

Problème de l'étiquetage des unités de bas niveau de formalité dans la lexicographie bilingue Français-Russe et Français-Ukrainien

Natalya Shevchenko
Université Lumière Lyon 2
42, rue des Girondins,
69007 Lyon
France
N.Shevchenko@univ-lyon2.fr
Nat-mail@mail.ru

Abstract

The ultimate aim of the present research is the creation of a French-Ukrainian dictionary of unconventional language. This presentation focuses on the problem of stylistic labelling in dictionaries. The existence of a problem appeared with an analysis of four monolingual French dictionaries and two bilingual French-Russian ones. The absence of labelling criteria common to lexicographers causes subjectivity and therefore disagreement between authors in the usage labelling of lexical items. In order to remedy this, the author resorted to sociolinguistic methods, in particular to a survey by questionnaire.

1 Processus de démocratisation de la langue littéraire en France

1.1 Causes

Depuis ces dernières vingt-cinq années on observe un processus de démocratisation du dictionnaire français, c'est-à-dire un passage d'unités de bas niveau de formalité dans la norme littéraire. Ceci est motivé autant par des facteurs intra-linguistiques, en particulier par une variation du langage des garants de la norme littéraire (médias, enseignants, personnels politiques et de la culture), que, par exemple, par les changements ethnographiques dans le pays.

1.2. Conséquences

Une des conséquences de ce phénomène sociolinguistique est un effacement des frontières entre différents niveaux de langue et l'utilisation massive d'unités sub-standard dans le langage courant, puis dans la littérature et ensuite, leur apparition dans les dictionnaires généraux. La vitesse de ce processus et son développement très autonome sont à l'origine de difficultés pour le lexicographe, dont certaines sont étudiées ci-dessous.

2 L'état actuel de la lexicographie ukrainienne dans le domaine du sub-standard

Le présent article évoque la question de l'étiquetage des unités sub-standard dans les dictionnaires français monolingues et par conséquent, dans les dictionnaires bilingues, en parti-

culier français-russes. Pour la lexicographie ukrainienne, ce problème ne se pose pas encore, en raison du manque de dictionnaires bilingues français-ukrainiens, la lexicographie bilingue étant tournée vers l'anglais et le russe. De plus, il faut noter une présence très faible de dictionnaires du lexique ukrainien sub-standard sur le marché. A ce jour, il n'existe que trois ouvrages monolingues et aucun dictionnaire bilingue – et encore, sur ces trois dictionnaires, un seul est un travail de professionnel: *A Short Dictionary of Ukrainian Slang* de L. Stavys'tka (2003, 2005), les deux autres étant plutôt oeuvres d'amateurs: *Peršyj slovnyk ukrajins'kogo molodižnogo slengu* («Le premier dictionnaire ukrainien de l'argot des jeunes») de S. Pyrkalo (1999) et *Slovnyk žargonu zločenciv* («Dictionnaire de l'argot des malfaiteurs») de O. Povčenko (1996).

3 Analyse lexicographique

3.1 Dictionnaires généraux

Étant donné que la lexicographie ukrainienne se réfère toujours à sa homologe russe plus puissante et, sans doute plus avancée et mieux développée, ma recherche est passée par l'étude du problème de l'étiquetage dans la lexicographie russe. Pour ce faire, une analyse de six dictionnaires dont quatre sont français monolingues et deux sont bilingues français-russe a été menée. Voici la liste des dictionnaires sélectionnés pour cette étude: *Le Nouveau Petit Robert* (2000); le *Dictionnaire de la Langue Française Lexis* de Larousse (1999); *Le Dictionnaire Hachette Langue Française* (2000); *TLFi* (2002); Gak et Ganchina, *Nouveau dictionnaire français-russe* (2000); Griniova et Gromova, le *Dictionnaire du français familier et populaire* (1997).

Afin d'effectuer ce travail, 144 lexèmes sub-standard¹ répartis sur tout l'alphabet et choisis de façon aléatoire ont été sélectionnés. Les mots ont été recueillis auprès d'informateurs de différents âges et des deux sexes mais qui appartenaient toujours à des milieux cultivés. Les items recueillis ont été divisés en 42 séries synonymiques avec un mot ou une expression neutre en tête de chaque série. Ensuite, les étiquettes stylistiques de chaque unité de l'échantillon ont été relevées dans les dictionnaires sélectionnés.

La première déficience constatée pendant l'analyse est la discordance entre les marques stylistiques qui règne dans les dictionnaires. Ce qui démontre le problème de la «catégorisation» en ce sens où pour la même unité lexicale plusieurs étiquettes sont attribuées dans les différents dictionnaires parus simultanément. Ainsi, le mot *emmerder* est un exemple «prototypique», une même unité étant étiquetée de trois façons différentes selon les auteurs: FAM (*Hachette* et le *Petit Robert*), POP (chez Griniova et Gromova et dans le *Lexis* de Larousse) et VULG (dans le dictionnaire de Gak et dans le *TLFi*). Pour la suite de la recherche, j'ai dû faire une synthèse des données de six dictionnaires afin d'établir une étiquette commune qui soit unique pour chaque item (Tableau 1).² L'exemple de *fichu* est plutôt exceptionnel par l'una-

¹ Les unités étiquetées *grossier* ou *vulgaire* ainsi que les injures, ont été éliminées afin de ne pas mettre l'enquêtrice en difficulté.

² Juste une petite partie du tableau initial est présentée ici.

nimité de tous les lexicographes, et on observe que dans la plupart des cas cette homogénéité est absente. Or ici, avec une étiquette dominante la règle de la majorité a été appliquée (Tableau 1): ARG pour *blé*, POP pour *flouze*, FAM pour *godillot*, etc. En cas de domination de deux, voire trois étiquettes, celles-ci étaient toutes gardées, comme par exemple pour les mots *foutu*, *godasse*, *bidon*, etc.

La deuxième difficulté formelle rencontrée lors de cette étude concerne la correspondance entre l'étiquetage et le niveau de langue. En effet, parmi les dictionnaires français, seul le *Petit Robert* propose dans l'introduction les définitions explicites des étiquettes et des niveaux de langue auxquels elles sont rattachées. Les autres lexicographes ne le font pas, se limitant à la liste d'abréviations et aux brèves définitions qu'on peut trouver dans le corpus du dictionnaire. Pour les ouvrages bilingues, chez Gak et Ganchina (2000) on trouve les explications de la terminologie relative aux niveaux sub-standard adoptée dans l'ouvrage ainsi que des critères d'attribution des étiquettes de registre de langue. Il faut rappeler que ce dictionnaire est en quelque sorte une synthèse de plusieurs dictionnaires français, en particulier: le *Grand Robert de la langue française* en neuf volumes (1985), le *Dictionnaire du français non conventionnel* (1991) de J. Cellard et A. Rey, le *Petit Larousse illustré* (1992). D'où une certaine objectivité de ses auteurs dans l'étiquetage des unités sub-standard. Comme ils le mentionnent dans la préface, en cas de qualification hétérogène du même mot par des lexicographes français, les auteurs russes ont retenu la marque du niveau le plus élevé, c'est-à-dire, la marque POP à l'exclusion de l'argot, celle de FAM au lieu de POP et aucune marque³ si FAM se trouvait absente de l'un des dictionnaires français consultés (Gak 2000: XII). Cette méthode permet de bien suivre les changements lexicaux qui se produisent constamment en français et qui sont dus, entre autres, à la démocratisation de la langue.

3.2 Dictionnaires spécialisés

Evidemment, on peut trouver les définitions plus complètes de différents registres de langue dans les ouvrages de terminologie linguistique, tels que, par exemple, le *Dictionnaire de linguistique* (2001) de J. Dubois ou *Bol'shoj enciklopedičeskij slovar': Jazykoznanije* («Grand dictionnaire encyclopédique: Linguistique») de B. Jarceva (2000). Cependant, l'absence de cette information dans la plupart des dictionnaires français complique la perception du spectre stylistique de la langue non seulement pour un lecteur étranger, mais aussi pour un spécialiste de langue.

³ En tel cas, le mot est présenté comme stylistiquement neutre.

MOTS	GRITOVA, GROMOVA Dictionnaire du français familier et populaire 1997	GAK, GANCHINA Nouveau dictionnaire français-russe 2000	LAROUSSE Dictionnaire de la langue française LeNis 1999	Dictionnaire HACHETTE langue française 2000	Le nouveau PETIT ROBERT 2000	TLF ⁴ 2002	ÉTIQUETTE COMMUNE
<u>Appelé à disparaître</u>							
ficht	FAM	FAM	FAM	FAM	FAM	FAM	FAM
fautu	POP	VULG	POP	FAM	FAM	VULG	FAM/POP/VULG
<u>Argent</u>							
bk	POP	ARG	ARG	ARG	FAM	FIG ARG	ARG
flouze	POP	POP	POP	ARG	POP	ARG	POP
fric	POP	ARG	POP	POP	FAM	POP	POP
ossille	POP	POP	POP	POP	FAM	POP	POP
péze	POP	ARG	ARG	ARG	ARG	ARG, POP	ARG
pognon	POP	POP	POP	POP	FAM	POP et POP	POP
thune	ARG	ARG	ARG	ARG	ARG MOD	ARG, POP	ARG
<u>Chausure</u>							
godasse	POP	FAM	POP	FAM	FAM	POP	FAM/POP
godillot	FAM	FAM	FAM ou POP	FAM	FAM	POP	FAM
grolle	POP	FAM	POP	POP	FAM	ARG et POP	POP
pompe	POP	POP	POP	POP	FIG POP	ARG, POP	POP
<u>Chveux</u>							
gouffe	-	-	-	-	-	-	-
tl's	POP	POP	POP	FAM	FAM	ARG, POP	POP
<u>Chose factice</u>							
bidon	POP	FAM	POP	FAM	FAM	ARG et POP	FAM/POP

Tableau 1. Attribution d'une étiquette commune.

⁴ Ce mot est absent des dictionnaires mais signalé par des locuteurs comme usité dans le milieu des jeunes.

4 La notion de *populaire* dans les dictionnaires analysés

La troisième observation porte sur la notion de *populaire* dans les ouvrages lexicographiques analysés. Ainsi, ce terme est défini comme «propre au peuple, destiné au peuple» dans le *Hachette* (2000: 910) et «conforme aux goûts du peuple, répandu dans le peuple» dans le *Lexis* (1999: 1391). Seul le *Petit Robert* (2000: XXVIII) donne une définition plus explicite: «*Populaire*: qualifie un mot ou un sens courant dans la langue parlée des milieux populaires (souvent argot ancien répandu), qui ne s'emploierait pas dans un milieu social élevé».

Néanmoins, il apparaît clairement que ces définitions ne sont pas établies sur la base de critères communs au domaine de la lexicologie. L'absence de ces critères déterminés une fois pour toutes et reconnus par tous les lexicographes est le problème le plus important de l'étiquetage. Il en résulte un désaccord remarquable entre les différents dictionnaires. Ainsi, par exemple, la notion de «bas» niveau social qui transparait dans la définition de l'étiquette POP ne correspond plus à la réalité sociale d'aujourd'hui.

5 L'enquête de terrain

5.1 Composition du questionnaire

J'ai tenté de résoudre ce problème de détermination des critères d'inclusion en menant une enquête de terrain à Lyon en décembre 2002 – mars 2003. Pour ce faire, j'ai établi un questionnaire toujours sur la base de mes 144 mots sub-standard appartenant à quatre niveaux de langue différents. Les sujets devaient indiquer les circonstances dans lesquelles ils employaient le mot en choisissant une des six réponses proposées:

- 0 – Je ne connais pas ce mot (ou cette expression) ou bien je ne le connais pas dans ce sens.
- 1 – Je le connais, mais je ne l'emploie jamais.
- 2 – Je pourrais l'employer, seulement avec des amis proches et du même sexe que moi.
- 3 – Je pourrais l'employer avec des amis, des inconnus de mon âge.
- 4 – Je pourrais l'employer avec des amis, des inconnus de mon âge, des inconnus plus âgés.
- 5 – Je pourrais l'employer avec des amis, des inconnus de mon âge, des inconnus plus âgés, mes professeurs ou mes supérieurs (directeur, chef de service...).

5.2 L'échantillon des sujets

Les sujets avaient été sélectionnés selon trois critères binaires: âge, sexe et niveau d'instruction. Le premier oppose les locuteurs de 18-29 ans à ceux ayant plus de 30 ans, sans limite supérieure. Le dernier repose sur les études supérieures du sujet, manifestées par sa profession. Ainsi, la catégorie des «instruits supérieurs» est représentée dans sa majorité par des professeurs de l'enseignement secondaire, des universitaires et des étudiants d'université, tandis que la catégorie des «instruits non supérieurs» est formée en grande partie par des ouvriers de trois usines rhône-alpines (l'une, Valscius, producteur de carton à Saint-Etienne, les deux autres, Renault Véhicules Industriels, usines Renault de Vénissieux et de Saint-Priest) et des employés (en général, des femmes travaillant au service clientèle) de la Compagnie Générale des Eaux à Lyon.

5.3. Le déroulement de l'enquête

Outre le temps et la faisabilité, le souhait d'éviter au maximum le phénomène de l'*hypercorrection*⁵ a motivé, entre autres, le choix d'un questionnaire sans présence de l'intervieweur.

5.4. Les résultats

168 exemplaires du questionnaire de 6 pages ont été distribués; 162 exemplaires correctement remplis ont été dépouillés. Les résultats de cette enquête m'ont permis d'évaluer deux aspects:

D'une part, le degré de familiarité observé chez 162 sujets *hommes* et *femmes* appartenant à deux classes d'âge (de 18-29 ans et de 30 et plus) et présentant des niveaux d'instruction différents.

D'autre part, la correspondance des étiquettes par rapport à l'emploi effectif des mots.

L'analyse des résultats de l'enquête a révélé:

D'un côté que le facteur *sexe* a une incidence forte ($2(1)=206,65$; $p<0,001$) sur le degré de familiarité observé chez les sujets: les sujets masculins déclarant une plus grande familiarité que leurs homologues féminins, notamment les *hommes instruits supérieurs de 30 ans et plus*. Etant donné que la majorité des sujets de cette catégorie est représentée dans cette enquête par des enseignants, on peut en conclure que chez les universitaires l'*hypercorrection* est quasi inexistante.

De l'autre côté, l'analyse de l'emploi par les informateurs *instruits supérieurs* et *instruits non supérieurs* des mots marqués *populaire* de l'échantillon a montré que leur usage est comparable d'une classe sociale à l'autre. Pour cette analyse tous les items étiquetés POP dans les dictionnaires consultés ont été prélevés de l'échantillon. Ainsi, 61 unités ont été dégagées. Comme la marque *populaire* est censée refléter, selon les définitions données ci-dessus, l'appartenance sociale des mots (page 5), c'est leur emploi par des personnes *instruites supérieures* ou *instruites non supérieures* qui pourrait confirmer cette appartenance. Ainsi, les réponses des sujets de ces deux catégories ont été comparées. La différence non significative ($t(60)=0,9$; $p<0,05$) de leurs réponses (23,66% chez les *instruits supérieurs* contre 23,97% chez leurs opposants) révèle le nivellement des usages du lexique populaire par ces catégories sociales. De ce fait, on peut conclure que les locuteurs francophones ne font plus la différence entre ces deux niveaux de langue ou encore que la notion linguistique de *populaire* n'est plus fondée sur une caractéristique d'ordre social. L'étiquette POP est donc inadaptee. Par conséquent, elle ne sera donc pas appliquée dans mon dictionnaire.

Cependant, il est à souligner que deux caractéristiques de cette enquête, son caractère local et la dimension restreinte de l'échantillon du lexique sub-standard incitent à une grande prudence avant de généraliser. Une étude plus large qui prendrait en compte d'autres unités

⁵ Le terme d'*hypercorrection*, introduit par W. Labov, reflète une envie de s'exprimer plus correctement à l'écrit, lors d'un questionnaire ou d'une interview ou encore en présence d'un interlocuteur d'une langue dominante.

lexicales et d'autres catégories socioprofessionnelles pourrait aboutir à des résultats différents.

6 Pourquoi ne pas employer des méthodes de la linguistique de corpus?

Evidemment, l'application des méthodes de la linguistique de corpus pourrait aussi résoudre le problème de catégorisation du lexique sub-standard et réduire la subjectivité des auteurs des dictionnaires toujours présente lors de l'étiquetage. Pourtant, actuellement, les bases de données françaises, dont la plus grande est *Frantext*, ne sont pas vraiment adaptées à un travail sur le lexique de bas niveau de formalité, et ceci pour deux raisons:

Premièrement, une forte prédominance des sources écrites et en particulier des textes de la littérature classique par rapport aux sources orales dans ces corpus.

Deuxièmement, concernant les bases phonétiques conçues à partir d'enregistrements, on peut remarquer que leurs tailles, malgré une extension régulière, restent assez restreintes.

Pour le russe, le corpus le plus important *Nacional'nyj korpus russkogo jazyka* («Corpus national de la langue russe») est en phase d'élaboration. Et pour l'ukrainien, d'après O. Dems'ka-Kulčyc'ka (2003), spécialiste de linguistique de corpus, les aspects théoriques de ce domaine de la recherche lexicologique ainsi que ses outils pratiques sont encore peu connus en Ukraine.

7 Conclusion

Or, il faudra attendre encore quelques années avant que la méthodologie de corpus soit applicable à l'étude de l'étiquetage stylistique dans le domaine de la lexicographie bilingue français-russe et français-ukrainien. En attendant, il sera utile d'effectuer une recherche sociolinguistique comprenant des enquêtes de terrain en France, comparable à celle décrite ci-dessus mais au niveau national et avec le plus grand nombre possible d'unités françaises sub-standard.

Bibliographie

A. Dictionnaires

Dubois, J. (dir.) (1999), *Dictionnaire de la Langue Française Lexis*. Paris, Larousse-Bordas.

Gak, V. G., Ganchina, K. A. (2000), *Nouveau dictionnaire français-russe*. Moscou, Rousski Yazyk.

Griniova, E. F., Gromova, T. N. (1997), *Dictionnaire du français familier et populaire*. Moscou, Tsitadel.

Fouquet, E. (dir.) (2000), *Le Dictionnaire Hachette Langue Française*. Paris, Hachette Livre.

Rey-Debove, J., Rey, A. (dirs.) (2000), *Le Nouveau Petit Robert*. Paris, Dictionnaires Le Robert.

Trésor de la Langue Française informatisé (2002), <http://atilf.atilf.fr/tlf.htm>.

B. Autres

Dems'ka-Kulčyc'ka, O. (2003), 'Bazovi ponjattja korpusnoji lingvistyky', *Ukrajins'ka mova* 1(6), pp. 40-43.