

2010, à paraître dans les *Actes des Huitièmes journées scientifiques du réseau Lexicologie, terminologie, traduction*, 15-17 octobre 2009, Lisbonne.

Le Dico-FRAR : base de données lexicographiques bilingue français-arabe

Samia Bouchaddakh

Centre d'analyse et de traitement informatique du français québécois (Catifq),
Université de Sherbrooke

Observatoire de linguistique Sens-Texte (OLST), Université de Montréal
samia.bouchaddakh@usherbrooke.ca

Résumé

Cet article présente le Dico-FRAR, une base de données lexicographiques bilingue français-arabe. Cette base de données a été conçue en vue de proposer à l'utilisateur francophone qui désire encoder en arabe, non seulement la traduction des unités françaises, mais également une description globale des unités lexicales dans les deux langues mises en parallèle. Les équivalents fournis, qui sont le résultat de cette mise en parallèle, sont décrits de manière systématique sous leurs différentes facettes : sémantique, syntaxique et lexico-combinatoire. La finalité de la base Dico-FRAR est multiple et offre, sur les plans de son contenu informationnel et de sa conception, un intérêt pour l'utilisateur apprenant de la langue arabe, pour le traducteur et pour le lexicographe. Cet article comprend trois sections principales : problématique générale de notre travail et mise en contexte, description de la base Dico-FRAR et exemple de traitement lexicographique bilingue tel qu'il est proposé dans cette base.

Mots-clés

Base de données relationnelle; lexicographie bilingue; lexicologie explicative et combinatoire; traduction; divergences interlinguistiques; encodage

1. Introduction : problématique et contexte

Le traitement des vocables proposé dans les dictionnaires bilingues français-arabe n'est pas toujours satisfaisant, notamment pour les usagers francophones qui consultent ces ouvrages en vue de produire un texte dans la langue cible qu'ils ne maîtrisent pas encore parfaitement. En effet, dans une perspective d'encodage, les usagers recourent à ces dictionnaires bilingues dits « de thème » pour connaître certes l'équivalent traductif arabe à une unité lexicale française, mais également toutes les informations sémantiques, syntaxiques ainsi que la combinatoire lexicale qui lui sont rattachées, afin de pouvoir

l'employer correctement dans la langue cible. Ces usagers ont aussi besoin d'avoir plus de données sur les divergences lexicales, sémantiques, stylistiques, etc. qui existent entre les deux langues mises en présence, afin de prévenir les éventuelles erreurs et de réussir efficacement leurs tâches d'encodage et de traduction.

Cependant, les nombreuses lacunes et faiblesses dont souffrent les dictionnaires bilingues français-arabe représentent un handicap pour la réussite de telles tâches. Parmi ces faiblesses, citons la multiplicité des équivalents proposés, le manque de précision dans les données sémantiques et syntaxiques consignées, la non-identification des collocations, l'absence d'informations sur les divergences interlinguistiques qui existent entre les deux langues, le manque de structure dans l'information contenue dans les articles, etc.

L'insatisfaction des usagers de ces dictionnaires est fortement ressentie du fait que la langue arabe, très différente du français, leur est difficilement accessible. En effet, en cas d'absence ou manque d'informations, l'utilisateur ne peut facilement aller chercher un complément d'information dans un dictionnaire unilingue. La consultation d'un dictionnaire arabe n'est pas une tâche facile pour un locuteur non arabophone. Les entrées sont très souvent structurées selon la racine et non la forme du lexème. Tous les lexèmes, excepté quelques unités grammaticales, sont ramenés à une racine généralement trilitère représentant un contenu sémantique commun. Les lexèmes dérivés à partir de cette racine, dont la forme est purement consonantique, sont inclus au sein de cet article de dictionnaire¹. Cette tradition lexicographique arabe, qui pose la dérivation comme principe d'organisation des entrées dans le dictionnaire, suppose un savoir préalable et pose problème pour les locuteurs non arabophones.

Par ailleurs, l'on assiste depuis plusieurs années au développement d'une variété de base de données lexicales dans différentes langues, ce qui a permis de réaliser des avancées notables dans le domaine lexicographique unilingue (par exemple, WordNet [Fellbaum 1998], FrameNet [Baker et al. 1998, Fillmore et al. 2001], DiCo [Polguère 2000, 2003], Dafles [Selva et al. 2002, Verlinde et al. 2005]). En revanche, en lexicographie bilingue, c'est au cours de la dernière décennie que des travaux intéressants ont commencé à se multiplier (par exemple, Fontenelle 1997, Alonso Ramos 2001). Toutefois, la majorité de ces travaux sont concentrés, notamment, sur le traitement des langues indo-européennes, comme l'anglais et le français. Dans la paire de langues français-arabe, en particulier, il n'existe pas encore, à notre connaissance, de ressource lexicographique satisfaisante susceptible d'être mise à la disposition des traducteurs et des apprenants de l'arabe langue étrangère.

Dans ce contexte, nous avons examiné les problèmes qui se posent dans les dictionnaires bilingues français-arabe et développé une base de données lexicales visant à améliorer les descriptions qu'on y trouve, particulièrement dans ceux dits « de thème » ou

¹ Nous pouvons citer pour illustration, l'exemple canonique de la racine KTB qui donne par dérivation le verbe [*kataba*] 'écrire', [*kitâb*] 'livre', [*kuttâb*] 'école islamique', [*kâtib*] 'écrivain', [*maktûb*] 'écrit ou lettre', etc. Toutes ces unités sont décrites sous l'entrée KTB

« d’encodage » [Bouchaddakh 2009, en préparation]. Cette base de données (BD), que nous proposons de présenter dans les pages qui suivent, est baptisée « Dico-FRAR ».

Notre exposé est organisé de la façon suivante. Nous présentons d’abord les propriétés générales de notre BD (section 2.1). Ensuite, nous définissons le contenu informationnel que renferment nos articles de dictionnaire bilingues (section 2.2), avant d’introduire, dans ses grandes lignes, la structure générale de la BD réalisée et de montrer son intérêt, aussi bien pour l’usager que pour le lexicographe (section 2.3). Finalement, dans la troisième et dernière section, nous donnons un exemple d’article bilingue français-arabe, ce qui permet d’illustrer, par le fait même, le traitement que nous préconisons dans une base de données lexicographiques bilingue.

2. Dico-FRAR

2.1 Caractéristiques générales

La base de données lexicographiques Dico-FRAR repose sur les principes de la *lexicologie explicative et combinatoire* (LEC) [Mel’čuk *et al.* 1995] dont le modèle lexicographique permet de combler, à plus d’un égard, les lacunes et les faiblesses des dictionnaires bilingues traditionnels.

Conçue principalement pour l’apprenant francophone de la langue arabe qui est de niveau intermédiaire ou avancé, ainsi que pour le traducteur et pour tous ceux qui s’intéressent à la production d’énoncés en arabe, la base Dico-FRAR est composée des trois parties suivantes. La première est réservée à la description unilingue des unités françaises, la deuxième prend en charge la description des unités lexicales arabes et la troisième a pour objet la mise en relation de ces unités françaises et arabes dans une optique d’encodage.

L’unité de description dans cette base est la *lexie*². Une fois que les différentes acceptions du vocable ont été distinguées et ordonnées, chaque *lexie* est décrite sur les plans sémantique, syntaxique et lexico-combinatoire avec une illustration de son emploi dans différents exemples.

Pour les unités arabes, les articles de dictionnaire sont élaborés dans un métalangage français, ce qui rend l’accès à l’information plus aisée pour un usager francophone et facilite aussi la mise en correspondance des deux langues (le français et l’arabe), ainsi que la sélection des meilleurs équivalents. Dans Bouchaddakh [2008], nous avons montré que la structuration explicite des définitions rédigées dans un même métalangage, selon les principes de la LEC, permettait de sélectionner adéquatement les équivalents à proposer dans un dictionnaire bilingue. L’article bilingue, qui résulte de cette mise en correspondance, fournit à l’usager toutes les informations qui lui sont nécessaires dans sa tâche d’encodage ou de traduction. Il propose ainsi à ce dernier des équivalents plus appropriés, souligne les divergences interlinguistiques (sémantiques, stylistiques, syntaxiques, etc.) entre deux *lexies* dites « équivalentes » et, enfin, présente des solutions

² Dans la terminologie de la LEC, une *lexie* est soit un lexème soit une locution associé(e) à un sens donné. Un regroupement de *lexies* qui ont le même signifiant et un lien sémantique évident est appelé un *vocable*.

en cas d'absence d'équivalent et des mises en garde contre les éventuelles erreurs d'encodage. Cela étant, passons maintenant à la présentation du contenu informationnel d'un tel article bilingue.

2.2 Contenu informationnel de l'article bilingue

Afin d'offrir un traitement lexicographique rigoureux, l'article bilingue, qui est ainsi la synthèse de la mise en parallèle des descriptions des unités lexicales françaises et arabes, doit renfermer un certain nombre d'informations que nous définissons comme suit.

2.2.1 Informations liées à l'identification de la lexie-vedette

- L'entrée lexicale qui est le nom du vocable français.
- Le numéro du vocable français pour distinguer les homonymes s'il y a lieu.
- La partie du discours (désormais PDD) du vocable français.
- Le numéro d'acception de la lexie à traduire. Ce numéro est identique à celui qui a été proposé dans l'article unilingue français de la même lexie. Par souci d'uniformité, nous gardons dans l'article bilingue le même dégroupement des sens des vocables que celui proposé dans nos articles unilingues. La consultation des différents articles n'en sera que facilitée pour l'utilisateur.
- La forme propositionnelle (ou structure actantielle) de la lexie française qui permet d'explicitier le nombre de ses actants. Par exemple, pour la lexie *châtier*^I (cf. section 3), la forme propositionnelle est *X châtie Y pour Z par W*. On pourrait également caractériser (ou « typer » dans la terminologie de la LEC) les différentes variables X, Y, etc. dans la forme propositionnelle, ce qui nous donnerait quelque chose comme : *l'individu X châtie l'individu Y pour l'acte Z par le mal W*.

2.2.2 Informations liées à l'identification de l'équivalent principal proposé

- L'équivalent traductif arabe qui est, bien entendu, une lexie et non un vocable.
- La racine de l'unité arabe, s'il y a lieu. La notion de 'racine' revêt, en effet, une importance particulière dans l'apprentissage de la langue arabe.
- Un *nota bene* pour préciser la nature de l'équivalent fourni (comme le montre l'article proposé dans la section 3) ou pour renvoyer l'utilisateur à une variante ou à un autre équivalent possible, selon le contexte. C'est également dans cette section qu'on attire l'attention de l'utilisateur sur la forme du féminin ou du pluriel non régulière pour les noms et les adjectifs.
- La PDD de l'équivalent. Généralement cette PDD correspond à celle de la lexie française, mais comme il n'y a pas toujours une équivalence parfaite entre les PDD françaises et arabes, l'équivalent peut parfois appartenir à une PDD différente ou particulière à la langue arabe.
- La forme propositionnelle de la lexie arabe.

2.2.3 Informations sémantiques

Les informations d'ordre sémantique sont d'une grande utilité pour l'utilisateur, d'autant plus que la majorité des équivalents traductifs proposés sont des quasi-équivalents ou des équivalents partiels. Ces informations seront classées dans deux

sections séparées. La première reprend la définition de la lexie telle qu'elle apparaît dans l'article unilingue arabe. La deuxième porte sur la ou les divergences sémantiques entre les deux lexies française et arabe utiles pour que l'utilisateur prenne réellement conscience du sens véhiculé par tel ou tel équivalent et puisse décider, avec discernement, de sa pertinence ou non dans un contexte donné.

2.2.4 Informations liées à la combinatoire syntaxique

- Une description du régime syntaxique de l'équivalent.
- Une mise en garde contre les éventuelles erreurs dues aux divergences, dans le comportement syntaxique, entre la lexie française et son équivalent arabe. Par exemple, les alternances syntaxiques peuvent être différentes d'une langue à une autre. C'est le cas de *châtier***I** et de son équivalent arabe (cf. section 3).

2.2.5 Informations liées à la combinatoire lexicale

Ces informations sont de deux types :

- Le premier type concerne les lexies reliées à l'équivalent par des liens paradigmatiques. Ces liens relèvent de la dérivation sémantique comme les synonymes, les converses, les antonymes, ainsi que les différents noms actanciels typiques de la lexie, la contrepartie verbale d'une lexie nominale ou, au contraire, la contrepartie nominale d'un verbe ou d'un adjectif, etc.
- Le deuxième type porte sur les lexies reliées à l'équivalent par des liens syntagmatiques. Il est ici question, bien entendu, des collocations (expressions semi-figées bipartites composées de deux éléments : la base (la lexie-vedette) et le collocatif [Hausman 1979, Mel'čuk 1996]). Contrairement au premier type d'informations lexicales, il ne s'agit pas, dans ce cas-ci, de répertorier simplement les lexies qui sont en lien syntagmatique privilégié avec la lexie arabe (les collocatifs), mais également de préciser leur sens et leur emploi, et de proposer des équivalents traductifs aux collocations françaises, le tout illustré d'exemples traduits en français.

Dans l'approche de la LEC, la modélisation de ces deux types de relations est réalisée au moyen des Fonctions Lexicales (FL). Faute d'espace, nous ne nous attarderons pas, ici, à la description du mécanisme qui les sous-tend; celui-ci est, à notre avis, suffisamment connu (Mel'čuk 1996, Mel'čuk *et al.* 1995 et Mel'čuk et Polguère 2007). Disons simplement que les FL spécifient l'ensemble des possibilités de combinatoire paradigmatiques (les dérivations sémantiques) et syntagmatiques (les collocations ou les cooccurrences lexicales restreintes) d'une unité lexicale donnée. Par exemple, pour *remercier*, la FL S_0 , qui renvoie à la nominalisation du verbe, a comme valeur *remerciement* et la FL Magn, qui dénote l'intensification, a comme valeur *énormément* et *infiniment*. Notons que, dans nos articles de dictionnaire, les FL qui sont jugées difficiles à saisir pour un usager non familier avec ce type d'encodage, sont vulgarisées par des formules qui en explicitent le sens. Par exemple, la FL Magn est explicitée par *De façon sévère* dans notre description de la lexie *châtier***I** (section 3).

2.2.6 Illustration par des exemples

Les exemples, dans l'article bilingue, sont deux sortes :

- Des exemples illustrant l'emploi des équivalents proposés dans la langue cible. Ceux-ci sont traduits en français, souvent littéralement, pour mieux aider l'utilisateur à saisir l'emploi des unités arabes.
- Un ou deux exemples de traduction avec l'équivalent proposé. Contrairement aux premiers, il s'agit ici d'exemples français traduits en arabe avec l'emploi de l'équivalent fourni.

Les figures 2 et 4 présentées dans la section 2.3.2 montrent de façon concrète comment est organisé le contenu informationnel de notre article bilingue. Il est à noter que celui-ci concerne surtout l'article bilingue avec équivalent direct en arabe, c'est-à-dire avec un équivalent qui est une lexie arabe décrite sur les plans sémantique, syntaxique et lexical. Or comme certaines lexies françaises n'ont pas d'équivalent direct en arabe, mais peuvent correspondre plutôt à des collocations ou à des expressions libres dans la langue cible, nous avons prévu un deuxième type d'article bilingue, simplifié, qui contiendra les données suivantes :

- Les informations liées à l'identification de la lexie-vedette française, c'est-à-dire le nom du vocable, son numéro en cas d'homonymie, sa PDD, le numéro de l'acception à traduire et sa forme propositionnelle qui permet d'explicitier le nombre de ses actants.
- L'équivalent traductif arabe qui sera, si la langue en dispose, une collocation ou une expression libre donnée comme solution au problème d'absence d'équivalence directe entre les deux langues.
- Un nota bene pour préciser la nature de l'équivalent fourni ou renvoyer l'utilisateur à l'article unilingue des lexies qui composent la collocation ou l'expression proposée comme équivalent.
- La forme propositionnelle de cet équivalent, lorsqu'il s'agit d'une collocation. Dans un article bilingue, il est en effet nécessaire d'explicitier la forme propositionnelle et le régime des collocations, lorsqu'elles sont proposées comme équivalent traductif. Cette précision est utile, non seulement parce que les collocations appartiennent à la langue que l'utilisateur ne maîtrise pas, mais aussi parce que la structure actantielle de la collocation est rarement déductible du régime de la base. Ainsi, les collocations verbales (Verbe + Nom), par exemple, ne prennent pas toujours les mêmes actants que la base nominale et peuvent même hériter des propriétés syntaxiques du collocatif.
- Le tableau de régime, s'il y a lieu, avec une collocation verbale par exemple.³
- Un ou deux exemples de traduction.

³ Dans ce deuxième type d'article bilingue, les champs réservés au tableau de régime de l'équivalent arabe, à sa forme propositionnelle et, même, à l'équivalent arabe peuvent rester vides. Dans un tel cas, on indiquera NA pour « non applicable ».

En résumé, notre BD comporte quatre types d'articles : des articles unilingues français, des articles unilingues arabes, des articles bilingues avec équivalent direct et des articles bilingues sans équivalent direct. Dans la section suivante, nous verrons comment cette BD est structurée.

2.3 Structure du Dico-FRAR

Le Dico-FRAR est une base de données dictionnaire. Chaque article qu'elle contient présente les données sous forme de texte. Cela dit, le texte lexicographique n'a pas besoin d'être compilé en un seul bloc. Avec l'outil informatique, il est possible de structurer l'article de dictionnaire sous différents blocs plus ou moins indépendants, ce qui facilite la consultation pour l'utilisateur et la gestion des données pour le lexicographe. L'outil informatique permet, par exemple, d'afficher plus aisément les relations sémantiques et lexicales entre les diverses unités décrites et de réviser rigoureusement le contenu informationnel fourni.

Ainsi, pour concevoir notre base de données français-arabe, notre choix s'est arrêté sur une base relationnelle⁴ gérée par le logiciel Access, qui fait partie de la suite bureautique de Microsoft Office. Pour chacun des quatre types d'articles que renferme notre BD, nous avons conçu des tables pour contenir les diverses données. Cette façon de faire, c'est-à-dire de créer des tables séparées, permet de gérer facilement la base et de l'enrichir, au besoin, en ajoutant des données ou des champs, sans avoir à la réaménager à chaque fois. Au total, notre BD est structurée en seize tables distinctes dont certaines sont reliées entre elles pour former nos quatre types d'articles. Nous nous limitons, dans ce texte, à la description de la structure de l'article bilingue avec équivalent direct.

2.3.1 Organisation des tables de l'article bilingue avec équivalent direct

Pour cet article dictionnaire, nous avons créé quatre tables réservées chacune à un type d'information particulier, à savoir : *Entree_bilingue_ED*⁵, *Relations_pardigmatiques_ar*, *Collocations_fr_ar* et *Vulgarisations_fl*.

La première, *Entree_bilingue_ED*, est la table principale de l'article bilingue avec équivalent direct. Les différents champs qu'elle renferme présentent les données caractéristiques essentielles de la lexie-vedette française à traduire et de son équivalent arabe comme le nom du vocable, le numéro de la lexie, la PDD, etc. Ces données sont directement importées des tables liées aux articles unilingues français et arabes. La table contient également les champs relatifs aux données sémantiques et syntaxiques associés à la lexie arabe, ainsi qu'aux exemples qui illustrent l'emploi de cette lexie dans un contexte de la langue cible.

⁴ Une base de données relationnelle est une base dans laquelle les données sont stockées dans des tables distinctes, selon le type d'information à enregistrer, mais qui sont reliées entre elles et peuvent être regroupées selon les besoins.

⁵ Pour éviter tout problème avec Access, les accents n'ont pas été utilisés dans les noms des tables et des champs créés.

La table *Relations_paradigmatiques_ar* renferme, comme son nom l'indique, les informations liées aux relations lexicales paradigmatiques de l'équivalent arabe fourni. Ces informations étant également déjà stockées dans une des tables de l'article unilingue arabe, nous les avons simplement sélectionnées et importées directement depuis la table en question.

La table *Collocations_fr_ar* est destinée à contenir les données relatives aux relations lexicales syntagmatiques de la lexie française et de son équivalent arabe.

Enfin, la table *Vulgarisations_fl* a été créée pour nous permettre de regrouper différentes valeurs d'une même fonction lexicale donnée.

Pour mieux expliquer l'utilité de cette dernière table, rappelons que l'application d'une seule fonction lexicale ou d'une seule vulgarisation de fonction lexicale à une lexie peut donner plusieurs valeurs qui sont plus ou moins différentes sur le plan sémantique. À titre d'exemple, les collocations suivantes *punition méritée/ juste/ équitable* sont toutes encodées, dans le formalisme de la LEC, avec la même fonction lexicale $Ver_{(punition)}$ ou sa vulgarisation *Que l'acte Z justifie*, mais séparées avec un point virgule. Toutefois, dans un article bilingue, il importe, selon nous, de distinguer plus clairement ces collocations pour proposer à l'usager les bons équivalents qu'il pourra employer sans équivoque. C'est pourquoi, bien qu'elles elles soient décrites avec la même fonction lexicale ou la même vulgarisation de fonction lexicale, nous avons départagé ces collocatifs en les citant comme des enregistrements distincts dans la table *Collocations_fr_ar* (voir figure 1). Avec la table *Vulgarisations_fl*, nous pouvons faire cette distinction sans perdre de vue le lien entre les différentes données.

Nom_vocable_fr	Num_lexie_fr	Nom_vocable_ar	Num_lexie_ar	Vulgarisation_fl	Collocatifs_arabes	Collocatifs_français
PUNITION	1	عقاب [^c <i>iqâb</i>]	0	Que l'acte Z justifie	عادل [^c <i>âdil</i>]	équitable, juste
PUNITION	1	عقاب [^c <i>iqâb</i>]	0	Que l'acte Z justifie	مستحق [<i>mustahḥaq</i>]	méritée

Figure 1. Extrait de la table *Collocations_fr_ar*⁶

Les diverses tables présentées ci-dessus sont reliées entre elles via les champs communs qu'elles contiennent, ce qui permet de garder l'unité de l'article dictionnaire et la cohérence des données qu'il renferme.

2.3.2 La mise en page

Après avoir identifié le contenu informationnel de nos différents articles et structuré à partir de là notre BD en diverses tables reliées entre elles, nous avons conçu un formulaire de saisie pour chacun de nos quatre types d'article. Ce formulaire, réalisé grâce aux liens établis entre les tables, permet d'enregistrer nos données de façon

⁶ Pour faciliter la lecture des unités lexicales arabes au lecteur de cet article, nous utilisons le système de translittération de l'arabe DIN-31635. Les lexèmes translittérés sont mis en italique entre deux crochets.

conviviale, en ayant toujours sous les yeux les informations qui sont réparties de manière automatique dans les différentes tables sous-jacentes. Le formulaire permet aussi de corriger ou modifier efficacement, s'il y a lieu, les données consignées sans avoir à ouvrir nécessairement toutes les tables où des modifications doivent être apportées, au risque d'en oublier quelques-unes. La figure 2 donne un aperçu de ce formulaire pour l'article bilingue avec équivalent direct.

Nom_Vocab	Num_Lexie	Pdd	Fp
CHÂTIER	ا	Verbe	X châtie Y pour Z par W

Nom_Vocab	كَب	Racine	كَب
Num_Lexie	2	NB	L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde.
Pdd	Verbe فعل		
Fp	كَب X ي Z ب W		

Informations sémantiques

Déf

L'individu X tenant l'individu Y responsable de l'acte Z qui va à l'encontre de certains principes auxquels X se réfère || X, qui a une autorité sur Y, fait subir un mal W à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.

Mise en garde

Prendre note que la composante sémantique liée à l'amélioration de la conduite de Y est plus forte avec la lexie arabe qu'avec la lexie française. De plus, le mal W ou la punition infligée avec "châtier" est sévère, ce qui n'est pas forcément le cas avec كَب2. C'est pourquoi.

Informations syntaxiques

Tr

X = I = N nominatif
Y = II = N accusatif oblig.
Z = III = N على N ل
W = IV = N ب

Mise en garde

Contrairement à "châtier", كَب2 a un seul régime syntaxique. La présence de l'actant Y (1er complément) est obligatoire dans un énoncé.

Exp.

كَبه المعلم على تصرفه السيئ تأديبا صارما.
كَبها لخروجها المتكرر دون إذن.
كَب الأب أولاده بالعتكرب.

Mise en garde

L'instituteur l'a sévèrement châtié pour son mauvais comportement.
Il l'a châtié pour ses fréquentes sorties sans permission.
Le père a châtié ses enfants avec une punition corporelle.

Informations lexicales

Dérivés sémantiques

Nom_Fl	Vulgarisation	Valeur
A0		تأديبي
A2	[Y] que X ~	تحت التأديب

Enr : 1 sur 8

Collocations

Vulgarisation de FL: De façon justifiée

Collocatifs français	Collocatifs arabes	Remarques et commentaires	Exemples
bien, justement	0	Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe كَب2 avec une collocation composée de l'adjectif possessif.	كَبه المعلم أحسن تأديب. Lit. 'L'instituteur l'a châtié d'un meilleur châtiment'

Enr : 1 sur 1

Enr : 1 sur 2

Exemple de traduction

1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes.
- من المفروض أن لا تؤكَب الأطفال قبل أن يتبين لهم أخطاءهم.

2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés.
- عَشْرًا: الفتناء محظورة و جريمة تُعاقب عليها بصفة قاسية في جميع المجتمعات.

Figure 2. Formulaire de saisie de l'article bilingue avec équivalent direct

Comme nous l'avons signalé ci-dessus, certaines données de l'article bilingue sont directement importées des tables relatives aux articles unilingues et ce, après la mise en équivalence des lexies françaises et arabes.

Pour procéder à cette mise en équivalence des unités lexicales des deux langues, nous avons créé un deuxième type de formulaire que nous avons appelé *formulaire de mise en correspondance des lexies fr-ar*. Celui-ci se présente comme suit :

Figure 3. Formulaire de mise en correspondance des lexies fr-ar

Avec ce formulaire, qui comprend une partie réservée à l'entrée française et une autre à l'entrée arabe, nous avons accès de manière parallèle aux diverses lexies décrites dans les deux langues. Pour bien identifier les lexies en cause, le formulaire indique le nom du vocable, son numéro, sa PDD, le numéro de la lexie et sa forme propositionnelle. Pour obtenir de plus amples détails sur la lexie française ou arabe choisies, le formulaire offre en outre un accès direct aux articles unilingues (par le biais des onglets *Consultation*

dictionnaire unilingue français et *Consultation dictionnaire unilingue arabe* en haut du formulaire dans la figure 3).

Le formulaire contient aussi d'autres types d'informations dont nous avons besoin pour établir la mise en correspondance entre les lexies françaises et arabes. Nous avons ainsi, à côté du numéro de la lexie, le champ *Selection* présenté ici sous forme de case dans laquelle on coche pour sélectionner les lexies françaises et arabes équivalentes. Il y a aussi le champ *Ajout* (noté *Entrée inscrite dans le dictionnaire bilingue* dans la fig. 3), présenté également sous forme de case, qui sera cochée automatiquement lorsque la lexie française sélectionnée sera ajoutée dans le dictionnaire bilingue, ce qui permet de distinguer les lexies traitées et celles qui ne le sont pas encore⁷. Cette modification automatique dans le champ *Ajout* est effectuée via une requête de Mise à jour que l'on exécute avec un des deux boutons de commande nommés *Mettre à jour le dictionnaire bilingue* (selon que le type d'article bilingue choisi est avec ou sans équivalent direct). Une fois que nos deux lexies équivalentes ont été sélectionnées et ajoutées à la table *Entree_bilingue* correspondante, deux autres boutons de commande permettent d'accéder au formulaire de saisie de l'article bilingue lié à cette nouvelle entrée pour compléter la rédaction de l'article (voir en bas du formulaire dans la figure 3).

Les formulaires présentés plus haut sont, comme leur nom l'indique, des formulaires de saisie et de mise en correspondance. Ils sont donc destinés au lexicographe/rédacteur et non à l'utilisateur (l'apprenant ou le traducteur) de la BD. Par conséquent, pour une consultation des articles sans risque de modification des données enregistrées, des pages web ont été créées. Ces pages permettent d'avoir accès aux fiches lexicographiques contenues dans la base, via le web ou l'intranet.

Pour chaque type d'article de dictionnaire (unilingue arabe, unilingue français, bilingue français-arabe), nous avons conçu diverses pages d'accès aux données, à savoir une page d'accès pour l'article unilingue arabe, une autre pour l'article unilingue français, deux pages pour l'article bilingue avec équivalent direct (version complète et version abrégée) et, finalement, une page pour l'article bilingue sans équivalent direct. Toutes ces pages peuvent être consultées à partir d'une page d'accueil qui sert de point d'entrée à Dico-FRAR. L'utilisateur peut ainsi choisir, selon ses besoins, l'un ou l'autre des types d'article de dictionnaire auquel il veut accéder, ainsi que le niveau de détail souhaité pour l'affichage des données. Celles-ci, classées en diverses rubriques, ne s'affichent que si l'utilisateur le désire.

Les figures suivantes présentent un exemple de la page d'accès de l'article bilingue avec équivalent direct dans sa version complète (fig. 4) et sa version abrégée (fig. 5). Le contenu de cet article est présenté en détail à la section 3.

⁷ Rappelons que notre traitement lexicographique bilingue est unidirectionnel du français vers l'arabe. Donc, une lexie française doit être traitée une seule fois. Par contre, une lexie arabe peut être proposée comme équivalent possible à plusieurs lexies françaises.

Article bilingue avec équivalent direct - version complète

Nom_vocable_fr:	Num_vocable_fr:	Pdd_fr:	
CHÂTIER	0	Verbe	
Num_lexie_fr:	Fp_fr:	Breve_def_fr:	Ex_fr:
1	X châtie Y pour Z par W	X... fait subir un mal sévère à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.	On craint d'être châtiés pour avoir commis ce péché.
Équivalent arabe			
Nom_vocable_ar:	Racine:	Nb:	
أَذَبَ	أَذَبَ	L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde.	
Num_vocable_ar:			
0			
Num_lexie_ar:			
2			
Pdd_ar:			
"Verbe" فعل			
Fp_ar:			
أَذَبَ Y X لـ Z بـ W			
Informations sémantiques			
Def_ar:	Mise_en_garde_1:		
L'individu X tenant l'individu Y responsable de l'acte Z qui va à l'encontre de certains principes auxquels X se réfère X, qui a une autorité sur Y, fait subir un mal W à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.	Prendre note que la composante sémantique liée à l'amélioration de la conduite de Y' est plus forte avec la lexie arabe qu'avec la lexie française. De plus, le mal W ou la punition infligée avec "châtier", est sévère, ce qui n'est pas forcément le cas avec أَذَبَ. C'est pourquoi, dans les contextes où châtier est employé pour punir un coupable, un criminel pour un méfait grave, il est préférable d'employer le verbe عاقب qui a un sens plus général et qui ne vise pas l'amélioration de la conduite de		
Informations syntaxiques			
Tr_ar:	Mise_en_garde_2:		
X = I = N nominatif Y = II = N accusatif oblig. Z = III = N على, N لـ W = IV = N بـ	Contrairement à "châtier", أَذَبَ a un seul régime syntaxique. La présence de l'actant Y (1er complément) est obligatoire dans un énoncé.		
Ex:			
أَذَبَ المعلم على تصرفه السيئ تأديبا صارما.	L'instituteur l'a sévèrement châtié pour son mauvais comportement.		
أَذَبَهَا لترحولها المتكرر دون إذن.	Il l'a châtié pour ses fréquentes sorties sans permission.		
أَذَبَ الأب أولاده بالقرص.	Le père a châtié ses enfants avec une punition corporelle.		
Informations lexicales : dérivés sémantiques			
Nom_fl_ar:	Vulgarisation_fl_ar:	Valeur_fl_ar:	
QSyn		عاقب 'punir'	
Informations lexicales : collocations			
Vulgarisation_de_Fl			
De façon sévère			
Collocatifs_français	Collocatifs_arabes	Remarques_et_commentaires	Exemples
durement, lourdement, rigoureusement, sévèrement	0	Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe أَذَبَ soit avec une expression libre composée de la préposition بـ + le masdar سدة.	أَذَبَ المعلم على تصرفه السيئ تأديبا صارما . Lit. 'L'instituteur l'a châtié d'un châtiment sévère pour son mauvais
Exemple de traduction			
1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes. - من المفروض أن لا يُؤذَب الأطفال قبل أن يتبين لهم أخطاءهم .			
2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés. - عَشْيَانِ المَحَارِم مَحظور و جريمة يعاقب عليها بشدة في جميع المجتمعات.			

Figure 4. Page d'accès à l'article bilingue avec équivalent direct. Version complète

La version détaillée présente l'article bilingue complet avec toutes les informations sémantiques, syntaxiques et lexicales. La version abrégée présente en gros l'équivalent arabe, sa PDD, sa forme propositionnelle, les collocations arabes de l'équivalent, ainsi que les équivalents des collocations françaises et l'exemple de traduction.

Article bilingue avec équivalent direct - version abrégée

Nom_vocable_fr		Num_vocable_fr		Pdd_fr	
CHÂTIER		0		Verbe	
Num_lexie_fr		Fp_fr		Breve_def_fr	
I.		X châtie Y pour Z par W		X... fait subir un mal sévère à Y afin que Y ne refasse pas Z et améliore sa	
				On craint d'être châtiés pour avoir commis ce péché.	
Nom_vocable_ar		Num_vocable_ar		Racine	
أَدَب		0		أَدَب	
Num_lexie_ar		Pdd_ar		Fp_ar	
2		فعل 'Verbe'		أَدَبَ يَأْدِبُ لِـ Z بِـ W	
				L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde.	
Information lexicale : collocations					
Vulgarisation_de_FL					
De façon justifiée					
Collocatifs_français		Collocatifs_arabes		Remarques_et_commentaires	
bien, justement		0		Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe أَدَبَ avec une collocation composée de l'adjectif comparatif أَحْسَنُ (mis à l'accusatif) + le masdar تَأْدِيبُ. On peut aussi employer le verbe أَدَبَ avec une expression libre comme, par exemple, "كما يجب" qui signifie "comme il se doit".	
				Exemples	
				أَدَبَهُ الْعَلِمُ أَحْسَنَ تَأْدِيبٍ . Lit. "L'instituteur l'a châtié d'un meilleur 'châtiment'".	
Exemple de traduction					
1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes.				- من المفروض أن لا تؤدب الأطفال قبل أن يتسنى لهم أخطاؤهم .	
2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés.				- عَشِيْرَاتُ الْعَرَبِ مَحْظُورٌ وَجَزَاءُ بِعَاقِبِ عَلَيْهَا بِشَدَّةٍ فِي جَمِيعِ الْمَجْتَمَعَاتِ.	

Figure 5. Page d'accès à l'article bilingue avec équivalent direct.
Version abrégée

3. Exemple d'article de dictionnaire bilingue dans le Dico-FRAR

L'exemple d'article de dictionnaire bilingue avec équivalent direct qui va suivre permettra au lecteur de se faire une idée concrète du traitement lexicographique que nous proposons dans le Dico-FRAR et la pertinence des informations qui y sont données⁸. La lexie-vedette sélectionnée pour cet exemple est *châtier***I**.

⁸ Comme il a été mentionné à la note 6, le recours à la translittération de l'arabe dans notre exemple d'article de dictionnaire n'a pour objectif que d'aider le lecteur à lire le présent travail.

ChâtierI Verbe

Forme propositionnelle

X châtie Y pour Z par W

أَدَّبَ ['addaba]

'Verbe' فعل

Racine

أَدَّبَ ['db]

Forme propositionnelle

أَدَّبَ X لِي Y بِ Z W

Informations sémantiques

Définition

L'individu X tenant l'individu Y responsable de l'acte Z qui va à l'encontre de certains principes auxquels X se réfère || X, qui a une autorité sur Y, fait subir un mal W à Y afin que Y ne refasse pas Z et améliore sa conduite en respectant les principes de X.

NB

L'équivalent proposé n'est pas un équivalent parfait. Voir la définition et la mise en garde ci-dessous.

Mise en garde

Prendre note que la composante sémantique liée à l'amélioration de la conduite de Y est plus forte avec la lexie arabe qu'avec la lexie française. De plus, le mal W ou la punition infligée avec "châtierI" est sévère, ce qui n'est pas forcément le cas avec أَدَّبَ. C'est pourquoi, dans les contextes où "châtierI" est employé pour punir un coupable, un criminel pour un méfait grave, il est préférable d'employer le verbe عاقب, qui a un sens plus général et qui ne vise pas l'amélioration de la conduite de Y. Voir l'exemple de traduction 2.

Informations syntaxiques

Tableau de régime

X = I = N nominatif

Y = II = N accusatif oblig.

Z = III = N على [°alâ], N لِي [li]

W = IV = N بِ [bi]

Mise en garde

Contrairement à "châtierI", أَدَّبَ a un seul régime syntaxique. La présence de l'actant Y (1er complément) est obligatoire dans un énoncé.

Exemples

أدبه المعلم على تصرفه السيئ تأديبا صارما.

[*'addaba-hu al-mu^oallimu ^oalâ tašarrufi-hi as-sayyi' ta'dîban šâriman.*]

أدبها لخروجها المنتكر دون إذن.

[*'addaba-hâ li-hurûji-hâ al-mutakarrir dûna 'idn.*]

أدب الأب أولاده بالضرب.

[*'addaba al-'abu 'awlâda-hu bi-aḡ-ḡarbi.*]

'L'instituteur l'a sévèrement châtié pour son mauvais comportement.'

'Il l'a châtié pour ses fréquentes sorties sans permission.'

'Le père a châtié ses enfants avec une punition corporelle.'

Informations lexicales

Dérivés sémantiques

Nom_FI	Vulgarisation_FI	Valeur_FI
QSyn		عاقب 'punir' ^o <i>[âqaba]</i> اقتص [Y=N] 'infliger à qqn un mal identique ou égal à celui causé' جزى [Y=N] 'donner ce qu'une personne mérite, donner son dû à qqn.' [<i>'iqtašša min N=Y</i>] [<i>jazâ</i>]
QAnti		سامح, صفح [Y=N] عفا [<i>sâmaḡa, šafaḡa ^oan</i>] عفا [Y=N], غفر [Y=N] 'pardonner' [<i>'afâ ^oan, ḡafara l</i>] أثاب, كافأ 'récompenser' [<i>'atâba, kâfa'a</i>]
S0		تأديب 2 [<i>ta'dîb</i>]
S3	Nom de Z	خطأ 'faute' [<i>ḡaḡa</i>], ذنب 'mal, méfait' [<i>danb</i>] مخالفة 'infraction' [<i>muḡâlafâ</i>]
S4/Sinstr	Nom de W	عقاب, عقوبة 'peine, punition' [<i>'iqâb, ^ouqûba</i>] تأديبي [<i>ta'dîbi</i>]
A0		تحت التأديب [<i>taḡta at-ta'dîb</i>]
A2	[Y] que X ~	
Able2	[Y] que l'on pourrait ~	يؤدب [<i>yu'addab</i>]

Qual2

[Y] que l'on est susceptible de ~

مخطئ 'fautif' [muḥṭi]; مخالف 'contrevenant' [muḥâliḥ]; جاني, مذنب 'coupable' [jâni, mudnib]

Collocations

Vulgarisation_FI	Collocatifs français	Collocatifs arabes	Remarques et commentaires	Exemples
De façon sévère	durement, lourdement, rigoureusement, sévèrement	0	Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe 2أدب [addaba] soit avec une expression libre composée de la préposition ب [b]+ le masdar شِدَّة/ صرامة / قسوة [šidda/šarâma/qaswa], soit avec une collocation composée du masdar 2تأديبا [ta'diban] (mis à l'accusatif indéfini parce qu'il fonctionne en tant qu'objet interne) + l'un des adjectifs suivants (mis également à l'accusatif indéfini) : شديدا/ صارما / قاسيا : [šadīdan/šariman/qâsiyan]. Pour plus d'informations, consulter le dictionnaire unilingue arabe.	أدبه المعلم على تصرفه السيئ تأديبا صارما. [addaba-hu al-mu'allimu ʿalâ tašarrufi-hi as-sayyi' ta'diban šariman.] Lit. 'L'instituteur l'a châtié d'un châtiment sévère pour son mauvais comportement'.
De façon justifiée	bien, justement	0	Ces collocatifs adverbiaux n'ont pas d'équivalents directs en arabe. On peut employer le verbe 2أدب avec une collocation composée de l'adjectif comparatif أحسن [aḥsana] (mis à l'accusatif) + le masdar 2تأديبا [ta'dib]. On peut aussi employer le verbe 2أدب avec une expression libre comme, par exemple, "كما يجب" [kamâ yajib] qui signifie 'comme il se doit'.	أدبه المعلم أحسن تأديبا. [addaba-hu al-mu'allimu 'aḥsana ta'dib.] Lit. 'L'instituteur l'a châtié d'un meilleur châtiment'.

Exemple de traduction

1- On ne doit pas châtier les enfants avant qu'ils puissent connaître leurs fautes.

من المفروض أن لا تؤدب الأطفال قبل أن نبيّن لهم أخطاءهم. [min al-mafrûḍ 'an lâ nu'addib al-'atfâl qabla 'an nubayyin lahum 'aḥṭâ'ahum]

2- L'inceste est universellement prohibé et même sévèrement châtié dans toutes les sociétés.

غشيانُ المَحارمِ مَحظور و جريمة يُعاقب عليها بشدّة في جميع المجتمعات. [gîšyânu al-maḥâhim maḥẓûr wa jarîma yu'âqib ʿalayha bi-šidda fî jamîʿ al-mujtamaʿât.]

4. Conclusion

Dans cet article, nous avons présenté la BD lexicographiques bilingue Dico-FRAR qui peut être proposée comme modèle pour le développement d'un dictionnaire informatisé français-arabe d'encodage. La finalité de cette base de données est multiple. Elle offre à l'utilisateur un traitement lexicographique qui répond, autant que faire se peut, à ses attentes en matière d'encodage ou de traduction. Elle propose également à celui-ci la possibilité de consulter non seulement les articles bilingues français-arabe, mais aussi les articles unilingues français et les articles unilingues arabes rédigés dans un métalangage français. De plus, le Dico-FRAR offre au lexicographe un environnement de travail agréable et convivial pour qu'il arrive à consulter ces articles unilingues, à comparer et à trier les données afin de rédiger des articles bilingues satisfaisants. Enfin, avec la variété et la richesse des données sémantiques, syntaxiques et lexicales des vocables qui seront décrits de façon complète dans les deux langues mises en parallèle, cette base de données, une fois enrichie avec un nombre élevé de vocables, pourra, en outre, être exploitée à des fins de recherche dans divers domaines. Nous pensons particulièrement aux domaines de la lexicographie, de la linguistique comparée, voire de l'enseignement et l'apprentissage d'une langue étrangère.

Remerciements

Je tiens à remercier Gaétane Dostie et Alain Polguère pour avoir lu et commenté une première version de cet article.

Références

- Alonso Ramos, M. (2001) « Construction d'une base de données des collocations bilingue français-espagnol », *Langages*, 143, pp. 5-27.
- Baker, C. F., *et al.* (1988) « The Berkeley FrameNet Project », in : *Proceedings of COLING-ACL'98*, Montréal, pp. 86-90.
- Bouchaddakh, S. (2009, en préparation) *Application des principes de la lexicologie explicative et combinatoire à la construction d'un dictionnaire bilingue informatisé français-arabe*, thèse de doctorat dirigée par G. Dostie et A. Polguère, Université de Sherbrooke.
- Bouchaddakh, S. (2008) « La définition dans les dictionnaires bilingues : problèmes de polysémie et d'équivalence interlangues », in : *Proceedings of the XIII EURALEX International Congress : EURALEX'2008*, Barcelone, pp. 807-812.
- Fellbaum, C. (ed.) (1998) *WordNet : An Electronic Lexical Database*, Cambridge (MA), MIT Press.
- Fillmore, C. *et al.* (2001) « Building a Large Lexical Databank Which Provides Deep Semantics », in : *Proceedings of the Pacific Asian Conference on Language, Information and Computation*, Hong Kong, pp. 3-26.
- Fontenelle, T. (1997) *Turning a Bilingual Dictionary into a Lexical Semantic Database*, Tübingen, Max Niemeyer Verlag.

- Hausmann, F. J. (1979) « Un dictionnaire des collocations est-il possible? », *Travaux de linguistique et de littérature*, 17(1), pp.197-195.
- Mel'čuk, I. A. *et al.* (1995) *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Duculot.
- Mel'čuk, I. A. (1996) « Lexical Functions : A Tool for the Description of Lexical Relations in the Lexicon », in : Wanner L. (ed.), *Lexical Functions in The Lexicography and Natural Language Processings*, Amsterdam/ Philadelphia, Benjamins, pp. 37-102.
- Mel'čuk, I. A. et Polguère (2007) *Lexique actif du français : L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, coll. "Champs linguistiques", Bruxelles, De Boeck & Larcier.
- Selva, T. *et al.* (2002) « Le DAFLES, un nouveau dictionnaire pour apprenants du français », *Proceedings of the Tenth EURALEX International Congress : EURALEX'2002*, Copenhague, Vol.1, pp. 199-208.
- Polguère, A. (2000) « Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French », *Proceedings of the Ninth EURALEX International Congress : EURALEX'2000*, Stuttgart, pp. 517-527.
- Polguère, A. (2003) « Étiquetage sémantique des lexies dans la base de données DiCo », *Traitement Automatique des Langues (T.A.L.)*, vol. 44, no 2. pp. 39-68.
- Verlinde, S. *et al.* (2005) « Dictionnaires électroniques et environnement d'apprentissage du lexique », *Revue de linguistique appliquée*, vol.10, no 2, pp. 19-30.

Dictionnaires consultés

- Al-Manhal : dictionnaire français-arabe* (2001) par S. Idris, Beirut, Dar al-Adab.
- Al-Mawrid : a modern English-Arabic dictionary* (1982) par M. Al-Baalbaki, Beirut, Dar al-ilm lil-malain.
- Dictionnaire arabe-français* (1860) par Kasimirski, Beyrouth, Librairie du Liban.
- Dictionnaire arabe-français, français-arabe : As-Sabil* (1983) par D. Reig, Paris, Larousse.
- Dictionnaire français-arabe* (1970) par M. Al-Najjari Bek, Beirut, Librairie du Liban.
- Dictionnaire arabe-français, français-arabe* (1999) par D. Reig, Paris, Larousse.
- Dictionnaire arabe-français-anglais : langue classique et moderne* (1967) par R. Blachère, M. Chouemi, et C. Denizeau, Paris, G.-P. Maisonneuve et Larose.
- Nouveau dictionnaire français-arabe* (1952) par E. Torkia, Caire, Bureau de la Publicité.