

Méthode sociolinguistique d'étiquetage du niveau de langue dans les dictionnaires bilingues (sur l'exemple d'un dictionnaire français-ukrainien)

Natalya Shevchenko
Université Lumière Lyon 2

This article describes a new sociolinguistic method in the labelling of unconventional units in bilingual dictionaries. This study was undertaken as part of the preparation of a French-Ukrainian dictionary of unconventional language.

1. Réflexions théoriques

La pratique des marques d'usage dans les dictionnaires français et ukrainiens remonte au XVII^e siècle.¹ Cependant, elle a pris de l'ampleur au début du XX^e siècle seulement et ce n'est que depuis une trentaine d'années que lexicographes et utilisateurs se sont penchés sur ce problème sérieusement.

Il est à noter que toutes les marques d'usage ne créent pas de difficultés. Ainsi, les étiquettes VIEILLI, DIAL. (dialectal), ANGL. AMER. (anglais américain), CUIS. (terme de cuisine), etc. sont univoques. Par contre, FAM. (familier), POP (populaire), ARG (mot d'argot) et autres marques de niveau de langue sont parfois contradictoires d'un ouvrage à un autre et ne reflètent pas forcément la réalité langagière.

La question essentielle est le choix des critères d'attribution de ces étiquettes/marques. Ainsi, Corbin (1980b, 1989) et Béjoint (1981, 1982) traitent déjà de ce problème dans certains de leurs travaux des années 1980. Ensuite, Paulin (1997), dans son étude sur la notion de non-standard dans les dictionnaires bilingues français-anglais suggère quelques propositions pour tenter d'améliorer la situation. Cependant, la question que posait Béjoint en 1981 sur la manière d'attribuer ces étiquettes, reste, 27 ans après, sans réponse précise: "[...] s'agit-il d'une décision individuelle ou collégiale, basée sur une étude quantitative ou non, utilisant l'avis de linguistes, d'experts, ou de simples informateurs, etc.?" (1981: 7). Les critères d'attribution des marques d'usage ne sont toujours pas déterminés tant dans les dictionnaires français ou anglais que dans les ouvrages russes ou ukrainiens.

L'absence ou la pénurie d'informations explicites à ce sujet dans les préfaces et les introductions des dictionnaires semble montrer le peu d'importance accordée à ces mentions par les lexicographes. Pourquoi les garder alors? Certains lexicographes ont effectivement essayé sans succès de les enlever de leurs dictionnaires², d'autres ont proposé leurs propres systèmes d'étiquetage des unités sub-standard (désormais SS), avec des astérisques ou encore des vignettes.

Ainsi, le système d'astérisques du *Grand Robert & Collins* (2003) va de un à trois en avertissant le lecteur étranger quant à l'usage réservé à telle ou telle unité. Tandis que les vignettes (□■, ■□, ■■) dans le *Dictionnaire du français non conventionnel* de Cellard et Rey (1991) "constituent indirectement, pour un lecteur mal familiarisé avec le français populaire, un 'mode d'emploi' du mot" (p. XVIII). Il est évident que dans les deux cas, le public visé est celui des locuteurs étrangers. Ce choix lexicographique est certainement basé sur le fait qu'un locuteur

¹ Voir les travaux de Horec'kyj (1963), Rey-Debove (1982), Matoré (1968), Glatigny (1990), etc.

² Ainsi, Gove décida en 1961 d'enlever un grand nombre de marques d'usage du *Webster's Third New International Dictionary* en motivant cette mesure par la suffisance des citations permettant reconstituer le caractère sub-standard des mots. Ceci valut à cette édition du dictionnaire beaucoup de critiques, en particulier dans la presse populaire. [Follett, cité dans Béjoint (1981: 8)].

natif n'ira pas chercher le registre du mot dans un dictionnaire sauf dans un but didactique. Néanmoins, ces deux systèmes ont une même faille: ils spécifient seulement le contexte d'emploi du mot tandis que le caractère social de la langue demeure totalement négligé. Le passage des mots d'un registre à un autre est visualisé par des vignettes (□■, ■□) mais reste obscur avec les astérisques.

Une autre remarque concerne la subjectivité du lexicographe et le caractère prescriptif des étiquettes dans ces deux ouvrages: si l'orientation prescriptive ou descriptive du dictionnaire et par conséquent de son métalangage est un choix méthodologique, la subjectivité reste un problème à réduire au minimum. L'objectif de la recherche décrite ci-dessous est de proposer une alternative empirique à cet étiquetage subjectif des unités SS dans les dictionnaires.

Il existe actuellement plusieurs méthodes de conception, d'élaboration et de production lexicographiques dont les trois principales sont:

- Travail avec des fiches lexicales formées des extraits de textes de la littérature, de la presse générale et spécialisée, en s'en remettant totalement à l'intuition linguistique et aux connaissances personnelles du lexicographe, avec l'aide d'informateurs natifs.
- Elaboration et exploration de vastes bases de données informatisées. Utilisation de corpus non seulement basés sur des sources écrites mais aussi sur des retranscriptions d'émissions radiodiffusées et d'oral spontané.
- Méthode sociolinguistique: enquêtes de terrain dans les deux langues.

La première méthode tombe en désuétude ou plutôt est remplacée par la linguistique de corpus pour la plupart des langues européennes. Cependant, elle reste toujours la plus appliquée en Ukraine.

L'utilisation d'un corpus, en particulier un corpus d'oral, ouvre au lexicographe beaucoup de possibilités inconnues auparavant et lui permet, entre autres, de rendre compte d'une langue véritablement naturelle et vivante. Cependant, l'état actuellement peu avancé des corpus de ce genre en français et surtout leurs tailles assez restreintes imposent des limites à une étude lexicographique. D'autant plus qu'en ukrainien ce domaine de recherche en est seulement au stade initial de développement, à savoir qu'à nos jours, il existe déjà un corpus électronique de la langue ukrainienne mais accessible seulement à ses créateurs: l'équipe de la Fondation Ukrainienne en Linguistique et Informatique auprès de l'Académie Nationale des Sciences.³

Or, ne pouvant disposer d'une base de données sur le langage SS ni en français ni en ukrainien, j'ai opté pour la troisième méthode, qui seule permet au lexicographe d'être au plus près du locuteur natif et de sa perception de la langue.

2. Propositions pratiques

Les deux objectifs principaux pour les enquêtes menées en France et en Ukraine étaient la réduction de la subjectivité lors de l'étiquetage et la récolte des néologismes SS inconnus des dictionnaires ukrainiens. Seul le premier fait l'objet du présent article.

Premièrement, j'ai dû construire mon corpus. Pour cela j'ai pris un dictionnaire monolingue du lexique général en français, le *Petit Robert* (PR, 2001) dans sa version électronique et j'en ai extrait toutes les unités marquées comme SS. Le choix du dictionnaire est justifié par les résultats des recherches antérieures (Shevchenko 2006). La seconde extraction a été effectuée de façon aléatoire pour obtenir à peu près 1000 items, nombre jugé maniable dans le cadre d'une thèse. Pour ce faire, un mot sur cinq a été retiré de la liste initiale de 5237 unités établie lors de la première extraction. Certaines unités signalées comme *vieillies* par des locuteurs natifs pendant les entretiens semi-directifs ont été supprimées. Les homonymes qui n'ont pas été retenus par la sélection aléatoire ont été rajoutés. En conséquence, la liste finale comporte 1047 items. Ce sont ces mots qui ont constitué par la suite la nomenclature de mon dictionnaire bilingue français-

³ Український мовно-інформаційний фонд НАН України.

ukrainien. Collocations, expressions et exemples composant chaque article du dictionnaire ont été également recensés ce qui a amené le nombre d'unités à 2179.

L'étape suivante était la composition d'un questionnaire: toutes ces unités, mots et expressions mélangés, y ont été incluses. Le but de ce questionnaire était obtenir les témoignages des Français sur l'étiquetage des unités SS de la liste. La terminologie des niveaux de langue a été expliquée au début du questionnaire pour éviter l'ambiguïté dans la compréhension des termes pour certains participants. Pour étiqueter les unités du questionnaire, le choix suivant de marques a été proposé aux sujets: *neutre*, *familier*, *grossier/vulgaire*, *argot*, *autre* et *inconnu*. L'étiquette *populaire* n'y figure pas, étant apparue comme inadaptée lors des recherches précédentes (Shevchenko 2006). La décision d'effectuer une enquête sur la totalité des unités de l'échantillon est motivée par la nécessité d'avoir un jugement de la part des locuteurs natifs sur chaque mot/expression de la liste et non pas sur un échantillon aléatoire ou autre (stratifié, d'une ou de plusieurs lettres, etc.). Cette nécessité résulte du fait qu'on ne pourrait jamais appliquer par analogie les étiquettes attribuées par des informateurs lors de l'enquête à des unités qui ne sont pas passées par le questionnaire.

L'enquête française a été menée principalement⁴ à Lyon et son agglomération entre novembre 2005 et mai 2006. Elle a été proposée aux sujets soit sous forme électronique soit sur papier et a pris plus de 6 mois étant donné la taille du questionnaire. Ce questionnaire très volumineux (60 pages) a été rempli par 31 personnes :⁵ 12 hommes et 19 femmes dont la plupart ont entre 20 et 35 ans, la moyenne d'âge pour les hommes étant de 41 ans et pour les femmes de 40 ans. Je n'ai pu éviter une certaine homogénéité dans les catégories sociales: presque tous mes sujets ont fait des études supérieures, la majorité étant du milieu universitaire.

A l'issue de cette enquête, chaque unité de l'échantillon s'est vu attribuer une étiquette ou deux selon que les opinions des sujets étaient divisées ou non. Ces étiquettes ont été comparées par la suite avec celles qu'on trouve dans le *PR*. Ainsi, sur 2179 mots et expressions du questionnaire, l'étiquetage du *PR* est conforme⁶ aux témoignages des sujets seulement dans la moitié des cas (54,8 %), le reste des étiquettes (45,2 %) étant en désaccord avec les attestations de terrain.

L'analyse détaillée de cette enquête ne peut pas être exposée dans cet article faute de place mais figurera dans ma thèse de doctorat, à paraître fin 2008. Je vais tout de même mentionner ci-dessous les trois cas les plus fréquents où le décalage entre le dictionnaire et les réponses des sujets manifeste tantôt le retard du *PR* par rapport à l'évolution de la langue, tantôt son anticipation. Ainsi, 440 mots et expressions (20%) signalés comme *neutres* par les locuteurs sont encore attestés comme *familier* dans le dictionnaire, tandis que 195 items (9%) encore *vulgaires* pour la plupart des sujets sont déjà étiquetés comme *familiers* dans le *PR*. L'étiquette *inconnu* attribuée à 163 unités (7,5%) marquées comme *familier* témoigne de la présence dans le dictionnaire de mots *démodés* ou *vieillis*, apparemment inconnus des locuteurs.

Cependant, il ne faut pas oublier que dans ce genre d'enquête, la représentativité quantitative et sociale de l'échantillon reste discutable. Le minimum statistique des réponses ainsi que la localisation sur un public universitaire me permettent de faire quelques observations importantes mais nécessitent tout de même un approfondissement et un élargissement de l'enquête.

Parallèlement à cette enquête, j'ai trouvé pour chaque unité du questionnaire français un équivalent neutre en ukrainien. C'est avec ces mots et expressions standard que le questionnaire ukrainien a été composé. Les sujets ukrainophones ont dû, par la suite, produire des équivalents SS des unités du questionnaire, leur registre ayant été préalablement déterminé par des locuteurs français. Bien entendu, les mots marqués comme *neutres* ou *inconnus* par les sujets

⁴ J'ai également eu quelques sujets de Toulouse et de Paris.

⁵ Le minimum nécessaire pour le domaine linguistique étant de 30: Javeau, C. (1985). *L'enquête par questionnaire: Manuel à l'usage du praticien*. Bruxelles: Éd. de l'Université de Bruxelles, Paris: les Éd. de l'Organisation.

⁶ Sur la conformité des étiquettes dans le *Micro Robert* (1971) voir l'article de Corbin (1980a).

francophones n'ont pas été inclus dans le questionnaire ukrainien. Quelques exemples du passage du français vers l'ukrainien via les témoignages des locuteurs natifs sont présentés dans le Tableau 1.

Unité SS en FR	Etiquette attribuée par des locuteurs français	Equivalent/définition neutre en FR (tiré(e) du <i>PR</i>)	Equivalent neutre en UKR	Equivalent SS en UKR trouvé par des locuteurs ukrainien
Crotte de bique	FAM/VULG	Interjection par laquelle on manifeste son impatience, son dépit	І треба ж такому (трапитись)!	Ні хрена собі! Ні фіга собі!
	FAM/VULG	Une chose sans valeur	Малоцінна річ, пусте	Фігня, хуйня
(Doubler) un gros-cul	FAM/VULG	Poids lourd	Велика вантажівка	Фура, тір
Être, marcher à voile et à vapeur	ARG/FAM	Être bisexuel	Бути бісексуалом	Бі, гей, грати на два фронти
Il a trop pompé	FAM	Il a trop bu (de l'alcool)	Напиватися (алкоголю)	Набратися, набухатися
Thune	ARG	Argent	Гроші	Філки, бабло, бабки
Tire	ARG	Automobile	Автомобіль	Тачка
Volaille	VULG	Groupe de femmes, de jeunes filles	Група жінок чи молодих дівчат	Шабаш, стадо
	ARG/VULG	Fille de mauvaise vie	Жінка легкої поведінки	Курва, повія, шльондра
	VULG	Femme, fille	Жінка, дівчина	Тьолка, діваха, кобіта

Tableau 1

Vu la taille de l'échantillon, il a été divisé en 3 parties ayant formé 3 questionnaires séparés en ukrainien: deux sur les unités FAM (chacun de 13 pages) et le troisième sur le reste du lexique (7 pages).

Cette enquête a été menée en Ukraine en juin – novembre 2006 à l'Université Nationale Ivan Franco à Lviv. Elle s'est déroulée, elle aussi, dans un milieu universitaire. La plupart des sujets, dont la majorité (88%) sont des filles de 20 ans en moyenne, étudient à la Faculté des Langues au Département de traduction anglais-ukrainien et de linguistique contrastive. Ainsi, 74 personnes ont rempli les 3 questionnaires dont 57 (33+24) sont des questionnaires sur le lexique FAM et seulement 17 questionnaires rendus sont sur l'autre lexique concerné (VULG, ARG, FAM/VULG, etc.).

Certes, ces résultats peu nombreux ainsi que l'homogénéité de l'échantillon des sujets ne permettent pas de tirer des conclusions statistiquement fiables. Cependant, le milieu des jeunes et surtout cette génération des jeunes Ukrainiens qui sont les premiers sur l'aire post-soviétique à parler leur langue maternelle dès la naissance, est très favorable à une étude sur le langage SS. Quant à l'aspect quantitatif des deux enquêtes (en France et en Ukraine), celles-ci pourraient être élargies si elles étaient menées au niveau national. Néanmoins, les résultats obtenus prouvent la faisabilité d'une telle recherche ainsi que son intérêt lexicographique afin de solutionner le problème de l'étiquetage.

3. Conclusion

Le recours aux témoignages de jeunes locuteurs natifs d'ukrainien sur place a aidé à compléter de quelques dizaines de néologismes SS les dictionnaires de la langue ukrainienne consultés.⁷ Cependant, ces mots inconnus des dictionnaires sont bien attestés dans les romans des auteurs ukrainiens contemporains.⁸

L'application de la méthode sociolinguistique m'a permis de proposer une des solutions éventuelles au problème de la catégorisation du lexique SS dans les dictionnaires. En outre, la taille de l'échantillon lexical analysé a permis d'appréhender les problèmes réels d'une étude identique qui serait réalisée au niveau national, avec un vocabulaire plus étendu dans les deux langues.

⁷ Regarder la liste des dictionnaires ci-dessous.

⁸ Les romans de Andrukhovych, Dnistrovyy, Dereš, Žadan, Žoldak, Pyrkalo, Pokal'čuk, etc.

Références bibliographiques

Dictionnaires

- Cellard, J.; Rey, A. (1991). *Dictionnaire du français non conventionnel*. 2nd ed. Paris: Hachette.
- Le Grand Robert & Collins* (cd-rom). Paris: Dictionnaires Le Robert, 2003.
- [PR]. *Le Nouveau Petit Robert de la langue française* (cd-rom). Paris: Dictionnaires Le Robert, 2001.
- Stavyts'ka, L. (2003). *A Short Dictionary of Ukrainian Slang*. Kyiv: Krytyka.
- Stavyts'ka, L. (2005). *A Dictionary of Ukrainian Slang*. Kyiv: Krytyka.
- Білоноженко, В. М. (et al.). (1999). Фразеологічний словник української мови. [Dictionnaire phraséologique de la langue ukrainienne]. 2nd ed. Київ: Наукова думка.
- Бусел, В. Т. (2004). Великий тлумачний словник сучасної української мови. [Grand dictionnaire de la langue ukrainienne moderne]. 2nd ed. Київ-Ірпінь: Перун.
- Венгренивська, М. А., Венгреновська, Г. Ф., Оратовський, Т. Б. (2000). Українсько-французький і французько-український фразеологічний словник. [Dictionnaire phraséologique ukrainien-français et français-ukrainien]. Київ: Генеза.
- Караванський, С. (2000). Практичний словник синонімів української мови. [Dictionnaire pratique des synonymes de la langue ukrainienne]. Київ: Українська книга.

Autres

- Béjoint, H. (1981). "Variété de langue et marques d'usage dans les dictionnaires". *Grazer Linguistische Studien* 15. 7-16.
- Béjoint, H. (1982). "Les marques d'usage dans les dictionnaires de la langue anglaise". *Trema* 7. 69-74.
- Corbin, D. & P. (1980a). "Le monde étrange des dictionnaires. Les "marques d'usage" dans le Micro Robert". *Bulletin du Centre d'Analyse du discours de l'Université de Lille III* 4, "Synonymies". 237-324.
- Corbin, P. (1980b). "Niveaux de langue: pèlerinage chez un archétype". *Bulletin du Centre d'Analyse du discours de l'Université de Lille III* 4, "Synonymies". 325-352.
- Corbin, P. (1989). "Les marques stylistiques/diastratiques dans le dictionnaire monolingue". Dans Hausmann, F. J. et al. (eds.). *Wörterbücher / Dictionaries / Dictionnaires. Ein internationales Handbuch zur Lexikographie / An International Encyclopedia of Lexicography / Enciclopèdie internationale de lexicographie* 1. Berlin, New York: Walter de Gruyter. 673-680.
- Glatigny, M. (dir.). (1990). *Les marques d'usage dans les dictionnaires (XVIIe – XVIIIe siècles)*. Lille: Presses Universitaires de Lille. Lexique 9.
- Hausmann, F. J. et al. *Dictionnaires. Encyclopédie Internationale de lexicographie* 1. Berlin, New York: Walter de Gruyter. 673-680.
- Норес'куї, Р. І. – Горещький, П. Й. (1963). *Історія української лексикографії*. Київ: Видавництво Академії наук Української РСР.
- Matoré, A. (1968). *Histoire des dictionnaires français*. Paris: Larousse.
- Paulin, A. (1997). "Analyse de la notion de non-standard dans les dictionnaires bilingues français-anglais". *Meta* 42 (1). <http://www.erudit.org/revue/meta/1997/v42/n1/>.
- Rey-Debove, J. (1982). "Le métalangage dans les dictionnaires du XVIIe siècle (Richelet, Furetière, Académie)". Dans *La lexicographie française du XVIe au XVIIIe siècle. Actes du Colloque International de Lexicographie* (9-11 octobre 1979). Wolfenbüttel: Herzog Auguste Bibliothek. 137-148.
- Shevchenko, N. (2006). "Problème de l'étiquetage des unités de bas niveau de formalité dans la lexicographie bilingue français-russe et français-ukrainien". Dans Corino, E.; Marellò, C.; Onesti, C. (eds.). *Proceedings of the XIIIth Euralex International Congress*. Torino, 6-9 September 2006. Alessandria: Edizioni dell'Orso. 575-581.