique /champ des concepts-clés/ est de grande importance, car non seulement les idiotismes particuliers sont plus rapidement retrouvés, mais encore ils passent plus facilement dans le vocabulaire actif s'ils sont associés à d'autres, analogues. L'établissement de résaux d'af-

finités à l'intérieur de la langue favorise la reproduction et offre une aide précieuse à la mémoire.

c/ L'ensemble des concepts-clés et de leurs synonymes devrait se cristalliser et dessiner le système des champs onomasiologiques du français — ou de toute autre langue —, encore insuffisamment décrit.

3. Conclusions

3.1. En résumé, nous pouvons donc dire que notre logiciel de traitement des idiotismes — que nous avons baptisé IDIOMAT — permettra de traiter, d'analyser sur Commodore 64 une base de données comprenant jusqu'à 2000 idiotismes de n'importe quelle langue à caractères latins.

Il est vrai que le fonds phraséologique d'une langue dépasse largement les 2000 unités — on parle en général de 8000 a 10 000 unités —, mais, sur ces dix mille, le nombre de celles qu'on utilisera fréquemment et spontanément ne dépassera certainement pas 2000 même dans notre langue maternelle et à plus forte raison dans une langue seconde. Ainsi le nombre limite des idiotismes imposé par la capacité de l'ordinateur individuel ne constituera pas vraiment une restriction et un inconvénient sensibles pour l'utilisateur. Reste évidemment le problème de l'arbitraire dans le choix des idiotismes dont nous avons parlé sous 1.2.1.

3.2. A notre avis, IDIOMAT aura une utilité certaine en premier lieu dans les recherches linguistiques, en particulier phraséologiques, en mobilisant rapidement une base de données importante selon des points de vue variés et combinés et en permettant de faire également des analyses statistiques.

Ceci n'empêchera pas pour autant son utilisation dans l'enseignement d'une langue étrangère. Cependant, dans ce cas, il devra être plutôt un outil complémentaire à la disposition du professeur qui pourra s'en servir pour préparer ses cours ou éventuellement pendant ses cours. IDIOMAT sera moins apte à être utilisé directement dans l'apprentissage individuel de la langue, A cette fin, l'apprenant devra utiliser notre didacticiel EXERCOMAT-IDIOTISMES, en cours de préparation, qui s'appuyera sur le corpus du dictionnaire décrit sous 1.2.

3.3. Nous travaillons également sur la possibilité d'un traitement parallele des idiotismes de deux langues dont l'utilité est évidente aussi bien pour les recherches de linguistique contrastive que pour l'apprentissage des langues.

3.4. Il serait aussi imaginable de demander à l'ordinateur de lister tous les idiotismes traités par exemple d'après leur premier mot-clé, d'après leur deuxième mot-clé, d'après les niveaux de langue ou encore dans un regroupement selon les concept-clés

etc. Ainsi l'utilisateur du progiciel pourrait avoir en supplément une série de petits cahiers-manuels à sa disposition qu'il consulterait à son gré.

Notes

1 Pour les détails concernant les caractéristiques des locutions et leur utilité dans l'apprentissage d'une langue étrangère voir entre autres: Bally, Ch.: Précis de stylistique. Genève, Eggimann, 1905, chap. IV. — Bárdosi, Vilmos: Les locutions françaises en 150 exercices. Budapest, Tankönyvkiadó, 1983, Egységes jegyzet et De fil en aiguille. Les locutions francaises: recueil thématique et livre d'exercices. Budapest, Tankönyvkiadó, 1986. — Burger, H. — Buhofer, A. — Sialm, A.: Handbuch der Phraseologie. Berlin - New York, Walter de Gruyter, 1982. — Casares, I.: Introducción a la lexicografía moderna. Madrid, 1950. — Coseriu, E.: Structure lexicale et enseignement du vocabulaire. Actes du 1er Colloque International de Linguistique Appliquée. Nancy, 1964, pp. 175-217. — Galisson, R.: Des mots pour communiquer. Paris, Clé International, 1983, chap. III. "Pour une méthodologie de l'accès aux locutions figuratives en français, langue maternelle et étrangère". — Hessky, Regina: Deutsch-ungari-

sche phraseologische Sammlung. Budapest, Tankönyvkiadó, 1982. — Juhász, József: A frazeológia mint nyelvészeti diszciplina, in: Tanulmányok a mai magyar nyelv szókészlettana és jelentéstana köréből. Budapest, Tankönyvkiadó, 1980, pp. 79-97. — O. Nagy Gábor: Mi a szólás? in: Magyar Nyelv, 50 /1954/, n° 3-4, pp. 110-126 et 396-408. — Nazarjan, A.G.: Frazeologija sovremennovo frantsuskovo jazika. Moscou, Vichaia Skola, 1976. — Rey, A.: Le lexique: images et modèles. Paris, A. Colin, 1977, chap. 8.

2 Burger, H.: Idiomatik des Deutschen. Tübingen, 1973.

3 Bárdosi, Vilmos — Longueau, Jean-Yves de: Petit dictionnaire thématique des locutions françaises et Bárdosi, Vilmos: De fil en aiguille. Les locutions françaises: recueil thématique et livre d'exercices à paraître chez Tankönyvkiadó à Budapest.

4 Bárdosi, Vilmos: Egy új típusú szólásszótár szükségességéről. Filológiai Közlöny, 1982/2-3, 344-355 et Les limites de l'utilisation des dictionnaires de locutions. Annales Universitatis Budapestinensis, Sectio Linguistica, 1985, 17-26.

Bárdosi Vilmos — Csink László

Le traitement des locutions idiomatiques par microordinateur c. cikkének rezüméje

A cikk az ELTE Francia Tanszékén folyó francia-magyar frazeológiai kutatásokról számol be. A szerzők összefoglalják az utóbbi 3-4 év elméleti és gyakorlati munkáját, az eddig született eredményeket, majd leírják a szólásanyag egy évvel ezelőtt elkezdődött személyi számítógépes feldolgozásának menetét. Tárgyalják a mintegy 1000 szólást tartalmazó adatbázis felépítését, az adatfelvitel módszereit, majd bemutatják a Magyarországon legelterjedtebb számítógépre, a Commodore 64-re készült IDIOMAT nevű francia szólásokat elemző-visszakereső programjuk működési elvét, mely azonban az adatbázis előzetes megváltoztatásával tetszőleges latin ábécés nyelvre alkalmazható. Végül vázolják a program felhasználási lehetőségeit és fejlesztési elképzeléseiket.