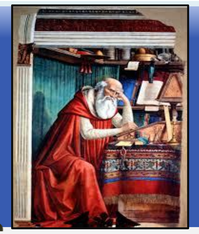


Outils informatiques – de simples instruments de recherche ?

Olga Inkova, chargée de cours, Unité du russe, Olga.Inkova@unige.ch, **Nataliia Popkova**, doctorante, Nataliia.Popkova@etu.unige.ch

Les ***connecteurs*** sont des mots qui établissent *un lien logique ou chronologique* entre *deux états de choses*. Plusieurs mots peuvent remplir cette fonction (ex. *et*, *mais*, *de même que*) aux niveaux propositionnel, illocutoire ou textuel.

Base de données des connecteurs (**BDC**) est un outil informatique créé à partir du sous-corpus parallèle du Corpus National de la langue russe NKRJA www.ruscorpora.ru. Cet outil est conçu spécialement pour étudier les connecteurs russes et français:

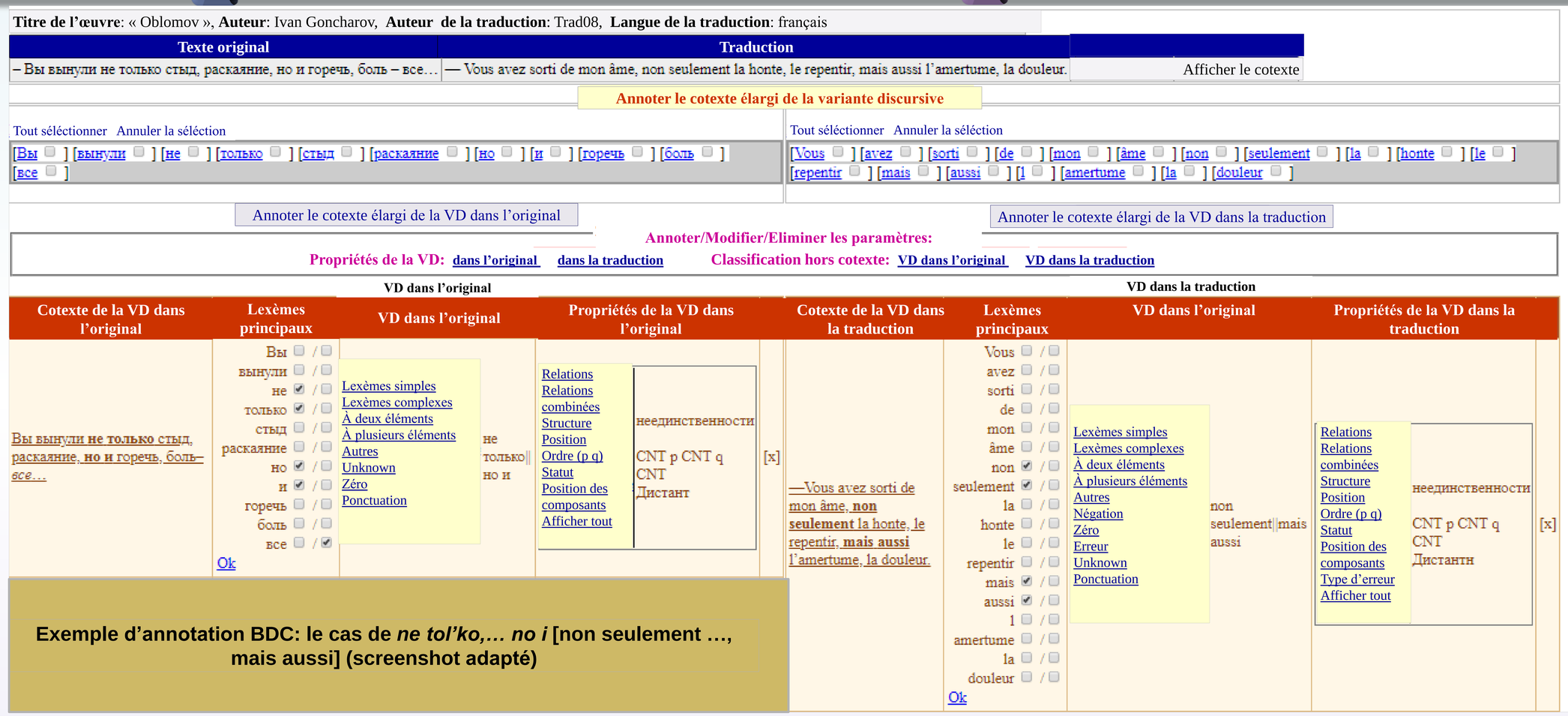
http://a179.ipi.ac.ru/PublicLingvoConnectors/main.aspx.

Projet de recherche russe-suisse «Corpus-based contrastive study of connectors in Russian» (FNS n° IZLRZ1_164059)» est le fruit de la coopération entre le Centre de recherches «Sciences informatiques et gestion» de l'Académie des Sciences de Russie) et de l'Unité de Russe de la Faculté des Lettres.

Variante discursive (VD) est la forme dans laquelle le connecteur apparaît dans le texte. Elle est annotée avec son **cotexte** dans l'original et son équivalent dans la traduction. Lors de l'annotation manuelle la VD et son cotexte sont attribués **à deux types de clusters** (classification à facettes).

1er type : la VD est classée selon sa structure formelle. Cela permet de fixer toutes les combinaisons possibles de connecteurs et des éléments à l'intérieur d'un connecteur sans préjuger de leurs statuts dans le système de la langue.

2e type : les propriétés du cotexte de la VD sont annotées selon les paramètres élaborés pour l'étude du fonctionnement des connecteurs.



Exemple d'annotation BDC: le cas de *ne tol'ko,… no i* [non seulement …, mais aussi] (screenshot adapté)

Conclusions

Base de données des connecteurs:

- ❑ constitue non seulement un outil d'analyse digitale, mais contribue aussi à produire, classer et enregistrer de nouvelles informations sur les connecteurs;
- ❑ aide à résoudre certains problèmes théoriques et contradictions de traitement liés à la variabilité formelle des connecteurs et à la polyfonctionnalité des items capables de remplir la fonction de connecteur.
- ❑ contribue à décrire la taxonomie de relations logico-sémantiques marquées par les connecteurs et leur possible co-occurrence.

- ✓ Les données statistiques représentatives disponibles dans la base de données permettent de vérifier les hypothèses des linguistes.
- ✓ Les liens hypertextuels permettent d'accéder directement aux données en question.

Références

- ❑ Kruzhkov M. *Supracorpora Databases as Corpus-based Superstructure for Manual Annotation of Parallel Corpora*, Moreno Ortiz A. and Pérez-Hernández C. (eds), CILC2016. 8th International Conference on Corpus Linguistics. EPiC Series in Language and Linguistics, 2016. Vol 1. P. 236-248.
- ❑ Inkova O., Popkova N. *Statistical data as information source for linguistic analysis of Russian connectors*, Informatics and its applications, 3, 2017, pp. 123-131.
- ❑ Zatsman I., Nuriev V., Inkova O. *The Construction of Classification Schemes: Methods and Technologies of Expert Formation*, Automatic documentation and mathematical linguistics, 51/1, 2017, pp. 27-41.