AFCP

ATALA

Journées d'Études sur la Parole
Traitement Automatique des Langues Naturelles
Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues

PARIS Inalco du 4 au 8 juillet 2016

Organisé par les laboratoires franciliens

https://jep-taln2016.limsi.fr

Conférenciers invités:

Christian Chiarcos (Goethe-Universität, Frankfurt.)
Mark Liberman (University of Pennsylvania, Philadelphia)

Coordinateurs comités d'organisation

Nicolas Audibert et Sophie Rosset (JEP)
Laurence Danlos & Thierry Hamon (TALN)
Damien Nouvel & Ilaine Wang (RECITAL)
Philippe Boula de Mareuil, Sarra El Ayari & Cyril Grouin (Ateliers)

Table des matières

Origines des erreurs en Traduction Spécialisée : différentiation textométrique grâce aux corpus de textes cibles annotés

Natalie Kübler[1], Maria Zimina[1], Serge Fleury[2]
(1) CLILLAC-ARP EA 3967, Université Paris Diderot-Paris 7, France
(2) CLESTHIA EA 7345, Sorbonne Nouvelle-Paris 3, Paris, France
nkubler@eila.univ-paris-diderot.fr, mzimina@eila.univ-paris-diderot.fr, serge.fleury@univ-paris3.fr

RESUME

L'étude présente une analyse quantitative de traductions annotées selon une typologie d'erreurs, en vue de l'amélioration des méthodologies d'enseignement de la traduction spécialisée (TS). Les productions annotées sont alignées avec les textes originaux au niveau de la phrase. Les spécificités morpho-syntaxiques sur les contextes sources regroupés par types d'erreurs de traduction permettent de récolter des indices sur les éléments complexes des discours spécialisés qui génèrent des erreurs lors du processus de transfert du sens. La visualisation contextuelle de ces éléments via la Lecture Textométrique Différentielle (LTD) ouvre des perspectives pour la conception des modules de prise en charge des difficultés caractéristiques des apprenants en TS.

ABSTRACT

Origins of errors in Specialized Translation: textometric differentiation through annotated target text corpora

This study focuses on the quantitative analysis of translations annotated according to an error typology. The purpose of the study is to improve current methodologies of Specialized Translation (SP) teaching. We conduct a quantitative analysis of annotated translations aligned with their original texts. Characteristic elements are computed on morpho-syntactic level of analysis in source contexts grouped according to translation errors in the target text. They reveal specific linguistic elements that are complex in terms of meaning transfer in specialized languages. Contextual visualization of these results through Differential Textometric Browsing (DTB) opens up new horizons for the development of language modules based on the types of difficulties that learners face in SP.

MOTS-CLES : Annotation d'erreurs, corpus alignés, enseignement de la Traduction Spécialisée (TS), textométrie, Lecture Textométrique Différentielle (LTD), méthode des spécificités, visualisation.

KEYWORDS: Aligned corpora, characteristic elements computation, Differential Textometric Reading (DTR), annotation, Specialized Translation teaching, textometric analysis, visualisation.

Natalie Kübler, Maria Zimina, Serge Fleury

1 Introduction

1.1 Contexte de l'étude

Récemment, les approches centrées sur le processus de traduction et sur l'évaluation des méthodes d'enseignement ont fait l'objet de plusieurs travaux de recherche (Bowker, Bennison, 2003 ; Pearson, 2003 ; Castagnoli et al., 2011 ; Loock et al., 2014 ; Frankenberg-Garcia, 2015). En Traduction Spécialisée (TS), beaucoup de chantiers restent encore à explorer pour mieux cerner les difficultés de transfert de sens mobilisant des compétences linguistiques et cognitives très variées face à la complexité des textes en langues de spécialité (Kübler, 2011 ; Froeliger, 2013). Dans ce contexte, la recherche expérimentale à base de corpus offre la possibilité d'observer de manière systématique les stratégies employées par les traducteurs qui peuvent passer inaperçues avec d'autres approches (Granger et al., 2002 ; Frérot, 2010).

Notre travail s'appuie sur une méthodologie d'enseignement de la traduction spécialisée qui permet une identification systématique des problèmes de traduction par le biais d'une analyse textométrique de corpus de traduction annotés selon la typologie d'erreurs MeLLANGE (Castagnoli et al., 2011), avec l'aide du logiciel *Le Trameur* (Fleury, Zimina, 2014). Ce cadre méthodologique est mis en place depuis septembre 2013 à l'université Paris Diderot dans le cadre du Master 1 Industrie des langues et Traduction spécialisée (ILTS).[1]

Dans un premier temps, la méthodologie employée nous a permis de distinguer les catégories d'erreurs les plus saillantes et les schémas morphosyntaxiques qui les caractérisent en contexte, à savoir les problèmes touchant les termes composés et complexes, mais également les collocations, les prépositions, les verbes, etc. Les analyses faites sur les productions des apprenants de ces deux dernières années (2013-2015) nous ont aussi amenés à intégrer à l'enseignement de la TS une approche plus ciblée des problèmes posés par la traduction des termes spécialisés et des GN complexes (Kübler et al., 2016).

Dans cette nouvelle étude, nous nous intéressons aux origines des erreurs en traduction spécialisée par profilage des contextes sources à partir des annotations des erreurs dans les productions en langue cible. Nous pensons que les corpus de traductions alignées avec les textes sources peuvent aider à récolter des indices sur les éléments complexes des discours spécialisés qui sont à l'origine des erreurs récurrentes. La découverte de ce type d'indices peut ensuite alimenter la réflexion sur le développement des modules spécifiques qui ciblent les problèmes récurrents des apprenants.

1.2 Objectifs

L'objectif de cette expérimentation est la création de cours spécifiquement adaptés aux problèmes identifiés dans les textes sources. Ce travail vise à contribuer au développement de la méthodologie d'enseignement de la TS qui mêle par ailleurs des compétences transversales à plusieurs niveaux. Pour les étudiants, il s'agit de découvrir et de se former à l'approche de la traduction basée sur le

[1] http://www.eila.univ-paris-diderot.fr/formations-pro/masterpro/ilts/index

corpus, au travail sur les discours spécialisés en collaboration étroite avec les experts du domaine, et à l'analyse terminologique et notionnelle préalable au processus de traduction.

La prise en compte de l'alignement des textes originaux et des traductions annotées selon la typologie d'erreurs MeLLANGE vise à élaborer des propositions méthodologiques à base d'exemples concrets à l'origine des erreurs de traduction. Cet ancrage dans le texte source constitue le trait caractéristique de ce volet de l'étude.

Notre approche est exploratoire : les corpus sont exploités pour détecter des éléments complexes dans les contextes sources à partir des profils d'erreurs repérées dans la traduction. Sur ce plan, on peut constater des similitudes entre les objectifs de cette étude et des travaux sur le profilage d'erreurs de la traduction automatique (Kübler et al., 2013 ; Wisniewski et al., 2014). Dans les deux cas, il s'agit de recueillir des informations sur les origines des erreurs et d'en tenir compte dans les nouvelles productions.

2 Corpus aligné et annoté ER-TRAD-SP1 (anglais-français)

La présente étude exploite les traductions de l'anglais vers le français réalisées par les étudiants M1 en 2014-2015. Il s'agit de 55 extraits de 14 articles scientifiques en Sciences de la Terre (37 324 occurrences de formes graphiques au total). Ce corpus est subdivisé en deux sous-corpus : *ER-TRAD-SP1* (15 311 occurrences de formes graphiques) constitué de traductions réalisées sans accès aux corpus, et *ER-TRAD-SP2* (22 013 occurrences de formes graphiques) constitué de traductions réalisées avec l'aide du corpus. Les traductions portent sur des extraits d'articles scientifiques avec une très haute densité terminologique et le registre de langue caractéristique du discours scientifique. Les problèmes de traduction qu'affrontent les apprentis traducteurs sont nombreux et variés. L'annotation des erreurs selon la typologie MeLLANGE permet toutefois une catégorisation fine des différents problèmes rencontrés dans ce type de discours. Au total, le sous-corpus *ER-TRAD-SP1* compte 886 annotations ; le sous-corpus *ER-TRAD-SP2* compte 893 annotations (Kübler et al., 2016).

Actuellement, le corpus *ER-TRAD-SP1* (traductions sans accès au corpus) a été aligné au niveau de la phrase avec les textes originaux. Ce travail a été réalisé à l'aide d'une série de scripts en s'appuyant sur les alignements phrastiques initialement proposés par les étudiants au cours de la traduction. Nous avons également fait appel aux fonctions du programme *MkAlign*[2] pour la vérification et synchronisation finale de l'alignement. Chaque volet du corpus a été étiqueté par *TreeTagger*[3] intégré dans *Le Trameur*[4] et converti en une base textométrique au format XML dans laquelle l'annotation des erreurs de traduction est renseignée pour le volet français.

[2] http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/
[3] http://www.tal.univ-paris3.fr/mkAlign/
[4] http://www.tal.univ-paris3.fr/trameur/

La Figure 1 montre deux segments issus de l'alignement des phrases dans la base textométrique *ER-TRAD-SP1* au format lu par *Le Trameur*. Cette base comporte 6 niveaux d'annotation. Pour chaque *item* dénombré (type forme ou délimiteur), on retrouve :

1. <f> sa forme graphique </f>
2. <c> sa catégorie morpho-syntaxique </c>
3. <l> son lemme </l>
4. <a> son type d'erreur MeLLANGE (ou son absence)
5. <a> le commentaire éventuel de l'annotateur
6. <a> l'indication sur la présence d'erreur tous types confondus (ou son absence)

```
<item type="forme" pos="6"><f>would</f><c>MD</c><l>would</l><a>-</a><a>-</a><a>-
</a></item>
<item type="delim" pos="7"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-
</a></item>
<item type="forme" pos="8"><f>have</f><c>VH</c><l>have</l><a>-</a><a>-</a><a>-
</a></item>
<item type="delim" pos="9"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-
</a></item>
<item type="forme" pos="10"><f>only</f><c>JJ</c><l>only</l><a>-</a><a>-</a><a>-
</a></item>
<item type="delim" pos="11"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-
</a></item>
<item type="forme" pos="12"><f>a</f><c>DT</c><l>a</l><a>-</a><a>-</a><a>-</a></item>
<item type="delim" pos="13"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-
</a></item>
<item type="forme" pos="14"><f>small</f><c>JJ</c><l>small</l><a>-</a><a>-</a><a>-
</a></item>
<item type="delim" pos="15"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>-
</a></item>
<item type="forme" pos="16"><f>effect</f><c>NN</c><l>effect</l><a>-</a><a>-</a><a>-
</a></item>                                                                    original
                                                                               traduction
<item type="forme" pos="26884"><f>aura</f><c>VER_futu</c><l>avoir</l><a>
Transfert-contenu</a><a>Indicatif ou conditionnel, pour "would have".</a><a>Erreur
</a></item>
<item type="delim" pos="26885"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>
-</a></item>
<item type="forme" pos="26886"><f>peu</f><c>ADV</c><l>peu</l><a>-</a><a>-</a><a>-
</a></item>
<item type="delim" pos="26887"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>
-</a></item>
<item type="forme" pos="26888"><f>de</f><c>PRP</c><l>de</l><a>-</a><a>-</a><a>-
</a></item>
<item type="delim" pos="26889"><f> </f><c>DELIM</c><l>BLANK</l><a>-</a><a>-</a><a>
-</a></item>
<item type="forme" pos="26890"><f>conséquences</f><c>NOM</c><l>conséquence</l><a>-
</a><a>-</a><a>-</a></item>
```

FIGURE 1 : Base textométrique alignée ER-TRAD-SP1 (extraits)

3 Méthodes : analyse des origines d'erreurs en TS par la différentiation textométrique

3.1 Diagnostics de spécificités sur contextes alignés

La *textométrie multilingue* (Fleury, Zimina, 2014 ; Zimina, Fleury, 2015) propose des méthodes quantitatives adaptées à l'observation des variations de fréquence d'unités textuelles (formes, lemmes, catégories, etc.) en contextes alignés. En suivant cette approche, nous mobilisons les profils d'erreurs en contextes cibles pour amorcer l'analyse quantitative des contextes sources correspondants. Les contextes d'erreurs de traduction (phrases alignées) sont analysés en termes de *spécificités* avec *Le Trameur*. Cette méthode permet de mesurer les variations de la fréquence dans un corpus découpé en parties (Lebart, Salem, 1994). Dans notre cas, le contexte d'erreur correspond à la phrase et la comparaison s'effectue entre deux parties : les phrases avec et sans erreurs de traduction en langue cible. Les empans d'erreurs (calculés en nombre d'occurrences de formes graphiques) varient selon les types d'erreurs.[5]

La *méthode des spécificités* (Lafon, 1984) met en évidence pour chaque unité de décompte les parties de corpus dans lesquelles l'unité possède de nombreuses occurrences (spécificités positives) ainsi que celles où son effectif est au contraire anormalement faible (*spécificités négatives*). On calcule le diagnostic de spécificité relatif à l'effectif constaté à base des paramètres suivants : k_{ij} - sous-fréquence de l'unité dans la partie, F_i - fréquence de l'unité dans l'ensemble du corpus, T_j - nombre des unités dans la partie, T - nombre total des unités du corpus. Un calcul probabiliste permet de porter un jugement sur l'effectif analysé (k_{ij}) compte tenu des trois autres nombres (F_i, T_j, T). Si l'effectif k_{ij} se situe dans les limites de ce que le calcul permettait d'espérer, la répartition constatée est considérée « banale ». Si ce n'est pas le cas, on calcule un *indice de spécificité* de l'unité. Le diagnostic est fourni sous la forme $\pm xx$ où le signe (+ ou -) indique un sur-emploi ou un sous-emploi de l'unité dans la ou les partie(s) sélectionnée(s) par rapport à l'ensemble du corpus ; *xx* est un indice de spécificité qui est d'autant plus élevé que la sous-fréquence analysée s'écarte d'une répartition « neutre » qui est sous-jacente au modèle des *spécificités*.[6]

Sur la Figure 2, les diagnostics de *spécificités* sont présentés sous forme de listes de catégories morpho-syntaxiques caractéristiques des contextes d'erreurs répertoriées dans la traduction. Pour chaque erreur, les catégories surreprésentées dans les contextes sources sont triées par la valeur d'*indice de spécificité* (du plus haut vers le plus bas). Seuls les profils sources d'erreurs fréquentes sont représentés (lorsque la zone de couverture de l'annotation cible est supérieure à 50 occurrences de formes graphiques annotées). Cette liste des *spécificités* sert de point d'entrée pour explorer les profils d'erreurs à l'aide des fonctionnalités disponibles dans *Le Trameur* : analyse des sections alignées, concordances, graphiques de ventilation, cartographie différentielle sur corpus parallèle, etc. (Zimina, Fleury, 2015).

[5] Sur le calcul des longueurs moyennes des empans d'erreurs, consulter Kübler et al. (2016).

[6] Le modèle probabiliste utilisé ici pour évaluer la répartition est le modèle hypergéométrique, cf. Lafon (1984, pp. 54-68). Sur la pratique du calcul des *spécificités* on consultera également Lebart, Salem (1994, pp. 172-176).

Erreur MeLLANGE (Nb. occ. formes annotées)	Eléments caractéristiques des contextes sources : *spécificités positives* sur catégories *TreeTagger* (extraits d'articles en Sciences de la Terre)	Indice de spécificité (seuil de 10)
Distorsion (395 occ. formes annotées)	Verb, past participle (*considered*, *derived, inspected*) Modal verb (*can*, *could, may, must, should, would*)	+3 +2
Formulation maladroite (366 occ. formes annotées)	Complementizer *that* Verb, gerund/participle (*melting, bearing*) *Wh*-abverb (*when, where*) Verb *be*, present non-3rd p. (*are*)	+3 +3 +2 +2
Trop littérale (286 occ. formes annotées)	Verb, past participle (*compared, formed, shown*) List marker (*1, 2, b, d*) Modal verb (*might, should, would*) Adverb, comparative (*more*) Adverb (*essentially, respectively, slightly*) Verb *be*, base form Verb, base form (*generate, induce, melt*)	+3 +3 +3 +2 +2 +2 +2
Terme traduit par non terme (150 occ. formes annotées)	Noun singular or massive (*mantle*, *olivine, subduction*) Verb *be*, past (*was, were*) Verb, past participle (*analysed*, *derived, used*) Verb *be*, present non-3rd p. (*are*)	+3 +3 +2 +2
Choix incorrect (128 occ. formes annotées)	*Wh*-determiner (*which*) Modal verb (*can*, *may, should*) Determiner (*the*) Verb, past tense (*observed*, *increased*) Verb, present, non-3rd p. (*conclude, suggest*)	+3 +2 +2 +2 +2
Omission (98 occ. formes annotées)	Verb *have* present, non-3rd p. (*have*) Verb, past participle (*formed*, *recorded*, *correlated*) Verb *be*, present 3rd p. sing (*is*)	+3 +2 +2
Collocation incorrecte (93 occ. formes annotées)	Verb *be*, past (*was, were*) Cardinal number (*300, 500, two*) Verb *have* present, non-3rd p. (*have*) Verb, past participle (*reported, based*)	+5 +3 +2 +2
Syntaxe (70 occ. formes annotées)	Modal verb (*can, cannot, may*) Noun singular or massive (*buoyancy*, *crust*, *pressure*) Noun plural (*data, lavas, measurements*) Verb, present, non-3rd p. (*affect, conclude, denote*)	+3 +3 +2 +2

TABLE 1: Profilage des contextes sources d'erreurs de traduction

Origines des erreurs en TS : différentiation textométrique grâce aux corpus de textes cibles annotés

On remarque que les profils des contextes sources de certaines erreurs sont proches. Dans ce cas, il s'agit le plus souvent des erreurs en relation de cooccurrence. Par exemple, les erreurs type « Distorsion » (77 contextes) et celles qui relèvent des problèmes de « Formulation maladroite » (72 contextes) partagent 15 contextes (phrases), par exemple :

> *(Original) The current melting points are up to 1000 K higher than the melting points obtained by the observation of surface motion of a laser-heated sample (8).§*
>
> *(Traduction) : Les points de fusion actuels ont une* ***température_supérieure, _jusqu'à_1000_K,_aux_points_de_fusion_observés****[Formulation-maladroite_LA-ST-AW] sur* ***la_surface_en_mouvement****[Distorsion_TR-DI] de l'échantillon chauffé par laser.§*

3.2 Analyse des résultats

Tous types d'erreurs confondus, les catégories les plus spécifiques des contextes sources problématiques sont :

1. les participes passés (indice de spécificité : +4, fréquence totale dans l'original : F_i=322, fréquence locale dans les phrases comportant des erreurs dans la traduction en français : k_{ij}=277)
2. l'auxiliaire *be* au passé (+3, F_i=64, k_{ij}=59)
3. les modaux (+2, F_i=88, k_{ij}=77)
4. les prépositions (+2, F_i =1 620, k_{ij} =1 316)
5. les noms au pluriel (+2, F_i =915, k_{ij}=746)
6. la préposition *to* (+2, F_i=238, k_{ij}=198)
7. les gérondifs et participes présents (+2, F_i=175, k_{ij}=149).

Par exemple :

> *(Original) Fluid* ***inclusions*** *in the vein* ***minerals*** *representative of first breakdown* ***fluids*** *and fluid* ***inclusions*** *in olivine-enstatite* ***representing*** *final breakdown* ***fluids, were analysed*** *by* ***crushing*** *the* ***samples*** *in vacuum. §*
>
> *(Traduction) Les échantillons d'inclusions fluides dans les veines minérales représentatives de* ***la****[Type-annotateur_UD] première rupture des fluides, et de celles représentant* ***la_rupture****[Terme-traduit-par-non-terme] finale, ont été brisés* ***dans_la_chambre_pour_analyse****[Distorsion]. §*

De façon générale, les diagnostics de *spécificités* indiquent qu'il y a moins d'erreurs dans la traduction des énoncés comportant des séquences de nombres, unités de mesure, symboles, références, noms propres (-4). Les phrases avec les verbes (avec ou sans auxiliaire) au présent à la 3ème personne du singulier posent également moins de problèmes de traduction (-3), par exemple :

> *(Original) Quartz **is** generally fine-grained and calcite **occurs** mostly as cement in the matrix (**Fig**. **2a**).§*
>
> *(Traduction) Le quartz est généralement composé de grains fins et la calcite se trouve essentiellement comme ciment dans la matrice (Fig. 2a). §*

Au total, les *spécificités* morpho-syntaxiques des contextes sources calculées au seuil fixé à 10 couvrent 2 699 occurrences de formes annotées. Les erreurs relevées dans l'annotation des productions totalisent 2 277 occurrences de formes annotées.

On note que les correspondances type *spécificités* sources/erreurs de traduction ne constituent pas des alignements parfaits au niveau sous-phrastique mais attirent l'attention sur des éléments caractéristiques de la langue source qui seraient à l'origine des problèmes de traduction. Par ailleurs, toutes les occurrences spécifiques (en gras ci-dessous) ne déclenchent pas systématiquement des erreurs ; on constate encore plusieurs types d'erreurs en co-occurrence qui partagent les mêmes contextes :

> *(Original) In **addition**, the **most** intense **mass** peaks of the second **category** correspond mainly to dioxygenated molecules and **exhibit** a more distinctive **preference** of even **number** of **carbon** over odd **number** than those corresponding to other extended **recurring** molecular series.§*
>
> *(Traduction)*
> ***En_outre,_cette_seconde_catégorie_présente_des_pics_de_masses_dont_les_plus_intenses**[Type-annotateur_TR-UD]* correspondent à des molécules dioxygénées, montrant une préférence plus marquée pour **un nombre** [Omission_TR-OM]** pair **de**[Formulation-maladroite_LA-ST-AW]*** carbones que d'autres séries moléculaires récurrentes d'une large étendue.§*
>
> *(Commentaires de l'annotateur)*
> **Vérifiez si le sens de votre traduction est bien le même que dans "the most intense mass peaks of the second category correspond mainly"*
> ***Il n'est pas inutile de rester précis dans ce type de texte et de traduire aussi "over odd number"*
> **** Il faut rendre la comparaison plus claire (cf. "than those")*

Pour analyser ce type de résultats dans une perspective contrastive, on mobilise la visualisation différentielle en contexte disponible dans *Le Trameur*.

3.3 Aides visuelles au repérage de la complexité en contexte

Dans *Le Trameur*, la *Lecture Textométrique Différentielle* (LTD) fournit des aides à la lecture contrastive de textes comparés appuyées par l'affichage synchrone des résultats de leur analyse textométrique parallèle (Patin et al., 2016). Les deux ensembles textuels sont affichés simultanément à l'écran pour faciliter les comparaisons. Les éléments caractéristiques sélectionnés par seuillage dans les contextes sources et cibles sont rendus « visibles » au fil des textes par un système de surlignage. Cette visualisation différentielle permet de cerner les facteurs qui sont à

l'origine de chaque type d'erreur de traduction, par l'examen contextuel des différentes causes et processus qui y sont liés.

La Figure 2 montre un extrait de la LTD générée par *Le Trameur* sur les phrases correspondant aux erreurs de syntaxe. Les résultats du calcul des *spécificités* morpho-syntaxiques dans les contextes sources couvrent 280 occurrences de formes graphique annotées. Les erreurs de syntaxe (70 occurrences de formes graphiques annotées) et les spécificités morpho-syntaxiques calculées sur les contextes originaux (280 occurrences de formes graphiques) sont surlignées en jaune. La visualisation attire l'attention sur la structure complexe des GN en anglais qui intègrent de nombreux termes composés (reflétée par la surreprésentation caractéristique des noms au singulier et au pluriel, cf. Table 1). La traduction en français nécessite dans ce cas la maîtrise des stratégies de traduction qui rétablissent les relations explicites entre les constituants dans la phrase au niveau micro-syntaxique (pour les apprenants, ces types de relations sont souvent peu explicites en anglais de spécialité). Ces diagnostics relevant de la complexité du texte scientifique en anglais peuvent déclencher des requêtes et vérifications dans les corpus comparables du domaine de spécialité, selon la méthodologie mise en place dans l'expérimentation (Kübler et al. 2016).

VOLET : EN	VOLET : FR
Thus, H 2 generation may be episodic depending on the rates of formation and destabilization of mineral surface layers during progressive waterrock interaction in an open system.§	Conclusion, la production de H 2 peut être réalisée en plusieurs stades, selon les taux de formation et de déstabilisation de couches superficielles minérales, pendant une interaction eau-roche progressive, au sein dun circuit ouvert. §
Therefore, the extraction of continental crust from this already-depleted reservoir (the EDR) cannot have greatly increased the Sm/Nd ratio of the MORB source; otherwise its 143Nd/144Nd value would have evolved to higher values than those observed for any terrestrial rock.§	Par conséquent, lextraction de croûte terrestre de ce réservoir, déjà appauvrit, ne peux avoir augmenté le rapport Sm/Nd de la source de basalte de dorsale médio-océanique, de façon conséquente. Sans quoi sa valeur en 143Nd/144Nd aurait évolué de manière plus importante que les valeurs observées sur nimporte quelle roche terrestre.§
the volume of mantle from which the continental crust was extracted must be large.§	le volume de croûte terrestre extrait du manteau se doit être inférieur à celui-ci.§
These mechanisms require a positive volume change of dehydration; otherwise, elevated pore pressures will not occur.§	Ces mécanismes requièrent un changement de volume positif de la déshydration; sans quoi il ne pourra pas y avoir de pressions interstitielles élevées.§
Above 6.5 GPa, the antigorite dehydration reaction had a negative Clapeyron slope and volume change of reaction.§	Au-dessus de 6,5 GPa, la réaction de la déshydratation de l'antigorite avait une pente de Clapeyron négative et un changement de volume de réaction négatif. §
Owing to the near-edge spectral characteristics (peak position, structure) for potential candidate (hydr)oxides (for example, ferrihydrite, haematite and goethite), we cannot uniquely identify the Fe(iii)bearing phase and henceforth use the term Fe(iii)-(hydr)oxides to indicate Fe bound to O and/or OH in a variety of crystal structures.§	En raison des caractéristiques spectrales près du seuil d'absorption (position maximale, structure) pour les candidats potentiels d'(hydr-) oxydes (tels que le ferrihydrite, l'hématite et la goethite), nous ne pouvons pas identifier définitivement la phase contenant du Fe(III) et utiliserons donc le terme (hydr-) oxydes de Fe(III) pour signaler l'existance d'un lien entre Fe et O et/ ou HO dans des diverses structures de cristaux. §
Hydrothermal organic reactions affect petroleum formation, degradation, and composition (2, 3), provide energy and carbon sources for deep microbial communities (4, 5), and may be important in the origin of life (6, 7).§	Les réactions organiques hydrothermales ont un effet sur la formation, la dégradation et la composition du pétrole, elles pourvoient de l'énergie et des sources en carbone pour des communautés microbiennes profondes et peuvent avoir leur importance dans l'origine de la vie. §
Basaltic lavas erupted at some oceanic intraplate hotspot volcanoes are thought to sample ancient subducted crustal materials.§	Des laves basaltiques en provenance de certains volcans à point chaud situés entre deux plaques océaniques constitueraient des échantillons d'anciens matériaux crustaux subduits. §
Here we report anomalous sulphur isotope signatures indicating mass-independent fractionation (MIF) in olivine-hosted sulphides from 20-million-year-old ocean island basalts from Mangaia, Cook Islands (Polynesia), which have been suggested to sample recycled oceanic crust.§	Dans cet article nous rapportons l'existence de signatures isotopiques atypiques du soufre présentes dans des sulfures hébergés dans des olivines en provenance de basaltes d'îles océaniques de Mangaia, Îles Cook (Polynésie), datant de 20 millions d'années. Ces signatures atypiques relèvent d'un phénomène de fractionnement indépendant de la masse (MIF) et constitueraient un échantillon de croute océanique recyclée.§
Terrestrial MIF sulphur isotope signatures (in which the amount of fractionation does not scale in proportion with the difference in the masses of the isotopes) were generated exclusively through atmospheric photochemical reactions until about 2.45 billion years ago.§	Les signatures isotopiques terrestres du sulfure produites par MIF (phénomène dans lequel le fractionnement n'est pas proportionnel à la différence de masse entre les isotopes) ont été générées exclusivement par réactions photochimiques atmosphériques qui ont eu lieu jusqu'il y a 2,45 milliards d'années. §
The RZ is composed of quartz, wollastonite (Wo, grey in Fig. 1a), Ca-rich garnet (Grt) and CM, and lacks calcite and phengite.§	La ZR est constituée de quartz, wollastonite (Wo, en gris dans la Fig. 1a), grenat roche en carbone (Grt), et matière carbonée. Elle ne contient pas de calcite, ni phengite. §
Using laser-heated diamond anvil cells, we constructed the solidus curve of a natural fertile peridotite between 36 and 140 gigapascals. §	À l'aide de cellules à enclume de diamants chauffées par laser, nous avons construit la courbe solidus d'une péridotite fertile et naturelle sous une pression allant de 36 à 140 Gigapascals. §

FIGURE 2: Spécificités sur contextes d'erreurs de syntaxe (export généré par *Le Trameur*)

4 Conclusion et perspectives

Nous avons avancé les premières propositions pour le profilage des erreurs de traduction en contextes alignés. La détection des difficultés d'apprentis traducteurs face aux textes originaux a mobilisé les annotations des productions selon la typologie d'erreurs MeLLANGE (Kübler et al., 2016) et la méthode des *spécificités* (Lafon, 1984 ; Lebart, Salem, 1994) avec la *Lecture Textométrique Différentielle* (LTD) sur contextes parallèles (Patin et al., 2016). Le repérage des éléments caractéristiques des contextes sources correspondant à un certain type d'erreur de traduction a permis de récolter des indices quantitatifs sur les constructions potentiellement complexes qui sont à l'origine des difficultés des apprenants (transfert de contenu, erreurs de langue, etc.).

Ces diagnostics peuvent être exploités de plusieurs façons. Ils peuvent alerter les apprenants eux-mêmes et donner lieu à des vérifications en corpus comparable du domaine de spécialité, mais aussi être utilisés par les enseignants pour proposer des exercices ciblés constitués à base de corpus. Cette voie ouvre des perspectives pour le développement des supports de cours informatisés qui allient la typologie d'erreurs issue de l'annotation des productions et l'exploration dynamique des contextes alignés et profilés par type d'erreurs.

Dans les expérimentations à venir, nous envisageons de comparer les erreurs de traductions dans les productions réalisées avec et sans apport de corpus (corpus *ER-TRAD-SP1* et *ER-TRAD-SP2*) afin d'observer si les *spécificités* des contextes sources liées aux erreurs changent en fonction des conditions de production des traductions. Ces nouveaux chantiers visent à alimenter la réflexion sur les méthodes d'enseignement qui amènent la diminution du nombre d'erreurs de traduction amorcée dans la première phrase d'expérimentation (Kübler et al., 2016).

Remerciements

Les auteurs remercient tous les membres de l'équipe CLILLAC-ARP (Paris 7), notamment Alexandra Mestivier (Volanschi) et Mojca Pecman, qui ont participé à l'annotation des erreurs de traduction et à la création des corpus utilisés dans ce volet de l'étude.

Références

BOWKER L., BENNISON P. (2003). Student Translation Archive and Student Translation Tracking System. Design, Development and Application. In Zanettin F., Bernardini S. and Stewart D. editors, *Corpora in translator education*. Manchester: St. Jerome Publishing.

CASTAGNOLI, S., CIOBANU D., KÜBLER N., KUNZ K., VOLANSCHI A. (2011). Designing a Learner Translator Corpus for Training Purposes. In Kübler N. editor, *Corpora, Language, Teaching, and Resources: From Theory to Practice*. Bern: Peter Lang.

Origines des erreurs en TS : différentiation textométrique grâce aux corpus de textes cibles annotés

FLEURY S., ZIMINA M. (2014). Trameur: A Framework for Annotated Text Corpora Exploration. Proc. of *COLING 2014 (the 25th International Conference on Computational Linguistics: System Demonstrations)*, August 2014, Dublin, Ireland, 57-61.

FRANKENBERG-GARCIA, A. (2015). Training translators to use corpora hands-on: challenges and reactions by a group of 13 students at a UK university. *Corpora*, 10/2, 351-380.

FRÉROT C. (2010). Outils d'aide à la traduction : pour une intégration des corpus et des outils d'analyse de corpus dans l'enseignement de la traduction et la formation des traducteurs. *Les Cahiers du GEPE* 2/2010, Outils de traduction - outils du traducteur ?

FROELIGER N. (2013). *Les Noces de l'analogique et du numérique - De la traduction pragmatique*. Paris : Les Belles lettres (collection Traductologiques).

GRANGER S., HUNG J., PETCH-TYSON S. (eds.) (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam and Philadelphia: John Benjamins.

KÜBLER, N. (2011). Working with different corpora in translation teaching. In Frankenberg-Garcia A., Flowerdew L., and Aston G. editors, *New Trends in Corpora and Language Learning*. London: Continuum.

KÜBLER N, YVON F., WISNIEWSKI G. (2013). Human Errors and Automatic Errors in Machine Translations. What are the Differences? *Errare Workshop*, Ermenonville 2013.

KÜBLER N., MESTIVIER A., PECMAN M., ZIMINA M. (2016). Exploitation quantitative de corpus de traductions annotés selon la typologie d'erreurs pour améliorer les méthodes d'enseignement de la traduction spécialisée. Actes des *13èmes Journées internationales d'analyse statistique des données textuelles (JADT 2016)*, 7-10 juin, Nice, France.

LAFON P. (1984). *Dépouillements et statistiques en lexicométrie*. Slatkine-Champion, Genève-Paris.

LEBART L., SALEM A. (1994). *Statistique textuelle*. Dunod.

LOOCK R., MARIAULE M., OSTER C. (2014). Traductologie de corpus et qualité : étude de cas, Actes du colloque *Tralogy II*, CNRS, 17-18 janvier, Paris, France.

PATIN S., ZIMINA M., FLEURY S. (2016). Lecture Textométrique Différentielle (LTD) de textes législatifs comparables de l'Union européenne. Actes des *13èmes Journées internationales d'analyse statistique des données textuelles (JADT 2016)*, 7-10 juin, Nice, France.

PEARSON J. (2003). Using parallel texts in the Translator Training Environment. In Zanettin F., Bernardini S. and Stewart D. editors, *Corpora in Translator Education*. Manchester: St Jerome Publishing.

WISNIEWSKI G., KÜBLER N., YVON F. (2014). A Corpora of Machine Translation Errors Extracted from Translation Students Exercises. Proc. of the *Ninth Language Resources and Evaluation Conference (LREC 2014)*, 26-31 May, Reykjavik, Iceland.

ZIMINA M. ET 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* 11(1).