

# BASE DES DONNÉES DES MARQUEURS DISCURSIFS COMME INSTRUMENT DE L'ANALYSE LINGUISTIQUE (SUR L'EXEMPLE DE *ZNATCHIT*)

NADEZHDA BUNTMAN

Université d'État Lomonossov de Moscou, Russie

**Résumé.** Notre recherche est centrée sur le fonctionnement des marqueurs discursifs (MD) dans les textes écrits. La sémantique du même MD peut varier en fonction de son statut morphologique, qui souvent est difficile à déterminer, et son rôle dans une énonciation concrète. Les principes de la base de données des MD ont été élaborés lors du projet de recherche sous la direction de Anna A. Zalizniak en 2016-2018. Le corpus de cette base de données contient les textes du sous-corpus parallèle russe-français et français-russe qui fait partie du Corpus National de la langue russe. Les textes littéraires des trois siècles (XIX<sup>e</sup>, XX<sup>e</sup> et XXI<sup>e</sup>) traduits et corrigés par les spécialistes représentent un riche matériel pour une analyse linguistique des unités langagières, notamment des marqueurs discursifs. Diverses variantes de traduction précédées par le travail d'interprétation qu'effectue le traducteur littéraire professionnel aide à préciser la sémantique et l'emploi du MD dans la phrase. Le MD, en tant qu'unité de langue, est annoté en deux langues ; langue-source et langue-cible. Les paramètres d'annotation qui semblent pertinents pour cette analyse sont, entre autres, la position initiale, intermédiaire ou finale du MD, la négation, la réplique dialogique. Aussi, des MD peuvent être attachés à des mots adjacents (*tak znatchit* or *a znatchit*) qui se trouvent tagués dans une annotation. Cette méthode où l'analyse sémantique s'appuie sur les possibilités du corpus ouvre, à notre avis, de nouvelles perspectives pour la recherche des phénomènes linguistiques.

**Mots-clés :** marqueurs discursifs, base de données, corpus parallèle, analyse sémantique

## INTRODUCTION

Cet article a pour but de montrer les possibilités d'une base de données en tant qu'instrument pour l'analyse sémantique des MD. L'objet de notre recherche est le fonctionnement des marqueurs discursifs (MD) dans le texte écrit sur l'exemple de *znatchit*. En fonction de leur statut morphologique - qui souvent est difficile à déterminer - et de leur rôle dans l'énonciation, la sémantique des

mêmes MD peut varier. Les principes de la base des données des MD russes qui sera notre instrument de recherche étaient élaborés dans le cadre du projet sous la direction de Anna A. Zaluzniak en 2016-2018.

Le corpus de cette base contient les textes du sous-corpus parallèle russe-français et français-russe du corpus national de la langue russe. Le sous corpus russo-français contient en ce moment vingt-deux textes intégraux de volume différent, dont la majorité sont des romans (F. M. Dostoïevski. *Crime et châtiment* (1866), I. S. Tourguenev. *Premier amour* (1860), M. A. Boulgakov. *La Garde blanche* (1924), S. A. Alexievitch. *La fin de l'homme rouge* – parties 1, 2) (2013) et autres), y sont aussi inclus des récits (A. S. Pouchkine. *La fille du capitaine* (1836), L. I. Oulitskaïa. *Sonietchka* (1992) et des nouvelles (A. P. Tchekhov. *Une banale histoire* (1889). Pour harmoniser le corpus y ont été rajoutés les textes russes non-fiction traduits en français tels que *la Constitution de la Fédération de Russie* (1993), un essai de Berdiaev (*De suicide* (1931)) et un roman policier (A. Vaïner, G. Vaïner. 38, *Rue Petrovka* (1975)).

Le sous-corpus franco-russe contient 15 oeuvres avec presque la même proportion de belles lettres (de H. de Balzac. *Le colonel Chabert* (1832) à F. Beigbeder. *99 francs* (2000) et d'essais (G. Genette : *Figures I, II* (1959-1965/1969), M.-P. Rey *L'effroyable tragédie : une nouvelle histoire de la campagne de Russie* (2012)).

Ces textes littéraires de langue russe datant de trois siècles (XIX, XX et XXI), traduits et corrigés par des spécialistes – traducteurs, rédacteurs et correcteurs – ensuite alignés par les étudiants du département français de la faculté des langues étrangères de l'Université Lomonossov de Moscou représentent un riche matériel pour l'analyse linguistique de toutes les unités de langue y compris les marqueurs discursifs. Les variantes de traduction du même MD dans divers contextes, le travail préliminaire de l'interprétation du traducteur professionnel qui précède le choix d'une unité de langue, permettent de préciser la sémantique et l'emploi des MD dans la phrase.

Il est à noter que dans la base des données supra corpus (BDSC) il existe des originaux pour lesquels il y a plusieurs traductions en français (*Oblomov* de Gontcharov 2 traductions, *Le Nez* de Gogol 4 traductions, *Le Manteau* de Gogol 4 traductions, *La mort d'Ivan Ilitch* de Tolstoï 2 traductions) et 2 traductions en russe de *La difficulté d'être* de Jean Cocteau.

## ASPECT THEORIQUE

L'intérêt pour les MD dans la communauté linguistique se fait voir à partir des années 90, ce qui s'explique par de nombreuses recherches antérieures ciblées sur l'analyse du discours (Redeker, 1991 ; Fraser, 1996 ; Lenk, 1998).

Les marqueurs discursifs sont d'une grand utilité pour identifier l'idéolecte, pour rendre dans la traduction le style individuel d'un sujet parlant, pour préciser le sens de l'énonciation. Ainsi, ont été entrepris les tentatives de cataloguer les

MD russes (Kiseleva et Paillard, 2003 ; Baranov, Plungyan et Rakhilina, 1993) et de se focaliser sur le fonctionnement des MD particuliers.

## METHODE

Le sous corpus russo-français et franco-russe est toujours en train de s'enrichir par de nouveaux textes, c'est pourquoi il nous semble pertinent d'indiquer le nombre de tokens dans chaque sens (tableau 1). Le fait que le nombre de tokens dans les traductions dépasse celui des originaux s'explique par l'existence de plusieurs traductions pour le même original.

Tableau 1 Données sur la quantité des tokens de la BDSC pour 25.05.19

Sens	Tokens dans les textes originaux	Tokens dans les textes traduits	total
Russe-français	1659239	2284898	3944137
Français-russe	649245	509033	1158278

Quelle est donc la procédure de tri des données textuelles et l'analyse qui s'ensuit ? D'abord, il est à signaler que l'unité lexicale qui est l'objet de notre étude est polyvalente. Ainsi, le lexème *znatchit* en tant qu'une forme de la 3 personne du singulier du présent du verbe, c'est-à-dire, jouant le rôle du prédicat dans une phrase avec un sujet (dans la plupart des cas exprimé par des pronoms démonstratifs neutres *eto*, *to*) n'était pas pris en considération. Formellement, le tri préalable se fait par le token, *znatchit* quand le lexème est précédé par une virgule ce qui élimine l'existence d'un substantif ou des pronoms *eto* ou *to*.

La deuxième étape prévoit la constitution des monoéquivalences (ME). Le terme *monoéquivalence* signifie une annotation bilingue d'un MD russe et de sa traduction comprenant plusieurs éléments : la variante discursive (VD) dans l'original et son cotexte immédiat, les propriétés du MD dans l'original, le MD dans la traduction et son cotexte, les propriétés du MD dans la traduction.

Tableau 2 Exemple d'une monoéquivalence dans le sens russe-français

ДПН 1866	Заперта; и дверь окрашена заново; отдается, <b>значит</b> , внаем».	<b>значит</b> < Parenth > < direct speech >	Il est fermé; et la porte a été repeinte, il est <b>donc</b> à louer. »	<b>donc</b> < Pres > < direct speech >
-------------	---	---	---	--

La polyéquivalence est un groupement de deux ME ou plus qui a une seule annotation dans l'original (MD dans l'original, ses propriétés et son cotexte) et un nombre respectif d'annotations dans la traduction. Le statut intermédiaire de *znatchit* entre connecteur logique et MD est visible dans la perception différente de la phrase par deux traducteurs, le premier a compris le lexème comme un MD

et a choisi de l'omettre sans porter préjudice au texte. L'autre le comprend comme connecteur et garde la même relation logique 'cause-conséquence' entre deux propositions.

Tableau 3 Exemple d'une polyéquivalence dans le sens russe-français

N	Cotexte de l'original	Propriétés de MD			
			№	Traduction	MD en traduction
17737	уши зачешутся – <b>значит</b> , к ДОЖДЮ,	значит < direct speech >  < Parenth >			
			9131	les oreilles grattent avant la pluie,	<b>zero</b>
			9141	Quand ce sont les oreilles, <b>alors</b> il pleuvra;	<b>alors</b>

Une des possibilités de lever l'ambiguïté du statut du MD est d'élargir le cotexte de l'énonciation pour prélever dans l'entourage linguistique des indications explicites qui puissent préciser la sémantique de MD. Ainsi, dans la BDSC, il existe une option de l'extension du cotexte. Ci-dessous sont présentées deux phases de l'extension possible du cotexte de MD

N	Texte original	traduction	
95	– <b>В полицию</b> , значит, зовут, <b>в контору</b> .	– On vous convoque à la police, au bureau...	<u>montrer le cotexte</u>

#### Cotexte élargi (1)

N	original	traduction
94	– Из какой конторы?..	– De quel bureau?
95	– В полицию, значит, зовут, в контору.	– On vous convoque à la police, au bureau...
96	Известно, какая контора.	On sait bien de quel bureau il s'agit.

#### Cotexte élargi (2)

N	original	traduction
92	Тот молча протянул ему серую, сложенную вдвое бумажку, запечатанную бутылочным сургучом.	Celui-ci lui tendit en silence un papier gris plié en deux et cacheté avec de la cire à bouteilles.
93	– Повестка, из конторы, – проговорил он, подавая бумагу.	– Un avis, du bureau, prononça-t-il en lui remettant le papier.

N	<b>original</b>	<b>traduction</b>
94	– Из какой конторы?..	– De quel bureau?
95	– В полицию, значит, зовут, в контору.	– On vous convoque à la police, au bureau...
96	Известно, какая контора.	On sait bien de quel bureau il s'agit.
97	– В полицию!.. Зачем?..	– À la police!... Pourquoi?
98	– А мне почему знать.	– Est-ce que je sais, moi!

La combinatoire des paramètres qui sont établis en fonction du but de la recherche et dont le nombre peut augmenter, vise à répondre à la demande des chercheurs en linguistique et des traducteurs. Ainsi, pour l'analyse sémantique du MD, il est important de connaître sa position dans la phrase (initiale, finale ou intermédiaire), s'il est employé dans une phrase exclamative, interrogative ou affirmative, s'il est accompagné de négation ou d'un autre MD ou de marqueurs de temps et de modalité. Le tableau 4 représente les options de recherche qui peuvent être combinés dans les deux langues et selon plusieurs paramètres. Ainsi, pour trouver les variantes de la traduction de *znatchit* qui ne figurent pas dans les dictionnaires bilingues, il suffit d'afficher les traductions évidentes et puis d'appuyer sur *Exclure*.

Tableau 4 Rubriques de la base des données des MD

Œuvres	Traductions	MD dans l'original <input type="checkbox"/> Exclure	MD dans la traduction <input type="checkbox"/> Exclure	Propriétés de la MD dans l'original <input type="checkbox"/> Exclure	Propriétés de la MD dans la traduction <input type="checkbox"/> Exclure
<u>Propriétés du CM</u> <input type="checkbox"/> Exclure	<u>Éléments de la MD dans l'original</u> <input type="checkbox"/> Exclure	<u>Éléments de la MD dans la traduction</u> <input type="checkbox"/> Exclure	Lexème dans l'original <input type="text"/> Lexème dans la traduction <input type="text"/> <u>Aide</u>	Fragment du cotexte de la MD dans l'original <input type="text"/> <input type="checkbox"/> Exclure	Fragment du cotexte de la MD dans la traduction <input type="text"/> <input type="checkbox"/> Exclure

Aujourd'hui, dans le sens russe-français sont annotées 265 monoéquivalences avec le MD *znatchit*, dans le sens français-russe, 138. L'unité linguistique – dans notre cas, le mot *znatchit* – est annotée dans la langue cible et dans la langue-source selon les paramètres que nous considérons importants.

## LE STATUT DE MD DANS LES DICTIONNAIRES

Le mot *znatchit* est traité dans divers dictionnaires de la langue russe comme conjonction de coordination exprimant la conséquence ou la conclusion d'énoncés immédiatement antécédents (valeur logique ; introduit une phrase, un membre de phrase ou une proposition). Dans le dictionnaire de

Ozhegov, (Ozhegov, 1949-1992) seule la troisième acception précise le statut morphologique de ce lexème, les deux premiers étant difficile à déterminer :

1. ВВОДН. Следовательно, стало быть, выходит. *Вещи собраны, ты, з., уезжаешь?* 2. **значит** (**значило, значило бы, будет значить**). Употр. в знач. связки «это», «это есть». *Простить з. забыть.* 3. союз. И поэтому, следовательно. *Тучи собираются, з. будет дождь. Ты сердиться, з. ты неправ.* • **А значит (и значит)**, союз, то же, что **значит** в 3 значении). (Ozhegov, 1949-1992)

Le dictionnaire de Ushakov met l'accent sur le rôle de *znatchit* en tant que MD, en précisant qu'il s'emploie plutôt dans le style parlé et qu'il assume généralement la fonction de l'incise dans une phrase (Ushakov, 1935-1940).

Tatiana Efrémova, auteur du dictionnaire, le qualifie comme conjonction ou unité incise en y rajoutant des synonymes *sledovatelno, stalo byt, poetomou, vledstviie etogo, v silou tgo tchto* ne donne comme traduction que *donc* (Scerba, 2004), tandis que celui de Gak (Gak, 2004), bien que ce soit dans un autre sens (français-russe) et on est obligé de voir la sémantique de MD à partir du lexème *donc*, en fait ressortir plusieurs nuances, entre autres, le retour à la pensée interrompue, méfiance, surprise.

## POSITION DES MD

Pour voir les différentes variantes de position de *znatchit* il suffit de choisir une rubrique de Position. Dans la position initiale (PrePos) sur 265 occurrences dans les originaux russes nous avons repéré 157 MD PrePos, or, souvent, dans les traductions le lexème correspondant se place après le sujet et le prédicat.

Пороки были свойственны гениальным людям в такой же мере, как и добродетели ... – <b>Значит</b> , ты наполовину гений, – соглашалась моя жена, – <i>ибо</i> пороков у тебя достаточно...	<b>значит</b> < Parenth > < PrePos > < direct speech >	Le vice a toujours été le propre des génies au même titre que la vertu. – Tu es <b>donc</b> à moitié génial, convenait ma femme, car ce ne sont pas les vices qui te manquent...	<b>donc</b> < direct speech
--	--	---	-----------------------------------

Dans le sens français-russe, il s'avère aussi que la position initiale prédomine :  
 Sur 138 occurrences de *znatchit* 128 sont en PrePos :

J'irais <b>donc</b> en enfer, mon fils, si tu n'accomplissais pas mes volontés.	<b>donc</b> < Pers1 > < direct speech >	<b>Значит</b> , я пойду в ад, сын мой, если ты не исполнишь моей воли.	<b>значит</b> < Fut > < Parenth > < PrePos > < Pers 1 > < direct speech >
---	--	---	--

Pour ce qui est de l'emplacement de *znatchit* en position finale (PostPos), sur 265 dans le sens russe-français il n'y a que 14 occurrences. En position finale se rapproche en sa fonction dans le discours du tic verbal, ce qui explique le fait que parmi ces occurrences se trouvent quatre ZERO, c'est-à-dire, quatre cas d'omission de l'unité dans la traduction. MD PostPos sert à noter un discours affectif ou expressif dans une situation donnée, exprimée ou suggérée par le contexte.

10027	ВЭМ 1975	– Ну, может быть, вы хоть момент этот запомнили, про что говорилось, когда с ведром- то выходили? – спросил с надеждой Жеглов. Сосед пожал плечами: – Да он уже кончился, матч, <b>значит</b> .	<b>значит</b> < Parenth > < PostPos > < direct speech >	– Peut-être que vous vous souvenez de ce que le journaliste disait quand vous êtes sorti avec le seau, hein? demanda Jeglov, plein d'espoir. Le voisin haussa les épaules: – Ah oui, il était déjà fini, ce match.	<b>ZERO</b>
-------	-------------	---	---	---	-------------

Souvent, MD PostPos est signe d'un idéolecte d'un personnage, ce qui permet au traducteur d'être plus libre avec la traduction du MD. Dans l'exemple ci-dessous, l'équivalent de MD se déplace en PrePos, se transforme en une phrase avec sujet et prédicat et à la fin de la proposition est rajouté le point d'exclamation absent dans l'original :

ВЭМ 1975	Да мало ли что сделать можно, когда в кармане копейка живая шевелится!.. – Любишь деньги, <b>значит?</b> – прищурился горбун. – Люблю, – сказал я с вызовом.	<b>значит</b> < Parenth > < Interrog > < PostPos > < direct speech >	Je pourrai me marier, acheter une vache, une paire de cochons à engraisser. D'ailleurs, on peut faire pas mal de choses quand on a des espèces sonnantes et trébuchantes en poche! – <b>Je vois que</b> t'aime le fric?! dit le bossu en plissant les yeux. – Je l'aime, répondis-je avec défi.	<b>voir</b> < Pres > < Interrog > < Exclam > < direct speech >
-------------	---	---	---	---

Dans la position intermédiaire dans le sens russe-français sont chiffrés 49 occurrences :

9164	БММ 1940	– Доктор, –[...] он, <b>значит</b> , действительно болен?	<b>значит</b>	– Docteur, [...] il est <b>donc</b> réellement malade? <b>donc</b>
			< Parenth >	
			< Interrog >	
			< PrePos2 >	
			< direct speech >	

Et dans le sens français-russe en sont recensées neuf :

V80 1872	«Mais vous êtes <b>donc</b> bien pressé d'arriver à Hong-Kong?»	<b>donc</b>	<b>значит</b>
		< Interrog >	< Parenth >
		< Pers2 >	< Interrog >
		< direct speech >	< direct speech >
			< Pers 2 >

## PERSPECTIVES DE RECHERCHE

Dans la position intermédiaire du MD il serait intéressant de voir le statut morphologique des mots qui encadrent le MD pour analyser ensuite leur corrélation.

Plusieurs chercheurs proposent d'établir une liste, un répertoire des MD, pourtant, il n'en existe pas de liste exhaustive, puisque, d'un côté, les MD varient selon leur emploi dans le discours oral ou écrit, et, de l'autre côté, leur emploi varie selon l'auteur, le style et la prise de position du traducteur russe.

Aussi, certains MD sont liés aux mots-satellites (*tak znatchit* ou *znatchit tak*) qui sont marqués en italique dans l'annotation.

32	ДПН 1866	<i>Так, значит</i> , решено уж окончательно:	<b>значит</b>	<b>Ainsi donc</b> ,	<b>donc</b>
			< Parenth >	tout est décidé définitivement;	
170	БММ 1940	– <i>Так, значит</i> туда? – спросил мастер,	<b>значит</b>	– <b>Alors</b> , je dois aller là-bas? dit le Maître <b>alors</b>	
			< Interrog >		
			< Parenth >		
			< Complex >		

La méthode qui comprend une analyse sémantique et les possibilités du corpus ouvrirait, à notre avis, de nouvelles perspectives de l'analyse des phénomènes linguistiques.

## BIBLIOGRAPHIE

- Aijmer, K. (2002) *English Discourse Particles: Evidence from a Corpus*. Amsterdam: J. Benjamins.
- Baranov A. N., Plungyan V. A., Rakhilina E. V. (1993) *Putevoditel po diskursivnym slovam russkogo yazyka*. Moscow: Pomovsky and partners (ed.).
- Fraser, B. (1996) Pragmatic markers. *Pragmatics*, 6 (2): 167–190.
- Fraser, B. (2005) Towards a theory of discourse markers. In K. Fisher (ed.) *Approaches to Discourse Particles* (pp. 189–204). Amsterdam: Elsevier Press.
- Kiseleva, K. L., Paillard, D. (2003) *Diskursivnyye slova russkogo yazyka: kontekstnoye varirovaniye i semanticheskoye edinstvo*. Moscow: Azbukovnik.
- Lenk, U. (1998) Discourse markers and global coherence in conversation. *Journal of Pragmatics*, 30: 245–257.
- Redeker, G. (1991) Linguistic markers of discourse structure. *Linguistics*, 29: 1139–1172.
- Viktorova, E. Y. (2014) Pragmaticheskaya polifunktionalnost diskursivnykh slov. *Vestnik Voronezhskogo universiteta. Lingvistika i mezhkulturnaya kommunikatsiya*, 2: 18–22.

## DICTIONNAIRES

- Efremova, T. F. (2000) *Tolkovyy slovar*. Disponible sur <http://rus-yaz.niv.ru/doc/dictionary-efremova/index.htm> [consulté le 6 novembre 2019].
- Gak, V. G., Ganshina, K. A. (2004) *Novyy frantsuzsko-russkiy slovar*. Moscow: Russkiy yazyk-Media.
- Ozhegov, S. I. et Shvedova, N. Y. (1949-1992) *Tolkovyy slovar*. Disponible sur <https://ozhegov.slovaronline.com/> [consulté le 6 novembre 2019].
- Scerba, L.V., Matusevich, M. I., Vorontsova, T. P. (2004) *Bolshoy russko-frantsuzskiy slovar*. Moscow: Russkiy yazyk-Media.
- Ushakov, D. N. (1935-1940) *Tolkovyy slovar*. Disponible sur <https://ushakovdictionary.ru/> [consulté le 6 novembre 2019].

## DATABASE OF DISCURSIVE MARKERS AS TOOLS OF LINGUISTIC ANALYSIS

**Abstract.** The subject of our research is the functioning of discursive markers (MD) in written text. Depending on their morphological status – which is often difficult to determine – and their role in the enunciation, the semantics of the same MD can vary. The principles of the Russian MD database were developed as part of the project led by Anna A. Zalizniak in 2016-2018. The corpus of this database contains the texts of the parallel Russian-French and French-Russian sub-corpus of the Russian National Corpus. These literary texts of three centuries (the 19<sup>th</sup>, 20<sup>th</sup> and 21<sup>st</sup>), translated and corrected by specialists, represent a rich material for the linguistic analysis of such language units as discourse markers. The different translation variants preceded by the work of the professional translator's interpretation make it possible to specify the semantics and the use of MD in the sentence. The language unit is annotated in the target language and in the source language according to the parameters that we consider important. Among these parameters are: initial, intermediate or final position of MD, negation, dialogic replica and others. Also, some MDs are linked to adjacent words (*tak znatchit* or *a znatchit*) that are

marked in the annotation. The method which includes a semantic analysis and the possibilities of the corpus would, in our opinion, open up new perspectives in the analysis of linguistic phenomena.

**Key words:** discursive markers, database, parallel corpus, semantic analysis

**Nadia Buntman** (maître de conference, traductrice) Université Lomonossov de Moscou, faculté des langues étrangères. Elle s'intéresse aux corpus linguistiques, analyse du texte littéraire, analyse du discours et la littérature française contemporaine. Email: nabunt@hotmail.com