Institut National des Langues et Civilisations Orientales

Département Textes, Informatique, Multilinguisme

Évaluation de la qualité de la TAN français-chinois fondée sur une typologie d'erreurs : le cas de la traduction littéraire

MASTER

TRAITEMENT AUTOMATIQUE DES LANGUES

Parcours :

Technologie de la Traduction et Traitement des Données Multilingues

par

Xinyi ZHONG

Directeur de mémoire :

Ilaine WANG

Encadrant :

Victorien VILLIERS

Année universitaire 2021/2022

Contents

List of Figures

List of Tables

Remerciement

Je voudrais remercier les enseignants de l'école, en particulier mon superviseur M. Victorien Villiers qui a été très prodigue de conseils et m'a encouragé à mettre en pratique mes idées pendant mon stage et mon mémoire.

Je remercie Mme Ilaine Wang pour sa grande patiente, sa bienveillance et la rigueur avec laquelle elle m'a aider à corriger mon mémoire.

Je tiens à remercier mon maître de stage, Mme Caroline Rossi pour l'occasion qu'elle m'a donnée d'approfondir ma compréhension de la traduction automatique et de la recherche sur les erreurs.

Je remercie ensuite mes camarades de classe, en particulier Liu Lufei, qui m'a beaucoup aidé à combler mes lacunes de compréhension du français.

Je remercie He Huan pour m'avoir donné des conseils sur l'annotation des erreurs et sur mon mémoire.

Je remercie mes collègues (Liu Xingyu, Dong Ruoran, Minh-Châu NGUYÊN, Zhou Yongxin, Fiorella, Sannara Ek, Eric Le Ferrand, Aidan M., Soline) de l'équipe GETALP du laboratoire d'informatique de Grenoble pour leur soutien au déroulement de mon stage.

Enfin, Je tiens à remercier mes parents et mon mari Geoffrey Williams Stücklin pour leur soutien de tous les instants.

RÉSUMÉ

La naissance de la technologie des réseaux neuronaux permet non seulement de rendre plus rapide la traduction automatique (noté TA ci-dessous), mais aussi d'en améliorer la qualité. Néanmoins, la TA souffre encore de nombreuses erreurs qui varient largement d'une paire de langues à l'autre. Pour les paires de langues qui ne sont pas apparentées (par exemple, français-chinois), la qualité de la TA souffre encore de tares importantes. L'analyse des erreurs est donc nécessaire, que ce soit pour sensibiliser les utilisateurs aux imperfections de la TA ou pour tenter d'affiner cette technologie. De ce fait, l'évaluation de la qualité et l'analyse des erreurs de traduction automatique restent des sujets de recherche importants dans le domaine du traitement automatique des langues.

Dans le contexte de démocratisation des moteurs de traduction automatique neuronale et d'attention croissante accordée à l'analyse des erreurs, nous explorerons les manières de développer une typologie d'erreurs pour la TA du français vers le chinois et les modalité de construction d'un corpus parallèle d'annotations d'erreurs.

Nous nous intéresserons dans un premier temps à l'histoire de la traduction automatique, aux différentes méthodes d'évaluation de la qualité de la traduction automatique, aux types d'erreurs ainsi qu'à l'annotation d'erreurs. Ensuite, nous présentons des expériences qui comprennent la préparation de jeux de donnée, l'application des pré-traitements et l'annotation des erreurs. Finalement, nous effectuerons une analyse quantitative fondée sur des statistiques des erreurs d'adéquation et de fluidité annotées. Nous discutons également des résultats sur des causes des erreurs et des difficultés d'annotation pour une analyse qualitative.

Mots clés : évaluation humaine, traduction automatique neuronale, typologie d'erreurs, langue chinoise, annotation des erreurs de traduction

INTRODUCTION

Présentation générale

À l'heure du foisonnement touts azimuts de l'information à l'échelle planétaire, la demande de traduction rapide de contenus met à rude épreuve la capacité des traducteurs. Aussi, la demande croissante de ressources multilingues a-t-elle fortement stimulé le développement de l'industrie mondiale de la traduction automatique (noté TA ci-dessous) [Way, 2018]. Or, la naissance de la technologie des réseaux neuronaux permet non seulement de rendre plus rapide la traduction automatique, mais aussi d'en améliorer la qualité. En vue de réduire les coûts et d'accroître l'efficacité de la traduction, nombreux sont ceux qui utilisent des moteurs de traduction automatique neuronale (TAN) librement accessibles en ligne. Signe d'une adoption rapide des outils de TAN par le marché de la traduction en France, la chambre syndicale des agences de traductions [1], prédit que 78 % des agences de traduction envisage d'initier ou d'augmenter le recours à la TA ou à la post-édition de TA (SFT [2], 2022).

Néanmoins, la TA souffre encore de nombreuses erreurs qui varient largement d'une paire de langues à l'autre. Si, pour les paires de langues proches (par exemple français-anglais), la TA peut produire une qualité avoisinant la traduction humaine, ses failles, loin d'être inexistantes, sont d'autant plus ténues [Sharou and Specia, 2022]. Pour les paires de langues qui ne sont pas apparentées (par exemple, français-chinois), la qualité de la TA souffre encore de tares importantes. L'analyse des erreurs est donc nécessaire, que ce soit pour sensibiliser les utilisateurs aux imperfections de la TA ou pour tenter d'affiner cette technologie. De ce fait, l'évaluation de la qualité et l'analyse des erreurs de traduction automatique restent des sujets de recherche importants dans le domaine du traitement automatique des langues. De nombreux projets récents, visant à développer des méthodes et outils à même de rendre les systèmes de TAN explicables, se sont concentrés sur les caractéristiques et les causes des erreurs de la TA, l'annotation des erreurs du corpus parallèle, et les typologies des erreurs.

C'est dans ce contexte de démocratisation des moteurs de TAN et d'attention croissante accordée à l'analyse des erreurs que s'inscrivent les travaux présentés dans ce mémoire. Nous explorerons les manières de développer une typologie d'erreurs pour la TA du français vers le chinois et les modalité de construction d'un corpus parallèle d'annotations d'erreurs.

Problématique

Les questions auxquelles ce mémoire tente de répondre sont les suivantes :

1. Quels types d'erreurs se présentent dans la TAN du français au chinois ?

1. https://cnetfrance.org/
2. Société française des traducteurs, https://www.sft.fr/fr/nos-marches

2. Comment annoter ces erreurs dans un corpus parallèle français-chinois ?
3. Parmi les types d'erreurs annotés, lesquels sont les plus fréquents ?
4. Pourquoi ces erreurs se produisent-elles ?

Plan de lecture

Ce mémoire est constitué des chapitres suivants :

- Dans l'introduction, nous présenterons le contexte, la problématique traitée et le plan suivi dans le mémoire.
- Dans le chapitre 1 (état de l'art), nous nous intéresserons à l'histoire de la traduction automatique, aux différentes méthodes d'évaluation de la qualité de la traduction automatique, aux types d'erreurs ainsi qu'à l'annotation d'erreurs.
- Le chapitre 2 (corpus) présente des expériences qui comprennent la préparation de jeux de donnée et l'application des pré-traitements.
- Dans le chapitre 3 (annotation), nous expliquons la typologies d'erreurs employée, l'outil d'annotation utilisé et les démarches d'annotation des erreurs.
- Dans le chapitre 4 (résultats et discussion), nous présentons tout d'abord une analyse quantitative fondée sur des statistiques des erreurs d'adéquation et de fluidité annotées. Ensuite, nous discutons des résultats sur des causes des erreurs et des difficultés d'annotation pour une analyse qualitative.
- Dans le chapitre 5 (conclusion générale), nous dresserons la conclusion de notre mémoire.
- Dans le chapitre 6 (perspectives), nous discutons des limites de notre mémoire et des travaux futurs.

CHAPTER 1

ÉTAT DE L'ART

Sommaire

1.1 Introduction

Dans ce chapitre, nous commençons par un bref aperçu de l'histoire de la TA et des principales approches. Nous présentons ensuite les méthodes et les métriques d'évaluations humaine et automatique. Enfin, nous aborderons les typologies d'erreurs et les outils d'annotation des erreurs.

1.2 Traduction automatique (TA)

1.2.1 Historique

L'histoire du développement de la TA peut être divisée en trois périodes : les années 1950-1960, les années 1970-1980, les années 1990-2013 et de 2013 à aujourd'hui.

Les années 1950-1960

En 1949, Warren Weaver a élaboré un mémorandum sur la traduction intitulé « Translation »[1]. Cet événement emblématique est considéré comme le début des recherches du domaine de la TA. Dans ce mémorandum, l'auteur a proposé d'utiliser

1. https://repositorio.ul.pt/bitstream/10451/10945/2/ulfl155512_tm_2.pdf

la cryptographie au service de la TA. Il a également déclaré dans une lettre à Norbert Wiener que la traduction pouvait se résoudre à un processus de déchiffrage. Ces explorations ont jeté les bases de la « traduction automatique ».

En 1954, l'université de Georgetown et IBM ont mené ensemble une démonstration importante de traduction automatique [Hutchins, 2004] conçue par Leon Dostert et Paul Garvin, chercheurs en linguistique à Georgetown, et Cuthbert Hurd et Peter Sheridan, membres du personnel d'IBM, à l'aide de l'ordinateur IBM 701. Les chercheurs ont programmé six règles grammaticales pour traduire 60 phrases (soit 250 mots) du russes vers l'anglais. Ce fut la première démonstration publique d'un système de traduction automatique.

Cependant, les recherches sur la TA qui ont suivi ne disposaient pas d'une analyse linguistique approfondie et de nombreuses questions essentielles sont restées sans réponses. En 1966, la commission *Automatic Language Processing Advisory Committee* (noté ALPAC ci-dessous) publie un rapport célèbre intitulé « Language and Machines »[2]. Le rapport a affirmé que la TA peine à surmonter les "barrières sémantiques" et que la qualité de TA est nettement inférieure à celle des traductions humaines. Ce rapport a infirmé la faisabilité de la TA et a préconisé aux grandes organisations de mettre un terme aux recherches et investissements dans ce domaine.

Les années 1970-1980

Le rapport de l'ALPAC a freiné la recherche sur la TA aux États-Unis. Cependant, elle s'est poursuivie dans d'autres pays.

Au Canada, le projet TAUM (Traduction Automatique de l'Université de Montréal) a développé le système intitulé « Météo » qui a permis de traduire automatiquement des bulletins météorologiques [Hutchins, 2001].

Par la suite, en Europe, les besoins en traduction deviennent de plus en plus importants. De nombreux moteurs de TA ont émergé au cours de cette période. Le système Systran est non seulement mis en place à la Commission européenne en 1976, mais aussi adopté et installé au sein d'autres organismes ou entreprises tels que l'OTAN ou Xérox [Piggott, 1992]. Par ailleurs, d'autres concurrents apparaissent, comme Logos (système développé par Bernard E. Scott) [Scott, 1989] ou METAL (système développé au Texas) [Little, 1989].

La fin des années 1980 a été marquée par une résurgence et un renouveau de la TA. À cette époque, l'avènement des microprocesseurs a permis un véritable bond qualitatif de la puissance des ordinateurs. En outre, nombre de travaux fondamentaux en informatique et en linguistique ont été menés, comme en témoigne les algorithmes essentiels, et les analyses lexicale et syntaxique.

Dès années 1990 à 2013

Depuis la publication de l'article [Brown et al., 1988], l'approche de recherche de la traduction automatique a également évolué, passant d'une approche à base de règles à une approche fondée sur les statistiques qui repose sur des ensembles de textes pour construire un modèle de traduction. Les différentes générations de systèmes de traduction automatique statistique utilisait des approches distinctes fondées sur des mots ou des phrases ([Brown et al., 1993] ; [Koehn et al., 2003]).

2. Disponible librement en ligne : https://nap.nationalacademies.org/resource/alpac_lm/ARC000005.pdf

Par ailleurs, les années 1990 ont été marquées par l'apparition des outils de traduction assistée par ordinateur (TAO) [Barbin, 2020], en particulier au sein des agences commerciales, des services des administrations publiques et des entreprises multinationales, où les traductions sont produites à grande échelle.

L'énorme demande du marché a contribué au développement de la traduction automatique statistique. De nombreuses entreprises ont également participé à la recherche dans ce domaine. En avril 2006, Google a lancé son service de traduction gratuit, un moteur de SMT fondé sur des phrases. Google sera ensuite rapidement suivi par Microsoft et Baidu. En 2007, Moses, un moteur de traduction automatique statistique librement accessible, a été rendu publique [Koehn et al., 2007]. Le décodeur de Moses peut être utilisé pour entraîner des modèles statistiques de traduction de textes d'une langue source vers une langue cible.

Depuis 2013 à aujourd'hui

En 2013, une nouvelle approche de TA a été introduite dans l'article « *Recurrent Continuous Translation Models* » (modèles de traduction continus récurrents) [Kalchbrenner and Blunsom, 2013]. Cet article a présenté une classe de modèles probabilistes de traduction continue, nommés « modèles de traduction continue récurrents », qui sont purement basés sur des représentations continues pour les mots, les phrases et les expressions et ne reposent pas sur des alignements ou des unités de traduction syntagmatique. Cette initiative a débouché sur des recherches dans le domaine de la traduction automatique neuronale (TAN).

En septembre 2016, l'équipe Google Brain a publié un blog montrant qu'elle avait remplacé la traduction basée sur les phrases par la TAN dans la traduction chinois-anglais de son produit Google Translate [3]. À peine un an plus tard, le Facebook AI Institute (FAIR) a annoncé son approche de la TAN à l'aide d'un modèle de phrase convolutif (CNN), une méthode capable d'atteindre des performances similaires à la TAN basée sur un modèle de langue récurrent (RNN), mais fonctionnant 9 fois plus vite que cette dernière. En juin de la même année, Google a publié le modèle de TAN basé uniquement sur le mécanisme de « *self-attention* », qui n'utilise ni CNN ni RNN [Vaswani et al., 2017]. Cette nouvelle architecture appelée *Transformer* prend en compte le sens des mots en contexte, ce qui a considérablement amélioré la qualité de la traduction.

Désormais, la majorité des moteurs de traduction automatique en ligne sont des systèmes neuronaux, notamment Google Translate et DeepL (ce dernier est utilisé par la suite).

1.2.2 Approches principales

Dans l'histoire, la TA a présenté trois approches différentes : l'approche basée sur les règles, l'approche basée sur les corpus, et l'approche basée sur l'apprentissage profond.

Les systèmes basés sur les règles utilisent des dictionnaires bilingues et monolingues, des grammaires et des règles de transfert pour créer des traductions [Castilho et al., 2017]. Les problèmes d'un tel système comprennent le manque de fluidité, des difficultés de gérer les exceptions aux règles, les coûts élevés de développement et de personnalisation.

3. https://ai.googleblog.com/2016/09/a-neural-network-for-machine.html

Les systèmes basés sur les corpus sont divisés en systèmes de traduction statistiques et en systèmes basés sur des exemples. L'avantage de ces systèmes par rapport aux systèmes sur les règles est leur performance robuste lors de la sélection du lexique. De plus, ils nécessitent moins d'efforts humains pour être entraînés. Cependant, la qualité de traduction demeure imprévisible. Les traductions qu'ils produisent sont parfois mal structurées ou comportent des erreurs grammaticales, sans compter la difficulté de trouver des corpus pour certains domaines thématiques ou certaines paires de langues.

Les progrès récents des systèmes de TA viennent néanmoins bousculer ce ressenti, faisant apparaître la traduction machine comme une alternative crédible à la traduction humaine [Peraldi, 2016].

1.3 Évaluation de TA

1.3.1 Évaluation humaine

En ce qui concerne l'évaluation humaine, elle s'articule traditionnellement autour des critères d'adéquation et de fluidité [Rivera-Trigueros, 2021]. L'adéquation évalue la qualité sémantique, autrement dit, si l'information a été correctement transmise ou non, ce qui nécessite une comparaison avec des traductions de référence (monolingue) ou avec le texte source (bilingue). Le critère de fluidité vise à apprécier la qualité syntaxique de la traduction. Pour ce faire, la comparaison avec le texte source n'est pas indispensable et l'évaluation se fait dans un mode monolingue. Nous analyserons ci-dessous la typologie d'erreurs de TA sur la base de ces deux volets. Notons qu'il existe d'autres méthodes d'évaluation qui se fonde sur des critères de lisibilité, de compréhension, de convivialité et d'acceptabilité des traductions [Castilho et al., 2018]. Toutefois, ces critères ne seront pas abordés dans le présent mémoire.

Les méthodes d'évaluation les plus courantes sont l'échelle de Likert, la correction d'erreurs et le remplissage de textes à trous [Chatzikoumi, 2019]. Il est important de faire remarquer que l'identification, l'annotation et la classification des erreurs est une autre méthode d'évaluation humaine largement utilisée [Rivera-Trigueros, 2021], et qui sera par ailleurs utilisée par la suite dans le présent mémoire.

L'évaluation humaine a longtemps été considérée comme le critère d'évaluation indépassable de la qualité des traductions automatiques. Toutefois, l'appréciation humaine présente de nombreux inconvénients, comme le fait d'être chronophage, coûteuse, non reproductible et, dans de nombreux cas, incohérente (subjective) entre les évaluateurs humains. Les évaluations automatiques revêtent par conséquent une nécessité à la fois technique et pratique.

1.3.2 Évaluation automatique

En général, les métriques d'évaluation automatique comparent la sortie d'un système de TA avec une ou plusieurs traductions de référence [Rivera-Trigueros, 2021]. Elles calculent principalement la similarité entre la sortie de la TA et la traduction de référence afin d'évaluer la qualité de la TA [Han, 2022].

De même que l'évaluation humaine, l'évaluation automatique peut être classée en catégories syntaxiques et sémantiques. Les métriques syntaxiques se concentrent sur les informations syntaxiques relatives aux caractéristiques des mots, des syntagmes et de la structure des phrases. Les métriques sémantiques se penchent sur

les significations de la langue, notamment les significations lexicales, grammaticales et pragmatiques [Han, 2022]. Plus précisément, d'après un rapport publié par cette société en 2022 [4], les outils d'évaluation automatique peuvent évaluer la qualité des TA en comparant la similarité syntaxique ou la similarité sémantique entre les traductions automatiques et humaines (Tableau 1.1) .

Métriques syntaxique	Métriques sémantique
TER [Snover et al., 2006]	BERTScore [Zhang et al., 2019]
hLEPOR [Han et al., 2013]	PRISM [Thompson and Post, 2020]
SacreBLEU [Post, 2018]	COMET [Rei et al., 2020]

Table 1.1: Métriques principales d'évaluation automatique

La métrique TER (*Translation Error Rate*) permet de quantifier les opérations d'édition qu'une TA requiert pour correspondre à une traduction de référence. Elle tente de résoudre le problème soulevé par le WER (*Word Error Rate*) en interdisant le réagencement des mots. La métrique WER, une des premières métriques utilisées, se base sur la distance de Levenshtein ou distance d'édition ([Levenshtein, 1965] ; [Nießen et al., 2000]).

Nous souhaitons préciser que la métrique SacreBLEU permet de calculer sans problème des scores BLEU partageables, comparables et reproductibles [5]. Le score BLEU est une métrique d'évaluation automatique qui compare la similitude entre une TA et une (ou plusieurs) traduction(s) de référence, produisant un score plus élevé lorsque le TA se rapproche de la production humaine [Papineni et al., 2002]. Il a été conçu à l'origine pour calculer la similitude au niveau du document plutôt qu'au niveau de la phrase.

Enfin, nous abordons le BERTscore que nous utilisons dans le présent mémoire, qui est basé sur BERT et les modèles pré-entraînés analogues qui permettent l'intégration de connaissances contextuelles. Il réalise un plongement de mots basé sur le contexte pour chaque token de la TA et fait de même pour la traduction de référence, avant d'effectuer une comparaison par paires et un calcul du degré de similitude Cosinus pour chaque token de la TA et de la traduction de référence. Il ne pénalise ni les paraphrases ni les synonymes. Comme le corpus que nous utilisons relève du domaine littéraire (section 2.3.1), domaine caractérisé par une diversité de formulations et de syntaxes plus marquée que dans les textes techniques, une métrique sémantique semble être la plus indiquée.

Si l'évaluation automatique présente de nombreux avantages (le faible coût, la rapidité, la reproductibilité et la possibilité d'ajuster et d'optimiser les paramètres des modèles de traduction automatique [Han, 2022]), les résultats des métrique d'évaluation sont représentés par un nombre entre zéro et un. Ainsi, ces résultats ne donnent aucunes précisions sur les problèmes spécifiques d'un système de TA.

4. Disponible librement en ligne : https://inten.to/machine-translation-report-2022/
5. https://github.com/mjpost/sacrebleu

1.4 Typologies et annotation d'erreurs

1.4.1 Typologies d'erreurs

Il existe de nombreuses recherches sur les types d'erreurs de TA dans différentes paires de langues. Nous présentons ci-dessous les types les plus représentatifs ou les plus récents.

Nous abordons d'abord la typologie proposée par [Vilar et al., 2006]. Cette dernière utilise un schéma hiérarchique et divise les erreurs de TA en 5 grandes catégories : mots manquants, ordre des mots, mots incorrects, mots inconnus et ponctuation (cf. l'arborescence Figure 1.1). Les quatre premiers types ont tous des sous-catégories.

Les chercheurs David Vilar et al. se concentrent sur trois directions de traduction : espagnol vers anglais, anglais vers espagnol et chinois vers anglais. Certaines catégories d'erreurs ont été modifiées pour tenir compte des erreurs de TA anglais-chinois. Ils utilisent un système de traduction automatique statistique de l'Université technique de Rhénanie-Westphalie (RWTH) qui est basé sur une combinaison de sept modèles différents, les plus importants étant des modèles basés sur des phrases dans les deux sens source-cible et cible-source et un modèle de langue cible.

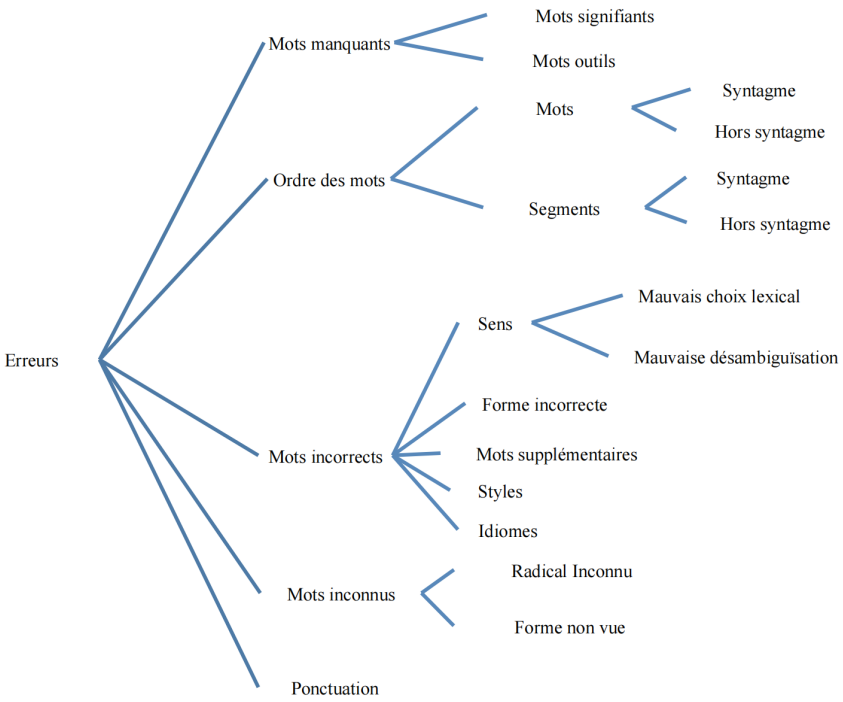

Figure 1.1: Typologie d'erreurs de Vilar et al. [Esperança-Rodier, 2018]

Deuxièmement, nous présentons la typologie d'erreurs de SCATE (*Smart Computer-aided Translation Environment*) [Tezcan et al., 2016]. Comme nous pouvons constater dans le tableau Figure 1.2, cette topologie est conçue également de manière hiérarchique et contient différentes sous-catégories fondée sur les notions traditionnelles d'adéquation et de fluidité. L'identification des erreurs d'adéquation nécessite une lecture bilingue du texte, car il est impératif de tenir compte à la fois du texte source et de la TA.

Les chercheurs Arda Tezcan et al. élaborent non seulement une définition claire de chaque type d'erreur, mais fournissent aussi des indications sur l'annotation des erreurs de TA, par exemple l'outil utilisé, les étapes à suivre et les précautions [6]. L'une des étapes les plus intrigantes est la suivante. Ils montrent comment définir les limites des erreurs d'adéquation dans la TA et les lier aux éléments de texte source correspondants, ce qui nous permet de mieux comprendre la nature des erreurs de TA et de repérer les éléments de texte source qui échappe à l'appréhension d'un système de TA. Par exemple, dans le cas d'une erreur sur un nom, les chercheurs préconisent de l'annoter avec son article afin d'harmoniser la méthodologie d'annotation et réduire ainsi l'influence de la subjectivité humaine.

Les caractéristiques de SCATE ci-dessus ont répondu aux besoins de la présente recherche. C'est sur la base de cette méthode que la typologie d'erreurs a été construite.

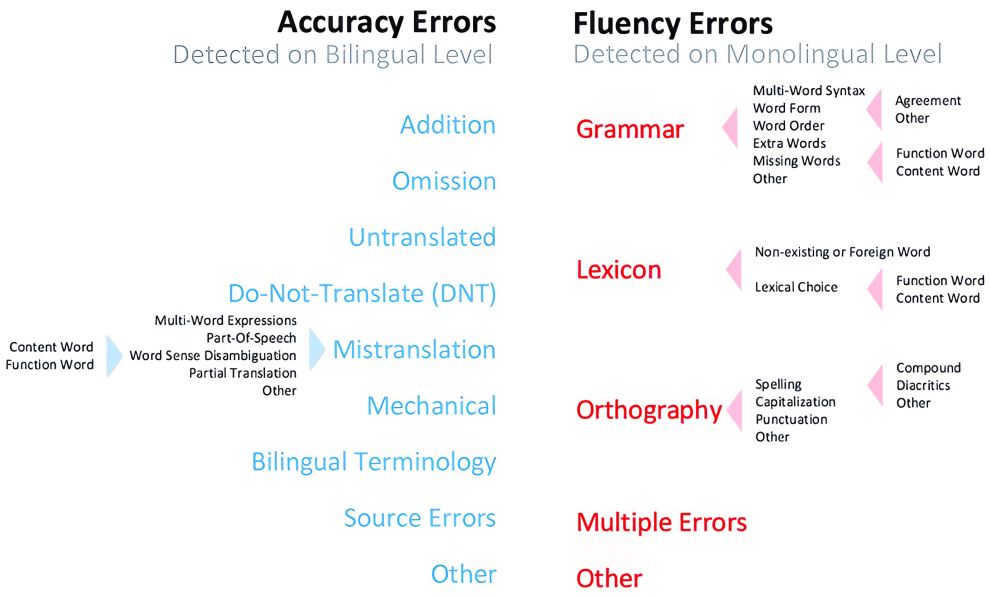

Figure 1.2: Typologie d'erreurs de SCATE [Tezcan et al., 2016]

La troisième typologie principale est celle élaboré par [Lommel et al., 2014] qui ont fait appel aux concepts d'adéquation et de fluidité pour établir la typologie d'erreurs de TA dans les métriques multidimensionnelles de qualité (MQM [7]). La plus grande différence entre cette typologie et les deux précédentes est qu'elle introduit la notion de spécifications de traduction basées sur un ensemble de paramètres de traduction dont les valeurs dépendent principalement de facteurs externes au texte source, tels que le lecteur et son objectif [Mariana, 2014], par exemple le style, le registre et la convention locale.

Cette typologie a également inspiré notre recherche. Le texte source que nous avons utilisé est tiré d'un roman, écrit dans un style léger et humoristique. Les caractéristiques linguistiques du français et du chinois (mandarin) étant très différentes, il serait difficile de reproduire le style du texte source si la traduction ne tenait pas compte du ressenti des lecteurs de la langue cible et ne recherchait que la fidélité au texte source.

6. Disponible librement en ligne : https://users.ugent.be/~atezcan/
7. Multidimensional Quality Metrics, http://themqm.info/typology/

Hormis ces trois typologie d'erreur, notons qu'il existe des recherches qui tiennent en compte de la pondération de la gravité d'erreurs, ou qui se limitent aux erreurs critiques, par exemple [Toudic et al., 2014], [Peraldi, 2016] et [Sharou and Specia, 2022] (Figure 1.3).

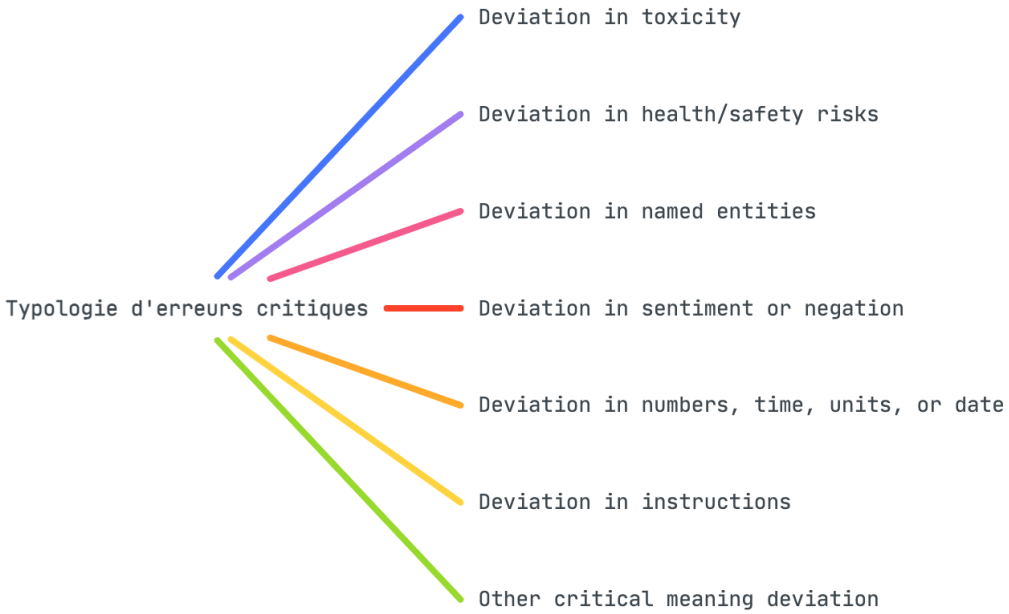

Figure 1.3: Typologie d'erreurs critiques

1.4.2 Outils d'annotation d'erreurs

Nous avons mentionné ci-dessus que les systèmes de TA sont uniquement évalués quantitativement par l'utilisation de métriques (section 1.3.2). Ces dernières ne nous permettent pas d'identifier les erreurs précises d'un système de traduction automatique. Il est donc recommandé de recourir à l'analyse des erreurs par des évaluateurs humains, qui identifient et annotent les erreurs dans les phrases traduites par la machine. Pour ce faire, nous devons utiliser l'outil d'annotation des erreurs.

Dans la partie suivante, nous présentons trois outils d'annotation : BLAST [8], ACCOLÉ [9] et INCEpTION [10].

BLAST est un outil graphique permettant d'effectuer une analyse des erreurs humaines, à partir de tous systèmes de TA et toutes paires de langues (Figure 1.4). Il dispose de trois modes de travail pour la gestion des annotations d'erreur : l'ajout de nouvelles annotations, la modification des annotations existantes et la recherche parmi les annotations. BLAST fournit deux types d'annotations : les annotations d'erreur et les annotations de support. Le premier type d'annotation est fondé sur une typologie hiérarchique d'erreurs et est créé par les utilisateurs, tandis que le second type d'annotation est créé automatiquement par la plate-forme mais peut être modifié [Stymne, 2011].

ACCOLÉ est une plate-forme collaborative d'annotation d'erreurs pour des corpus alignés. Elle permet l'annotation d'erreurs de traduction selon des typologies

8. https://cl.lingfil.uu.se/~sara/blast/
9. https://lig-membres.imag.fr/fbrunet/accole-plateforme-pour-ledition-collaborative-derreurs
10. https://inception-project.github.io/

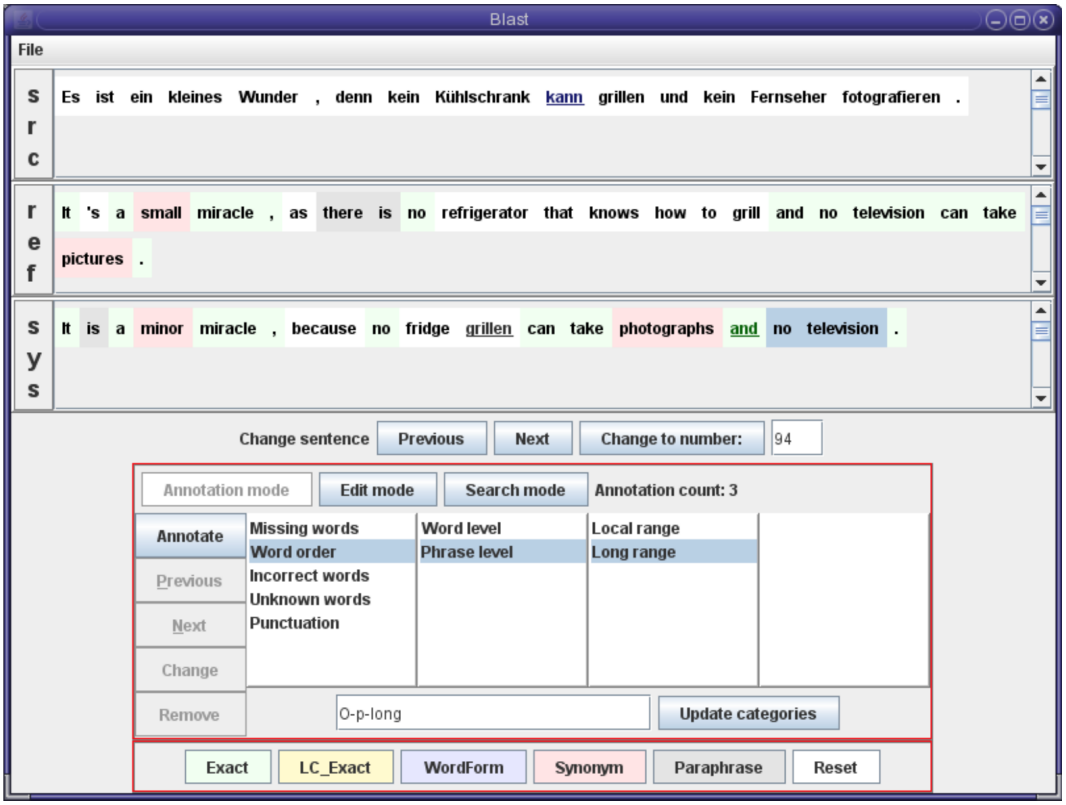

Figure 1.4: Capture d'écran d'interface de BLAST [Stymne, 2011]

d'erreurs intégrées ou téléchargées, sur plusieurs corpus annotés de textes différents, traduits par différents systèmes de TA statistique ou neuronale (Figure 1.5). Cette plate-forme propose 15 corpus, 19 projets comprenant 248 784 mots et 23 525 annotations [Brunet-Manquat and Esperança-Rodier, 2018].

INCEpTION est une nouvelle plate-forme d'annotation pour des tâches comprenant l'annotation interactive et sémantique [Klie et al., 2018b]. Elle intègre des fonctionnalités d'apprentissage automatique qui aident et guident automatiquement les annotateurs dans ces tâches très chronophages et laborieuses (Figure 1.6).

Nous utilisons la plate-forme INCEpTION pour deux raisons. Premièrement, nous avons déjà utilisé cette plate-forme pour d'autres projets et nous en connaissons le fonctionnement. Deuxièmement, notre guide d'annotation s'inspire de SCATE [11], qui utilise un outil d'annotation appelé BRAT [12], intégré dans INCEpTION. Les méthodes d'annotation sont donc similaires entre les deux [Tezcan et al., 2016].

1.5 Conclusion

En résumé, nous avons appris à connaître les caractéristiques des technologies de traduction automatique à différents stades de développement. Dans le contexte de l'utilisation de plus en plus courante de la traduction automatique, l'évaluation de la qualité de la traduction automatique est devenue particulièrement importante. L'évaluation humaine traditionnelle et l'évaluation automatiques émergentes ont leurs propres avantages et inconvénients. Bien que les outils automatiques soient

11. Disponible librement en ligne : https://users.ugent.be/atezcan/
12. https://brat.nlplab.org/

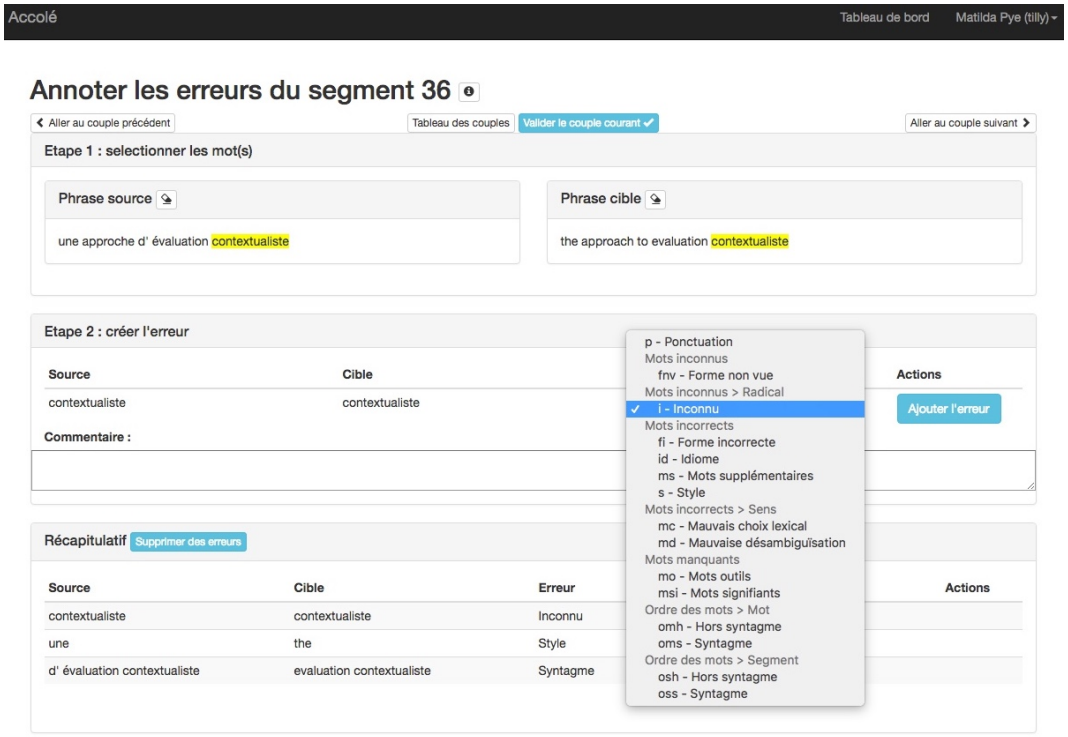

Figure 1.5: Capture d'écran d'interface d'ACCOLÉ

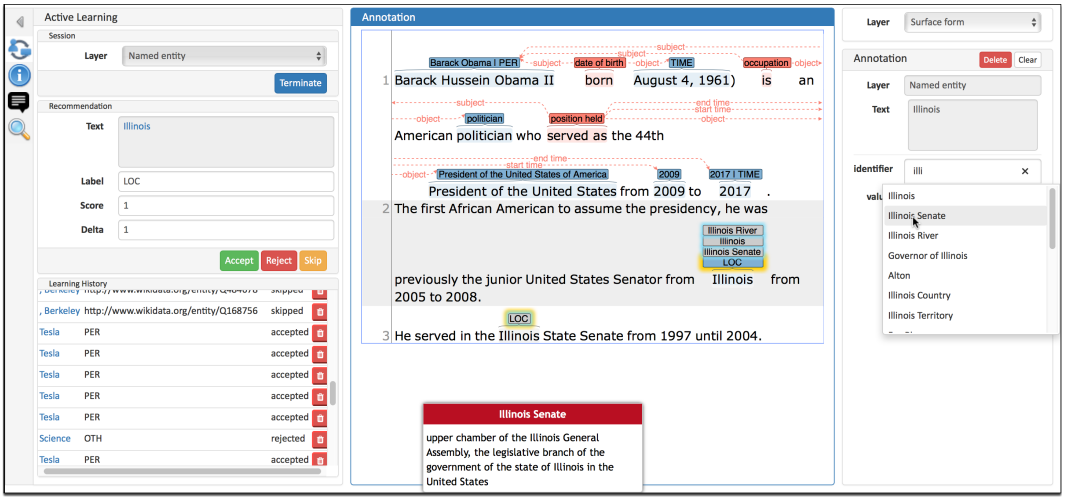

Figure 1.6: Capture d'écran d'interface d'INCEpTION

largement utilisés, les résultats ne sont que des scores et l'utilisateur n'a aucun moyen de savoir exactement ce qui ne va pas dans la traduction. En construisant des typologies d'erreurs et des annotations d'erreurs, nous pouvons savoir exactement quelles sont les erreurs de traduction automatique pour une paire de langues spécifique et contribuer à améliorer la technologie de la traduction automatique.

CHAPTER 2

CORPUS

Sommaire

2.1 Introduction

Dans ce chapitre, nous allons décrire, dans un premier temps, les concepts clés, la préparation de nos données, les caractéristiques du texte source et la génération de la TA. Ensuite, nous présenterons les pré-traitements appliqués pour construire les corpus. Enfin, nous allons passer en revue les étapes de l'annotation.

2.2 Présentation des concepts clés

1) « paragraphe » : un ensemble de segments séparé par un retour à la ligne.

2) « texte cible » : le texte écrit dans la langue source.

3) « texte cible » : le texte écrit dans la langue cible.

2.3 Préparation de jeux de donnée

2.3.1 Caractéristiques du texte source et traduction de référence

Notre texte original provient de *Encyclopédie du Savoir Relatif et Absolu des Chats* [1]. Ce roman français qui compte 160 pages a été écrit par Bernard Werber et publié

1. https://www.albin-michel.fr/encyclopedie-du-savoir-relatif-et-absolu-des-chats-978222644

Figure 2.1: Étiquettes de *Encyclopédie du Savoir Relatif et Absolu des Chats*

par l'éditeur Albin Michel [2] le 30 octobre 2019. Nous n'utilisons que le texte des paragraphes (sans la table des matières et les légendes), soit un total de 12 119 mots en 256 paragraphes. Il s'agit d'un roman encyclopédique sur les chats (Figure 2.1), qui se situe dans le domaine littéraire [3] (Tableau 2.1).

Genre	Sous-genre
narratif	Roman
poétique	Chanson
théâtral	Comédie
épistolaire	Lettre
argumentatif	L'essai

Table 2.1: 5 genres littéraires principaux

La traduction de référence est issue de 《一只猫写的"喵"百科》 [4], la version chinoise simplifiée publiée en Chine le 4 octobre 2022, traduite par l'auteur de ce mémoire. Le texte contient 256 paragraphes, soit 17 869 caractères et 188 mots non-chinois.

Nous avons obtenu l'autorisation d'utiliser le texte original et la traduction de référence dans le cadre de nos études.

2.3.2 Génération de la TA

Nous devons obtenir une TA du français vers le chinois simplifié (plus précisément, le mandarin) de l'ensemble du livre. Afin de fournir le plus de contexte possible et de l'obtenir automatiquement, nous envisageons une traduction paragraphe par paragraphe, plutôt que phrase par phrase, et utilisons un moteur de TA qui dispose d'une API. En outre, dans la pratique, lorsque les utilisateurs veulent traduire un texte, ils préfèrent souvent saisir l'ensemble du texte ou du paragraphe dans le moteur de traduction automatique, plutôt que phrase par phrase, puisque cela est chronophage.

À ce jour, 44 des moteurs de TA les plus populaires fournissent une API, dont 35 permettent la traduction du français vers le chinois [5]. Nous avons finalement choisi le moteur DeepL Translator, car il présente les caractéristiques suivantes :

1) un moteur de traduction automatique neuronale

2. https://www.albin-michel.fr/
3. https://www.amazon.fr/Encyclopédie-Savoir-Relatif-Absolu-Chats/dp/2226445552#detailBullets_feature_div
4. https://book.douban.com/subject/35534139/
5. https://machinetranslate.org/engines

2) une version dédiée dans la paire français > chinois simplifié

3) une version d'essai librement accessible en ligne

4) une API permettant de récupérer les traductions paragraphes par paragraphe

En outre, lorsque le chinois est la langue cible, la qualité de la TA de DeepL est généralement plus élevée[6], ce qui signifie qu'il y a moins d'erreurs à annoter. Ce moteur était donc le choix le plus raisonnable étant donné le temps limité disponible pour l'étude.

2.3.3 Caractéristiques du corpus fr-zh (TA)

Nous avons construit un corpus parallèle, où le texte source est le texte du livre français mentionné ci-dessus et la traduction est une TA en chinois simplifié générée par DeepL Translator.

Locales/adminlang	français > chinois simplifié
date de création	27 Sep 2022
nombre de paragraphes	256
nombre de mots/caractères	12 119 mots pour le texte source ; 19 204 caractères pour la TA
outil d'alignement	omegaT + modification manuelle

Table 2.2: Description du corpus

2.4 Application des pré-traitements

2.4.1 Segmentation en phrases

Nous devons tout d'abord segmenter les paragraphes dans des textes source et cible en phrases. Pour ce faire, nous avons testé les outils de segmentation en phrases (NLTK et spaCy) avec un échantillon de 50 paragraphes du texte source et 50 paragraphes du texte cible, ensuite nous avons choisi celui qui fonctionnait le mieux pour segmenter l'ensemble du corpus. Le test comporte trois étapes suivantes :

1) segmenter manuellement l'échontillon en phrases ;

2) utiliser NLTK et spaCy pour découper les paragraphes en français et de spaCy pour découper ceux en chinois, puisque NLTK ne permet pas de segmenter les paragraphes pour la langue chinoise ;

3) comparer les résultats de deux méthodes de segmentation en phrases.

Nous pouvons observer dans le tableau 2.3 que pour les paragraphes en français, le résultat de NLTK (213 phrases) est plus proche de celui de la segmentation manuelle (214 phrases). Alors que pour l'échantillon en chinois, l'outil spaCy a obtenu le même résultat que la segmentation manuelle.

Selon ces conclusions, nous avons utilisé NLTK pour segmenter le texte source et spaCy pour la TA (Tableau 2.4).

6. https://www.deepl.com/en/blog/20200319

échantillon	segmentation manuelle	NLTK	spaCy
src_fr_sample	214	213	219
TA_zh_sample	214	s.o.	214

Table 2.3: Résultats de l'échantillonnage de segmentation en phrases

texte	outil	nb de phrases
source	NLTK	713
TA	spaCy	713

Table 2.4: Résultats de la segmentation en phrases du corpus entier

2.4.2 Tokenization

La tokenization est de segmenter un texte en « unités minimales »[7]. Il permet de repérer les limites des mots, c'est-à-dire le début et la fin d'un mot [Palmer, 2007]. La tokenization est considérée comme une phase de pré-traitement relativement triviale pour des langues comme le français et l'anglais[8][9], car les mots sont souvent séparés par des espaces. Cependant, le chinois est une langue isolante dont le système d'écriture est fondé sur le sinogramme et ne marque pas explicitement les frontières entre mots typographiques ("这几天天天天气不好"). La tokenization présent donc un défi pour le traitement des textes écrits en chinois.

Dans cette étude, l'objectif principal de la tokenization est de faciliter la sélection des mots correspondant au texte original lors de l'annotation. Nous souhaitons utiliser le plate-forme d'annotation INCEpTION pour annoter des erreurs de la TA. De manière pragmatique, cet outil ne permet pas de sélectionner, dans un texte chinois, une unité plus petite qu'un mot typographique (= entre deux espaces). Par exemple dans la Figure 2.2, nous ne pouvions pas sélectionner que le caractère « 一» ou « 个», car « 一个» est reconnu comme étant un seul token, au lieu d'être segmenté en deux caractères. Par conséquent, nous avons testé cinq outils de tokenization[10] (Jieba[11], PkuSeg[12], SnowNLP[13], THULAC[14], HanLP[15]) en utilisant un échantillon de TA (*TA_zh_sample* dans le Tableau 2.3) qui a été segmenté en 214 phrases, soit 4650 tokens dont 4606 caractères chinois.

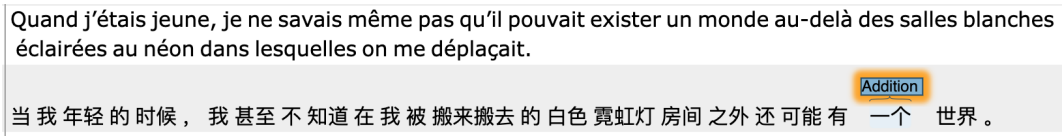

Figure 2.2: Exemple de sous-segmentation sur INCEpTION

7. https://damien.nouvels.net/cours/tal/2$_{M}orphologieTerminologieLexiques.pdf$
8. https://hal.archives-ouvertes.fr/hal-01807765/document
9. Chinese Word Segmentation with External Lexicons on Patent Claims
10. 五款中文分词工具在线PK: Jieba, SnowNLP, PkuSeg, THULAC, HanLP
11. https://github.com/fxsjy/jieba
12. https://github.com/lancopku/pkuseg-python
13. https://github.com/isnowfy/snownlp
14. https://github.com/thunlp/THULAC-Python
15. https://github.com/hankcs/HanLP

Pour ce faire, nous avons comparé le nombre d'espaces pour les différents résultats de tokenization, car plus le nombre d'espaces est élevé, plus la tokenization est fine, autrement dit, moins il existe de sous-segmentation (Tableau 2.5). Au vue de la taille restreinte du présent corpus, ce mémoire fait abstraction des écarts entre les temps d'exécution de ces outils, bien qu'ils soient importants, ne sont pas pris en compte.

Outil	Exemple [16]	nb d'espaces
Jieba	一个_物体	1
PkuSeg	一个_物体	1
SnowNLP	一个_物体	1
THULAC	一个_物体	1
HanLP	一_个_物体	2

Table 2.5: Exemple du nombre d'espaces

Outil	Temps d'exécution	nb d'espaces
Jieba	1.10s	2862
PkuSeg	3.20s	2919
SnowNLP	4.11s	3009
THULAC	1.12s	3056
HanLP	5.87s	3067

Table 2.6: Comparaison des résultats de la tokenization

Par la suite, nous avons analysé les résultats de plus près et avons constaté que la principale différence entre HanLP et les autres outils était la segmentation de la structure « 一(un ; une) + classificateur générique 个/gè/ ». En chinois, les classificateurs sont utilisés pour dénombrer ou désigner des objets, des notions abstraites ou le nombre d'occurrences d'une action [17]. Néanmoins, il existe un usage particulier où le classificateur est utilisé sans un numéral devant. Dans l'exemple « 我有个弟弟» (j'ai un petit frère), le numéral « 一» devant le classificateur « 个» est optionnel. Si la TA est « 我有一个弟弟», le numéral « 一» peut être annoté comme une erreur de fluidité. Nous souhaitions donc pouvoir annoter les numérals et les classificateurs séparément. Ainsi, Nous considérons que HanLP répond le mieux aux besoins de cette étude (Tableau 2.6).

2.4.3 Alignement

L'outil d'alignement que nous utilisons est omegaT [18]. Il permet à la fois un alignement automatique (Figure 2.3) et une modification manuelle (Figure 2.4). L'autre avantage de ce logiciel est que la mise en évidence des chiffres permet de repérer rapidement les phrases susceptibles d'être alignées. De plus, l'utilisateur peut visualiser le contenu du texte et l'ajuster manuellement en même temps dans la fenêtre du logiciel.

17. https://fr-academic.com/dic.nsf/frwiki/377811#cite_note-darrobers-0
18. file:///Applications/OmegaT.app/Contents/Java/docs/index.html

Après avoir importé les fichiers à aligner, nous obtenons le score moyen de l'alignement automatique (Figure 2.3 : "Average Score"). En principe, plus le score moyen est bas, meilleur est l'alignement. Il est possible de choisir entre deux modes, *Heapwise* et *Parawise*. En mode *Heapwise*, les textes sont évalués globalement. En mode *Parawise*, ils sont évalués par segment. En outre, deux algorithmes sont proposés au choix, *Viterbi* et *Forward-Backward*. Nous choisissons simplement celui qui donne les meilleurs résultats.

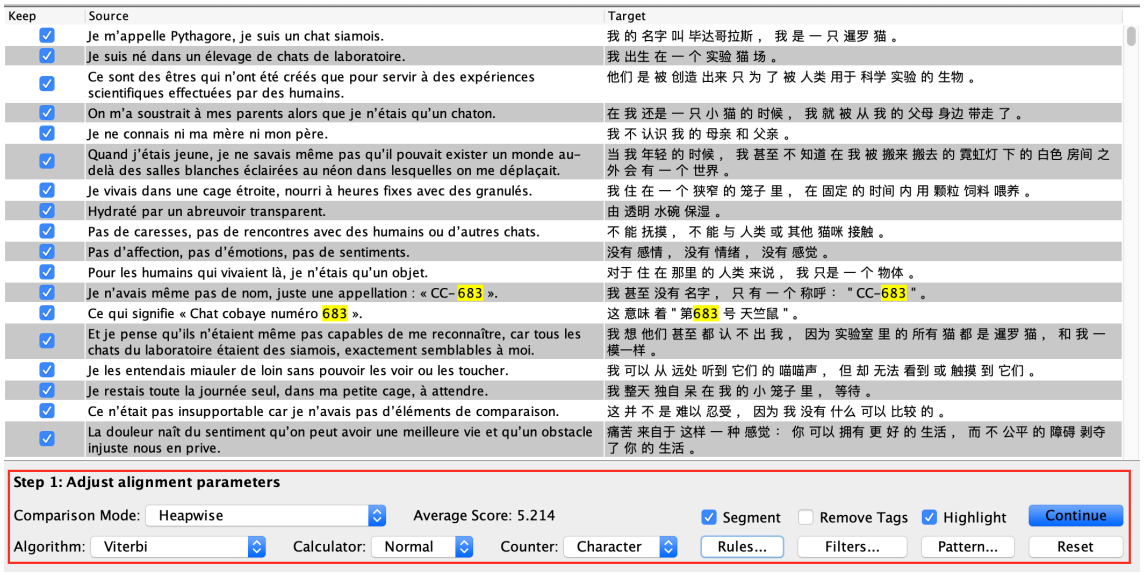

Figure 2.3: Alignement automatique sur omegaT

Après le traitement automatique, nous pouvons effectuer des ajustements manuels à l'aide de trois fonctions (Figure 2.4) : diviser ("Split"), fusionner ("Merge") et modifier ("Edit"). Le fichier de sortie est en format TMX.

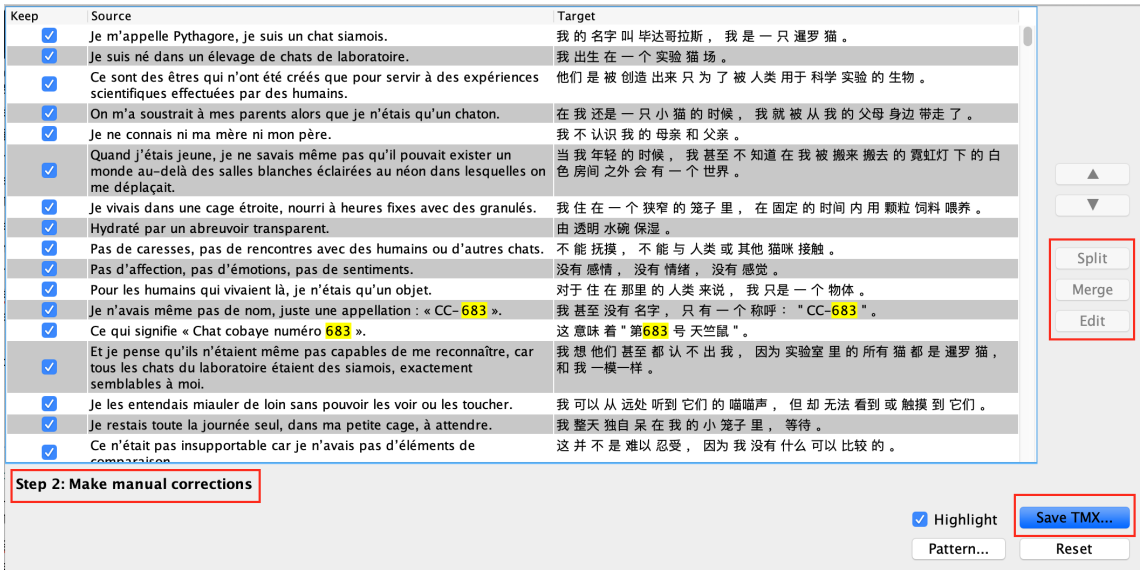

Figure 2.4: Modification manuelle sur omegaT

Nous avons constaté que les facteurs liés à la traduction automatique qui affectent le résultat de l'alignement sont les suivants :

1) plusieurs phrases du texte original ont été traduits en une phrase (Figure 2.5, le texte original comporte deux phrases se terminant par un point "." ; le texte traduit ne comporte qu'une seule phrase se terminant par un point "。") ;

2) le moteur de TA traduit deux fois la même phrase originale (Figure 2.6) ;

3) la ponctuation à la fin de la phrase originale a échappé à la TA (dans la Figure 2.7, il devrait y avoir un point d'exclamation à la fin de la phrase en chinois en colonne de droite).

Au final, nous avons créé un corpus parallèle de 704 segments français-chinois simplifié avec la TA comme le texte cible.

Il y eut une seconde épidémie de peste en 1 540. Là encore, la moitié de la population périt et, une nouvelle fois, les possesseurs de chats survivants furent accusés d'être responsables de ce malheur et systématiquement mis à mort.	在 1540 年 爆发 了 第二 次 鼠疫 ， 一半 的 人口 再次 死亡 ， 幸存 的 猫 主人 再次 被 指责 为 鼠疫 的 罪魁祸首 并 被 系统 地 处死 。

Figure 2.5: Deux phrases source en une seule phrase cible

Une fois repérées, il ne nous reste plus qu'à bondir pour les attraper.	一旦 定位 ， 我们 就 简单 地 扑捉 它们 。 一旦 它们 被 发现 ， 我们 要 做 的 就 是 扑 上去 抓住 它们 。

Figure 2.6: Une phrases source traduite deux fois

Ah, comme l'ignorance est confortable !	啊 ， 无知 是 多么 舒服 啊

Figure 2.7: Ponctuation manquante à la fin d'une phrase cible

2.5 Conclusion

Ce chapitre décrit les étapes et les outils disponibles pour le traitement des textes français et chinois, ainsi que les problèmes d'alignement liés à la TA. Nous finissons par construire un corpus parallèle qui facilite l'annotation des erreurs de TA.

Locales/adminlang	français > chinois simplifié
date de création	27 Sep 2022
nombre de segments	704
nombre de mots/caractères	12 119 mots pour le texte source ; 19 204 caractères pour la TA
outil de segmentation en phrase	NLTK pour le français ; spaCy pour le chinois
outil de tokenization	HanLP (seulement pour le chinois)
outil d'alignement	omegaT + modification manuelle

Table 2.7: Description du corpus à annoter

CHAPTER 3

ANNOTATION

Sommaire

3.1 Introduction

Dans ce chapitre, nous présenterons les deux types d'erreurs qui doivent être annotées. Nous expliquerons ensuite comment personnaliser la plate-forme d'annotation pour répondre aux besoins de l'annotation des erreurs. Enfin, nous aborderons les étapes de l'annotation.

3.2 Présentation des concepts clés

(1) « couche » : sur INCEpTION, toutes les annotations appartiennent à une couche d'annotation. Chaque couche a un type structurel qui définit si elle constitue un span, une relation ou une chaîne. Il définit également le fonctionnement des annotations ("Behaviors") ainsi que le type de caractéristiques ("Features") et d'étiquettes qu'elles portent ("Tagset") (Figure 3.1).

(2) « span » : un segment continu (Figure 3.2).

(3) « relation » : une relation binaire entre deux spans représentée par un arc entre les spans (Figure 3.3).

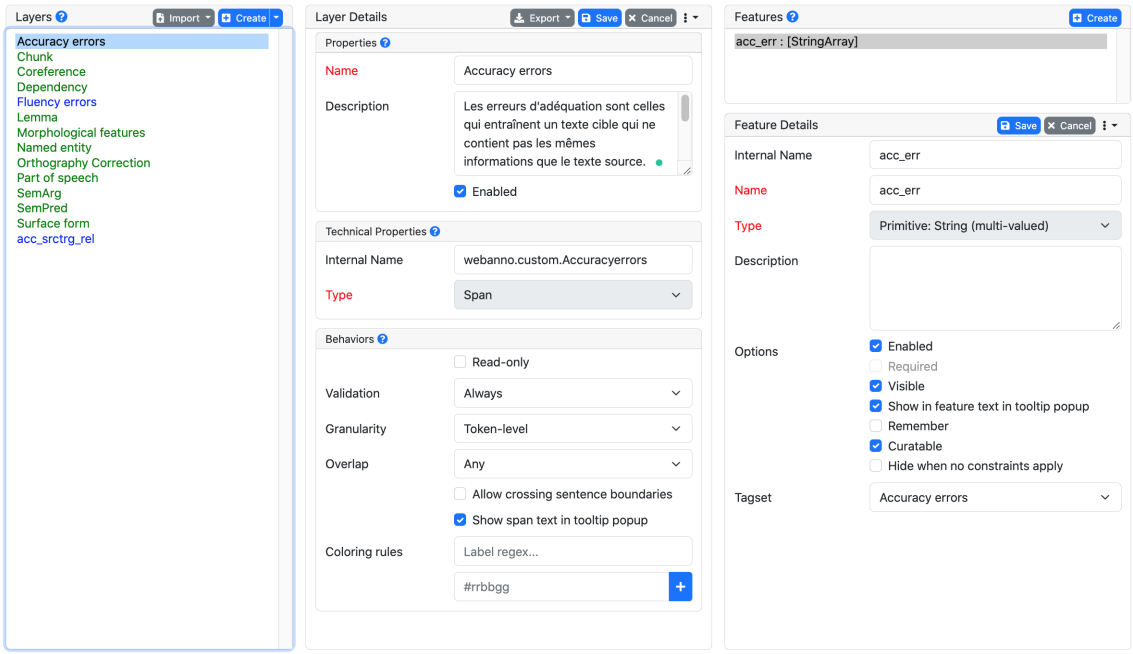

Figure 3.1: Interface d'une couche sur INCEpTION

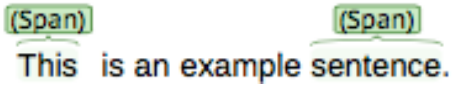

Figure 3.2: Exemple d'un span

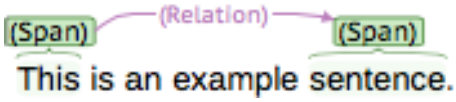

Figure 3.3: Exemple d'une relation

3.3 Typologie d'erreurs

Nous nous sommes principalement appuyés sur la typologie d'erreurs de SCATE [Tezcan et al., 2016] (Figure 1.2). Nous avons largement retenu les erreurs d'adéquation de cette typologie, mais avons supprimé certaines sous-catégories de la grande catégorie de fluidité, par exemple la capitalisation, des signes diacritiques et forme des mots, qui ne sont pas des caractéristiques linguistiques de la langue chinoise.

En outre, nous avons fait appel aux autres recherches sur les erreurs de TA avec le chinois comme la langue cible ([Qiu et al., 2020] ; [刘群et al., 2014]). Ces études nous ont aidés à prouver que les types d'erreurs mentionnés ci-dessus peuvent être retenus.

Enfin, nous faisons également référence aux métriques de qualité multidimensionnelles pour les types d'erreurs liées au style et à la convention locale (MQM)[1]. En effet, les éditeurs des livres traduits ont généralement des exigences en matière de style et de localisation. Mais nous avons effectué les ajustements nécessaires en raison de la paire de langues étudiée dans ce mémoire (français-chinois simplifié,

1. https://themqm.info/typology/

plus précisément le mandarin).

En résumé, la typologie d'erreurs que nous employons dans ce mémoire comprennent deux catégories principales : l'adéquation (Figure 3.4) et la fluidité (Figure 3.5). Les étiquettes dans les figures sont les noms de chaque type d'erreur affichés sur l'interface d'annotation de la plate-forme INCEpTION (section 3.4.2). Nous fournirons une description précise de chaque catégorie dans les sections suivantes.

	Types d'erreurs			Étiquettes
Erreurs d'adéquation	Ajout			ADD
	Omission			OMI
	Segment non-traduit			UNTRANS
	Segment à ne pas traduire			DNT
	Traduction erronée	Expressions à plusieurs mots		MISTRANS_MWE
		POS		MISTRANS_POS
		Désambiguïsation du sens des mots	Mots outils	MISTRANS_WSD_FW
			Mots signifiants	MISTRANS_WSD_CW
		Partiel		MISTRANS_PARTIAL
		Sémantiquement non lié		MISTRANS_WSD_SemUnrel
		Autres		MISTRANS_OTHER
	Mechanical	Ponctuation		MECH_PUNC
	Terminologie bilingue			BILING_TERM
	Erreur de source			SrcErr
	Autres			ACC_OTHER

Figure 3.4: Erreurs d'adéquation

	Types d'erreurs			Étiquettes
Erreurs de fluidité	Grammaire	Syntaxe des mots composés		GRAM_MWS
		Ordre des mots		GRAM_WO
		Mots supplémentaires	Répétition	GRAM_EW_REP
			Composé numéral-classifieur	GRAM_EW_NumCla
			Mots outils	GRAM_EW_FW
			Mots signifiants	GRAM_EW_CW
			Autres	GRAM_EX_OTHER
		Mots manquants	Mots outils	GRAM_MW_FW
			Mots signifiants	GRAM_MW_CW
		Autres		GRAM_OTHER
	Lexique	Mots inexistants		LEXI_NEW
		Choix lexical	Mots outils	LEXI_LC_FW
			Mots signifiants	LEXI_LC_CW
	Mechanical	Ponctuation		MECH_PUNC
	style	Registre		STY_REGIS
		Style étrange		STY_AWK
		Incohérence avec la référence externe		STY_IncExRef
	Convention locale	Chiffre		LOC_NUM
		Date		LOC_DATE
		Mesure		LOC_MESURE
	Erreurs multiples			MulErr
	Autres			FLU_OTHER

Figure 3.5: Erreurs de fluidité

3.3.1 Erreurs d'adéquation

(1) Ajout (ADD) : le texte cible est inexistant dans la source ;

(2) Omission (OMI) : le texte source est absent de la cible ;

(3) Segment non-traduit (UNTRANS) : le texte source n'est pas traduit (mais a été copié vers la cible) alors qu'il aurait dû être traduit en chinois ;

(4) Segment à ne pas traduire (DNT) : le contenu source est inutilement traduit dans la langue cible alors alors qu'il n'aurait pas dû être traduit ;

(5) Traduction erronée : le contenu source a été traduit (quand il devrait l'être) mais la traduction est incorrecte ;

   (a) Expressions à plusieurs mots (MISTRANS_MWE) : la traduction est incorrecte (et souvent trop littérale) parce que la phrase française contenait une expression à plusieurs mots telle qu'un idiome, un proverbe, une collocation, un composé ou un verbe à particule ;

   (b) POS (MISTRANS_POS) : la traduction représente une catégorie lexicale (Part-of-Speech) incorrecte du texte source correspondant ;

   (c) Désambiguïsation du sens des mots : le texte cible se réfère à un sens différent (et erroné) du texte source. Il comprend les mots signifiants (GRAM_EW_FW, par exemple les noms, les verbes, les adjectifs qualitatifs ainsi que les adverbes), et les mots outils (GRAM_EW_CW, par exemple les déterminants, les prépositions, les pronoms et les conjonctions de coordination et de subordination) ;

   (d) Partiel (MISTRANS_PARTIAL) : la traduction est incorrecte en raison de la traduction partielle ;

   (e) Sémantiquement non lié (MISTRANS_WSD_SemUnrel) : le sens du texte traduits n'a aucun rapport avec celui du texte source et n'a aucun sens dans le contexte.

   (f) Autres (MISTRANS_OTHER) : la traduction est incorrecte, mais ne peut pas être classée dans les catégories ci-dessus.

(6) Mechanical (MECH_PUNC) : ce type d'erreur n'est pas lié au contenu du texte, par exemple des ponctuations ;

(7) Terminologie bilingue (BILING_TERM) : la traduction ne correspond pas aux exigences de la terminologie bilingue prédéfinie ;

(8) Erreur de source (SrcErr) : Erreurs présentes dans le texte source.

(9) Autres (ACC_OTHER) : autres erreurs concernant les textes source et cible, qui n'appartiennent à aucune des catégories d'erreurs d'adéquation ci-dessus.

3.3.2 Erreurs de fluidité

(1) Grammaire : erreurs concernant les règles grammaticales de la langue chinoise ;

   (a) Syntaxe des mots composés (GRAM_MWS) : la syntaxe d'une expression à plusieurs mots est incorrecte, même si les choix de mots individuels sont corrects. ;

   (b) Ordre des mots (GRAM_WO) : les mots ne sont pas dans le bon ordre ;

   (c) Mots supplémentaires : les mots qui ne sont pas indispensables dans la traduction, comprenant les mots répétitifs (GRAM_EW_REP), les composés chiffre-classificateur (GRAM_EW_NumCla), les mots signifiants (GRAM_EW_FW) et les mots outils (GRAM_EW_CW) ;

(2) Lexique : le texte source est absent de la cible ;

   (a) Mots inexistants (LEXI_NEW) : le mot ne fait pas partie du lexique chinois ou est un mot inconnu. Cette erreur se produit souvent lorsque le ou les mots source ne sont pas traduits en chinois. D'autre part, le système de TA peut également générer des mots qui n'appartiennent ni à la langue source ni à la langue cible ;

   (b) Choix lexical : le mot fait partie du lexique chinois mais un autre mot devrait être utilisé pour générer une phrase correcte en chinois. Ce type d'erreurs comprend également les mots signifiants (LEXI_LC_CW) et les mots outils (LEXI_LC_FW) ;

(3) Mechanical (MECH_PUNC) : ce type d'erreur n'est pas lié au contenu du texte, par exemple des ponctuations ;

(4) Style : ce type d'erreurs se compose du registre (STY_REGIS), le style étrange (STY_AWK) et Incohérence avec la référence externe (STY_IncExRef) ;

(5) Convention locale : erreurs survenant lorsque la traduction ne respecte pas la teneur locale ou les formats des données, par exemple les chiffres (LOC_NUM), les dates (LOC_DATE) et la mesure (LOC_MESURE) ;

(6) Erreurs multiples (MulErr) : une combinaison d'erreurs qui rend difficile d'annoter séparément les erreurs de fluidité ;

(7) Autres (FLU_OTHER) : autres erreurs de fluidité, qui n'appartiennent à aucune des catégories d'erreurs de fluidité ci-dessus.

3.4 Outil d'annotation

3.4.1 Raison du choix

Comme nous l'avons mentionné précédemment, nous devons annoter deux catégories d'erreurs : l'adéquation et la fluidité. Pour les erreurs d'adéquation, il faut prendre en compte des éléments à la fois dans le texte source et dans le texte cible. Nous devons donc lier le texte source à la TA lors de l'annotation des erreurs d'adéquation, c'est-à-dire d'annoter la relation entre les deux. Par conséquent, nous avons besoin de configurer trois couches d'annotations : la première pour les erreurs d'adéquation, la

deuxième pour celles de fluidité, et la troisième pour la relation attachée aux erreurs d'adéquation.

De ce fait, nous utilisons la plate-forme d'annotation INCEpTION [Klie et al., 2018a]. Elle permet d'effectuer différents types de tâches d'annotation sur des textes écrits [2]. Nous avons choisi cette plate-forme puisqu'elle répondait aux besoins suivants pour cette étude.

1) la possibilité de travailler avec plusieurs langues ;

2) l'annotation des spans et des relations dans plusieurs couches d'annotation ;

3) la possibilité de personnaliser des couches et des relations .

3.4.2 Configurations

Une fois la typologie d'erreurs déterminé, nous pouvons paramétrer les étiquettes et les couches sur la plate-forme d'annotation. Tout d'abord, nous devons configurer des étiquettes pour chaque sous catégorie de ces deux grandes catégories d'erreurs (Figure 3.6 et Figure 3.7).

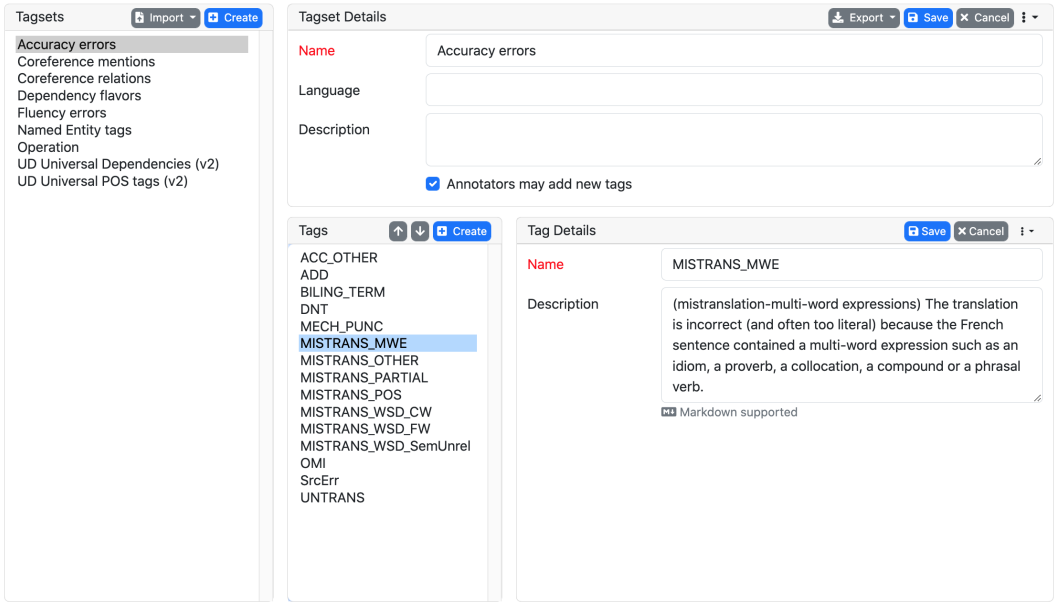

Figure 3.6: Configuration des étiquettes des erreurs d'adéquation

Ensuite, nous avons configuré séparément les couches d'annotation pour les erreurs d'adéquation et de fluidité (Figure 3.8 et Figure 3.9). Il est important de noter que les étiquettes doivent être sélectionnées ("Tagset" en bas à droite dans les figures). Or, nous ne pouvons pas sélectionner les sous-catégories lors de l'annotation.

Une autre option importante est le chevauchement ("Overlap"). Comme certaines spans peuvent être à la fois des erreurs d'adéquation et de fluidité, nous devons choisir "Any" qui nous permettra d'annoter les deux catégories.

Enfin, nous avons également mis en place une couche relationnelle pour les erreurs d'adéquation. Comme nous pouvons le voir dans la Figure 3.10, le type de cette couche d'annotation est « Relation », ce qui est différent des deux autres couches

2. https://inception-project.github.io/use-cases/

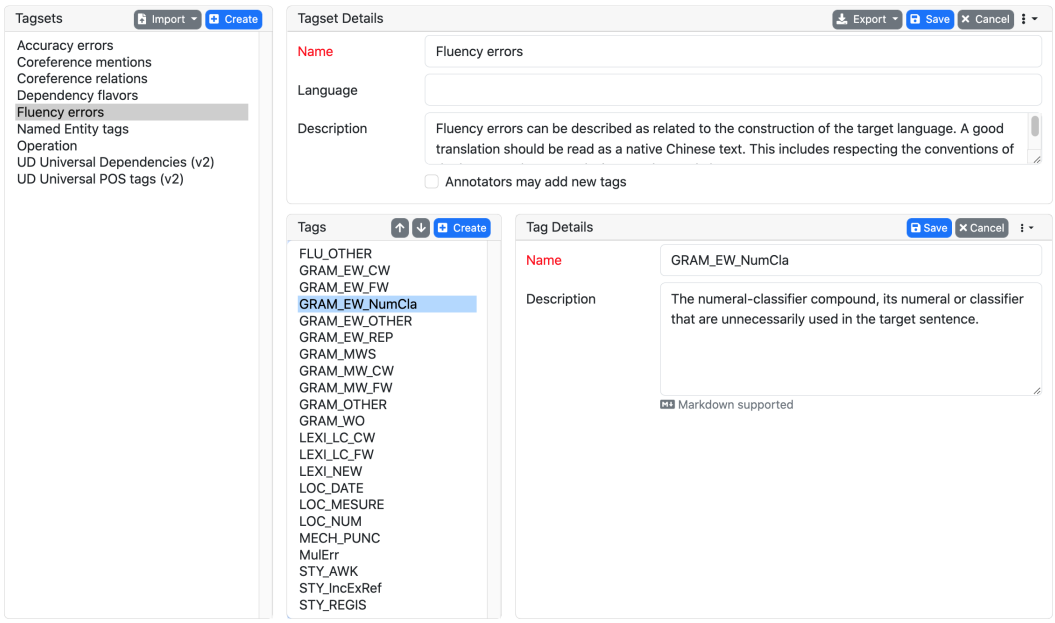

Figure 3.7: Configuration des étiquettes des erreurs de fluidité

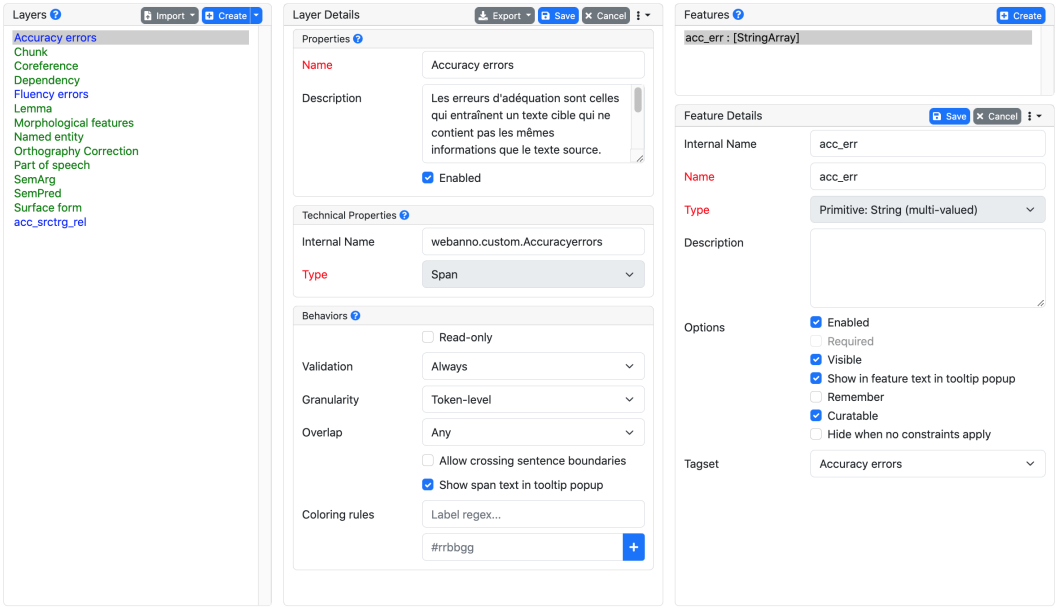

Figure 3.8: Configuration de la couche des erreurs d'adéquation

d'annotation dont le type est « Span ». Du fait que le corpus est une imbrication ligne par ligne du texte source et de la TA, nous avons choisi l'option "Allow crossing sentence boundaries" qui permet l'annotation du croisement des lignes (Figure 3.12).

3.5 Annotation des erreurs

3.5.1 Échantillonnage

En raison de contraintes de temps, nous avons sélectionné les phrases dont le score est inférieur à 0,9 (le plus bas est 0 et le plus haut est 1) à l'aide d'une métrique

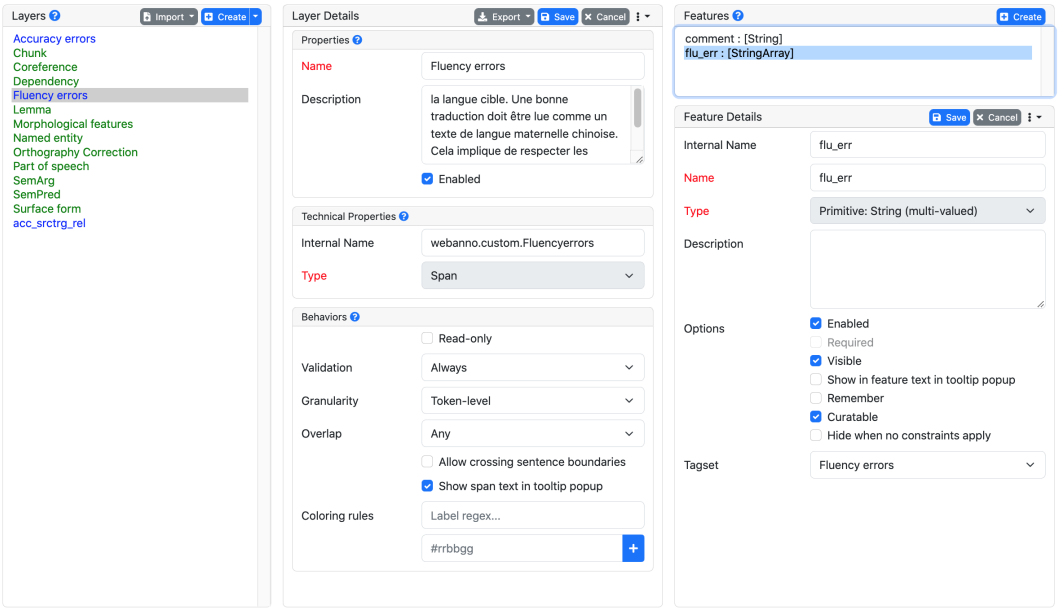

Figure 3.9: Configuration de la couche des erreurs de fluidité

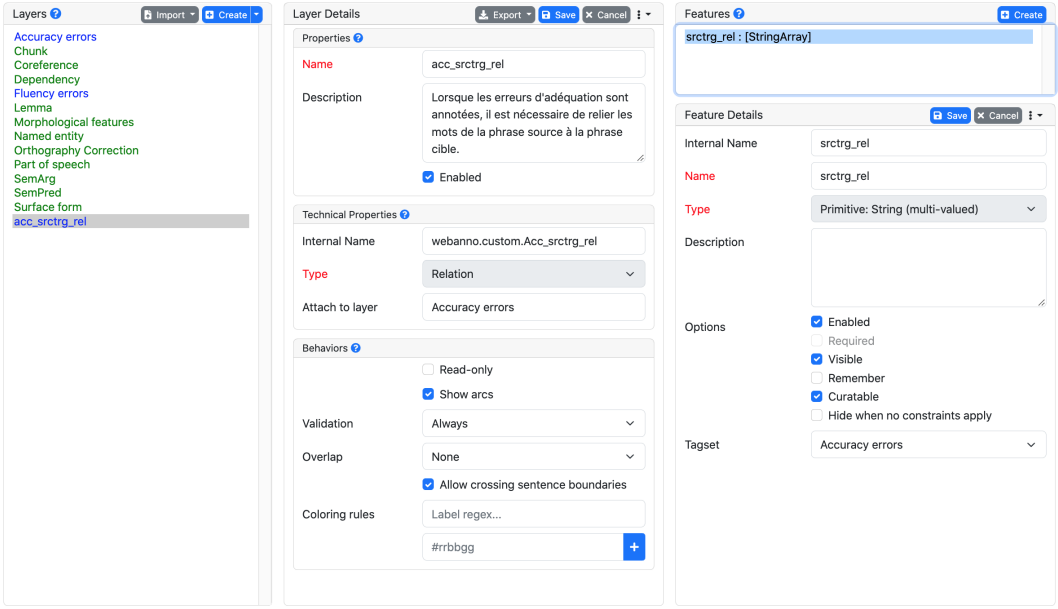

Figure 3.10: Configuration de la couche de relation

automatique BERTScore [Zhang et al., 2019], ensuite nous avons choisi au hasard 100 phrases à annoter.

La métrique BERTScore analyse des distances en cosinus entre les représentations BERT de la TA et la traduction de référence. Autrement dit, elle calcule la similarité sémantique. Nous pensons que cette métrique est appropriée pour évaluer la similarité entre la TA et la traduction de référence, car il existe généralement plus d'une traduction qui transmet le même sens.

3.5.2 Principes d'annotation

L'annotation des erreurs de TA peut également être considérée comme l'identification des erreurs de TA, alors que ce dernier fait partie de la post-édition [O'Brien, 2022]. Nous avons appliqué les principes de post-édition de TAUS [3] pour les raisons suivantes.

Premièrement, il encourage autant que possible la réutilisation des sorties de la TA brute. La volonté de réutiliser autant que possible le résultat brut de la TA est un aspect essentiel de la post-édition, car de nombreux traducteurs professionnels ont un préjugé défavorable à l'égard de la TA. Ils estiment que la TA est inférieure à la traduction humaine, de sorte qu'ils ont souvent tendance à modifier la TA, voire à retraduire.

Deuxièmement, il indique comment effectuer une post-édition complète, c'est-à-dire identifier toutes les erreurs de TA (la colonne à droite dans la Figure 3.11) d'un point de vue sémantique, syntaxique, terminologique et stylistique.

Light post-editing	Full post-editing
Aim for semantically correct translation.	Aim for grammatically, syntactically and semantically correct translation.
	Ensure that key terminology is correctly translated and that untranslated terms belong to the client's list of "Do Not Translate" terms.
Ensure that no information has been accidentally added or omitted.	Ensure that no information has been accidentally added or omitted.
Edit any offensive, inappropriate or culturally unacceptable content.	Edit any offensive, inappropriate or culturally unacceptable content.
Use as much of the raw MT output as possible.	Use as much of the raw MT output as possible.
Basic rules regarding spelling apply.	Apply basic rules regarding spelling, punctuation and hyphenation.
No need to implement corrections that are of a stylistic nature only.	
No need to restructure sentences solely to improve the natural flow of the text.	
	Ensure that formatting is correct.

Figure 3.11: Annotation de fluidité

3. Translation Automation User Society, https://info.taus.net/mt-post-editing-guidelines

3.5.3 Démarches

Dans un premier temps, l'annotation a été effectuée par une seule annotatrice (l'auteure de ce mémoire), de manière autonome, orientée par les principes d'annotation et la typologie des erreurs, ainsi que par le guide d'annotation. Elle a d'abord annoté les erreurs d'adéquation sur les textes source et cible ainsi que la relation (Figure 3.12), ensuite les erreurs de fluidité uniquement sur le texte cible (Figure 3.13). Cette démarche s'explique par le fait que la TA doit d'abord transmettre le sens du texte source, et ensuite chercher à l'exprimer de manière fluide. Il se peut que la TA soit fluide, mais qu'elle ne corresponde pas au sens du texte source.

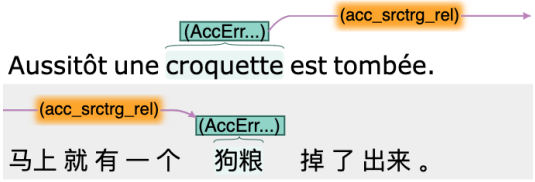

Figure 3.12: Annotation d'adéquation

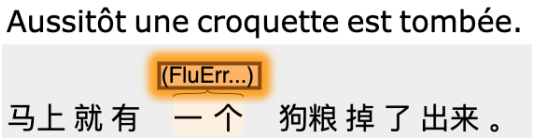

Figure 3.13: Annotation de fluidité

Ensuite, l'annotatrice a fait appel à une locutrice native de la langue chinoise ayant une expérience de la traduction pour déterminer les annotations. Cette locutrice a commencé par lire les textes source et cible pour identifier les éléments de la TA qui n'étaient pas fidèles au texte source. L'annotatrice lui a ensuite lu la TA et lui a demandé de préciser ce qui lui semblait bizarre ou peu naturel.

Enfin, elles ont discuté des erreurs qui avaient été annotées et ont déterminé en collaboration les annotations.

3.6 Conclusion

Dans ce chapitre, nous avons commencé par présenter les deux principales catégories de types d'erreurs, à savoir l'adéquation et la fluidité. Nous avons ensuite expliqué comment créer des étiquettes et mettre en place des couches d'annotation du type « span » et du type « relation » sur la plate-forme d'annotation. Enfin, nous avons présenté les principes et les étapes de l'annotation.

CHAPTER 4

RÉSULTATS ET DISCUSSION

Sommaire

4.1 Introduction

Dans ce chapitre, nous présentons tout d'abord une analyse quantitative fondée sur des statistiques des erreurs d'adéquation et de fluidité annotées. Ensuite, nous passons en revue les résultats sur des causes des erreurs et des difficultés d'annotation pour une analyse qualitative.

4.2 Analyse quantitative

Comme l'illustre le Tableau 4.1, nous avons annoté 225 erreurs d'adéquation (avec 106 relations) et 190 erreurs de fluidité dans 90 segments. Bien que nous ayons choisi 100 phrases à annoter, le processus d'annotation a révélé que 10 phrases étaient des titres. Elles étaient toutes très courtes et ne constituaient pas des phrases, nous avons donc décidé que ces phrases ne seraient pas prises en compte.

nb de tokens	2,781
nb de segments	90
nb d'erreurs d'adéquation	225
nb de relations d'adéquation	106
nb d'erreurs de fluidité	190

Table 4.1: Statistiques du corpus annoté

4.2.1 Erreurs d'adéquation annotées

Comme en témoigne la Figure 4.1, les traductions erronées (*Mistranlation*) représentent de loin la catégorie d'erreurs la plus récurrente dans l'échantillon annoté (192 erreurs, soit 85%), suivie des erreurs à caractère terminologique (14 erreurs, soit 6%) et celles liées au segment non-traduit (6 erreurs, 3%), puis celles dites « mechanical » (5 erreurs, soit 2%) et enfin les erreurs d'ajout et d'omission (respectivement 4 erreurs, soit 2%) (cf. section 3.3.1 pour les définitions des erreurs). Par contre, les erreurs liées aux segments à ne pas traduire (*Do-not-translate*) n'apparaissent pas dans le corpus annoté.

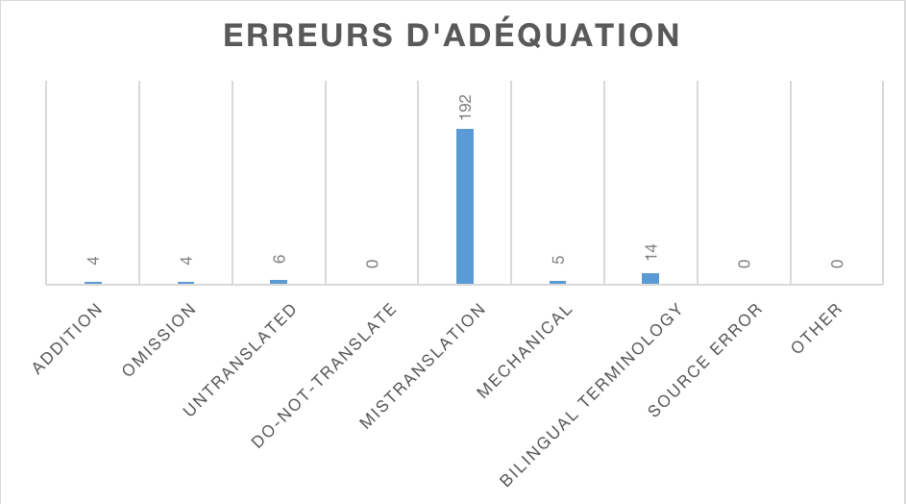

Figure 4.1: Résultat de l'annotation d'adéquation par catégories

Ces catégories d'erreur sont ensuite ventilées en sous-catégories, permettant une caractérisation plus granulaire. Ainsi, dans le tableau présenté dans la Figure 4.2, parmi les 192 erreurs que comporte la catégorie globale « traduction erronée », 119 ont été annotées comme « désambiguïsation du sens des mots » (section 3.3.1(5)(c)). Ce type d'erreur représente la majorité des sous-catégories de « traduction erronée ». Une analyse plus poussée révèle que 38,7% d'entre elles (87 erreurs) concernent les mots signifiants et que 14,2% (32 erreurs) concernent les mots outils.

Erreurs d'adéquation					
Types d'erreurs			Étiquettes	Nombre	%
Ajout			ADD	4	1.8%
Omission			OMI	4	1.8%
Segment non-traduit			UNTRANS	6	2.7%
Segment à ne pas traduire			DNT	0	0.0%
Traduction erronée	Expressions à plusieurs mots		MISTRANS_MWE	23	10.2%
	POS		MISTRANS_POS	4	1.8%
	Désambiguïsation du sens des mots	Mots outils	MISTRANS_WSD_FW	32	14.2%
		Mots signifiants	MISTRANS_WSD_CW	87	38.7%
	Partiel		MISTRANS_PARTIAL	25	11.1%
	Sémantiquement non		MISTRANS_WSD_SemUnrel	21	9.3%
	Autres		MISTRANS_OTHER	0	0.0%
Mechanical	Ponctuation		MECH_PUNC	5	2.2%
Terminologie bilingue			BILING_TERM	14	6.2%
Erreur de source			SrcErr	0	0.0%
Autres			ACC_OTHER	0	0.0%

Figure 4.2: Résultat de l'annotation d'adéquation par sous-catégories

Le type d'erreur « désambiguïsation du sens des mots » signifie que le texte source

a plusieurs sens, mais la traduction choisie par le moteur de traduction automatique ne correspond pas au sens envisagé par le texte source. En témoignent les deux exemples ci-dessous :

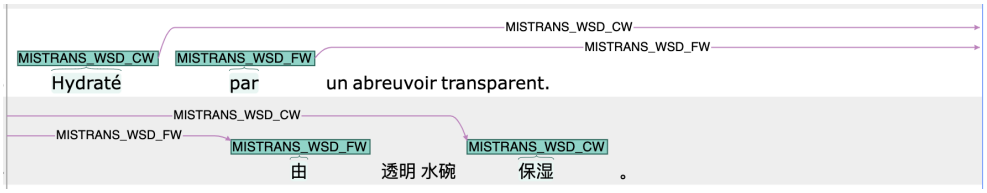

Figure 4.3: Exemple de la désambiguïsation du sens d'un mot signifiant

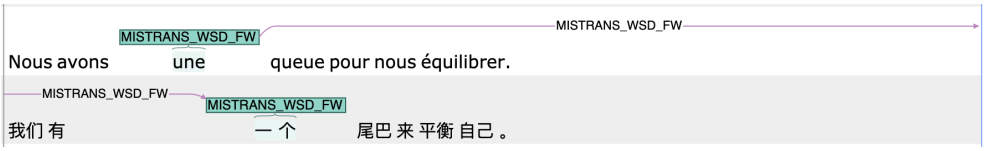

Figure 4.4: Exemple de la désambiguïsation du sens d'un mot outil

Dans la Figure 4.3, « hydraté » a en effet un sens de « combiner un corps avec de l'eau », qui correspond au mot chinois « 保湿». Cependant, dans le contexte, le mot « hydraté » doit être compris dans le sens de « 喝水» (boire de l'eau).

Dans la Figure 4.4, le numéral cardinal « une » a été traduit en « 一个». « 一(yī) » correspond à « une » ; « 个(gè) » est un classificateur générique rarement utilisé pour dénombrer une « queue ». La traduction correcte est « 一条(tiáo) », car « 条(tiáo) » est utilisé pour dénombrer des objets longs, fins et flexibles.

4.2.2 Erreurs de fluidité annotées

Nous constatons dans la Figure 4.5 (cf. section 3.3.2 pour les définitions des erreurs) que, les erreurs de grammaire représentent de loin la catégorie la plus récurrente dans l'échantillon annoté (93 erreurs, soit 49%), suivie des erreurs liées au style (46 erreurs, soit 26%), des erreurs lexiques (37 erreurs, 19%), de celles « mechanical » (13 erreurs, soit 7%) et enfin des erreurs de la convention locale (1 erreur).

S'agissant des sous-catégories (Figure 4.6), les erreurs les plus récurrentes sont liées aux mots supplémentaires qui comptent pour près de 37,9%, suivie des erreurs à caractère stylistique (24,2%) et enfin des erreurs de lexique (19,5%). Par ailleurs, certains types d'erreurs n'apparaissent pas dans le corpus, par exemple l'incohérence avec la référence externe du type de style.

La sous-catégorie « composé numéral-classifieur » dans la catégorie « mots supplémentaires » a une catégorie ajoutée par l'auteure de ce présent mémoire dans le processus d'annotation des erreurs, car ce type d'erreur se présente fréquemment (18 erreurs, soit 9,5%) et ne peut pas être classée dans d'autres catégories. Nous aborderons ce phénomène linguistique en chinois plus en détail dans la section consacrée à l'analyse quantitative.

Nous expliquons un exemple qui contient les types d'erreurs fréquents dans notre corpus (Figure 4.7). Nous pouvons constater que dans la phrase en français, la locution « être doué pour » est suivi par une virgule et une proposition subordonnée, et la traduction en chinois a gardé la même structure syntaxique. Cependant, le mot « 善于(shàn yú) », traduction du groupe de mots « être doué pour », doit obligatoirement

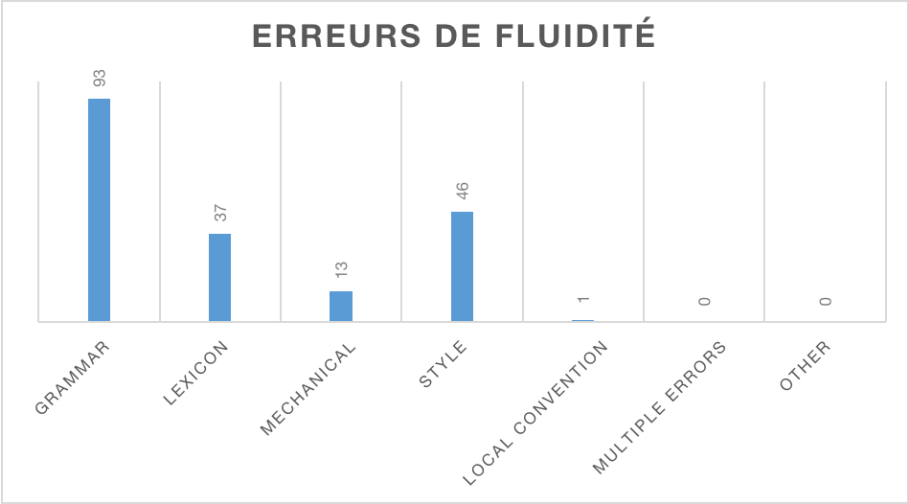

Figure 4.5: Résultat de l'annotation de fluidité par catégories

Erreurs de fluidité					
Types d'erreurs			Étiquettes	Nombre	%
Grammaire	Syntaxe des mots composés		GRAM_MWS	0	0.0%
	Ordre des mots		GRAM_WO	6	3.2%
	Mots supplémentaires	Répétition	GRAM_EW_REP	2	1.1%
		Composé numéral-classifieur	GRAM_EW_NumCla	18	9.5%
		Mots outils	GRAM_EW_FW	24	12.6%
		Mots signifiants	GRAM_EW_CW	28	14.7%
		Autres	GRAM_EX_OTHER	0	0.0%
	Mots manquants	Mots outils	GRAM_MW_FW	4	2.1%
		Mots signifiants	GRAM_MW_CW	11	5.8%
	Autres		GRAM_OTHER	0	0.0%
Lexique	Mots inexistants		LEXI_NEW	3	1.6%
	Choix lexical	Mots outils	LEXI_LC_FW	3	1.6%
		Mots signifiants	LEXI_LC_CW	31	16.3%
Mechanical	Ponctuation		MECH_PUNC	13	6.8%
style	Registre		STY_REGIS	3	1.6%
	Style étrange		STY_AWK	43	22.6%
	Incohérence avec la référence externe		STY_IncExRef	0	0.0%
Convention locale	Chiffre		LOC_NUM	0	0.0%
	Date		LOC_DATE	0	0.0%
	Mesure		LOC_MESURE	1	0.5%
Erreurs multiples			MulErr	0	0.0%
Autres			FLU_OTHER	0	0.0%

Figure 4.6: Résultat de l'annotation de fluidité par sous-catégories

être suivi d'un autre verbe, alors que « 每当(měi dāng) » (chaque fois) est un adverbe temporel qui n'a pas la même fonctionnalité que dans le texte source.

De plus, le pronom de la troisième personne du masculin pluriel « 他们(tā men) » (ils) n'est pas nécessaire dans la traduction, car il suffit d'avoir « 人类» (les humains) pour construire une phrase grammaticalement correcte et compréhensible. Par conséquent, le adverbe « 每当» est annoté comme un mot signifiant supplémentaire, alors que le pronom « 他们» un mot outil supplémentaire.

Dans le cas de l'erreur « 向前走三步，就退两步», cette expression n'est qu'une traduction littéral du texte source « faire trois pas en avant, en faire deux en arrière ». La TA n'est pas naturelle, car l'expression n'existe pas en chinois sous cette forme et doit être adaptée.

Enfin, nous attirons l'attention sur le fait que notre corpus compte 11 phrases

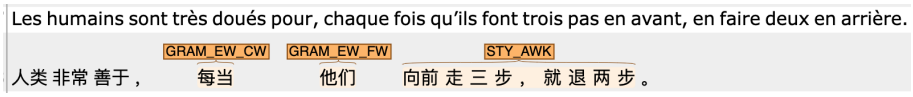

Figure 4.7: Exemple de sous-catégories des erreurs de fluidité

dont la traduction n'est pas annotée d'erreurs. Ce sont des phrases courtes, dont la plupart ne contiennent pas plus de 15 tokens.

4.3 Analyse qualitative

4.3.1 Causes des erreurs de la TA

Concernant les erreurs d'adéquation, comme nous pouvons le voir dans l'analyse des résultats précédente (section 4.2.1), la TA sélectionne parfois des mots qui sont grammaticalement corrects, mais qui ne correspondent pas au sens voulu par le texte source. La plupart de ces problèmes sont dus au fait qu'un mot a plusieurs significations dans une phrase de la langue source [刘群et al., 2014], et que la technologie actuelle de la traduction automatique neuronale n'est pas encore capable de comprendre réellement la phrase et de sélectionner le sens correct sur la base des informations relatives au domaine concerné et au contexte, ce qui entraîne une sélection incorrecte des mots pendant la traduction [Qiu et al., 2020].

Plus particulièrement, la traduction automatique traduit souvent les expressions idiomatiques et les expressions figées mot à mot, telles que les locutions, les proverbes et les dictons qui ont un sens souvent figuré ou imagé permettant d'asseoir une idée souvent bien définie. Cependant, les expressions ne peuvent pas être traduites littéralement dans une autre langue et doivent être adaptée afin de trouver une équivalence. Prenons par exemple « poser un lapin » qui veut dire « manquer un rendez-vous ». En chinois, la traduction est « 放鸽子» (fàng gē zi), littéralement « lâcher un pigeon ». Cependant, cette locution française peut aussi être entendue au sens littéral, par exemple, « le magicien pose un lapin dans un chapeau ». C'est toujours le contexte qui éclaire l'interprétation, littérale ou figurée, de l'expression. Ces informations contextuelles font souvent défaut à la traduction automatique, ce qui peut conduire à un choix de mots incorrect.

Pour la terminologie, la traductions de référence a adopté les traductions du dictionnaire *Names of The Worlds Peoples-A Comprehensive Dictionary of Names In Roman-Chinese* [Guo, 2007] conformément aux exigences de l'éditeur de la version chinoise. Par exemple, le nom de la déesse « Sekhmet » est traduit par « 塞赫迈特», et non « 塞克梅特» (version de la TA) ni « 塞赫麦特» (traduction sur wikipédia). Comme la traduction automatique neuronale n'est pas régi par un cadre stricte, il est impossible de se référer uniquement aux traductions d'un dictionnaire particulier.

S'agissant de la fluidité, nous constatons nombre de phrases traduites qui manquent de naturel. Certaines sont dues à une structure de phrase étrange et d'autres à des formulations redondantes. La TA a tendance à traduire mot à mot, ce qui rend les phrases longues et guindées. Une longue phrase française peut être exprimée en chinois avec un idiome de quatre caractères pour transmettre le même sens. Par exemple la phrase « il faut battre le fer tant qu'il est chaud ». En chinois, la traduction est « 趁热打铁» (chèn rè dǎ tiě). En outre, les structures de phrases sont très différentes entre le français et le chinois. Par exemple, pour exprimer les relations

de cause à effet, le français a tendance à mettre l'effet en premier et ensuite la cause, alors que le chinois fait l'inverse.

Nous pouvons envisager le développement des systèmes de traduction automatique avec un grand corpus monolingue pour entraîner des vecteurs de mots contextuels précis aux fins de la désambiguïsation du sens des mots, en exploitant au mieux les informations contextuelles pour remédier au problème de la sélection incorrecte des mots en raison de leur polysémie. En outre, nous pouvons également essayer d'introduire des bases de connaissances de domaines externes ou des graphes de connaissances afin d'utiliser pleinement les connaissances externes pour atténuer ce problème.

4.3.2 Points culminants de la TA

Malgré la grande marge d'amélioration de la TA, nous avons constaté qu'elle a produit de bons résultats dans certains cas.

Comme nous pouvons voir dans la figure 4.8, « par la force des choses » a été traduit en « 出于需要» (chū yú xū yào) par le système, littéralement « par nécessité ». Le système traduit le sens plutôt que le mot.

Par la force des choses, humains et chats se sont alors bien entendus.
出于 需要 ， 人类 和 猫咪 相处 得 很 好 。

Figure 4.8: Exemple 1 des bons résultats de la TA

Prenons un autre exemple montré dans la figure 4.9. Le texte source est une phrase continue sans ponctuations au milieu. Par contre, la TA contient trois segments qui sont séparés par deux points et une virgule. Cette structure avec des segments courts est plus proche de la syntaxe de la langue chinoise.

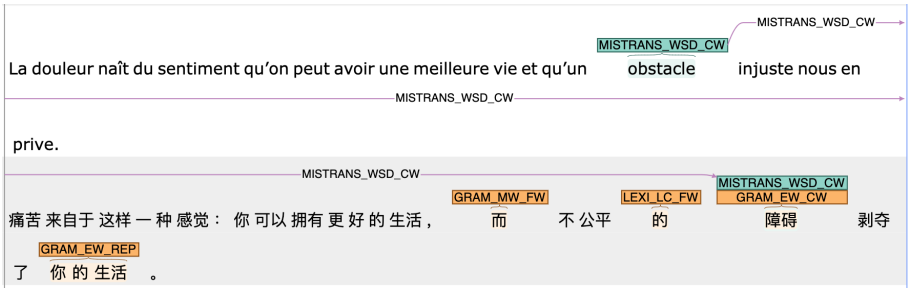

Figure 4.9: Exemple 2 des bons résultats de la TA

Il est également intéressant de discuter du fait que, même si les phrases du corpus annoté ne correspondaient pas à la traduction de référence lorsqu'elles ont été évaluées par la métrique, les annotateurs ne les considéraient pas comme erronées (Figure 4.8). Cela est particulièrement vrai pour les textes littéraires, pour lesquels de nombreuses versions de la traduction sont possibles. Cet exemple illustre donc également la difficulté des métriques pour évaluer la TA dans le domaine littéraire.

4.3.3 Difficultés d'annotation

Tout d'abord, lorsque nous avons à noter des erreurs d'adéquation, nous avons souvent eu besoin de nous référer au contexte. En effet, le corpus littéraire utilisé dans cette étude nécessite parfois non seulement un contexte mais aussi des savoirs de la

vie courante, afin de comprendre réellement le texte source. Par exemple, le mot « croquette » a été traduit en « 狗粮» (gǒu liáng) qui signifie « aliments pour chiens ». Bien que le système de TA comprenne que le mot « croquette » est lié aux aliments, selon le texte source, il devrait faire référence aux aliments pour les chats, et non pour les chiens.

Deuxièmement, nous avons des opinions plus différentes en ce qui concerne les erreurs de fluidité. La langue chinoise a des règles grammaticales flexibles, et certaines phrases sont structurées et formulées de nombreuses façons. Par conséquent, il est difficile d'harmoniser les orientations d'annotation sur la fluidité sans préciser l'objectif et le registre de la traduction. Une autre raison pour laquelle il nous a été difficile de nous mettre d'accord sur la fluidité est que les locuteurs natifs sont également influencés par la langue étrangère et s'habituent à un usage qui n'est pas conforme à leur langue maternelle.

Enfin, au cours de l'annotation, il a été nécessaire de se prémunir des biais cognitifs naturels du traducteur contre la machine. En effet, le traducteur a souvent tendance à sur-corriger une TA objectivement satisfaisante.

4.4 Conclusion

Ce chapitre nous a montré que les erreurs d'adéquation les plus fréquentes sont les erreurs de « traduction erronée » (particulièrement la désambiguïsation du sens des mots) et de « terminologie bilingue », alors que les erreurs de fluidité les plus fréquentes sont les erreurs de « grammaire » et de « style », notamment des mots supplémentaires et du style étrange.

Nous pouvons également constater que certains types d'erreurs n'apparaissent pas dans ce corpus. En effet, la distribution de fréquence des types d'erreurs n'est pas invariable et est susceptible de dépendre fortement du genre de texte. Par exemple, un texte dans un domaine spécialisé contient beaucoup de termes, ce qui augmente la probabilité d'erreurs terminologiques dans la TA. Or, les types d'erreurs comme « Autres » et « Erreurs multiples » ne sont probablement pas nécessaires, car ils sont peu favorable à la construction de types d'erreurs clairs et complets.

Ensuite, nous avons donné des exemples de cas où la TA donne de bons résultats. 11 phrases dans le corpus n'ont pas été considérés par les annotateurs comme erronées.

Finalement, nous avons abordé les difficultés d'annotation, espérant pouvoir éviter ces problèmes dans les études futures.

CHAPTER 5

CONCLUSION GÉNÉRALE

Dans ce mémoire, nous avons décrit l'historique de la traduction automatique et les approches principales. Cela nous donne une idée du développement technologique rapide et du nombre croissant d'utilisateurs de la TA. Ensuite, Nous avons abordé les évaluations humaines et automatiques sur les sorties de TA. Les deux méthodes d'évaluation présentent leurs propres avantages et inconvénients. Toutefois, une évaluation humaine est toujours nécessaire pour obtenir une compréhension contextualisée des erreurs. Par conséquent, nous avons montré plusieurs typologies d'erreurs répandues et des outils d'annotation d'erreurs qui peuvent faciliter l'identification des erreurs.

Au cours de la réalisation du mémoire, nous avons pu répondre aux questions posées dans l'introduction :

1. **Quels types d'erreurs se présentent dans la TAN du français au chinois ?**

   Comme nous pouvons constater dans les figures ci-dessous, la typologie d'erreurs de ce mémoire est divisée en deux catégories : l'adéquation (Figure 5.1) et la fluidité (Figure 5.2).

   Dans la catégorie d'adéquation, nous trouvons 9 sous-catégories : ajout, omission, segment non-traduit, segment à ne pas traduire, traduction erronée, mechanical, terminologie bilingue, erreur de source et autres. Deux d'entre elles contiennent également des sous-classes. Par exemple, la sous-catégories « traduction erronée » se divise également en 6 types d'erreurs : expression à plusieurs mots, POS, désambiguïsation du sens des mots, partiel, sémantiquement non lié et autres.

   Dans la catégorie de fluidité, nous trouvons 7 sous-catégories : grammaire, lexique, mechanical, style, convention locale, erreurs multiples et autres. Les premières deux sous-catégories ont encore des sous-classes.

2. **Comment annoter ces erreurs dans un corpus parallèle français-chinois ?**

   Dans ce mémoire, la construction d'un corpus parallèle annoté d'erreurs de traduction est divisée en deux parties principales : la préparation du corpus et l'annotation des erreurs (Figure 5.3).

   La première partie est constituée des étapes courantes du traitement du langage naturel, notamment la segmentation en phrases, la tokenization et

Erreurs d'adéquation	Ajout		
	Omission		
	Segment non-traduit		
	Segment à ne pas traduire		
	Traduction erronée	Expressions à plusieurs mots	
		POS	
		Désambiguïsation du sens des mots	Mots outils
			Mots signifiants
		Partiel	
		Sémantiquement non lié	
		Autres	
	Mechanical	Ponctuation	
	Terminologie bilingue		
	Erreur de source		
	Autres		

Figure 5.1: Typologie d'erreurs de la TAN fr-zh : adéquation

	Types d'erreurs		
Erreurs de fluidité	Grammaire	Syntaxe des mots composés	
		Ordre des mots	
		Mots supplémentaires	Répétition
			Composé numéral-classifieur
			Mots outils
			Mots signifiants
			Autres
		Mots manquants	Mots outils
			Mots signifiants
		Autres	
	Lexique	Mots inexistants	
		Choix lexical	Mots outils
			Mots signifiants
	Mechanical	Ponctuation	
	style	Registre	
		Style étrange	
		Incohérence avec la référence externe	
	Convention locale	Chiffre	
		Date	
		Mesure	
	Erreurs multiples		
	Autres		

Figure 5.2: Typologie d'erreurs de la TAN fr-zh : fluidité

l'alignement. Comme nous choisissons notre propre système de traduction automatique, nous devons d'abord obtenir la traduction nous-mêmes. Nous pouvons choisir de traduire phrase par phrase ou paragraphe par paragraphe. Pour être plus conforme à la pratique courante, nous avons choisi cette dernière méthode pour obtenir la traduction automatique.

La deuxième partie comprend les configurations d'étiquettes et des couches sur la plate-forme INCEpTION, et les étapes d'annotation des erreurs. Nous recommandons de configurer d'abord les étiquettes, puis les couches. Ainsi, lors de la mise en place des couches, il sera possible d'attacher les étiquettes aux couches correspondantes.

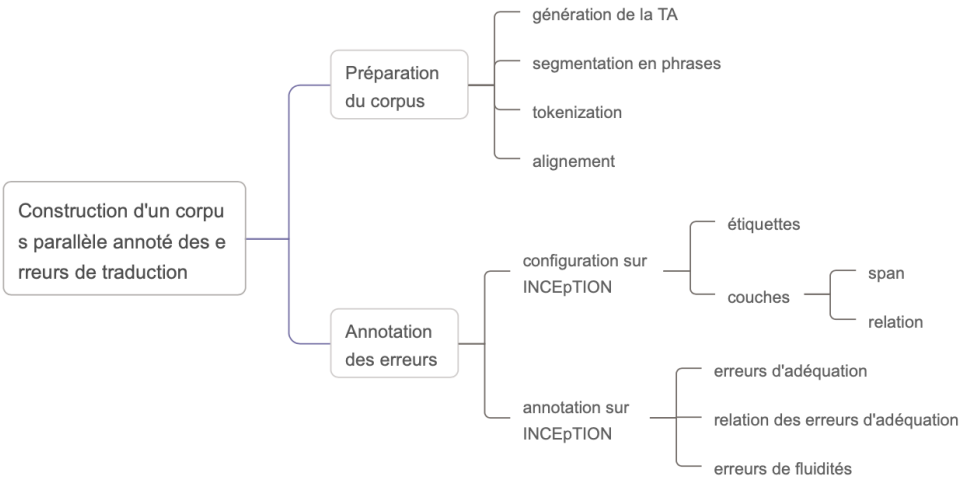

Figure 5.3: Schéma pour construire un corpus parallèle avec des erreurs annotés

3. **Parmi les types d'erreurs annotés, lesquels sont les plus fréquents ?**

   En termes d'adéquation, les deux sous-catégories présentant la plus grande fréquence d'erreurs sont les erreurs de « traduction erronée » (particulièrement la désambiguïsation du sens des mots) et de « terminologie bilingue ».

   En termes de fluidité, les deux sous-catégories présentant la plus grande fréquence d'erreurs sont les erreurs de « grammaire » et de « style », notamment des mots supplémentaires et du style étrange.

4. **Pourquoi ces erreurs se produisent-elles ?**

   La principale raison d'une traduction erronée est que les mots du texte source ont plusieurs significations et que le moteur de la machine est incapable de sélectionner le mot correct, faute de contexte. Deuxièmement, la traduction automatique neuronale n'est pas fondée sur des règles uniformisées et ne peut donc pas adopter la traduction d'un terme provenant d'une source particulière sans glossaire personnalisé.

   De même, les mots supplémentaires et les styles étranges sont ceux qui ont le plus d'impact sur la fluidité. La raison de ces erreurs est que la structure des phrases et l'usage des mots diffèrent considérablement entre le français et le chinois. Le français contient certains éléments grammaticaux qui ne sont pas forcément nécessaires en chinois, comme le coronaire. Les systèmes de TA ont tendance à traduire mot à mot et à respecter la structure des phrases du texte source. Cela peut entraîner un manque de concision et un phrasé peu naturel.

Dans le contexte d'une utilisation aussi répandue de la traduction automatique, l'étude des erreurs de traduction automatique revêt une importante signification pra-

tique. Tout d'abord, les utilisateurs ordinaires peuvent prendre conscience des lacunes de la traduction automatique et éviter ainsi de l'utiliser à mauvais escient. Par exemple, il est conseillé d'éviter d'utiliser la traduction automatique pour vérifier les traductions terminologiques, et d'utiliser la traduction automatique pour les phrases trop longues. Au contraire, il est conseillé de saisir des phrases complètes en incluant la ponctuation. [Loock, 2020] et [Rossi and Chevrot, 2019] ont déclaré qu'il est important de promouvoir des approches plus raisonnées et objectives par un usage critique de la traduction automatique, à l'image du concept anglophone de *MT literacy*.

En outre, pour les professionnels de l'industrie des services linguistiques, notamment les post-éditeurs de TA, comprendre les erreurs de la TA peut les aider à identifier rapidement les erreurs de TA et à améliorer leur efficacité.

CHAPTER 6

PERSPECTIVES

Dans ce dernière chapitre, nous présenterons les limites de notre travail et envisagerons des plans pour les travaux futurs.

Tout d'abord, nous pourrions affiner les types d'erreurs, par exemple en divisant la terminologie en noms de personnes, noms géographiques et termes de domaines spécialisés.

Deuxièmement, l'évaluation peut prendre en compte une combinaison d'autres dimensions, comme le niveau de gravité. D'après [Peraldi, 2016], le juge humain doit pouvoir pondérer son appréciation et notamment la gravité des erreurs rencontrées selon plusieurs critères (fidélité, grammaticalité, lisibilité, caractère idiomatique du texte, etc.) et en fonction du contexte d'utilisation du document (prise en compte du niveau de dissémination du document, etc.) et du niveau de qualité linguistique attendue. Autant d'éléments qui doivent par conséquent transparaître dans la typologie d'erreurs.

Troisièmement, nous pouvons essayer d'analyser les erreurs dans d'autres genres de textes, comme les textes spécialisés, pour voir quels sont les points communs et les différences entre les erreurs des différents genres de textes.

Quatrièmement, nous avons besoin que davantage d'annotateurs nous rejoignent pour rendre les résultats de l'annotation plus objectifs et plus complets. Selon le guide de TAUS pour évaluer la TA [1], au moins deux cents segments doivent être analysés.

Cinquièmement, nous devons annoter davantage de corpus afin d'obtenir des résultats plus représentatifs. De plus, si le corpus annoté est suffisamment grand, nous pourrons peut-être utiliser des techniques d'apprentissage actif pour améliorer l'annotation automatique de la plate-forme d'annotation d'INCEpTION (Annexe A.5) et réduire la charge de l'annotation manuelle.

1. Disponible librement en ligne : https://info.taus.net/quality-evaluation-using-adequacy-and-fluency-approaches

BIBLIOGRAPHIE

[Barbin, 2020] Barbin, F. (2020). La traduction automatique neuronale, un nouveau tournant ? *Palimpseste. Sciences, humanités, sociétés* , (4):51–53. – Cité page 15.

[Brown et al., 1988] Brown, P., Cocke, J., Pietra, S. D., Pietra, V. D., Jelinek, F., Mercer, R., and Roossin, P. (1988). A statistical approach to language translation. *Proceedings of the 12th conference on Computational linguistics -*. – Cité page 14.

[Brown et al., 1993] Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., and Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Comput. Linguist.*, 19(2):263–311. – Cité page 14.

[Brunet-Manquat and Esperança-Rodier, 2018] Brunet-Manquat, F. and Esperança-Rodier, E. (2018). ACCOLÉ : Annotation collaborative d'erreurs de traduction pour COrpus aLignÉs (ACCOLÉ: A collaborative platform of error annotation for aligned corpus). In *Actes de la Conférence TALN. Volume 2 - Démonstrations, articles des Rencontres Jeunes Chercheurs, ateliers DeFT*, pages 197–200, Rennes, France. ATALA. – Cité page 21.

[Castilho et al., 2018] Castilho, S., Doherty, S., Gaspari, F., and Moorkens, J. (2018). Approaches to human and machine translation quality assessment. *Machine Translation: Technologies and Applications*, page 9–38. – Cité page 16.

[Castilho et al., 2017] Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., and Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, 108:109 – 120. – Cité page 15.

[Chatzikoumi, 2019] Chatzikoumi, E. (2019). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, 26(2):137–161. – Cité page 16.

[Esperança-Rodier, 2018] Esperança-Rodier, E. (2018). Analyse de la qualité des traductions automatiques du français vers l'anglais, d'Expressions Poly-Lexicales (EPL) à partir d'un corpus parallèle-Quelles sont les erreurs les plus fréquentes par type d'EPL ? In *Conférence Lexicologie Terminologie Traduction 2018 (LTT 2018)*, Saint Martin d'hères, France. – Cité pages 5 et 18.

[Guo, 2007] Guo, g. (2007). *Names of the world's peoples: A comprehensive dictionary of names in Roman-Chinese*. China Translation amp; Publishing Corporation. – Cité page 45.

[Han, 2022] Han, L. (2022). An overview on machine translation evaluation. – Cité pages 16 et 17.

[Han et al., 2013] Han, L., Wong, D. F., Chao, L. S., He, L., Lu, Y., Xing, J., and Zeng, X. (2013). Language-independent model for machine translation evaluation with reinforced factors. In *MTSUMMIT*. – Cité page 17.

[Hutchins, 2001] Hutchins, W. J. (2001). Machine translation over fifty years. – Cité page 14.

[Hutchins, 2004] Hutchins, W. J. (2004). The georgetown-ibm experiment demonstrated in january 1954. *Machine Translation: From Real Users to Research*, page 102–114. – Cité page 14.

[Kalchbrenner and Blunsom, 2013] Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 1700–1709, Seattle, Washington, USA. Association for Computational Linguistics. – Cité page 15.

[Klie et al., 2018a] Klie, J.-C., Bugert, M., Boullosa, B., de Castilho, R. E., and Gurevych, I. (2018a). The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9. Association for Computational Linguistics. Event Title: The 27th International Conference on Computational Linguistics (COLING 2018). – Cité page 36.

[Klie et al., 2018b] Klie, J.-C., Bugert, M., Boullosa, B., Eckart de Castilho, R., and Gurevych, I. (2018b). The INCEpTION platform: Machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*, pages 5–9, Santa Fe, New Mexico. Association for Computational Linguistics. – Cité page 21.

[Koehn et al., 2007] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A., and Herbst, E. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, pages 177–180, Prague, Czech Republic. Association for Computational Linguistics. – Cité page 15.

[Koehn et al., 2003] Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 127–133. – Cité page 14.

[Levenshtein, 1965] Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics. Doklady*, 10:707–710. – Cité page 17.

[Little, 1989] Little, P. (1989). METAL - machine translation in practice. In *Proceedings of Translating and the Computer 11: Preparing for the next decade*, London, UK. Aslib. – Cité page 14.

[Lommel et al., 2014] Lommel, A., Uszkoreit, H., and Burchardt, A. (2014). Multidimensional quality metrics (mqm): A framework for declaring and describing translation quality metrics. *Tradumàtica: tecnologies de la traducció*, (12):455. – Cité page 19.

[Loock, 2020] Loock, R. (2020). Bowker, Lynne et Buitrago Ciro, Jairo (2019) : Machine Translation and Global Research : Towards Improved Machine Translation Literacy in the Scholarly Community. Bingley : Emerald Publishing, 111 p. – Cité page 52.

[Mariana, 2014] Mariana, V. R. (2014). The multidimensional quality metric (mqm) framework: A new framework for translation quality assessment. – Cité page 19.

[Nießen et al., 2000] Nießen, S., Och, F. J., Leusch, G., and Ney, H. (2000). An evaluation tool for machine translation: Fast evaluation for mt research. In *LREC*. – Cité page 17.

[O'Brien, 2022] O'Brien, S. (2022). How to deal with errors in machine translation: Post-editing. – Cité page 39.

[Palmer, 2007] Palmer, D. D. (2007). Chapter 2 : Tokenisation and sentence segmentation. – Cité page 26.

[Papineni et al., 2002] Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, page 311–318, USA. Association for Computational Linguistics. – Cité page 17.

[Peraldi, 2016] Peraldi, S. (2016). De la traduction automatique brute à la post-édition professionnelle évoluéenbsp;: Le cas de la traduction financière. *Revue française de linguistique appliquée*, Vol. XXI(1):67–90. – Cité pages 16, 20 et 53.

[Piggott, 1992] Piggott, I. M. (1992). *Systran development at the EC Commission: 1976 to 1992: A personal account*. Commission of the European Communities. – Cité page 14.

[Post, 2018] Post, M. (2018). A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics. – Cité page 17.

[Qiu et al., 2020] Qiu, B., Wang, M., Li, M., Chen, C., and Xu, F. (2020). "细粒度英汉机器翻译错误分析语料库"的构建与思考(construction of fine-grained error analysis corpus of English-Chinese machine translation and its implications). In *Proceedings of the 19th Chinese National Conference on Computational Linguistics*, pages 424–433, Haikou, China. Chinese Information Processing Society of China. – Cité pages 32 et 45.

[Rei et al., 2020] Rei, R., Stewart, C., Farinha, A. C., and Lavie, A. (2020). COMET: A neural framework for MT evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics. – Cité page 17.

[Rivera-Trigueros, 2021] Rivera-Trigueros, I. (2021). Machine translation systems and quality assessment: A systematic review. *Language Resources and Evaluation*, 56(2):593–619. – Cité page 16.

[Rossi and Chevrot, 2019] Rossi, C. and Chevrot, J.-P. (2019). Uses and perceptions of machine translation at the european commission. – Cité page 52.

[Scott, 1989] Scott, B. E. (1989). The LOGOS system. In *Proceedings of Machine Translation Summit II*, München, Germany. – Cité page 14.

[Sharou and Specia, 2022] Sharou, K. A. and Specia, L. (2022). A taxonomy and study of critical errors in machine translation. In *Proceedings of the 23rd Annual Conference of the European Association for Machine Translation*, pages 171–180, Ghent, Belgium. European Association for Machine Translation. – Cité pages 11 et 20.

[Snover et al., 2006] Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pages 223–231, Cambridge, Massachusetts, USA. Association for Machine Translation in the Americas. – Cité page 17.

[Stymne, 2011] Stymne, S. (2011). Blast: A tool for error analysis of machine translation output. In *Proceedings of the ACL-HLT 2011 System Demonstrations*, pages 56–61, Portland, Oregon. Association for Computational Linguistics. – Cité pages 5, 20 et 21.

[Tezcan et al., 2016] Tezcan, A., Hoste, V., and Macken, L. (2016). Scate taxonomy and corpus of machine translation errors. – Cité pages 5, 18, 19, 21 et 32.

[Thompson and Post, 2020] Thompson, B. and Post, M. (2020). Paraphrase generation as zero-shot multilingual translation: Disentangling semantic similarity from lexical and syntactic diversity. In *Proceedings of the Fifth Conference on Machine Translation (Volume 1: Research Papers)*, Online. Association for Computational Linguistics. – Cité page 17.

[Toudic et al., 2014] Toudic, D., Morin, K. H., Moreau, F., Barbin, F., and Phuez, G. (2014). Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée. *ILCEA: Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie* , (19). – Cité page 20.

[Vaswani et al., 2017] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. – Cité page 15.

[Vilar et al., 2006] Vilar, D., Xu, J., D'Haro, L. F., and Ney, H. (2006). Error analysis of statistical machine translation output. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA). – Cité page 18.

[Way, 2018] Way, A. (2018). *Quality Expectations of Machine Translation*, pages 159–178. Springer International Publishing, Cham. – Cité page 11.

[Zhang et al., 2019] Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., and Artzi, Y. (2019). Bertscore: Evaluating text generation with bert. – Cité pages 17 et 38.

[刘群et al., 2014] 刘群, Zhao, H., and Liu, Q. (2014). Common error analysis of machine translation output. – Cité pages 32 et 45.

APPENDIX A

DOCUMENTATION

A.1 Script pour la tokenization

```python
import time, datetime

start_time = time.time()

path_file = "/Users/zhongxinyi/TAL/trans/results/segmentation/cat_trg_zh_mt_220927_seg_final.txt"
text = open(path_file, "r")

import jieba
for t in text:
    jieba_list = jieba.cut(t, cut_all=False)
    result = " ".join(jieba_list).strip()
    print(result)

    path = "../results/tokenization/sample/cat_trg_zh_mt_sample.spacy.jieba"
    with open(path, "a") as output:
        output.write(result)
        output.write("\n")

from snownlp import SnowNLP
for t in text:
    result = " ".join(SnowNLP(t).words).strip()
    print(result)

    path = "../results/tokenization/sample/cat_trg_zh_mt_sample.spacy.snownlp"
    with open(path, "a") as output:
        output.write(result)
        output.write("\n")

# thulac
import thulac
thu_lac = thulac.thulac(seg_only=True)
for t in text:
    result = thu_lac.cut(t, text=True)
    print(result.strip())

    path = "../results/tokenization/sample/cat_trg_zh_mt_sample.spacy.thulac"
    with open(path, "a") as output:
        output.write(result)
        output.write("\n")

# pkuseg
import pkuseg
pku_seg = pkuseg.pkuseg()
for t in text:
    result = " ".join(pku_seg.cut(t))
    print(result)

    path = "../results/tokenization/sample/cat_trg_zh_mt_sample.spacy.pkuseg"
    with open(path, "a") as output:
        output.write(result)
        output.write("\n")

import hanlp
tok_fine = hanlp.load(hanlp.pretrained.tok.FINE_ELECTRA_SMALL_ZH)
for t in text:
    result = " ".join(tok_fine(t))
    print(result)

    path = "../results/tokenization/cat_trg_zh_mt_220927_seg_final.hanlp.txt"
    with open(path, "a") as output:
        output.write(result)
        output.write("\n")


# Impression du temps final
print(f"Time taken: {str(datetime.timedelta(seconds = time.time() - start_time))}")
```

Figure A.1: Script pour la tokenization

A.2 Les exemples du guide d'annotation

1. Addition (ADD)

Definition: Target text is not present in the source.

Annotation: Annotate the added target text which is not present in the source. You do not need to annotate any source text and therefore no linking is required either (since the text is not present in source). However, if you cannot separately annotate the target word(s) due to being a part of a compound or a specific phrase, select multiple words or the phrase. In this case annotate the source words that are covered in target annotation and link the source text to target with an arrow.

e.g.:

FR: Une fois repérées, il ne nous reste plus qu'à bondir pour les attraper.

MT: 一旦 定位 ， 我们 就 简单 地 扑捉 它们 。 一旦 它们 被 发现 ， 我们 要 做 的 就 是 扑 上去 抓住 它们 。

Figure A.2: Exemple du guide d'annotation

5.1. Multi-Word Expression (MISTRANS_MWE)

Definition: The translation is incorrect (and often too literal) because the French sentence contained a multi-word expression such as an idiom, a proverb, a collocation, a compound or a phrasal verb.

Annotation: Annotate the source multi-word expression that is incorrectly translated. Annotate the corresponding translation in target text. Link annotation in the source text to the annotation in the target text with an arrow.

e.g.:

FR: En revanche, je parvenais à visualiser l'écran et à déplacer la flèche du curseur sur sa surface telle qu'elle apparaissait dans mon esprit.

MT: 然而 ， 我 可以 想象 出 屏幕 的 样子 ， 并 在 屏幕 上 移动 光标 箭头 ， 因为 它 出现 在 我 的 脑海 中 。

Figure A.3: Exemple du guide d'annotation

A.3 Interface d'annotation d'INCEpTION

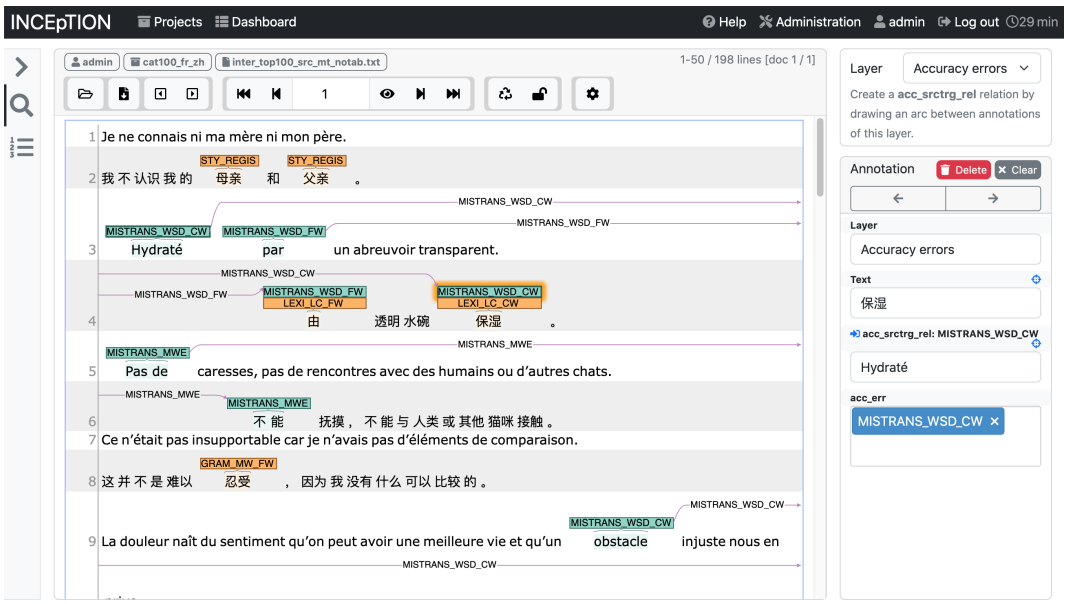

Figure A.4: Interface d'annotation manuelle des erreurs

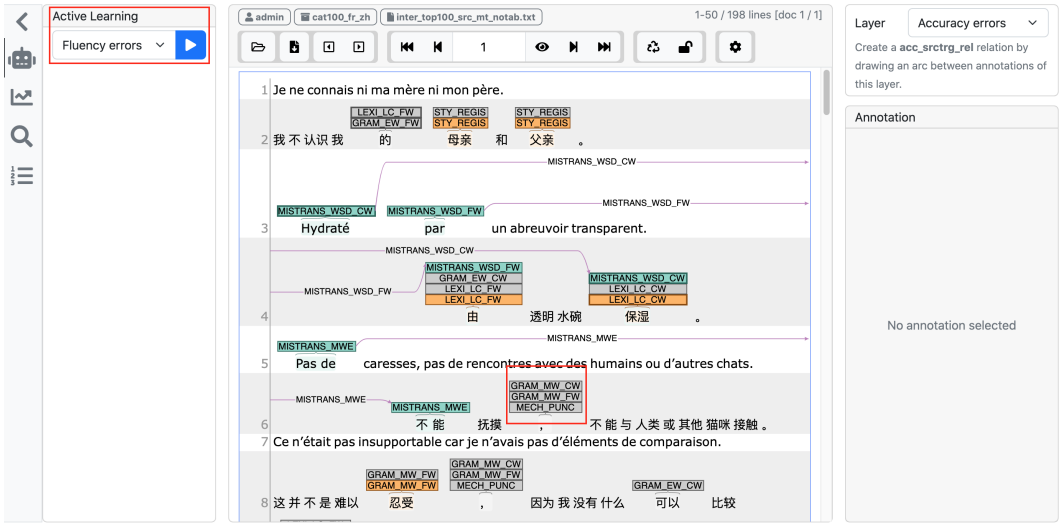

Figure A.5: Interface d'annotation automatique des erreurs