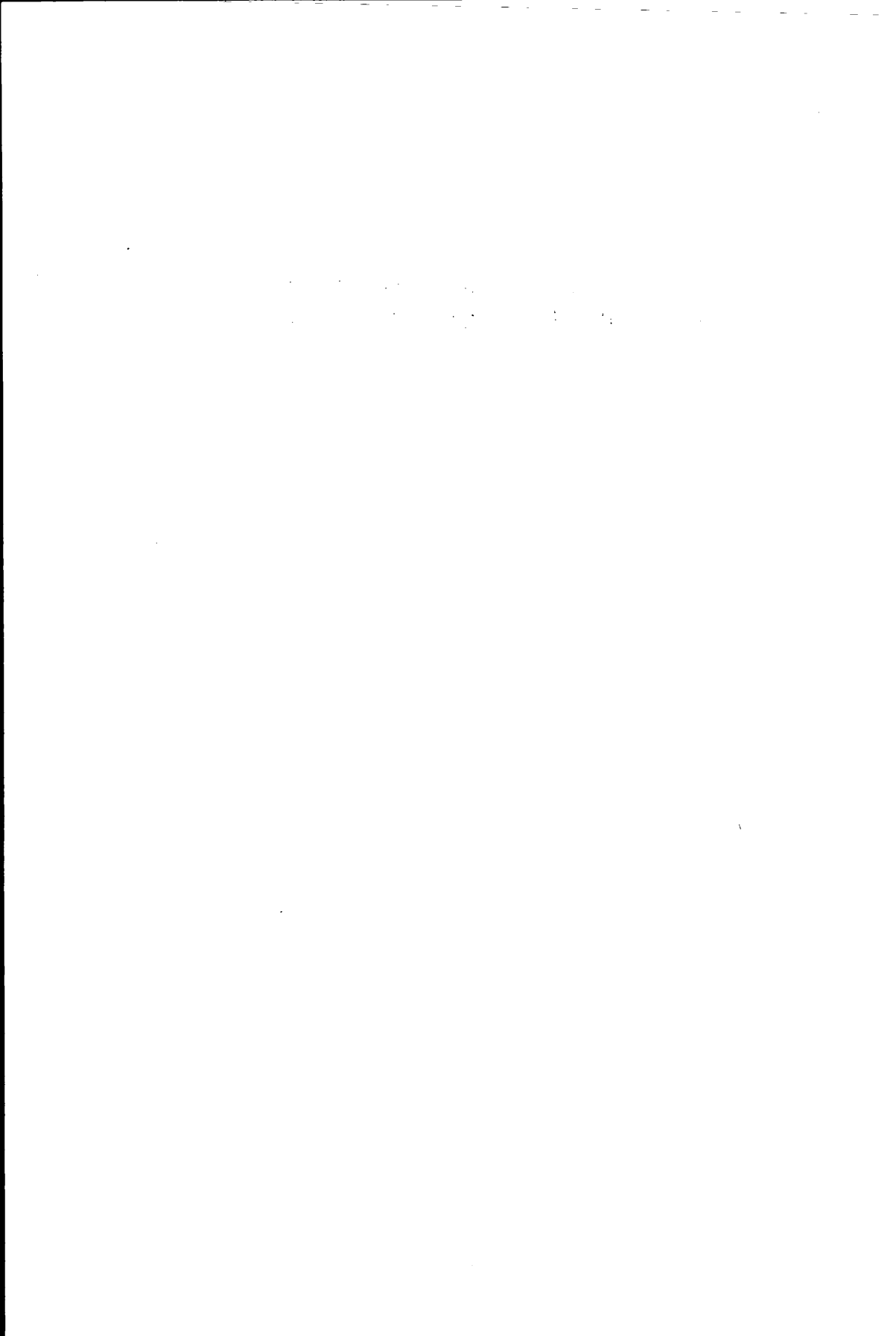


**Vers un nouveau dictionnaire français—  
hongrois / hongrois—français**



Tamás SZENDE  
Chantal PHILIPPE  
Membres de l'équipe française

## **Journées lexicographiques de décembre 1991**

Le Centre Interuniversitaire d'Etudes Hongroises (Université de la Sorbonne Nouvelle — Paris III) et le Centre Interuniversitaire d'Etudes Françaises (Université Loránd Eötvös — Budapest), ont pris l'initiative de lancer la préparation d'un nouveau dictionnaire hongrois—français / français—hongrois, parce que le seul outil lexicographique en usage actuellement est très dépassé dans sa conception.

Sous la direction scientifique du professeur Jean PERROT, directeur d'études à l'Ecole Pratique des Hautes Etudes, directeur du CIEH, deux équipes lexicographiques ont été constituées, l'une à Paris, l'autre à Budapest.

Il s'agit de doter l'ensemble des étudiants, enseignants et traducteurs concernés par le domaine français—hongrois, et dans une mesure plus large, toute personne engagée dans des relations entre le monde francophone et le monde hungarophone, d'un ouvrage de référence constituant une source privilégiée d'informations destinée aux deux communautés linguistiques.

Il est fondamental que le nouveau dictionnaire bilingue soit conçu en fonction de ses deux publics, en l'occurrence hungarophone et francophone, et tenant compte de leurs besoins respectifs: version et thème pour l'utilisateur de langue 1, thème et version pour l'utilisateur de langue 2. Centré sur le vocabulaire général, l'ouvrage enregistrera pour l'ensemble des deux parties environ 200 000 mots et locutions. Néanmoins il devra contenir les termes les plus usuels appartenant aux divers domaines scientifiques et techniques, notamment aux disciplines de pointe qui envahissent le parler quotidien. De même, il offrira un recensement complet des principaux éléments du vocabulaire des secteurs politique, économique et commercial.

Trop souvent négligée dans le passé, la langue parlée devra être mise en valeur, grâce notamment à des exemples empruntés aux médias et à des oeuvres littéraires reproduisant les langages de tous les jours. On s'efforcera de donner un maximum de renseignements sur les mots enregistrés et leur fonctionnement en faisant état d'exemples pertinents où le mot traité se trouve «en situation».

L'élaboration de ce dictionnaire a nécessité la mise en place d'une structure technique et informatique permettant que le manuscrit en gestation et plus tard l'ouvrage soient l'objet d'une mise à jour permanente. Aujourd'hui, la constitution d'un tel chantier est possible grâce à la lexicographie assistée par ordinateur.

Nous souhaitons que cet ouvrage soit publié dans le cadre d'une collaboration franco—hongroise associant des maisons d'édition que leur expérience qualifie pour cette réalisation.

C'est autour de ce projet que le CIEH a organisé des journées lexicographiques les 11, 12 et 13 décembre 1991, avec la participation de spécialistes français, finnois et hongrois, dont les interventions ont abordé différents aspects théoriques et pratiques de la lexicographie. Un des moments de ce colloque a été la démonstration de la base de données textuelles FRANTEXT sur laquelle est fondé le Trésor de la Langue Française.

Nous aimerions remercier ici tous les intervenants pour le concours qu'ils ont apporté au succès de cette rencontre. Nous tenons tout particulièrement à exprimer notre gratitude au professeur Jean PERROT qui a accepté de présider le colloque et en a animé les débats.

Nous ne sommes pas en mesure de réunir dans le cadre des *Cahiers* la totalité des actes du colloque, aussi nous limitons-nous à publier les contributions des principaux collaborateurs du futur dictionnaire et celle du professeur Lerat.

Jean PERROT

Ecole Pratique des Hautes Etudes — IV<sup>e</sup> section

### Soixante ans après: encore «enfiler des mots»

Le programme ambitieux qui a pour objet l'élaboration d'un nouveau dictionnaire bilingue associant français et hongrois se met en place à une date qui incite à le situer dans un développement historique auquel il est relié, celui des études hongroises en France.

1991: c'est le soixantième anniversaire de la création de la chaire des langues finno-ougriennes à l'Ecole des Langues Orientales, et à cette occasion il faut évoquer la mémoire d'Aurélien Sauvageot, qui a occupé cette chaire depuis sa création en 1931 et (avec une interruption de 1941 à 1943 en raison d'une mesure de suspension prise par le gouvernement de Vichy) jusqu'à sa retraite en 1967 et qui a aussi été l'auteur du premier grand dictionnaire français-hongrois et hongrois-français.<sup>1</sup> L'année 1992 sera marquée par un autre soixantième anniversaire important pour les hongrois: celui de la publication en Hongrie du premier volume, français-hongrois, de ce dictionnaire, paru à Budapest, aux éditions Dante, en 1932.

Soixante ans donc d'existence institutionnelle des études hongroises en France; c'est évidemment peu, et ce n'est pas la première fois que je souligne l'anomalie que constitue pour l'étude des langues finno-ougriennes ce démarrage tardif dans notre pays. D'autant plus que pendant les trente années qui ont suivi cette création, ces études n'ont eu qu'une présence très discrète: elles étaient certes illustrées brillamment par les travaux personnels de Sauvageot, mais elles ne sortaient pas de cette chère salle 5 de la rue de Lille où le maître les servait de tout son talent et de toute sa science. Ce n'est qu'à partir des années soixante que des progrès décisifs ont été faits, qui ont petit à petit implanté à l'Université et développé notamment dans la direction de la recherche ce domaine longtemps ignoré.

Or ce fut un événement tout à fait remarquable, extraordinaire, que cette apparition d'un gros dictionnaire français-hongrois dès 1932, l'année même où Sauvageot fut titularisé comme professeur dans cette chaire tout récem-

---

<sup>1</sup> Aurélien SAUVAGEOT avec la collaboration de Joseph BALASSA et Marcel BENEDEK, *Dictionnaire général français-hongrois et hongrois-français*, Budapest, Ed. Dante. Français-hongrois 1932; hongrois-français 1937.

ment créée, un Sauvageot qui avait soutenu ses deux thèses en juin 1929 et qui ne s'était attelé à l'énorme tâche du dictionnaire, le premier grand dictionnaire français—hongrois, que dans cette année 1929. Sauvageot nous a raconté l'histoire de son dictionnaire dans les *Souvenirs de ma vie hongroise*<sup>2</sup>:

«Un soir que j'avais été invité chez Marcel Benedek, celui-ci sur un ton compassé et uni, me demanda soudain: 'Mon cher ami, ne pensez-vous pas qu'il serait temps pour vous de rédiger le grand dictionnaire français—hongrois dont nous avons un si pressant besoin?' Je restai interloqué.» (164)

C'était un fait: il fallait «déplorer l'absence de tout instrument valable pour apprendre directement le hongrois à partir du français et réciproquement le français à partir du hongrois» (164) et Sauvageot était choqué de voir ses élèves hongrois utiliser comme intermédiaire un dictionnaire français-allemand.

Mais, dit-il, «je pensais que c'était aux spécialistes hongrois de produire des dictionnaires bilingues du français comme ils l'avaient fait pour l'allemand et même pour l'anglais. Je ne saisissais pas bien pourquoi, comme on disait vulgairement, je m'y collerais. Plus d'une fois, j'en avais touché un mot à Eckhardt qui m'avait simplement répondu qu'il envisageait de produire un jour ce dictionnaire français-hongrois que j'appelais de mes vœux, mais à son ton je comprenais que ce n'était pas pour lui un projet très pressant. Cela viendrait à son heure». (165)

Sauvageot n'était pas lui-même spécialement tenté: «Personnellement, je n'avais pas non plus mis à mon programme un ouvrage de ce genre. La lexicographie ne m'intéressait pas spécialement et j'étais même très méfiant à son sujet.» (165) Et de raconter qu'un savant suédois l'avait plutôt détourné de l'idée de faire de la lexicographie.

Donc il dit d'abord non à Benedek. Puis il se repentit et subit beaucoup de pressions, notamment de la part d'une femme qui lui avait ouvert beaucoup de portes en Hongrie.

«Elle me remontra que ce dictionnaire était indispensable pour la diffusion du français dans la Hongrie mutilée par le Traité de Trianon. Jusque-là, la classe sociale qui avait appris le français envoyait ses enfants soit en France, soit en Suisse, soit en Belgique pour les plonger dans un milieu d'expression française, et souvent ces mêmes enfants y avaient été préparés par des précepteurs ou, plus souvent encore, par des domestiques venues de France auxquelles ils étaient confiés à un âge si tendre qu'ils apprenaient le français en même temps que le hongrois, quand ce n'était pas avant. Les conditions de vie ayant changé, le public désireux de se familiariser d'amalieriser avec le français était contraint de recourir à d'autres méthodes. Le livre devenait indispensable». (166)

---

<sup>2</sup> Aurélien SAUVAGEOT, *Souvenirs de ma vie hongroise*, Budapest, Corvina, 1988.

Et Sauvageot pensait aussi à la langue hongroise:

«Je songeais aussi qu'il y avait le hongrois. C'était une belle langue, fruit d'un effort séculaire, moyen suprême d'expression d'une nation que nous avions ignorée et qui demandait d'être connue et reconnue. Pratiquement, cette langue n'était accessible que par le détour de l'allemand. Or, durant des siècles, les Hongrois avaient dépensé bien des efforts pour se soustraire à la domination de l'allemand. C'est contre lui qu'ils avaient rénové leur langue à plusieurs reprises.» (166)

Le terrain était d'ailleurs soigneusement préparé par Benedek, qui s'était entendu avec un éditeur pour confier à Sauvageot la rédaction d'un dictionnaire bilingue en deux gros volumes.

Alors Sauvageot s'est lancé: après tout, ses thèses étaient imprimées, il allait les soutenir, il était disponible. «Le tout était de se faire à l'idée de cette tâche, d'un genre nouveau pour moi. Il fallait se chercher une méthode pour la mener à biens dans les moindres délais. Car si l'éditeur semblait pressé, je ne l'étais pas moins que lui. Seulement, n'était — il pas follement téméraire de ma part de me lancer dans cette aventure?» (167)

L'entreprise était en effet difficile: elle était sans précédent par son ampleur et il n'était pas facile de l'aborder.<sup>3</sup> Les dictionnaires français unilingues étaient tous insuffisants, et pour le hongrois il n'y avait tout simplement pas de dictionnaire.

«En effet, il n'existait aucun dictionnaire hongrois-hongrois à l'usage des Hongrois eux-mêmes. Cette constatation m'avait stupéfié au tout début de mon étude de la langue». (168). Situation d'autant plus étonnante qu'il y avait une solide tradition philologique en Hongrie et que les philologues hongrois avaient entrepris la publication d'un dictionnaire étymologique!

Sauvageot devait alors s'assurer l'aide de Hongrois compétents: il eut celle de Benedek, écrivain, critique tririque, traducteur, mais non linguiste, complétée par celle d'un bon linguiste, Balassa, lui-même auteur d'un dictionnaire allemand-hongrois et hongrois-allemand. Eckhardt l'avertit des dangers:

«Vous n'avez pas l'air d'imaginer la difficulté d'une telle entreprise. Croyez bien que si personne ne l'a encore tentée, c'est que le risque est grand de s'y perdre ou tout au moins d'y perdre je ne sais combien d'années de travail». (171)

<sup>3</sup> Il avait existé plusieurs dictionnaires hongrois-français et français-hongrois dont certains étaient très estimables, notamment celui de KISS et KARÁDY qu'avait publié l'éditeur Heckenast à Pest et dont il faut en particulier citer la troisième édition «revue et considérablement augmentée», publiée en 1865 par BABOS Kálmán sous le titre de *Nouveau dictionnaire hongrois-français et français-hongrois d'après les ouvrages publiés [sic] jusqu'à ce jour*. Mais il était évidemment nécessaire, à l'époque où Sauvageot a décidé de se mettre à la tâche, de remplacer les ouvrages anciens par un grand dictionnaire de la langue contemporaine répondant aux exigences nouvelles.

Eh bien, trois ans plus tard, en 1932, les éditions Dante sortaient le dictionnaire français—hongrois: «178 pages grand format, avec des pages de deux colonnes où plus de 4000 mots français nouveaux étaient consignés pour la première fois (174): un «tour de force» aux yeux de Paul Boyer et d'Antoine Meillet.

Mais cela se passait dans les années 30: le volume II, plus gros encore, sorti en 1937. Le Sauvageot, plus d'un demi-siècle plus tard, malgré ses qualités incontestables, a été abandonné. Il a été remplacé après la deuxième guerre mondiale, dans les années cinquante, par un dictionnaire, en deux volumes encore plus nourris, mais au total moins bons, le dictionnaire d'Alexandre Eckhardt, fruit d'un effort remarquable pour intégrer toutes sortes d'informations sur les différents niveaux de langue, sur les termes techniques, mais encombré d'expressions inutiles, voire sans réalité linguistique, et souffrant cruellement d'une absence de contrôle par des francophones.

D'où la situation déplorable de manque où nous nous trouvons aujourd'hui. D'où la volonté de suivre le grand exemple de courage et d'efficacité que nous a laissé Sauvageot. La tâche reste aujourd'hui, assurément, très rude, elle est en un sens alourdie par des exigences accrues qui imposent la plus grande rigueur dans les procédures. L'existence de dictionnaires antérieurs ne constitue pas en soi une ressource, puisqu'il s'agit de faire du neuf; du moins existe-t-il aujourd'hui des dictionnaires unilingues, trésors, corpus pour l'une et l'autre Langue et toute une réflexion lexicographique qui permet de maîtriser les problèmes et d'orienter les choix méthodologiques.

Il y a l'obstacle du coût élevé de l'entreprise. J'ai pourtant le sentiment, fondé à la fois sur les soutiens dont nous avons déjà bénéficié et sur les manifestations multiples d'intérêt que nous constatons, que les institutions concernées, ici et là, sauront faire les efforts qui leur incombent.

Ce que dit Sauvageot dans ses *Souvenirs* après avoir raconté l'histoire de son dictionnaire a certes de quoi inquiéter ses successeurs.

Il avait bénéficié de soutiens très appréciables: du côté hongrois, du comte Klebelsberg, ministre de l'Instruction Publique; du côté français, de Louis-Edmond de Vienne, ministre plénipotentiaire de France, qui avait obtenu pour lui une subvention du gouvernement français. Mais l'épilogue est amer:

«Cette publication ne favorisa pas mes affaires auprès des linguistes français. Le grand et inoubliable arabiste qu'était William Marçais me reprocha très amicalement d'avoir perdu mon temps 'à faire du vocabulaire' au lieu de consacrer tous mes soins à quelque question importante de linguistique. Selon lui, on n'avait pas besoin d'être linguiste pour 'enfiler des mots'. Un autre confrère de la Société de Linguistique de Paris me remontra, moins amicalement cette fois, que j'avais fait fausse route. Il aurait mieux valu publier n'importe quoi d'autre qu'un ouvrage sans intérêt car, enfin, quelques

douzaines seulement de Français s'oublieraient jusqu'à faire du hongrois au cours des années qui viendraient. Seuls Paul Boyer et Antoine Meillet me félicitèrent de ce qu'ils considéraient comme un tour de force».

«( . . . ) L'épilogue de l'affaire, du côté français, fut que Meillet et son ami, mon maître Joseph Vendryès, furent battus lors du vote émis par la 4ème section de l'Ecole Pratique des Hautes Etudes lorsqu'ils voulurent, à l'occasion d'une vacance, obtenir qu'une direction d'études de langues finno-ougriennes soit créée et m'y soit attribuée. Louis-Edmond de Vienne, quelques années plus tard, n'allait pas être plus heureux lorsqu'il crut devoir demander ma nomination dans l'Ordre de la Légion d'Honneur en récompense de l'oeuvre accomplie. ( . . . ) La subvention obtenue par Louis-Edmond de Vienne fut la seule manifestation d'intérêt venue du côté français. Heureusement, le public hongrois eut une autre réaction. Une fois de plus, la Hongrie montrait son attachement à notre langue et à notre civilisation en face de l'indifférence française. Nous en restions toujours là, séparés par une énorme distance, comme si nous vivions dans deux mondes totalement distincts». (174-175)

On a, aujourd'hui, quelque peu progressé, et les deux mondes tendent de plus en plus à s'interpénétrer. L'équipe — les deux équipes — qui s'attaque à la rude tâche de faire un nouveau dictionnaire bénéficie d'une volonté mutuelle de resserrer les liens, et le désir de soutenir l'entreprise apparaît sincère dans les deux pays, où l'existence, dorénavant, de deux centres homologues qui sont des lieux institutionnels de rencontre et de coopération organisée, crée des conditions de travail particulièrement favorables.

Ce sera notre fierté d'apporter une contribution à la meilleure compréhension de ces deux mondes, que réunira dans le travail quotidien de ses équipes et dans la substance de ses articles le nouveau dictionnaire.

N'ayons pas peur, ensemble, d'«enfiler des mots».



Pierre LERAT  
Université de Paris—Nord

## Unités terminologiques et dictionnaire bilingue

Le dictionnaire de langue n'est jamais trop explicite. A plus forte raison, le bilingue ne fait que son travail s'il indique systématiquement la syntagmatique des mots, s'il renvoie de dénomination à dénomination logiquement associable et s'il décrit non pas tant les choses elles-mêmes que les rapports que l'on entretient avec elles dans une culture donnée. Ce traitement sémiotique classique (grammaire, sémantique, pragmatique) induit des conséquences que je voudrais montrer en prenant des exemples dans une base de données terminologiques en cours d'élaboration, limitée au droit mais dans un cadre multilingue (Diké). Mon propos se limitera au passage en revue des rubriques utilisées dans ce projet, avec l'espoir que ce qui paraît bon pour la lexicographie spécialisée le soit aussi pour la lexicographie non spécialisée.

### 1 — Entrées

En terminologie, la doctrine onomasiologique conduit à accorder une entrée à chaque notion. A ce compte, on distinguera par exemple *droit 1* (law), *droit 2* (right) et *droit 3* (duty). Dans un dictionnaire général, il serait bon de traiter ces trois valeurs comme des sous-entrées distinctes.

Il en va de même pour les dénominations longues, qui correspondent elles aussi à des concepts particuliers et qui gagnent à être traitées comme des unités de compte lexicographiques (par exemple, *abus de biens sociaux* ou *taxe sur la valeur ajoutée*).

### 2 — Domaines

L'attachement aux notions véhiculées par les termes, et non pas le souci exclusif d'une valeur moyenne des signifiés des mots, conduit à être exigeant en ce qui concerne le rattachement de ces notions à telle branche de la connaissance. Ainsi, *domaine* a quelque chose à voir avec *domestique* comme mot français et avec *domus* comme mot néo-latin, mais ce lien est purement génétique. Bien plus, il ne suffit pas de caractériser comme juridique l'emploi de *domaine* dans l'expression *domaine public*, car au vu des définitions juridiques il faut distinguer un terme *domaine public* en droit administratif et un autre en droit de la propriété intellectuelle et artistique; dans le premier cas, en effet, il s'agit de la partie de patrimoine des collectivités publiques qui est

inaliénable, dans le second ce qui est dénommé ainsi est le régime qui autorise l'exploitation libre et gratuite d'une invention ou d'une oeuvre à l'expiration d'un délai déterminé.

### 3 — Variantes

Si l'on est d'accord avec l'idée que «la valeur d'un dictionnaire comme outil de travail dépend beaucoup de la richesse de ses renvois» (Groffier et Reed: 63), il ne faut pas hésiter à multiplier les repérages de type «peu différent de». C'est le cas des abréviations: ainsi *action* peut désigner anaphoriquement aussi bien une action en justice qu'une action de société. De même, les quasi-synonymes permettent de nuancer ou de varier l'expression, si l'on a en vue non seulement un dictionnaire de décodage, mais aussi une aide à la rédaction; par exemple s'il est vrai que *droit financier* et *finances publiques* sont loin d'être synonymes, l'usage de l'un ou de l'autre n'en est pas moins indifférent après «en matière de», et cette substituabilité limitée fait partie de la connaissance du français, à sa façon.

### 4 — Grammaire

Les formes et les constructions usuelles gagnent à être circonscrites avec autant de précision que possible. A mon avis, on fait plus pour la défense d'une langue en la décrivant bien qu'en essayant d'endiguer l'emprunt ou en proposant des innovations par voie administrative. Ainsi, il ne suffit pas de dire que *déchéance* est un dérivé de *déchoir*: en droit, *déchéance* correspond seulement à *être déchu de*. La filiation historique est intéressante, le mode d'emploi ici et maintenant est indispensable.

### 5 — Phraséologie

En cette matière il y a lieu de distinguer entre des collocations plus ou moins désuètes, compilées et conservées pieusement même si leur pertinence fonctionnelle est douteuse, d'une part, et d'autre part la phraséologie vivante. Un critère terminologique est l'appartenance à une tradition professionnelle durable; ainsi, là où l'on parle de *décentralisation* l'on dit volontiers qu'elle *fait l'objet de mesures* du même nom, et le mode d'existence des *dépens* tient tout entier dans leur combinatoire lexicale: c'est quelque chose à quoi l'on *est condamné*.

### 6 — Définition

Sur le style de la définition, je n'ai rien à dire, si ce n'est que c'est un art. Sur le contenu, qu'il me soit permis de résumer des propositions que j'ai faites pour actualiser les notions traditionnelles de «genre prochain» et de «différence spécifique» (Lerat 1990).

Si l'on admet que la définition d'un objet technique n'est pertinente qu'à condition d'être assez précise pour ne pas altérer le contenu de connaissances correspondant, les propriétés logiques nécessaires et suffisantes sont en petit nombre: la relation générique immédiate, la relation «habere» au sens de F.

Kiefer (1974:35), c'est à dire le lien partitif entre l'englobant et l'englobé (quelque chose comme «relève de»), le verbe, adjectif ou nom prédicatif approprié et les objets typiques connexes, le tout d'après le témoignage des spécialistes des connaissances concernées. Ainsi, le domaine public au sens des droits d'auteur n'est pas une sorte de domaine, mais de régime juridique; il ne relève pas de l'immobilier, mais de la propriété intellectuelle; il s'applique à la création intellectuelle, non à la gestion de patrimoines; il vise des oeuvres de l'esprit, non des surfaces habitables et cultivables.

#### 7 — Sources

L'autorité d'un dictionnaire dépend beaucoup de ses sources. Ainsi, la réputation du Trésor de la langue française repose en bonne part sur l'importance du corpus littéraire dépouillé en amont, celle du Grand Robert sur la qualité de l'information scientifique et technique validée par des grands noms de la science. Cette validation ne s'applique pas seulement aux définitions, mais aussi à la phraséologie et aux remarques.

L'indication de sources a aussi pour fonction d'orienter la recherche d'informations complémentaires, ce qui est particulièrement utile quand l'espace typographique est très limité. Dans ces conditions, mieux valent des sources en petit nombre mais référencées de façon à la fois brève et claire, selon les normes ISO, que des abréviations ésotériques ou pas de sources du tout.

#### 8 — Remarques

On peut imaginer des dictionnaires électroniques où des données encyclopédiques (iconiques, géographiques, voire musicales) soient interrogeables à partir des entrées. Je m'en tiendrai plus modestement au contenu d'un dictionnaire de langue sur support papier, en évoquant quelques défauts d'information qui risquent d'être frustrants là où les mots ont valeur de termes.

Ex 1: à quoi bon une définition laborieuse de *fédéral* si l'on n'ajoute pas une liste indicative d'états fédéraux notoires?

Ex 2: on limite l'arbitraire de la dénomination *casier judiciaire* si l'on rappelle la pratique administrative prérévolutionnaire.

Ex 3: il est bon de savoir que l'expression *droits de la personne* est une conquête des féministes au Canada, mais qu'en Europe l'usage reste de dire *droits de l'homme*.

Ex 4: un complément de définition à portée pratique peut aussi avoir son utilité; ainsi, il n'est pas indifférent qu'une disposition juridique soit abrogée plutôt qu'annulée, car l'annulation complique les choses par son effet rétroactif.

### Conclusion

La lexicographie étant toujours une entreprise coûteuse, il est justifié de la préférer multifonctionnelle. En d'autres termes, il ne coûte pas beaucoup plus cher d'avoir en vue à la fois le thème et la version dans les deux langues, dans le cas du bilingue, et aussi la traduction spécialisée en même temps que la lecture des belles oeuvres.

En la matière, j'oserai dire que la qualité, en fin de compte, se mesure commercialement ou, si l'on préfère, à l'adéquation de l'offre à la demande. Il me semble en tout cas que le temps n'est plus où un programme de dictionnaire pouvait légitimement se limiter à la quête mythique de l'intralinguistique, en faisant comme si les mots ne servaient pas aussi à dénommer le réel. On ne sépare pas impunément la langue de la culture.

### Bibliographie

- Diké*, dictionnaire électronique du droit en cours d'élaboration, CELEX, Université de Paris XIII, avenue J. B. Clément, F. — 93430 Villetaneuse
- GROFFIER Ethel et REED David, *La lexicographie juridique. Principes et méthodes*, Cowansville (Québec), Ed. Yvon Blais, 1990
- KIEFER Ferenc, *Essais de sémantique générale*, Tours, Mame, 1974
- LERAT Pierre, «L'hyperonymie dans la structuration des terminologies», *Langages* 98, 1990, pp. 79—86.

Tamás SZENDE

Université de la Sorbonne Nouvelle — Paris III

### **Sélection des données lexicographiques: considérations méthodologiques à propos du nouveau dictionnaire hongrois-français**

La confection des dictionnaires bilingues est une pratique séculaire, mais contrairement aux usages du passé, ce n'est plus en ermite seulement que l'on arrive à construire une oeuvre originale. Aujourd'hui, un tel projet doit s'appuyer sur la collaboration étroite de linguistes, enseignants et traducteurs, locuteurs natifs de la langue d'arrivée et de la langue de départ. Non sans fierté, je peux vous annoncer que nous avons réussi à associer les détenteurs les plus éminents de ces types de savoirs. Le travail effectif de rédaction vient de commencer: en tant que responsable de l'équipe, je me suis permis d'assigner à chacun de mes collègues son travail, tenant compte de ses compétences et de ses préférences. Nous disposons désormais d'une bibliothèque adaptée aux besoins de la confection de cet ouvrage et d'un protocole de rédaction que nous avons élaboré avec les représentants de l'équipe hongroise.

La réalisation de ce projet nous a amenés à inventer un ensemble de programmation permettant de prendre en charge les exigences spécifiques d'un dictionnaire bilingue, susceptible d'assurer la cohérence et l'homogénéité de la nomenclature et d'automatiser au mieux chacune des démarches que comporte la rédaction d'un tel ouvrage. Il va sans dire que cette programmation est en constante évolution. Entrer en informatique, c'est en effet aborder une possibilité de renouvellement continu de nos perspectives, accepter d'envisager une remise en question permanente de nos habitudes. Si du côté français il existe un choix embarrassant de nomenclatures publiées et republiées régulièrement par des maisons d'édition aussi prestigieuses que *Le Robert* et *Larousse*, la description de la réalité linguistique hongroise souffre d'un retard de plusieurs dizaines d'années. Aussi, la constitution d'une base de données lexicales nous paraissait-elle inévitable.

En collaboration avec l'Institut de Linguistique de l'Académie Hongroise des Sciences nous mettons actuellement en mémoire une sélection de documents de langue hongroise publiés depuis 1960. On souhaite constituer ainsi une vaste banque de données textuelles qui s'enrichira au fur et à mesure des encodages réalisés et sera susceptible de répondre aux interrogations les

plus diverses. Notons que, pour l'instant, la langue écrite de l'époque contemporaine est représentée dans ce corpus uniquement par le discours littéraire en prose et que celui-ci comprend dès à présent des extraits d'oeuvres d'auteurs aussi variés que Abody, Boldizsár, Csoóri, Déry, Esterházy, Füst, etc. Le travail est loin d'être achevé. Notre corpus ne sera véritablement opérationnel que si on le complète par le traitement informatique de textes tirés des médias reproduisant massivement des expressions de tous les jours et des tournures de la langue parlée. L'utilisation de ce corpus permettra d'assurer une représentation équilibrée de certains domaines de spécialité et d'avoir des informations sur la naissance et la disparition des mots. Grâce à nos sources informatisées, on accumulera sur les vedettes un nombre considérable de contextes, donc un ensemble de données sémantiques et syntaxiques, ce qui permettra une description de plus en plus fidèle de leurs emplois.

Bien entendu, on ne devra pas se contenter de simples consultations pour voir si une réalité lexicale est attestée dans le corpus. Mémoire linguistique d'une époque, ce corpus nous réserve déjà bien des surprises en corrigeant nos jugements rapides et subjectifs ou en révélant l'existence de quantités de termes nouveaux et moins nouveaux que les dictionnaires disponibles actuellement omettent d'enregistrer. Ce n'est pas ici le lieu de s'étendre sur toutes ces questions, mais je voudrais simplement en souligner l'importance, étant donné que notre entreprise pourra se diversifier à l'avenir en une série de sous-projets visant à mettre à la disposition des chercheurs de domaines différents des glossaires et des classes de mots, des listes de fréquences, des descriptions quantitatives et d'autres instruments de travail complémentaires, autant de moyens d'observation lexicale et de réflexion sémantique aussi élaborés que possible dans le domaine français-hongrois.

Un dictionnaire bilingue met en parallèle deux langues en comparant à travers ses articles les données sémantiques et syntaxiques des mots. Destiné à la communauté linguistique hongroise et à la communauté linguistique française, il sera forcément à double direction; il fonctionnera donc pour le thème (locuteurs hongrois) et également pour la version (locuteurs français). Avec I. Mel'cuk, nous constatons une série de caractéristiques qui rendent les dictionnaires bilingues traditionnels insatisfaisants ou inadéquats, attachés surtout à réunir des équivalents, c'est-à-dire à traduire des mots par d'autres, sans se préoccuper des limites de leur équivalence. Sans chercher à en copier l'exhaustivité, nous nous référons dans notre méthodologie au *Dictionnaire Explicatif et Combinatoire (DEC)* qui met en oeuvre des procédures palliant ces insuffisances. L'application des principes du *DEC* à notre ensemble lexical bilingue permettra d'être exceptionnellement explicite en ce qui concerne la combinatoire lexicale des mots-et spécifiera non sporadiquement, mais si possible systématiquement les cooccurrences lexicales fréquentes, c'est-à-dire les combinaisons lexicales qui peuvent être attendues à un moment donné pour exprimer une idée donnée. Le recours méthodique aux fonctions lexicales permet de relever les expressions les plus fréquentes et les plus caractéristiques qui s'associent au mot. Ainsi, pour le mot 'tisztelet' /

'respect' on trouvera les combinaisons du type 'megkülönböztetett, őszinte, hálás ~ / profond, grand, respect / etc. (fonction Magn = très, intensément, à un degré élevé);' ~-et érez vki iránt' / avoir, éprouver du respect pour q' ou 'nagy ~-ben áll vki előtt' / 'jouir du respect de' (fonction Oper = verbe sémantiquement vide qui prend le nom du premier, deuxième, etc. actant de la situation comme son sujet grammatical et le mot comme son complément d'objet); á~ -et tanúsít vki iránt / faire preuve, témoigner de son respect envers q' (fonction CausManif = faire en sorte que quelque chose ait lieu; manifester) etc. D'autres exemples pour les fonctions citées: Magn: zaj ➔ pokoli / bruit infernal; köszön ➔ szívből / remercier chaleureusement; Oper: vágy ➔ érez / avoir envie; műtét ➔ végez / procéder à une opération; műtét ➔ átesik / subir une opération; CausManif: vélemény ➔ formál / émettre, exprimer une opinion; bocsánat ➔ kér / présenter ses excuses, etc.

Notre dictionnaire devra être orienté aussi bien vers la compréhension (vers le «que veut dire?») que vers la production (vers le «comment exprimer?», vers le «comment le placer dans le discours?»).

Comme le dit si bien Josette Rey-Debove, «le sens d'un mot, c'est son usage» et le mot isolé n'a généralement pas de sens. L'utilisateur de chaque dictionnaire a une double tâche de construction linguistique à effectuer: un premier choix sur l'axe paradigmatique et un deuxième choix sur l'axe syntagmatique. Ceci est d'autant plus important à souligner qu'il s'agit de deux langues génétiquement et typologiquement éloignées.

Il est notoire que le hongrois, aussi bien dans son lexique que dans sa structure, est très différent du français. Ceux qui apprennent ou enseignent le hongrois savent bien que le principal problème n'est pas d'ordre paradigmatique, même si les suffixes fonctionnels, par exemple, représentant pour la plupart une orientation dans l'espace et organisés selon un système tripartite, tendent à être des pièges aux locuteurs étrangers. On sait, en revanche, que ce qui est particulièrement problématique, c'est l'acquisition de l'aptitude à fabriquer des énoncés hongrois, à produire et à reproduire des structures syntaxiques correctes susceptibles de véhiculer des messages conformes aux normes. Notre dictionnaire essaiera donc de refléter fidèlement un système syntaxique extrêmement sensible aux déplacements, où l'accentuation d'un mot, voire d'un morphème, peut entraîner des changements de sens considérables.

Quant à la richesse de notre dictionnaire, nous ne la mesurerons pas au nombre d'entrées, ni au nombre d'équivalents, de semi-équivalents et d'équivalents-zéro (mots et expressions liés à une civilisation particulière). Certes, un dictionnaire doit aider le traducteur à diversifier son style, mais qu'il ne le fasse pas à travers une synonymie cumulative. Les bons articles fournissent des équivalents lexicaux, indiquent les principales variantes constructionnelles et font ressortir en même temps les identités ainsi que les différences structurales. Dans l'édification de l'article on tiendra compte à la fois des symétries et des asymétries; seul compte l'usage avec ses règles innombrables.

Toute négligence ou tout oubli temporaire de la réalité ne fait qu'augmenter la probabilité d'une traduction fautive. «Jó» veut dire probablement 'bon', mais attention! dès qu'on cherche à l'illustrer à travers des séquences usuelles, (p. ex.:

'légy jó, kisfiam' 'sois sage, mon petit';  
'légy oly(an) jó, írj, ha megérkeztél' 'sois gentil, écris à ton arrivée';  
'jó, hogy szólsz' 'heureusement que tu en parles';  
'jó meleg szoba' 'pièce bien chauffée'  
'mire jó ez?' 'à quoi cela sert-il?' etc, etc.)

on s'aperçoit que les seules compétences lexicale et grammaticale ne suffisent pas pour la production d'énoncés acceptables. Un dictionnaire bilingue doit introduire ses utilisateurs dans le vaste empire des usages et des conventions que l'on désigne généralement en français, par l'expression imagée, vague et quelque peu rebutante, «génie de la langue».

Notre hypothèse est la suivante: le traducteur (professionnel ou non) ne cherche pas seulement de simples équivalents et encore moins des phrases dans un dictionnaire bilingue; il est en réalité à la recherche de séquences susceptibles d'être insérées dans son discours et formant des énoncés naturels. Conséquence: plus un dictionnaire bilingue propose autour de ses vedettes des couples de syntagmes corrects et courants, structurellement symétriques ou asymétriques, plus il pourra être apprécié par l'utilisateur.

Notre méthode, présentée sommairement au cours de ce colloque mais qui est décrite sous tous ses angles dans le protocole de rédaction, va se perfectionner à l'usage. Nous l'avons voulue à la fois stricte et souple afin qu'elle soit valable pour le plus grand nombre de cas possibles. Mais, étant donné la variété infinie des types d'expressions d'une langue, on privilégiera, au sein de chaque article, la cohérence et la maniabilité sans trop se demander s'il existe un procédé classificatoire permettant d'embrasser toutes les catégories lexicales et constructionnelles.

## Problèmes posés par le traitement lexicographique des figés

1. Dans le cadre de ce rapide inventaire des principaux problèmes posés par le traitement lexicographique des figés il n'est guère possible de s'étendre longuement sur la problématique de la définition de *locution/expression figée*, ou sur celle de la délimitation de *collocation*, *locution figée* et *locution proverbiale*. En conséquence:

a) j'utiliserai ces termes techniques plutôt comme des termes intuitivement présents dans notre esprit;

b) je ne traiterai par la suite ni des collocations du type **diamétralement opposé**, qui sont d'une toute autre nature que les figées, ni des locutions proverbiales ayant un contenu de «sagesse» du type **A cheval donné on ne regarde pas la bouche** qui, posant grosso modo les mêmes types de problèmes que les figés, reçoivent en général un traitement très semblable à ceux-ci aussi bien du point de vue de leur description que de leur emplacement dans l'article de dictionnaire;

c) ainsi quand je dirai *figés*, j'entendrai par là un groupement de mots figé (GMF) de type et de figement variés à classer dans une section à part dans l'article de dictionnaire comme par exemple:

— les constructions verbales	<b>prendre part</b>
— les formules stéréotypées	<b>le troisième âge</b>
— les clichés situationnels	<b>un ange passe</b>
— la terminologie spécialisée	<b>le petit pont</b>
— les constructions parallèles	<b>sain et sauf</b>
— les comparaisons	<b>copains comme cochons</b>
— les kinégrammes	<b>s'arracher les cheveux</b>
— idiotismes traditionnels	<b>prendre la mouche; une poule mouillée</b>

2. Voici maintenant quelques problèmes généraux et spéciaux qui peuvent avoir une incidence plus ou moins importante sur la rédaction de la partie phraséologique des articles de dictionnaire bilingue. Puisqu'on est dans le bilingue, j'illustrerai mon propos par des exemples tantôt français tantôt hongrois.

## 2.1. La sélection des GMF

Le choix du corpus phraséologique d'un dictionnaire bilingue reste à l'heure actuelle malheureusement plutôt subjectif et intuitif dans la mesure où il n'y a pas eu d'études faites sur la fréquences des GMF du français ou du hongrois. Faire appel à la fois à tradition lexicographique et aux créations continuelles lexicographiquement encore non décrites et souvent éphémères du type **prendre la tête, inaugurer les chrysanthèmes, a labda gömbölyű, vki agyilag zokni**, toutes les deux soumises à une critique constante et sévère, semble être la solution de nécessité pour l'instant.

## 2.2. La délimitation et le classement des GMF dans la tradition lexicographique.

### 2.2.1. Les dictionnaires unilingues qui nous servent de point de départ, ont l'habitude de ne délimiter clairement — par des abréviations **Sz** ou **Szh** ou **loc.** — que les unités figées appelées traditionnellement idiomatiques. Voir les exemples de l'article **kanál** ou **cuiller** dans le *Magyar Nyelv Értelmező Szótára* (par la suite: *MNyÉS*) ou le *Petit Robert* (par la suite: *PR*): **megfojtaná egy kanál vízben; leteszi a kanalat; nagy kanállal eszik; úgy elverlek, hogy kanállal szednek majd fel — serrer la cuiller; faire qc. en deux coups de cuiller à pot; être à ramasser à la petite cuiller; ne pas y aller avec le dos de la cuiller.**

Dans un nombre de cas non négligeable, les dictionnaires unilingues «cachent» les GMF, sans les identifier en tant que tels, parmi les différents sens figurés d'un des éléments du figé. C'est le cas de **Cél: célba talál; célt téveszt; túllő a célon** ou de **But: atteinde, toucher le but; manquer le but; aller droit au but** dans les dictionnaires déjà cités.

A remarquer que le dictionnaire d'Eckhardt semble n'avoir en général aucun principe de délimitation et d'identification des GMF.

Trancher entre emploi figuré d'un lexème ou locution figée n'est pas toujours chose facile. Le test commutatif peut souvent s'avérer dans ces cas un outil d'identification efficace. Par exemple: **vki a szívére teszi a kezét vs \*vki a szívére helyezi/fekteti/ejti a kezét.**

### 2.2.2. C'est ici qu'il faut mentionner aussi le problème du choix des mots-clés des figés, choix qui a une influence sur leur classement lexicographique. En principe on a trois possibilités de classement:

a) On peut poser que le mot-clé sera toujours automatiquement le premier mot du figé. On aura du mal à respecter ce principe mécanique par exemple pour les figés hongrois dans lesquels l'ordre des éléments n'est pas tout à fait figé (**vki vkire nem nyit ajtót — vki vkire ajtót sem nyit**) ou dans lesquels le premier élément est facultatif (**vminek [nagy] ára van**). En plus, dans le cas des comparaisons figées du hongrois, ce principe s'avère complètement inopérant étant donné que le premier élément — **úgy, olyan** — est non seulement souvent facultatif, mais ferait regrouper sous un même mot un nombre impressionnant de figés, ce qui rendrait l'utilisation du dictionnaire pratiquement impossible.

b) Le mot-clé du figé pourrait aussi être l'élément dont le figement phraséologique est le plus grand, et dont l'apparition dans des figés est par conséquent le plus rare (par exemple **queue** dans le figé **tirer le diable par la queue**). Ceci ferait apparaître le figé dans le dictionnaire sous l'élément lexicographiquement le moins chargé permettant à l'utilisateur de trouver le plus rapidement possible le figé en question. Mais l'existence d'éléments facultatifs et le haut degré de subjectivité remettent en cause la réalisation conséquente de ce principe (pour **vki mehet Kukutyinba zabot hegyezni** le mot-clé serait-il **Kukutyin, zab** ou **hegyez?**).

c) Si ni le principe «mécanique», ni le principe de «fréquence négative» ne semblent être réalisables, on peut proposer d'utiliser le principe qui, sans être parfait, pose encore le moins de problèmes dans la pratique — la preuve en est qu'il est assez largement utilisé par les lexicographes — et qui fait appel aux catégories grammaticales. Les règles d'ailleurs assez automatiques de ce principe sont faciles à poser:

α) le mot-clé du figé est le premier substantif parmi les éléments constants du figé (**eau** dans **mettre de l'eau dans son vin**);

β) si la position du premier substantif est prise par des éléments alternatifs (**vki fűlig/nyakig ül az adósságban**), le premier dans l'alphabet (**fűl**) sera considéré comme mot-élé sous lequel apparaîtra le figé, l'autre (**nyak**) ne constituant qu'un élément de renvoi;

γ) si le figé ne contient pas de substantif, on considèrera comme mot-clé l'adjectif, faute d'adjectif le pronom, faute de pronom l'adverbe, faute d'adverbe le verbe. Le verbe vient en dernière position dans cette liste, même s'il figure dans l'immense majorité des figés, parce qu'il est souvent lexicographiquement trop chargé (**prendre, faire, être, rendre**, etc.) pour être opérant. En hongrois, on a en plus le problème des variantes avec préverbe, sans préverbe ou plusieurs préverbes (**vki vkinek a veséjébe lát** vs **vki vkinek belelát a veséjébe**; **vki előtt minden ajtó megnyílik** vs **kinyílik**; **vki vmit aprópénzre vált [fel]**) qui exclut pratiquement l'élection du verbe comme mot-clé.

δ) Seules les comparaisons échapperont à cette règle. Pour des raisons d'efficacité, il vaut mieux les classer sous l'élément *comparé*, forcément connus par l'utilisateur — l'adjectif (**triste comme un bonnet de nuit** sous **triste**) ou le verbe (**mentir comme un arracheur de dents** sous **mentir**) —, et non pas sous l'élément *comparant* de nature moins évidente.

En conclusion, une nomenclature essentiellement nominale, classée alphabétiquement semble s'imposer.

### 2.3. Les marques d'usage des GMF

J'appelle marques d'usage l'ensemble des informations complémentaires d'ordre sociolinguistique, grammatical et gestuel précisant les modalités d'emploi d'un GMF. Elles sont particulièrement précieuses pour tout utilisateur et pour ainsi dire inévitables pour les étrangers.

#### 2.3.1. Les jugements sociolinguistiques.

En ce qui concerne les jugements de type sociolinguistique, on peut constater qu'à propos de la qualification diachronique ou diastatique d'un

figé il y a des différences notables entre les dictionnaires disponibles sur le marché — ce qui paraît bizarrement presque normal — et souvent des inconséquences inquiétantes à l'intérieur d'un même dictionnaire. Voici quelques exemples concrets: **piquer un renard** est considéré par le *PR* comme *vieilli* seulement, alors que le *Lexis* le donne pour *vieux* et le *Dictionnaire du français (Hachette)* ne l'enregistre même pas, **danser devant le buffet** n'a pas de qualification dans le *PR*, par contre il est donné *pop.* dans le *Lexis*, **sucrer les fraises** est tantôt marqué *pop.* — sous *fraise* — tantôt *fam.* — sous *sucrer* — dans le *PR*.

Ces différences et inconséquences proviennent surtout du fait qu'il n'est pas toujours facile de constater si la qualification sociolinguistique se rapporte à un élément du figé (**haro** est bien un mot *archaïque* dans **crier haro**, **pofa** est indiscutablement *durva* dans **hogy nem sül le a bór a pofájárol**) ou au figé lui-même (pourquoi **kár a benzinért** en tant que tel serait-il *familier* ou **vki az anyja szoknyáján ül** *ironique*?). Dans ces deux derniers exemples, les qualifications *familier*, *ironique* ne se rapportent sûrement pas à un élément du figé, elle ne sont pas non plus quelque chose de constant du figé lui-même. Dans ces cas-là c'est plutôt la place respective des participants de la communication dans une hiérarchie sociale qui peut déterminer la valeur stylistique familière du figé ou un comportement humain concret (mollesse, veulerie) qui peut déclencher dans une situation de communication concrète une réaction verbale ironique.

Et n'oublions pas, pour clore ce sous-chapitre, que les problèmes posés par le choix de la qualification sociolinguistique auront toujours une influence décisive sur les équivalents dans la langue d'arrivée.

### 2.3.2. Les informations grammaticales

Une quantité importante d'emplois incorrects des figés commis surtout — mais non exclusivement — par les utilisateurs non francophones provient du fait que les dictionnaires fournissent peu d'informations, ou même aucune, sur les différentes restrictions d'ordre grammatical des figés. Je mentionnerai au passage cinq types de problèmes:

#### a) Restrictions sur le sujet du GMF

Tout d'abord, on ne trouve aucune indication dans les dictionnaires sur la nature animé ou non animé, l'âge ou le sexe du sujet des figés, ce qui peut entraîner des emplois fautifs de la part des étrangers. Il serait donc utile de signaler, chaque fois que c'est nécessaire, que par exemple **n'avoir ni queue ni tête**, **être cousu de fil blanc**, **nem ér egy lyukas garast sem** se construisent obligatoirement avec un sujet nom de chose, que **le marchand de sable est passé**, **rire aux anges** supposent normalement un sujet enfant-excepté bien sûr un emploi ironique-, que **se crêper le chignon**, **avoir des doigts de fée** ne se dit que de femmes, et qu'il n'y a que des individus de sexe masculin qui se **rincent l'oeil** en voyant une belle femme ou qui **sont comme un coq en pâte**.

b) Restrictions dans la forme du verbe

Il serait également d'une utilité certaine si les dictionnaires indiquaient que seule telle ou telle forme verbale et non telle autre est possible dans le figé en question. Quelquefois une forme lexicographique différente de l'infinitif l'indique déjà (**cela fera du bruit dans Landerneau**), mais il ne serait pas inutile de reprendre l'information entre crochets et préciser ainsi qu'il faut par exemple obligatoirement dire à l'impératif **occupe-toi de tes oignons**, une forme lexicographique infinitive **\*s'occuper de ses oignons** étant inimaginable, que **ne pas être dans son assiette** s'emploie presque exclusivement au négatif, que **vkít kemény fából faragtak** s'emploie toujours, **avoir maille à partir avec qn** surtout au passé, que **ne connaître qn ni d'Eve ni d'Adam** est dans la plupart des cas utilisé à la première personne du singulier et comme réponse à une question, etc.

c) Co-référence du possessif

La forme lexicographique traditionnelle des figés est en français, comme on le sait bien, l'infinitif où les formes possessives apparaissent à la troisième personne. Ainsi pour les non-francophones les figés **dire ses quatre vérités à qn** et **être dans tous ses états** sont identiques alors que du point de vue de leur fonctionnement ils sont loin de l'être: dans le premier le déterminant possessif «ses» est toujours invariable, dans le second il est variable. Et — une fois de plus — les dictionnaires ne donnent pas de renseignements de ce genre.

d) Transformation pronominale

Comment pourrait-on attendre qu'un étranger puisse individualiser l'utilisation d'un figé en faisant par exemple correctement, dans un contexte donné, la transformation pronominale de «de qn» du figé **apporter de l'eau au moulin de qn** si même les natifs hésitent là-dessus. Est-ce **il a apporté de l'eau à mon moulin** ou **il m'a apporté de l'eau au moulin** ou les deux, et si les deux formes coexistent, quelle est leur distribution? Le dictionnaire idéal que nous efforçons de rédiger devrait contenir aussi ce genre d'indication.

e) L'ordre des éléments constitutifs du figé

C'est un problème qui, pour des raisons évidentes, touche plutôt les figés hongrois. Pourrait-on dire que **vki vkinek bolhát tesz a fülébe** serait plutôt un GMF alors que **vki vkinek a fülébe bolhát tesz** resterait plutôt au niveau du concret? L'ordre des éléments pourrait-il être un signe de figement fiable? Autant de questions qui attendent encore d'être élucidées.

2.3.3. Un cas spécial: les GMF pouvant être accompagnées de gestes

Nombre de figés (**mon oeil, en avoir ras le bol**, etc.) peuvent être accompagnés de gestes. Il est bien évident qu'il est impossible de donner dans un dictionnaire bilingue une description détaillée ou une illustration du geste en question. Il serait par contre possible, voire souhaitable d'ajouter une marque conventionnelle < + geste > juste pour signaler à l'utilisateur qu'il se trouve en face d'un figé d'un emploi un peu particulier, quitte ensuite à se reporter éventuellement à un dictionnaire spécialisé.

## 2.4. La paraphrase des GMF

a) Dans la majorité des cas, il est relativement facile de donner une description sémantique, référentielle des figés (**prendre la mouche** = se mettre brusquement en colère) même s'il y a, malheureusement, quelquefois des différences sensibles entre les dictionnaires pour paraphraser un figé (**avoir pignon sur rue, tenir la dragée haute à qn**, etc.). Dans l'intérêt des utilisateurs, le dictionnaire idéal doit tout faire pour réduire au minimum les ambiguïtés dans les définitions.

Je ferai remarquer au passage que la description sémantique ne devrait jamais consister à donner comme équivalent un simple lexème ou encore moins un autre figé qui constituera un obstacle supplémentaire à la compréhension du figé de départ (cf. **tenir la dragée haute à qn** = «tenir tête à qn» dans le *PR*). Ainsi, autant que c'est possible, il faut s'efforcer de mettre dans la glose des éléments de précision d'ordre expressif qui peuvent nuancer l'emploi du figé en question en faisant apparaître le point de vue du locuteur. Ainsi par exemple dans la paraphrase du figé familier hongrois **vki eltűnik/eltávozik a balfenekén**, il faut souligner que quelqu'un disparaît, part *discrètement, sans être aperçu*.

b) Il faut distinguer ensuite un groupe de figés dont la paraphrase ne peut être donnée qu'à un niveau que j'appellerai pragmatique ou communicatif. En effet, des figés comme **un ange passe** ou **vmi alakul, mint púpos gyerek a prés alatt** ne peuvent pas être paraphrasés de la même façon que **prendre la mouche** et avoir comme définition pour le premier: \*«un silence gêné se fait soudain» ou pour le second: \*«vmi alakul». Il faudrait plutôt donner des commentaires du type: «remarque du locuteur lorsqu'il se produit dans une conversation un silence gêné et prolongé» ou «a beszélő az adott folyamattal, változással kapcsolatos elégedettségét fejezi ki». Ce genre de paraphrase devrait être distingué, même typographiquement par des chevrons doubles par exemple, des paraphrases dites sémantiques.

c) Il existe encore un troisième groupe de figés, numériquement peut-être moins important, dans le cas desquels on a besoin aussi bien d'une paraphrase sémantique que d'un commentaire pragmatique. Le figé hongrois **vki vkinek kitapossa a belét** peut avoir à la fois une description sémantique — «vki vkit súlyosan, durván bántalmaz» — et une description pragmatique au sens de «fenyegetés». Il en va de même pour le figé français **en faire voir à qn de toutes les couleurs**: «lui faire supporter toutes sortes de choses désagréables» et «menace».

## 3. Les équivalents en L2

Comme dernier point, je me contenterai de remarquer encore rapidement que les problèmes énumérés n'ont pas seulement une influence décisive sur la forme et le classement lexicographique des figés, mais qu'ils doivent évidemment être toujours considérés par les rédacteurs dans le choix des équivalents en L2. On n'insistera jamais assez sur le fait que, même dans les

cas apparemment les plus simples, un ou plusieurs des facteurs énumérés peuvent être là, de façon plus ou moins latente, et rendre difficile, le cas échéant même impossible, de trouver l'équivalent total — cas idéal, bien entendu — d'un figé dans une autre langue.

### Bibliographie d'orientation

BÁRDOSI, Vilmos (1986): *De fil en aiguille. Kalandozás a francia szólások világában*. Budapest, Tankönyvkiadó, (Előszó, 5—15).

BÁRDOSI, Vilmos — HESSKY, Regina: «Frazeológiai egységek szótári leírásának kérdései», à paraître dans *Filológiai Közlöny*.

ETTINGER, Stefan (1989): «Einige Probleme der lexikographischen Darstellung idiomatischer Einheiten (französisch — deutsch)», Gertrud Gréciano (Ed.): Europhras 88. Phraséologie contrastive. Actes du Colloque International, Klingenthal — Strasbourg, 12—16 mai 1988, in: *Collection Recherches Germaniques* n° 2, 95-115.

KROMANN, Hans-Peder (1987): «Zur Typologie und Darbietung der Phraseologismen in Übersetzungswörterbüchern», in: Jarmo Korhonen (Hrsg.): *Beiträge zur allgemeinen und germanistischen Phraseologieforschung. Veröffentlichungen des germanistischen Instituts 7. Universität Oulu*. Universität Oulu, 183-192.



## Une leçon à tirer de la rédaction de différents articles

Il existe en lexicographie bilingue un principe général souvent répété, souvent entendu (et qui est devenu presque un lieu commun), c'est que dans un dictionnaire bilingue, il y a en général autant de sens d'un mot en L1 qu'il a d'équivalents (qu'il y a de «traductions») en L2.

Pourtant il serait assez absurde de considérer toutes les traductions possibles comme des «sens». Par exemple, pour ce qui concerne les équivalents sémantiques du mot français **heure**, on peut imaginer plusieurs schémas:

**heure** *nf* ① *óra* *division de la journée*; *heure fixée* **quelle heure est-il?** hány óra (van)? *l'* ~ **exacte/juste** pontos idő; *mesure de durée* *l'* ~ **est subdivisée en 60 minutes** az óra 60 percből áll ② *moment* óra; időszak; napszak: ~ **de classe/de cours** tanóra; tanítási óra; **aux heures de repas** étkezési időben ◇ *pour l'* ~ pillanatnyilag . . . etc.

ou bien:

**heure** *nf* *óra* ① *division de la journée*; *heure fixée* **quelle heure est-il?** hány óra (van)? *l'* **exacte/juste** pontos idő; ② *mesure de durée* *l'* ~ **est subdivisée en 60 minutes** az óra 60 percből áll ③ *moment* óra; időszak; napszak: ~ **de classe/de cours** tanóra; tanítási óra; **aux heures de repas** étkezési időben  
◇ *pour l'* ~ pillanatnyilag . . . etc.

Mais je n'aurais jamais l'idée saugrenue de dresser un schéma comme par exemple ceci (confondant sous la même étiquette des choses tout à fait différentes, comme *1 heure de l'après-midi* et *une heure de classe*):

**heure** *nf* *óra* ① *division de la journée*; *heure fixée* **quelle heure est-il?** hány óra (van)? *mesure de durée* *l'* **n est subdivisée en 60 minutes** az óra 60 percből áll; *unité d'enseignement* ~ **de classe/de cours** tanóra; tanítási óra ② napszak; idő: *l'* ~ **exacte/juste** pontos idő; **aux heures de repas** étkezési időben  
et, horrible dictu:  
③ pillanat, parce que: *pour l'* ~ pillanatnyilag . . .

Cet exemple illustre bien le statut des «équivalents»: l'équivalent (la traduction) est très souvent accompagné d'indications d'ordre global ou général (*division de la journée, mesure de durée, etc.*), indications qui sont indispensables pour le fonctionnement de dictionnaire de «production». Sinon, comment un Français pourrait-il choisir entre les équivalents qu'on lui propose?

Que les équivalents soient accompagnés presque obligatoirement d'indications d'ordre général (surtout en cas de polysémie), veut dire en termes plus exacts que le choix des sens, la délimitation sémantique dépend, très souvent, d'une approche conceptuelle. Ce qui signifie que même dans un dictionnaire bilingue, les champs conceptuels ont une priorité structurelle par rapport aux champs sémantiques, surtout si le dictionnaire doit remplir non seulement la fonction de compréhension, mais aussi celle de production, les deux fonctions dans les deux langues, thème et version.

\* \* \*

Nous avons donc rapproché le sens du concept, au lieu de la traduction. C'est d'autant plus intéressant que, très souvent, à un seul sens ainsi défini, correspondent plusieurs traductions. Il serait assez bizarre de prendre toutes ces traductions comme autant de «sens». Voyons encore deux exemples pour cette répartition logico-sémantique:

**blanc, blanche I** *adj* ① *couleur*

② *qui n'est pas écrit*

③ *qui n'a pas son effet habituel: mariage ~ < formális, csak papíron létező házasság >; nuit blanche < át nem aludt éjszaka >; vers blancs < nem rímelő verssorok >; voix blanche < erőtlen/elszorult hang >*

Là, il y a un seul sens ③, bien que les «équivalents» soient très différents du point de vue strictement sémantique: c'est la définition qui les regroupe.

**blanc, blanche I** *adj*

**II** *n mf* (les Blancs)

**III nm** ① *couleur, peinture*

② *linge*

③ *espace vide*

④ *vin*

⑤ *partie blanche de qc ~ d'oeuf tojás fehérje; le ~ de l'oeil a szem fehérje; regarder qn dans le ~ des yeux a szemébe néz vkinek; ~ de poulet csirke mellehúsa*

Ici, les équivalents de III ⑤ ne sont pas aussi différents: il y a entre eux une certaine proximité sémantique aussi; pourtant, il y a autant de traductions que d'exemples.

\* \* \*

La question se pose maintenant de savoir ce qui justifie une telle démarche lexicographique. Je terminerai mon intervention par une réponse d'ordre épistémologique: est vrai, à mon avis, ce qui fonctionne bien. Et puisque le dictionnaire devient plus utilisable de cette façon (c'est à dire grâce à une structure plus claire), je crois qu'on peut considérer la démarche en question comme justifiée.



Júlia PAJZS

Institut de Linguistique de l'Académie

## **Le rôle de l'ordinateur dans la rédaction du nouveau dictionnaire Hongrois—Français/Français—Hongrois**

L'ordinateur se révèle utile à trois titres à toutes les étapes de la rédaction d'un dictionnaire:

- le corpus sur ordinateur, constitué de textes continus facilite la collecte des données. Les lexicographes peuvent y sélectionner des exemples adéquats, apprécier les différents sens et observer des sens nouveaux.
- La rédaction même des articles s'effectuant sur ordinateur, le dictionnaire ainsi sauvegardé est aisément et continuellement modifiable; il est en outre possible de le convertir directement sur bande de photocomposition.
- Enfin, et ce n'est pas négligeable, le dictionnaire électronique est aussi une base de données particulière, utilisable de nombreuses façons. Par exemple, il permet la rédaction du format de poche d'un dictionnaire ou de sa version accessible sur disquettes.

Ce sont ces aspects que je me propose de présenter en détail.

Le corpus utilisé par les rédacteurs de la partie hongrois-français a été collecté à l'origine pour le *Grand Dictionnaire Historique de la Langue Hongroise* à l'Institut de Linguistique de l'Académie des Sciences de Hongrie. L'objectif de ce projet est la préparation d'un dictionnaire de la langue hongroise de Gutenberg à nos jours — fondé en premier lieu sur la collecte électronique (voir Kiss L.—Pajzs J. 1989, Pajzs 1990, 1991). Dans ce but, nous réunissons des textes courts sur ordinateur, pour constituer un corpus qui comptera environ 20 millions de mots. Pour ce grand dictionnaire nous avons actuellement collecté environ 10 millions de mots, en majorité des textes du 19<sup>ème</sup> siècle. Dans le corpus du 20<sup>ème</sup> siècle, qui compte environ 2 millions de mots, nous avons sélectionné des textes qui nous ont semblé convenir au dictionnaire hongrois-français: des extraits de prose publiés après 1960. Ce corpus est constitué de 57 extraits totalisant actuellement plus de 200 000 mots. Nous allons l'étendre par l'apport de textes écrits après les années 60, qui viendront s'ajouter au corpus croissant du grand dictionnaire. En outre, nous voudrions également obtenir de temps en temps la publication complète sur un mois, de quotidiens hongrois.

Nous avons indexé les textes continus avec un logiciel simple (WORD-CRUNCHER) destiné à consulter le corpus. Nous avons ainsi établi la liste de toutes les occurrences des mots du corpus. Ce logiciel est très simple à utiliser: quand on récupère le corpus, tous les mots des textes s'inscrivent sur l'écran par ordre alphabétique avec le nombre de leurs occurrences. En positionnant le curseur sur le mot cherché on peut consulter l'ensemble des contextes de ce mot. En général, le contexte se compose de trois lignes mais on accède facilement à de plus larges extraits. On peut aussi retrouver des groupes de mots: les mots qui commencent ou finissent par les mêmes caractères, ce qui permet de rassembler toutes les occurrences d'une vedette. Il est aussi possible d'examiner les cooccurrences de différents mots. En recourant à toutes ces fonctions, les lexicographes peuvent d'une part chercher des exemples de la littérature susceptibles de figurer dans les articles du dictionnaire, ils peuvent d'autre part rassembler des informations sur les nuances de sens d'un mot, et éventuellement trouver des mots et sens nouveaux.

La saisie des articles sur ordinateur a pour principal avantage le fait qu'on puisse les récupérer aisément, les modifier, en imprimer la dernière version etc. Il est en principe possible d'enregistrer et d'imprimer les articles sous leur forme définitive avec les polices de caractères prescrites par le Protocole de rédaction; cependant, afin que ce dictionnaire ne soit pas seulement un joli fichier de traitement de texte, mais aussi une base de données lexicales d'une autre qualité, il faut que les diverses unités des articles soient marquées par des étiquettes spécifiques et non par les polices de caractères.

Les auditeurs ayant des expériences diverses dans le domaine informatique, je me permettrai d'éclaircir la différence entre un simple *amas de données* et une *base de données*.

- Un amas de données (par exemple un texte continu sans codes) n'est utilisable que comme fichier de traitement de texte: on peut le récupérer en entier, y chercher des chaînes de caractères, le modifier, l'imprimer etc. Il serait évidemment stupide de sauvegarder les données d'un annuaire téléphonique dans un texte continu parce que, dans ce cas, si on cherchait sur le minitel les coordonnées d'une personne nommée MICHEL, on obtiendrait toutes celles des personnes dont le nom, le prénom ou l'adresse comporte la chaîne MICHEL.
- Mais si on les sauvegarde dans une base de données, où le nom est contenu dans un champ, le prénom dans un autre et l'adresse dans un troisième, il est alors possible de limiter la recherche au nom et de déterminer pour chacun le prénom ou l'adresse etc.

De même, si on sauvegarde un dictionnaire comme un simple texte et non comme un base de données, on ne peut y chercher que les mots, tandis que dans une base de données on peut préciser si on cherche l'occurrence d'un mot parmi les vedettes, les équivalents, les exemples ou les locutions figées. La

combinaison de ces fonctions donne accès à des informations très utiles, comme par exemple la liste des vedettes ayant le même équivalent, ce qui garantit une meilleure cohérence au dictionnaire.

Le type le plus connu de base de données est celui qu'on appelle relationnel. L'annuaire téléphonique, le répertoire d'adresses ou l'inventaire d'une usine appartiennent à ce type dont la caractéristique est d'avoir les mêmes champs dans tous les articles: dans le cas de l'annuaire téléphonique il s'agit du nom, du prénom, de l'adresse et du numéro de téléphone. La longueur de chaque champ est plus ou moins constante (par exemple, 20 caractères suffisent pour le nom, 40 caractères pour l'adresse etc.). C'est pourquoi ces bases de données sont construites de telle manière que les champs sont définis par leur longueur maximale et leur ordre. Evidemment, le dictionnaire n'est pas de ce type. D'abord la quantité et la qualité des données figurant dans les articles sont très différentes: certains mots ont plus de dix sens alors que d'autres n'en ont qu'un; il se peut aussi qu'une vedette n'ait pas d'équivalent mais soit suivie d'exemples, etc. De surcroît, la longueur de chaque champ est très variable. Pour cette raison, il vaut mieux sauvegarder les dictionnaires comme des bases de données définies par une grammaire.

Le concept de base de données définie par une grammaire (Gonnet-Tompa, 1987), qui a été développé en traitant l'*Oxford English Dictionary (OED)*, par ordinateur, est fondé sur l'expérience qu'une base de données textuelle diffère par sa nature des autres bases de données, qui en majeure partie ne sont pas textuelles. L'objectif visé en forgeant ce concept est de pouvoir distinguer des champs de base de données tout en conservant le mieux possible la forme originale du texte. L'ordre et la structure des champs sont donnés par une grammaire, à laquelle les articles doivent être conformes pour faire partie de la base de données. Dans une base de données définie de cette façon on peut recourir aux opérations usuelles sur le contenu des champs.

La structure du nouveau dictionnaire hongrois-français/ français-hongrois peut être décrite par la grammaire suivante:

```

DIC := [ART1 .. ART1]
ART := ENT [BLS1 .. BLSn] LGG? IFS? COL
ENT := VDT CGR MAE? RCT?
MAE := (DDS? | NDL? | LIG?)
BLS := MAE? RCT? IDS? [EQV1, ..., EQVm] [EXP1 .. EXPk]
LFG := [EXP1 .. EXPj]
EQV := [IDS? mots CGR?]
EXP := [mots TRD]
mots := [a .. z A .. Z 0 .. 9.;]
COL := NOM DAT
    
```

Le début et la fin des champs sont codés — comme suit:

```

DIC := <DIC> ... </DIC>
ART := <ART> ... </ART>
    
```

DIC := <DIC> ... </DIC>  
ART := <ART> ... </ART>  
ENT := <ENT> ... </ENT>  
BLS := <BLS> ... </BLS>  
LFG := <LFG> ... </LFG>  
IFS := <IFS> ... </IFS>  
VDT := <VDT> ... </VDT>  
CGR := <CGR> ... </CGR>  
MAE := <MAE> ... </MAE>  
DDS := <DDS> ... </DDS>  
NDL := <NDL> ... </NDL>  
LIG := <LIG> ... </LIG>  
RCT := <RCT> ... </RCT>  
IDS := <IDS> ... </IDS>  
EQV := <EQV> ... </EQV>  
EXP := <EXP> ... </EXP>  
TRD := <TRD> ... </TRD>  
COL := <COL> ... </COL>  
NOM := <NOM> ... </NOM>  
DAT := <DAT> ... </DAT>

Les abréviations:

DIC dictionnaire

ART article

ENT entrée

BLS bloc sémantique

LFG locution figée

IFS informations supplémentaires

VDT vedette

CGR catégorie grammaticale

MAE marques d'emploi

DDS domaine de spécialité

NDL niveau de langue

LIG limitation géographique

RCT construction grammaticale spécifique

IDS indication de sens

EQV équivalent

EXP exemple

TRD traduction

COL saisie de l'article

NOM nom

DAT date de la dernière intervention

L'article du mot *óra* se présente sous cette forme:

```

<ART>
<ENT> <VDT> óra </VDT> <CGR> </ENT>
<BLS> 1 <IDS> (szerkezet) </IDS>
<EQV> <IDS> karóra </IDS> montre <CGR> f </CGR>; </EQV>
<EQV> <IDS> utcai </IDS> horloge <CGR> f </CGR>; </EQV>
<EQV> <IDS> fali </IDS> pendule <CGR> f </CGR>; </EQV>
<EQV> <IDS> asztali </IDS> pendulette <CGR> f </CGR>; </EQV>
<EQV> <IDS> gáz, villany </IDS> compteur <CGR> m </CGR>; </EQV>
<EXP> az pontosan jár <TRD> la montre marche bien; </TRD> </EXP>
<EXP> az ön órája siet/megállt <TRD> votre montre avance/
s'est arrêtée; </TRD> </EXP>
<EXP> az ön órája két percet késik
<TRD> votre montre retarde de deux minutes; </TRD> </EXP>
<EXP> beállítja az órát <TRD> mettre la montre à l'heure; </TRD> </
EXP>
<EXP> húzd fel az órát <TRD> remontre l'horloge; </TRD> </EXP>
<EXP> az ~ /delet ütött <TRD> midi vient de sonner </TRD> </EXP>
</BLS>
<BLS> 2 <IDS> idő </IDS>
<EQV> heure <CGR> f </CGR> </EQV>
<EXP> hány ~ (van)? <TRD> quelle heure est-il?; </TRD> </EXP>
<EXP> meg tudná mondani, hány ~ van? <TRD> avez-vous l'heure? </
TRD> </EXP>
...
</BLS>
...
</ART>

```

Cette forme est évidemment moins agréable à lire et moins immédiatement accessible que la forme non structurée enregistrée avec des caractères différents. Mais on peut facilement en établir une version typographiée tandis que l'inverse est impossible: on ne peut pas transformer les polices en délimitations des champs, parce qu'une même police peut apparaître dans des champs différents (par exemple on écrit en caractères normaux les équivalents, les traductions, certaines abréviations; en italique les catégories grammaticales, les marques d'emploi, les indications sémantiques etc.). En outre, l'utilisation d'un masque pour écrire les articles oblige les lexicographes à respecter cette grammaire. Celle-ci est évidemment modifiable selon la nécessité, mais il est essentiel qu'on sache à quelles règles obéit la structure d'un article afin d'écartier les ambiguïtés. Le logiciel qui transformera les codes en polices de caractères contrôlera d'abord la conformité des articles à la grammaire, ajoutera les codes éventuellement manquants, et si cela n'est pas possible, indiquera au rédacteur où l'article doit être corrigé. Après vérification de la conformité, il fournira une version typographiée de l'article que le lexicographe pourra consulter, et éventuellement modifier, imprimer etc. Mais la forme la plus pertinente de sauvegarde d'une base de données lexicales est celle où les différents champs sont marqués par les symboles du

*SGML (Standard Generalized Markup Language)*, norme internationale de sauvegarde de textes sur ordinateur. L'emploi en est judicieux, en particulier parce qu'elle facilite la mise en oeuvre des logiciels de recherche disponibles dans le commerce. L'un de ces logiciels — peut-être le meilleur — est le PAT (Gonnet 1987), établi pour le traitement informatique de l'*OED*, et qui permet d'en récupérer efficacement la forme SGML. Il donne accès à toutes les données d'un dictionnaire: par exemple on peut chercher les vedettes des articles où se trouve une locution figée, ou bien contrôler les constructions grammaticales spécifiques, ou les équivalents etc.

Enfin je voudrais mentionner quelques perspectives ouvertes par le traitement informatique. Dans la phase rédactionnelle, l'ordinateur permet de «retourner» les articles, c'est à dire que l'on peut transformer les articles français—hongrois en hongrois—français et inversement. Cette fonction bien sûr ne sert qu'à vérifier la cohérence de l'ensemble, mais elle constitue une aide efficace: d'une part on peut voir s'il y a des équivalents qui ne figurent pas comme vedette dans l'autre partie du dictionnaire et remédier à ces lacunes; d'autre part on peut éliminer les inconséquences.

Par ailleurs, la version électronique permet de réaliser la version de poche du dictionnaire. Il suffit de coder les vedettes et les parties des articles devant figurer dans la version abrégée. Avec un logiciel approprié on peut aussi contrôler les nouveaux renvois. Le manuscrit ainsi préparé pourra être publié après une nouvelle rédaction. Naturellement les rééditions remaniées du dictionnaire seront aisément réalisables à partir de la forme sauvegardée sur ordinateur.

Pour terminer, mais ce n'est pas le moins important, il est facile de produire une version de la base de données sur disquettes, ce qui serait très apprécié par une grande partie des usagers. Traducteurs et chercheurs redigent de plus en plus sur ordinateur, et dans ce cas, il leur est évidemment plus pratique de consulter un dictionnaire sur écran. La version électronique du dictionnaire présente plusieurs avantages sur la version imprimée: d'une part il serait possible de rechercher non seulement les vedettes, mais aussi d'autres éléments particuliers des articles (par exemple les constructions grammaticales spécifiques); d'autre part elle offrirait une possibilité inédite de recherche à partir des mots tels qu'ils figurent dans un texte (formes fléchies, suffixées, conjuguées) dans les deux langues. L'utilisateur entrerait le mot cherché sous la forme présente dans le texte et verrait apparaître à l'écran la liste des vedettes concernant cette forme. J'ai moi-même chaque jour l'occasion de constater quels grands services rendrait cette fonction, en particulier au début de l'apprentissage d'une langue. Les conditions techniques sont d'ores et déjà réunies: il existe un logiciel capable de distinguer les formes fléchies de la langue hongroise et leurs terminaisons (Prószéky—Tihanyi 1991), et pour les mots français, il y a un *Dictionnaire Electronique des formes fléchies (DE-LAF)* qui contient toutes les formes possibles et leurs vedettes (Courtois—Silberztein 1990). Ils pourraient être respectivement adaptés à la version

électronique du dictionnaire, ce qui nous permettrait de publier, vraisemblablement pour la première fois, un dictionnaire où il est possible de retrouver automatiquement les vedettes à partir des formes fléchies.

### Bibliographie

COURTOIS, B.—SILBERZTEIN, M.: (1990) *Dictionnaires électroniques du français*, Larousse, Paris.

GONNET, G.: (1987) *PAT—An efficient text searching system*, University of Waterloo Centre for the New Oxford English Dictionary

GONNET, G.—TOMPA, F.: (1987) *Mind Your Grammar: a New Approach to Modelling Text*: University of Waterloo Centre for the New Oxford English Dictionary

KISS, L.—PAJZS, J.: (1989) *A magyar irodalmi és köznyelv nagyszótára (1533—1990)* Magyar Nyelv 1989. évf. 2. szám., 129—136.

PAJZS, J.: (1990) *Számítógép és lexikográfia*, Doktori disszertáció Az MTA Nyelvtudományi Intézet kiadásában. Budapest.

PAJZS, J.: (1991) *Réalisation assistée par ordinateur de grands dictionnaires français et hongrois*, Cahiers d'Etudes Hongroises 3/91 Centre Interuniversitaire d'Etudes Hongroises, Université de Paris III, Institut Hongrois de Paris, 47—54.

PRÓSZÉKY, G.—TIHANYI, L.: (1991) *A Fast Morphological Analyzer and Spelling Checker for Agglutinating Languages*, Submitted to the 3rd Conference on Applied Natural Language Processing, Trento, Italy, 1992.