



HAL
open science

Mémoire de traduction [MT] : approche paradigmatique

Maria Zimina, Serge Fleury

► **To cite this version:**

Maria Zimina, Serge Fleury. Mémoire de traduction [MT] : approche paradigmatique. Équivalences : Revue de traduction et de traductologie, 2018, Des unités de traduction à l'unité de la traduction, 45 (1-2), pp. 259-278. hal-02514639

HAL Id: hal-02514639

<https://u-paris.hal.science/hal-02514639v1>

Submitted on 22 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Mémoire de traduction (MT) : approche paradigmatique

Maria Zimina

MCF, CLILLAC-ARP EA 3697, Université Paris Diderot, Sorbonne Paris Cité

Serge Fleury

MCF, CLESTHIA EA 7345, Université Sorbonne-Nouvelle

Résumé

En TA-TAO, la génération actuelle de MT (mémoires de traduction) est essentiellement axée sur l'utilisation de tables de correspondances phrastiques au format XML. Cette représentation statique a ses limites. La démarche textométrique permet une approche paradigmatique de la représentation de correspondances entre les textes source et cible. Cette approche peut être mobilisée au cours de la traduction, post-édition et révision afin de mieux cerner des constructions repérables au niveau sous-phrase et de pouvoir les intégrer dans le discours environnant. La textométrie repose principalement sur l'observation des variations de fréquences d'unités textuelles dans les différentes parties d'un ensemble de textes. Cette spécificité de la démarche textométrique permet d'utiliser un modèle de données particulier qui peut représenter des liens unissant les textes source et cible à plusieurs niveaux (mots, syntagmes, phrases, cadres discursifs, etc.). Dans ce travail exploratoire, nous montrons comment une MT dotée d'un moteur textométrique peut élargir les possibilités d'exploration de ressources traductionnelles, y compris dans le cas de textes source et cible comparables. C'est le concept de *ressource traductionnelle incrémentale* qui constitue le fil conducteur de cette réflexion : traduire signifie d'abord comprendre le système dans lequel s'inscrit le texte à traduire, puis construire son double, et c'est la démarche textométrique qui peut outiller ces processus.

Mots-clés : Mémoire de Traduction (MT), modèle de données *Trame/Cadre*, traduction assistée par ordinateur (TAO), textométrie, ressources traductionnelles incrémentales

Key words: Computer-Assisted Translation (CAT), incremental translational resources, textometrics, *Thread/Frame* data model, Translation Memory (TM).

1. Unités de traduction et informatisation

L'étude des mécanismes de la traduction humaine et l'analyse des propriétés de la notion d'équivalence traductionnelle sont fondamentales pour la recherche d'éléments formels qui se correspondent dans les textes source et cible. Les questions liées à la définition et la systématisation des unités de traduction sont largement étudiées en traductologie : « *Le concept d'unité est capital en traductologie car il repose sur les notions d'équivalence, d'analyse et de hiérarchisation des processus. Il est à la fois un instrument d'observation et d'analyse dans la pratique du commentaire de traduction et un moyen de rapporter à une grille structurée et évolutive les résultats dégagés grâce à ces investigations* » (Ballard 1993 : 260).

Au cours de son développement, la traductologie a développé un large éventail de méthodes qui permettent de repérer des *unités d'équivalence de traduction*. Un survol de la littérature consacrée à ce sujet montre qu'il s'agit d'un concept d'une grande complexité et il n'existe pas d'unanimité en ce qui concerne la définition ou la typologie d'unités (Zimina 2004a : 52-55). Le niveau d'analyse des unités exige généralement une très grande flexibilité. Il est conditionné par le type de discours, le genre de texte, le niveau et le registre de langue, le style de l'auteur. Dans le domaine de l'alignement automatique de corpus de textes traduits, le côté pratique résultant de l'exploitation des données bi-textuelles stockées sur support informatique entraîne une préférence pour les unités facilement repérables sur le plan de traitements automatisés, qui ne s'imposent pas forcément au plan théorique. On cherche à éviter le morcellement « inutile » de textes source afin de permettre à la machine de repérer correctement les traductions. Pour concilier les deux perspectives, la solution optimale vient d'une séparation conceptuelle très nette entre les unités textuelles dénombrables, issues des traitements informatisés, et l'identification des équivalences par analyses successives de corpus informatisés. La recherche de ces équivalences est guidée par le jugement humain et prend l'ancrage sur les unités textuelles qui circulent en corpus. Pour illustrer ces principes, nous nous intéresserons à l'utilisation de ressources traductionnelles existantes dans la production de nouvelles traductions.

2. Le concept de mémoire de traduction (MT)

Dans ce travail exploratoire, nous souhaitons enrichir la réflexion sur le concept de mémoire de traduction (MT), en anglais *Translation Memory (TM)*, au delà des correspondances phrastiques axées sur l'alignement des segments. A l'heure actuelle, EAGLES (Expert Advisory Group on Language Engineering Standards) propose la définition suivante de la MT : “*a multilingual text archive containing (segmented, aligned, parsed and classified) multilingual texts, allowing storage and retrieval of aligned multilingual text segments against various search conditions*” (EAGLES 1996 : 140). C'est la norme TMX (*Translation Memory eXchange*), basée sur le standard XML et gérée par LISA (*Localization Industry Standards Association*), qui est utilisée pour stocker et échanger les données collectées.

Intégrée au poste de travail du traducteur, une MT est une base de données linguistiques qui mémorise les segments extraits automatiquement de textes déjà traduits (phrases, éléments de liste, légendes, notes, titres, etc.) pour une réutilisation dans les nouvelles traductions. Le programme informatique qui génère la MT fait correspondre à

chaque phrase source un segment cible enregistré dans une table de correspondances. Ensuite, dans un système TA-TAO, la traduction enregistrée est systématiquement proposée à l'attention du traducteur dès qu'un segment identique ou proche apparaît dans le texte en cours de la traduction. Au-delà des correspondances exactes, c'est la recherche approximative (ou *fuzzy search* en anglais) qui est utilisée pour trouver des correspondances partielles avec un certain taux de recouvrement. Ce mécanisme est utilisé dans les systèmes dotés de plusieurs types de moteurs de traduction automatique (à base de règles, statistiques, hybrides, neuronaux, comme dans *Systran Enterprise Server 8*, cf. Figure 1).

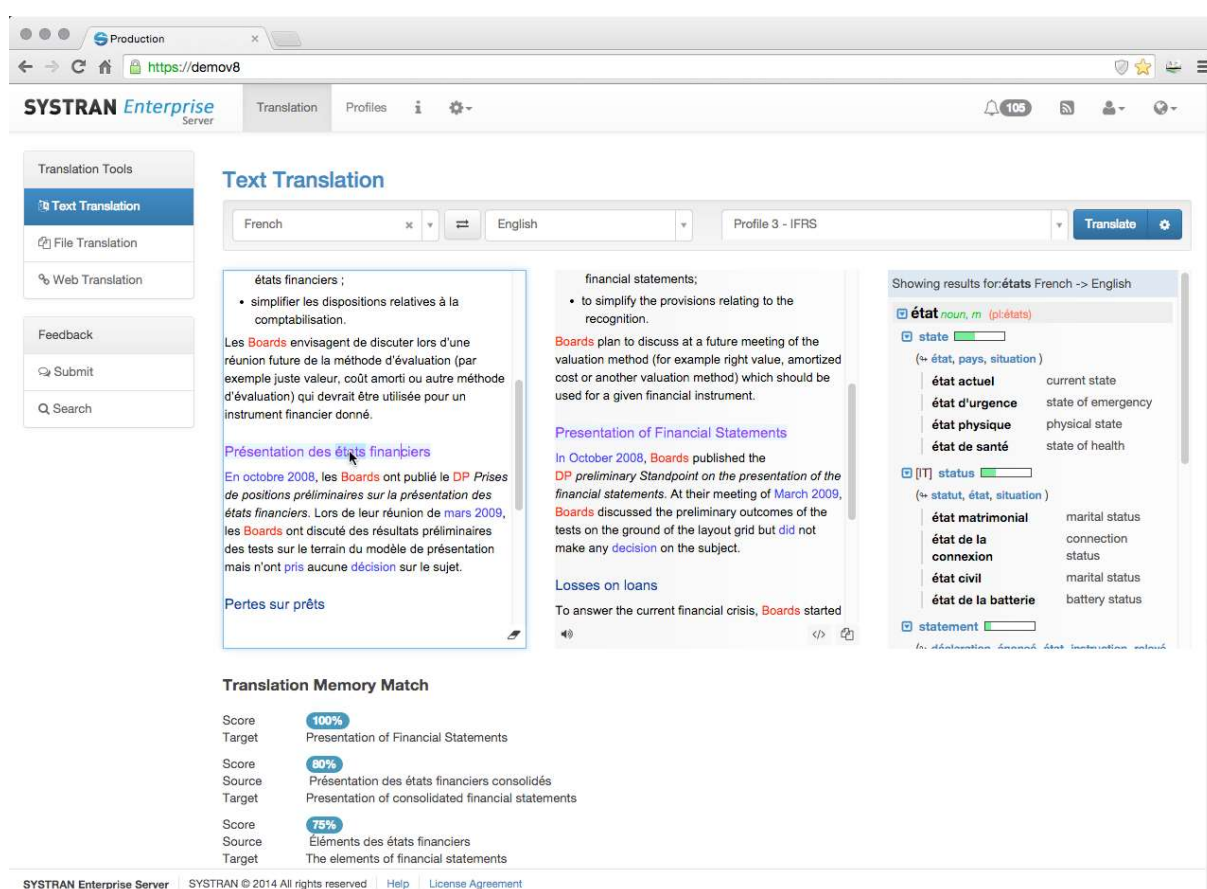


Figure 1 : Traducteur du texte *Systran* avec MT et Dictionnaire (capture d'écran disponible sur le site <https://www.translationsoftware4u.com/enterprise-global.php>)

Lorsque les MT sont axées sur les alignements phrastiques, les aides visuelles (surlignages, polices de couleurs différentes) présentes dans la plupart des systèmes de TA-TAO, orientent le traducteur vers une lecture tabulaire des segments textuels et privilégient le repérage des correspondances au niveau syntagmatique. Certes, le paradigme de *traduction automatique adaptative* utilisé, par exemple, dans *SDL Trados Studio 2017* (<https://www.sdl.com>), représente une amélioration. En apprenant des post-éditions réalisées, un service personnalisé de traduction automatique reposant sur l'auto-apprentissage améliore

la qualité des suggestions. Cependant, l'ensemble du réseau de correspondances qui reflètent les régularités de l'axe paradigmatique n'est pas pris en compte systématiquement dans les systèmes axés sur des représentations tabulaires des correspondances. Sur le plan cognitif, cette « segmentation excessive » est souvent problématique pour le traducteur (Rossi 2017). En effet, les récurrences qui composent le tissu discursif des textes sources et cibles sont difficiles à cerner dans les phrases prises individuellement et dissociées de leurs contextes discursifs. Ces récurrences reflètent pourtant des éléments constitutifs de la phraséologie spécialisée qui font partie de la « mémoire humaine » du traducteur qui exerce dans un domaine spécialisé : schémas lexicogrammaticaux, phrases prototypiques, etc. (Gledhill, Kübler 2015 ; Gledhill *et al.* 2017).

Depuis plusieurs années, la *textométrie* propose des méthodes qui se prêtent bien à l'identification des unités textuelles récurrentes dans les textes source et cible (segments répétés, cooccurrences et poly-cooccurrences, patrons, motifs, routines, etc.) (Née *et al.* 2017). Les développements récents ont aussi permis de concevoir un modèle de données appelé *Trame/Cadre*, implémenté dans les outils de textométrie *Le Trameur* et *iTrameur* (Fleury, Zimina 2014 ; Zimina, Fleury 2015 ; Fleury 2017). Ce modèle rend possible l'exploration et le stockage des liens entre les unités textuelles repérables sur l'axe paradigmatique sous forme de *sélections* automatiques créées à la demande de l'utilisateur (Patin *et al.* 2016). Sur ces bases, il est possible de mémoriser des éléments textuels qui composent des équivalences au niveau sous-phrastique même lorsque l'alignement des phrases est difficile, comme c'est le cas en traduction-adaptation. Une démonstration sur les textes comparables en anglais et russe est proposée dans la section qui suit afin d'illustrer l'intérêt d'une MT dotée d'un *moteur textométrique* pour l'exploration de ressources traductionnelles.

3. Une MT dotée d'un moteur textométrique : approche exploratoire

La réflexion sur la modélisation de l'alignement bi-textuel sur l'axe paradigmatique a été amorcée dans nos travaux récents (Zimina, Fleury 2015 ; Zimina 2017). Dans les exemples qui suivent, nous illustrons comment un moteur textométrique qui explore les données de textes source et cible permet de révéler des récurrences d'unités textuelles qui composent des équivalences de traduction sur l'axe paradigmatique. Pour proposer un scénario d'utilisation concret, nous partirons de la traduction automatique en russe de la séquence suivante des actualités diffusées en anglais par *BBC News* (Zimina 2017) : *Finnish scientists say certain bacteria [...] are found in greater abundance in [...]*. La traduction

automatique (*Systran PNMT* : <https://demo-pnmt.systran.net/production#/translation>, *Google*, etc.) propose la correspondance suivante : *Финские ученые говорят, что некоторые бактерии [...] находятся в большом количестве в [...]*. Dans la phase de post-édition, nous allons utiliser une MT composée des fils RSS de *BBC News* et de leurs traductions-adaptations en russes diffusées sur *Lenta.ru* (Klementiev, Roth 2006 ; Zimina, Fleury 2014). Les textes originaux en anglais comptent plus d'un million d'occurrences, tandis que leurs traductions-adaptations en russe sont, en moyenne, deux fois plus courtes (moins de 500 000 occurrences au total).

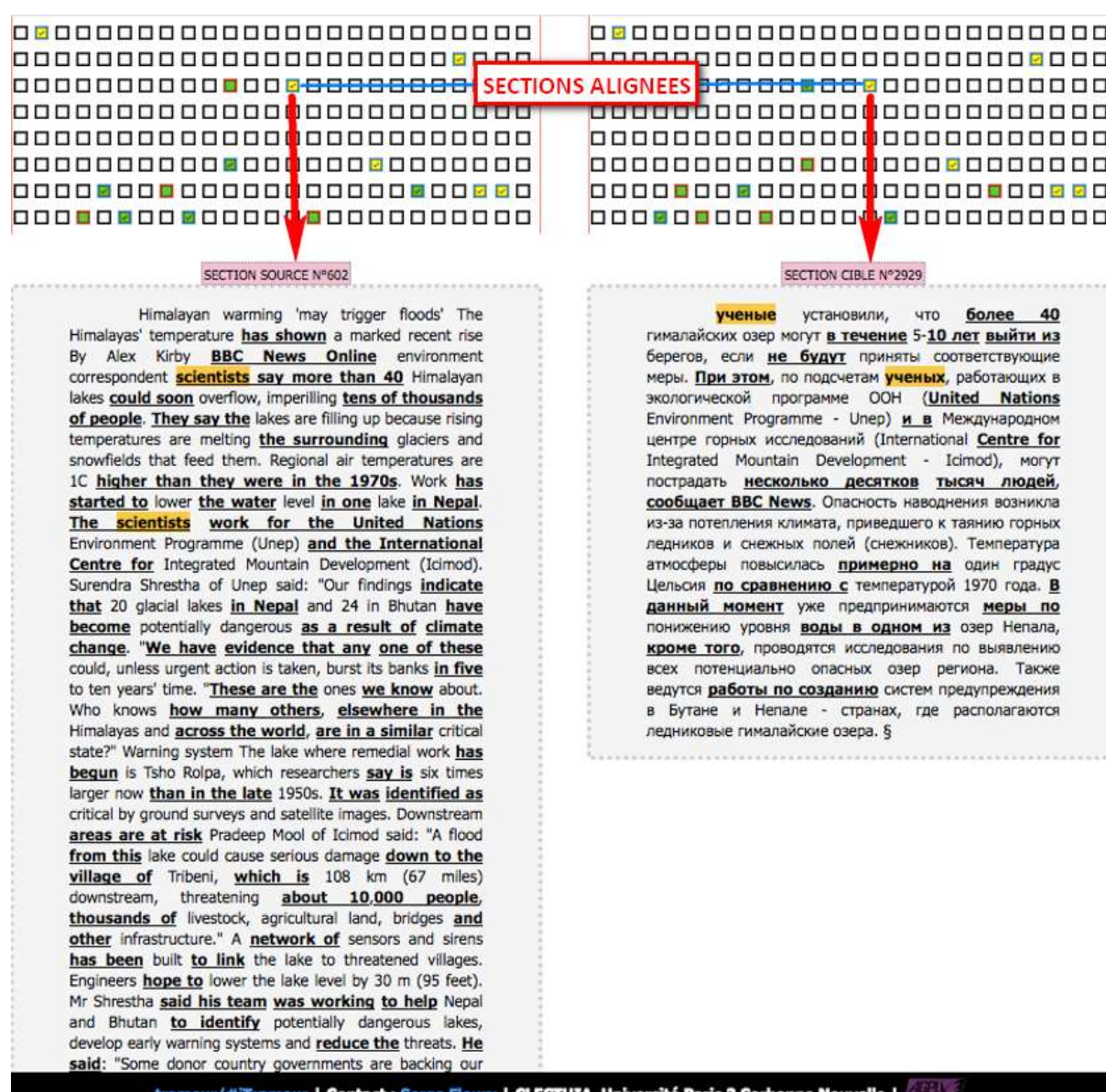


Figure 2 : MT dotée d'un moteur textométrique : *iTrameur* (<http://www.tal.univ-paris3.fr/trameur/#iTrameur>)

Légende : les récurrences sous-phrastiques (segments répétés) sont soulignées au fil du texte. La carte des sections en haut permet de comparer les distributions des pivots *scientists/ученые* dans les différents textes d'actualités alignés un-à-un (un carré correspond à un texte : l'alignement se lit de gauche à droite).

Le moteur textométrique que nous allons mobiliser pour explorer cette MT est accessible via le programme en ligne *iTrameur* (Fleury 2017). La visualisation des *segments répétés* (Salem 1987) au fil des textes (cf. Figure 2) montre bien que le segment *scientists say* fait partie des récurrences répertoriées dans le volet source. En revanche, aucune occurrence de *ученые говорят* n'est remontée par l'inventaire des segments répétés dans le volet cible. Cette absence caractéristique peut alerter le traducteur sur une erreur potentielle à corriger au cours de la traduction/post-édition/révision.

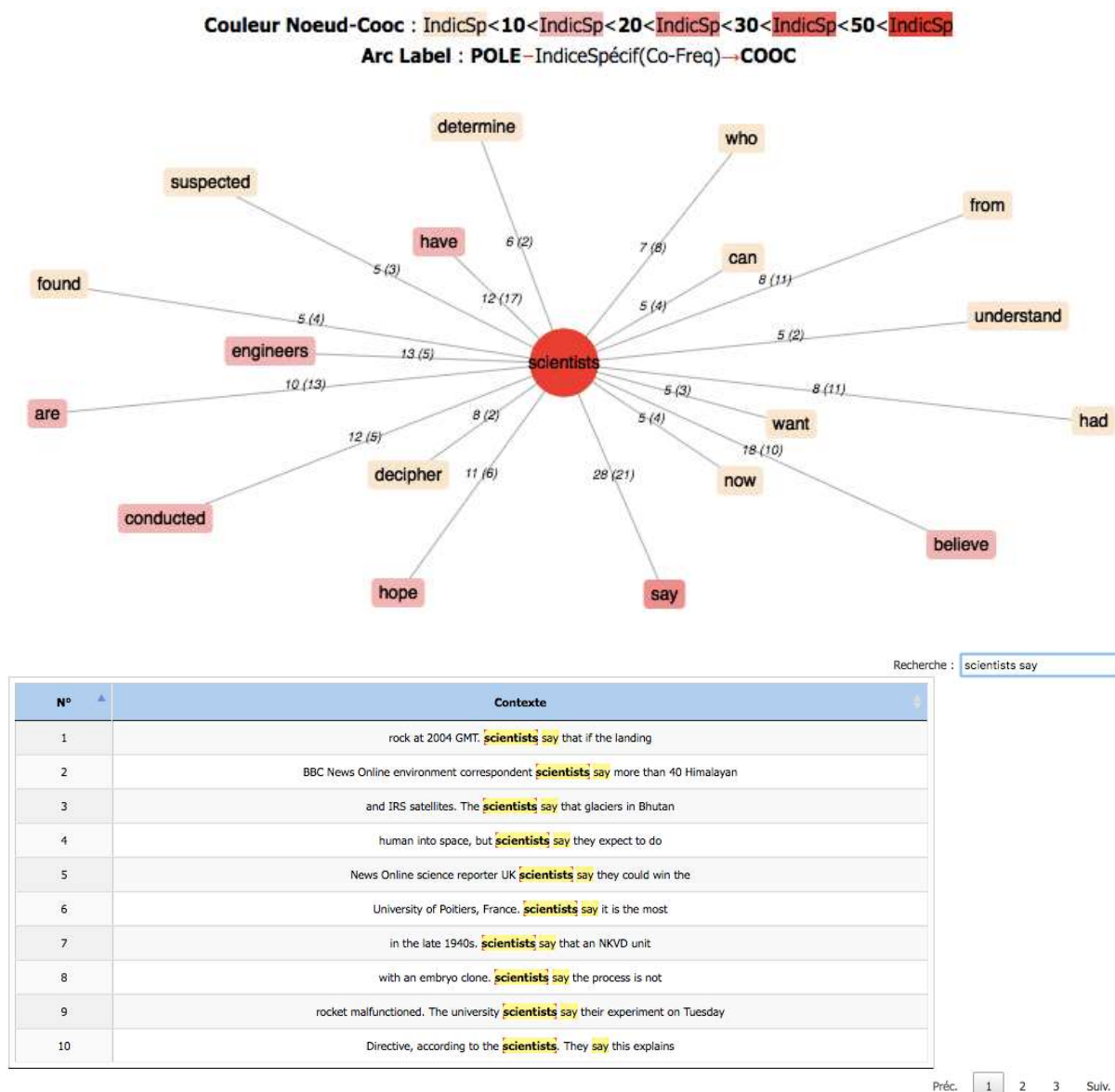


Figure 3 : réseau de co-occurrences généré avec *iTrameur* (axe paradigmatique du texte source en anglais, pôle *scientists*)

En poursuivant l'exploration, on découvre parallèlement deux schémas caractéristiques : *scientists* + *VVP* (verbe au présent) [Freq = 64] et *ученые* + *Vmip3p-a-e*

(verbe à l'aspect imperfectif) [Freq = 88] qui font partie des réseaux de cooccurrences des pôles *scientists* et *ученые* (cf. Figures 3-4). On identifie ensuite en contexte les réalisations possibles de ces schémas avec plusieurs verbes, par exemple : *scientists+say/believe/think* etc.) et *ученые+полагают/предполагают/считают/заявляют* etc.). Ces contextes bien ciblés peuvent être utilisés pour améliorer la traduction automatique *ученые говорят*.

On remplacera ainsi *говорят* par l'un des verbes du schéma identifié, en prenant en compte la fonction discursive correspondant au cadrage rhétorique de l'autorité scientifique, par exemple : *Финские ученые считают, что некоторые бактерии [...] находятся в большом количестве в [...]*, ou encore : *Финские ученые заявляют, что некоторые бактерии [...] находятся в большом количестве в [...]* (en fonction du contexte élargi).

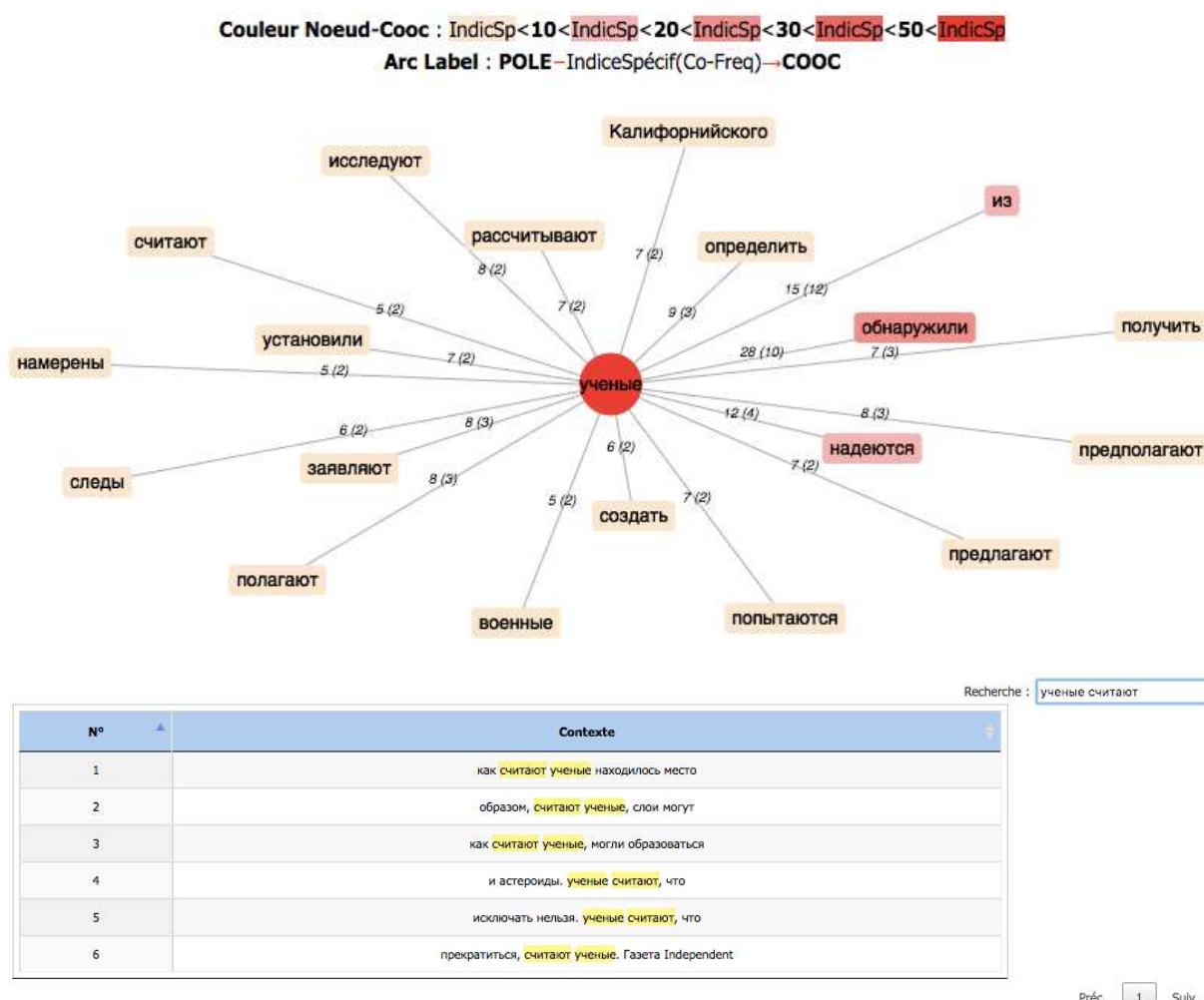


Figure 4 : Réseau de co-occurrences généré avec *iTrameur* (axe paradigmatique du texte cible en russe, pôle *ученые*)

4. Ressources traductionnelles incrémentales

Les *résonances textuelles* (Salem 2004 ; Zimina 2004b) qui ont fait émerger le pôle traductionnel *scientists/ученые* permettent d'induire successivement des emplois équivalents synonymiques, tels que *researchers/исследователи* en russe. On peut ensuite étudier des schémas récurrents qui caractérisent leurs voisinages contextuels. Les principes méthodologiques qui orientent une telle exploration *incrémentale* d'un corpus bi-textuel sont présentés en détail dans Zimina et Fleury 2014. A présent, ils régissent le fonctionnement des outils intégrés dans l'interface du programme *iTrameur* (Fleury 2017).

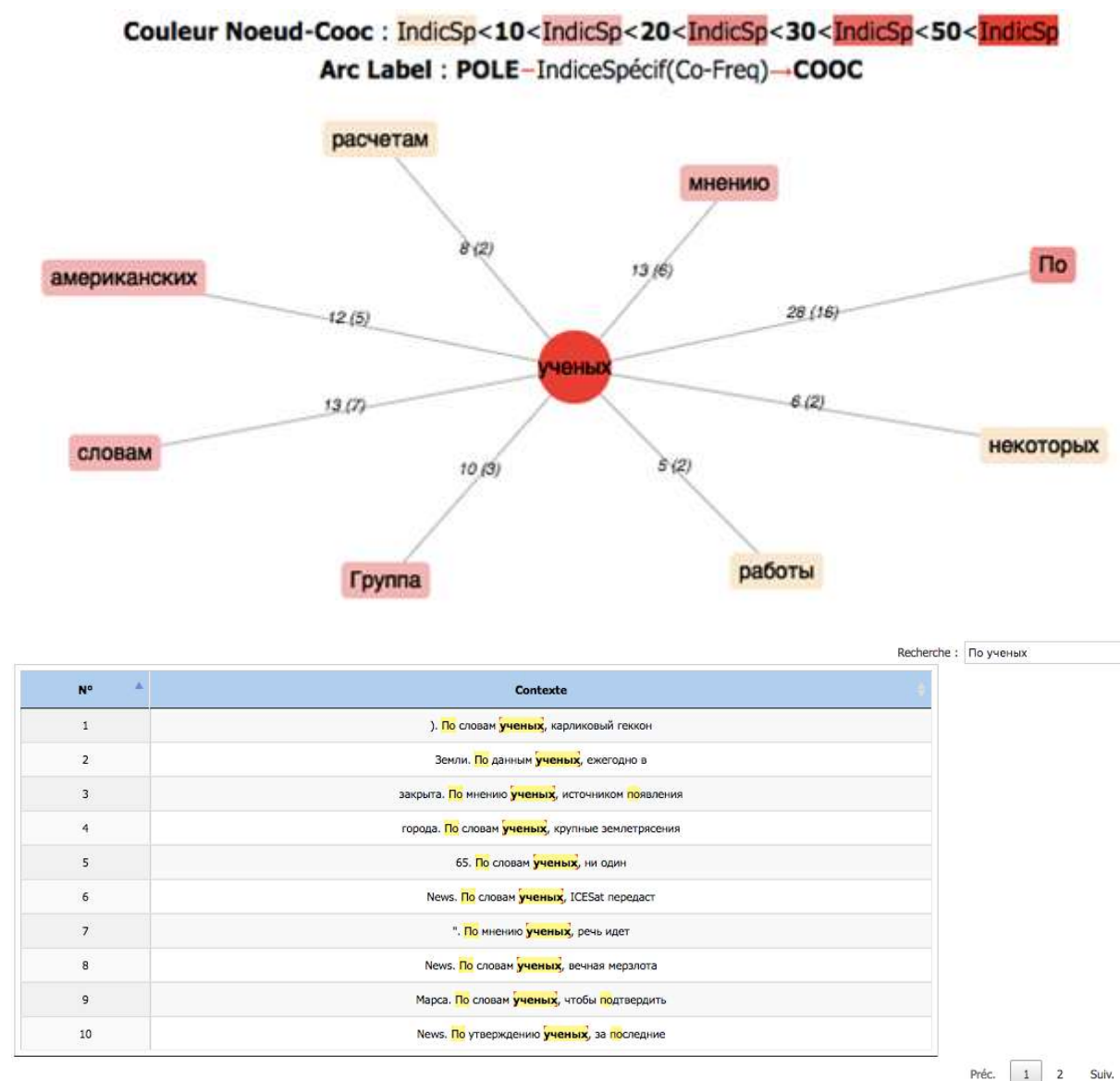


Figure 5 : réseau de co-occurrences généré avec *iTrameur* (axe paradigmatique du texte cible en russe, pôle *ученых*, la forme fléchie de *ученые*)

Par exemple, l'étude des contextes qui émergent en écartant les distributions où les schémas *scientists + VVP* et *ученые + VтипЗр-а-е* se correspondent, permet de découvrir la présence d'une forme fléchie *ученых* en russe, qui fait partie des correspondances de *scientists* dans le corpus. Ce nouveau pôle est soumis à son tour à l'analyse de co-occurrences (cf. Figure 5 ci-dessus) qui indique clairement la présence d'un autre schéma caractéristique : *По+N+учёных*. Son pivot est composé d'une préposition *по* (« selon » en français), qui peut éventuellement se trouver en début de la phrase, suivi d'un nom et de la forme fléchie *ученые*.


iTrameur Analyse textométrique de données	
21	<p>Himalayan warming 'may trigger floods' The Himalayas' temperature has shown a marked recent rise By Alex Kirby BBC News Online environment correspondent scientists say more than 40 Himalayan lakes could soon overflow, imperilling tens of thousands of people living around them. The lakes are filling up because rising temperature is melting the snowfields that feed them. Regionally, lake levels were in the 1970s. Work has started in Nepal. The scientists work for the United Nations Environment Programme (Unep) and the International Centre for Mountain Development (Icimod). Surendra Shrestha of Unep said: "Our findings indicate that 20 glacial lakes in Nepal and 24 in Bhutan have become potentially dangerous as a result of climate change. "We have evidence that any one of these could, unless urgent action is taken, burst its banks in five to ten years' time. "These are the ones we know about. Who knows how many others, elsewhere in the Himalayas and across the world, are in a similar critical state?" Warning system The lake where remedial work has begun is Tsho Rolpa, which researchers say is six times larger now than in the late 1950s. It was identified as critical by ground surveys and satellite images. Downstream areas are at risk Pradeep Mool of Icimod said: "A flood from this lake could cause serious damage down to the village of Tribeni, which is 108 km (67 miles) downstream, threatening about 10,000 people, thousands of livestock, agricultural land, bridges and other infrastructure." A network of sensors and sirens has been built to link the lake to threatened villages. Engineers hope to lower the lake level by 30 m (95 feet). Mr Shrestha said his team was working to help Nepal and Bhutan to identify potentially dangerous lakes, develop early warning systems and reduce the threats. He said: "Some donor country governments are backing our efforts, but much more aid is needed. Solving this problem is going to be costly, because glacial lakes are situated in remote areas which are difficult to reach." The UN has declared 2002 the International Year of the Mountains. Dr Klaus Toepfer, executive director of Unep, said: "Mountains were once considered indomitable, unchanging and impregnable. But we are learning that they are as vulnerable as the world's oceans, grasslands and forests. Drier future "It is not just the risk to human lives, agriculture and property that should worry us. Mountains are the world's water towers, feeding the rivers and lakes upon which all life depends. Glaciers are in retreat "If the glaciers continue to retreat at the rates being seen in places like the Himalayas, many rivers and freshwater systems could run dry." Unep says there is evidence that glacial lake outburst floods (glofs, as they are known) have been happening more often over the last thirty years. It says data from 49 Nepalese monitoring stations shows temperatures in the high Himalayas have been rising over that period by an average of 0.06C annually. Rapid melting The research into lakes at risk began in 1999, and is based on topographic maps, aerial photographs and images from Landsat, Spot and IRS satellites. The scientists say that glaciers in Bhutan are retreating at 30-40m annually. One glacier, Tradkarding, which feeds the Tsho Rolpa lake, retreated 100m last year. scientists from Bhutan and Nepal worked with Unep and Icimod on the survey. An international conference on protecting the world's mountains is being held in late October in Bishkek, the capital of Kyrgyzstan. §</p>
	<p>ученые установили, что более 40 гималайских озер могут в течение 5-10 лет выйти из берегов, если не будут приняты соответствующие меры. При этом, по подсчетам ученых, работающих в экологической программе ООН (United Nations Environment Programme - Unep) и в Международном центре горных исследований (International Centre for Integrated Mountain Development - Icimod), могут пострадать несколько десятков тысяч людей, сообщает BBC News. Опасность наводнения возникла из-за потепления климата, приведшего к таянию горных ледников и снежных полей (снежников). Температура атмосферы повысилась примерно на один градус Цельсия по сравнению с температурой 1970 года. В данный момент уже предпринимаются меры по понижению уровня воды в одном из озер Непала, кроме того, проводятся исследования по выявлению всех потенциально опасных озер региона. Также ведутся работы по созданию систем предупреждения в Бутане и Непале - странах, где располагаются ледниковые гималайские озера. §</p>
<p>trameur/#iTrameur Contact : Serge Fleury CLESTHIA, Université Paris 3 Sorbonne Nouvelle </p>	

Figure 6 : la contraction de l'information dans le texte cible en russe avec une utilisation caractéristique du schéma *По+Nom+учёных*

Comme le montrent les contextes présentés sur la Figure 6, ce schéma correspond à des emplois spécifiques (*По словам ученых* ou *по подсчетам ученых*) qui relèvent de la contraction de texte en adaptation (l'une des formules équivalentes en français serait « selon les études scientifiques »). En effet, on constate que le texte russe qui emploie le schéma répertorié est particulièrement raccourci comparé au texte original en anglais. Les détails des actualités présentées en anglais sont considérablement condensés ou même omis en russe (cf. Figure 6). L'analyse de contextes répertoriés montre que le schéma analysé a un rôle fonctionnel précis. Ce constat permet d'affiner l'analyse de la phraséologie du discours des actualités scientifiques en russe, en tenant compte des fonctions discursives spécifiques.

4. Résultats et perspectives

Les traces de la mise en correspondance des schémas équivalents relevés dans la MT anglais/russe peuvent être sauvegardées sous forme d'annotations sur sélections (Patin *et al.* 2016). Dans ce cas, après le rechargement de la MT dans l'interface utilisateur, les éléments constitutifs de ces équivalences seraient visibles parallèlement en contexte grâce à un système de surlignage automatique (Gledhill *et al.* 2017). On dispose ainsi d'une ressource textométrique complète au format *Trame/Cadre* (Zimina, Fleury 2016) qui offre plus de souplesse pour l'exploration de ressources traductionnelles. Par ailleurs, chaque étape de cette exploration est formalisée et reproductible. Pour y parvenir, la solution optimale consiste à accepter une séparation formelle très nette pendant le traitement automatique entre les apports des analyses quantitatives appliquées successivement à chaque volet du bi-texte par le moteur textométrique, et les hypothèses qualitatives, qui orientent l'exploration et amorcent des sélections pertinentes. C'est l'action commune de ces deux types d'apports distincts (le quantitatif et le qualitatif) qui permet de formaliser le concept de *ressource traductionnelle incrémentale* qui s'inspire largement des formalisations proposées dans les travaux antérieurs (Söze-Duval 2008 ; Zimina Fleury 2015). Nos expériences sur les données de traductions confirment la pertinence de l'utilisation de ces modèles pour permettre au traducteur de bénéficier des accès contextuels souples et de garder la main sur les processus déployés dans l'exploration de ressources textuelles disponibles.

5. Conclusions

L'accélération des rythmes de travail et l'accroissement des volumes de textes à traduire rend nécessaire le recyclage optimal des ressources de traduction existantes. Parallèlement, le débat sur l'évaluation de la qualité de la version finale de la traduction en

TA-TAO valorise une attention soutenue aux détails (Cronin 2013 : 128) : “*Quality is, in a sense, the return of the repressed translation detail. The careful, detailed attention to text, language, and meaning that is implicit in the act of translation re-emerges in the context of automation in the debates about the extent and role of post-editing, and about how to achieve acceptable quality in translation output.*”

Dans notre travail, nous avons montré comment une MT dotée d’un moteur textométrique peut élargir les possibilités d’exploration de textes traduits au-delà de la lecture tabulaire de segments alignés. C’est le concept de *ressource textuelle incrémentale* qui constitue pour nous le fil conducteur de la réflexion à mener pour changer certaines pratiques de travail avec les MT trop rigides. Les données de traductions existantes seraient alors systématiquement mobilisées en traduction, post-édition et révision pour « ...varier le paradigme (les éléments plus variables du schéma), afin d’obtenir des effets de style ou de mieux intégrer la construction dans le discours environnant. » (Gledhill 2008 : 44). L’étude de la dynamique fonctionnelle d’une telle approche exploratoire passe obligatoirement par des modélisations prototypiques qui sont déjà en partie implémentées dans *iTrameur* (Fleury, 2017). Dans les projets à venir, cette réflexion est amenée à se développer pour prendre en compte les questions de rapidité et d’ergonomie d’exploration des données textuelles multilingues.

6. Références

EAGLES 1996, “Evolution of Natural Language Processing Systems”, *Final Report*, EAG-EWG-PR.2, Version of October 1996. <http://www.issco.unige.ch/downloads/ewg96.pdf>

BALLARD (M.) 1993. « L’unité de traduction. Essai de redéfinition d’un concept. », éd. Ballard (M.), *La traduction à l’université. Recherches et propositions didactiques*. PUL, Lille : 223-262.

CRONIN (M.) 2013. *Translation in the Digital Age*, Milton Park, Abingdon, Oxon, Routledge.

FLEURY (S.) *iTrameur*. <http://www.tal.univ-paris3.fr/trameur/#iTrameur>

FLEURY (S.), ZIMINA (M.) 2014, “Trameur: A Framework for Annotated Text Corpora Exploration”, *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations*, Dublin, Dublin City University, 23-29 August 2014, ed. Tounsi (L.), Rafal (R.), ACL, 2014: 57-61. <http://www.aclweb.org/anthology/C14-2013.pdf>

GLEDHILL (C.) 2008, « Le signe et le syntagme : entre phraséologie et lexico-grammaire. Une synthèse du modèle systémique fonctionnel de Michael Halliday et de la théorie sémiotique de Charles S. Peirce ». *Habilitation à diriger des recherches*. Université de Bretagne Occidentale.

GLEDHILL (C.) 2008. *Portée, Pivot, Paradigme : trois termes pour faire le point sur les expressions verbo-nominales*. *Zeitschrift für Französische Sprache und Literatur Beiheft* 35: 59-76. Stuttgart : Franz Steiner Verlag.

GLEDHILL (C.), KÜBLER (N.) 2015, “How Trainee Translators Analyse Lexico-Grammatical Patterns”, *Journal of Social Sciences*, Special issue on Phraseology, Phraseodidactics and Construction Grammar(s), ed. González-Rey (M. I.), vol. 11, no 3, 2015: 162-178.

GLEDHILL (C.), PATIN (S.), ZIMINA (M.) 2017, « Lexico-grammaire et textométrie : identification et visualisation de schémas lexico-grammaticaux caractéristiques dans deux corpus juridiques comparables en français », *Corpus*, vol. 17 « segments et séquences textuelles : méthodologie et caractérisation ».

KLEMENTIEV (A.), ROTH (D.) 2006, “Weakly Supervised Named Entity Transliteration and Discovery from Multilingual Comparable Corpora”, *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, Sydney, 17 July 2006, ACL, 2006: 817-824.

NEE (É.) (dir.), FLEURY (S.), BARATS (C.), LEBLANC (J.-M.), SITRI (F.), VENIARD (M.) 2017, *Méthodes et outils informatiques pour l'analyse du discours*, Presses Universitaires de Rennes.

PATIN (S.), ZIMINA (M.), FLEURY (S.) 2016, « Lecture Textométrique Différentielle (LTD) de textes législatifs comparables de l'Union européenne », *Actes des 13es Journées internationales d'Analyse statistique des Données Textuelles*, Nice, Université Nice-Sophia-Antipolis, 7-10 juin 2016, éd. Mayaffre (D.), Poudat (C.), Vanni (L.), Magri (V.), Folette (P.), *JADT 2016 - Statistical Analysis of Textual Data*, vol. 2, Nice, Presses de FacImprimeur, 2016 : 743-753.

ROSSI (C.) 2017, De quoi la traduction automatique est-elle le nom ? *En attendant Nadeau. Journal de la littérature, des idées et des arts*. En ligne : <https://www.en-attendant-nadeau.fr/2017/07/25/traduction-automatique/>

SALEM (A.) 1987, *Pratique des segments répétés. Essai de statistique textuelle*. *Histoire & Mesure*, vol. 2, no 3-4, 1987 : 224-227.

SALEM (A.) 2004, « Introduction à la résonance textuelle », *Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles*, 10-12 mars 2004, Louvain-la-Neuve, Université catholique de Louvain, éd. Purnelle (G.), Fairon (C.), Dister (A.), Louvain-la-Neuve, Presses Universitaires de Louvain, vol. 2, 2004 : 986-992.

SÖZE-DUVAL (K.) 2008, « Pour une textométrie opérationnelle ». En ligne : <http://www.tal.univ-paris3.fr/trameur/RTI6provisoire.doc>

ZIMINA (M.) 2004, « Approches quantitatives de l'extraction de ressources traductionnelles à partir de corpus parallèles. » Thèse pour le Doctorat en Sciences du langage, Université de la Sorbonne nouvelle – Paris 3. En ligne : <http://tel.archives-ouvertes.fr/tel-00008311/fr>

ZIMINA (M.) 2004, « Alignement textométrique des unités lexicales à correspondances multiples dans les corpus parallèles », *Actes des 7es Journées internationales d'Analyse statistique des Données Textuelles*, 10-12 mars 2004, 10-12 mars 2004, Louvain-la-Neuve, Université catholique de Louvain, éd. Purnelle (G.), Fairon (C.), Dister (A.), Louvain-la-Neuve, Presses Universitaires de Louvain, vol. 2, 2004 : 1195-1202.

ZIMINA (M.), FLEURY (S.) 2014 « Approche systémique de la résonance textuelle multilingue », *Actes des 12es Journées Internationales d'Analyse Statistiques des Données Textuelles*, Paris, INaLCO, 3-6 juin 2014, éd. Née (É.), Daube (J-M.), Valette (M.), Fleury (S.), Paris, JADT.org : 717-728.

ZIMINA (M.), FLEURY (S.) 2015, « Perspectives de l'architecture Trame/Cadre pour les alignements multilingues », *Nouvelles perspectives en sciences sociales : revue internationale de systémique complexe et d'études relationnelles*, vol. 11, no 1, 2015.

ZIMINA (M.) 2017, « Vers une Mémoire de Traduction (MT) dynamique et multidimensionnelle », atelier « Textométrie et Traduction », *1er Congrès Mondial de Traductologie (CMT)*, Université Paris-Nanterre, 10-14 Avril 2017. <https://cmt.u-paris10.fr/wp-content/uploads/2016/02/624-ZIMINA-corrig%C3%A9.pdf>