



**Institut National des Langues et des Civilisations Orientales**

**Département Textes, Informatique, Multilinguisme**

---

**OPTIMISATION ET AMÉLIORATION DU PROCESS DE  
TRADUCTION DES RAPPORTS D'INTERVENTION POUR  
SERVICE EUROPE**

---

**Master Traitement Automatique Des Langues**

**Parcours : Traductique**

Par Svetlana KOVALEVA

Encadrement par l'école :

Cyril Grouin

Encadrement par l'entreprise :

Raul Perea Olivo, Charlotte De Truchis

Année universitaire 2014-2015

Novembre 2015

## **TABLE DE MATIÈRES**

<b>REMERCIEMENTS</b>	<b>3</b>
<b>RÉSUMÉ</b>	<b>4</b>
<b>INTRODUCTION</b>	<b>5</b>
<b>I. ÉTAT DE L'ART</b>	<b>10</b>
<b>II. MATÉRIEL ET MÉTHODE</b>	<b>16</b>
<b>III. PERSPECTIVES</b>	<b>31</b>
<b>CONCLUSION</b>	<b>33</b>
<b>BIBLIOGRAPHIE</b>	<b>34</b>
<b>ANNEXE</b>	<b>37</b>

## **Remerciements**

J'adresse mes remerciements aux personnes qui m'ont aidée dans la réalisation de ce mémoire.

Je remercie tout d'abord mon directeur de mémoire M. Grouin qui m'a guidée dans mon travail. Ses remarques critiques, ses suggestions et son encouragement m'ont permis d'avancer à chaque fois un peu plus dans la recherche que j'avais entreprise.

Ce travail n'aurait pu se faire sans le constant soutien de mon tuteur de stage à GE Healthcare Raul Perea Olivo, mon manager Olivier Bonnabel et mes collègues Charlotte De Truchis et Gabriel Gadani, qui m'ont accompagnée tout au long de mon stage et m'ont permis de mettre en avant et de développer certaines de mes qualités et d'apprendre de nouvelles choses.

Je remercie également toute l'équipe pédagogique de l'INALCO, ainsi que les différents intervenants professionnels de la formation d'Ingénierie linguistique de l'INALCO d'avoir assuré ma formation théorique.

## Résumé

Ce travail s'inscrit dans le cadre du projet CHU (Complaint Handling Unit) Translation Productivity au sein de GEHC à Buc, France. Le travail effectué consiste à optimiser l'extraction, le traitement des rapports d'intervention générés dans les systèmes de gestion de services dans plusieurs langues, ainsi que leur traduction réalisée à la demande de la FDA (Food and Drug Administration). Le projet a été réalisé entre l'équipe de Traduction, de Global Services et de Sourcing. Nous analysons donc le problème de traitement de rapports, ainsi que la difficulté de la traduction technique dans le domaine médical. La solution mise en œuvre consiste à gérer spécifiquement la concaténation et la segmentation des textes à traduire, en tenant compte de la mémoire de traduction et des portions déjà rédigées en anglais dans les textes sources. Le résultat de ce travail permet d'améliorer l'effet de levier et de réduire les coûts de traduction.

**Mots-clés :** Traduction automatique, optimisation de la traduction automatique, traduction automatique statistique, extraction d'information, extraction de données, prétraitement, normalisation de textes.

# INTRODUCTION

## Problématique de recherche

L'industrie médicale est complexe, l'expertise de l'industrie est donc essentielle pour une bonne traduction. La traduction médicale est une des spécialités les plus difficiles à maîtriser. D'autant plus que la médecine couvre différentes spécialités (oncologie, cardiologie, ophtalmologie, cancérologie etc.) : trouver une agence spécialisée en traduction médicale fiable, rapide, devient une priorité pour tous les acteurs de l'industrie médicale. C'est donc dire à quel point la terminologie médicale présente des embûches, des difficultés qui tiennent à la structure de la langue source, ainsi qu'à la spécialisation de cette langue ou de la langue cible. La langue peut être un obstacle pour consulter des informations pertinentes et de qualité sur la santé et pour prodiguer des soins appropriés. Il existe donc un besoin amplifié à l'échelle mondiale. L'anglais est depuis longtemps *la lingua franca* des scientifiques – notamment dans la santé publique [Bulletin de l'Organisation mondiale de la Santé, Vol.93, Numéro 6, juin 2015, p. 361- 436<sup>1</sup>].

*GE Healthcare*, fabricant d'équipements d'imagerie médicale, reçoit quotidiennement les rapports de techniciens intervenant dans les hôpitaux européens équipés de matériel GE. Les rapports étant rédigés dans la langue locale de chaque technicien, un besoin de traduction spécialisée du domaine médical est donc nécessaire. Le technicien effectue le suivi d'un problème dans sa langue locale et prévient s'il a été résolu ou non et s'il existe des risques pour la santé du patient suite à la demande de la FDA (Food and Drug Administration), service du gouvernement américain responsable de la pharmacovigilance et une autorité en charge de la régulation et supervision des dispositifs médicaux (DM) aux États-Unis.

---

<sup>1</sup> L'OMS <http://www.who.int/>

‘Nous analysons le problème de traitement de rapports, y compris la difficulté de la traduction technique dans le domaine médical. En raison du statut particulier du texte technique, qui se doit d’être un reflet de la réalité extralinguistique, lors de l’interprétation du texte original, le traducteur ne devra pas s’attacher à restituer « ce qu’a voulu dire » l’auteur, mais « ce qu’il aurait dû » écrire pour que le sens corresponde à cette réalité extralinguistique’ [Fontanet, 2005].

Chaque année, la FDA reçoit plusieurs centaines de milliers de rapports de dispositifs médicaux de décès suspects de périphériques associés, des blessures graves et des troubles dans le fonctionnement [FDA Guidance : Medical Device Reporting for User Facilities, April 1996, Department of Health and Human Services<sup>2</sup>].

Pour la surveillance post-commercialisation, la FDA utilise le système *Medical Device Reporting* (MDR) afin de contrôler les performances des appareils, de détecter les problèmes de sécurité potentiels liés à l'appareil, et de contribuer à l'évaluation bénéfice-risque de ces produits.

Les rapporteurs (les fabricants, les installations de l'utilisateur de l'appareil, et les importateurs) sont tenus de présenter certains types de rapports d'événements indésirables, ainsi que les problèmes liés aux produits à la FDA ( au sujet de dispositifs médicaux). En outre, la FDA encourage également les professionnels des soins de santé, les patients, les soignants et les consommateurs à soumettre des rapports volontaires sur les événements indésirables graves qui peuvent être associés à un dispositif médical, ainsi que les erreurs d'utilisation, les problèmes de qualité de produits, et les échecs thérapeutiques. Ces rapports, ainsi que des données provenant d'autres sources, peuvent fournir des informations cruciales qui contribuent à améliorer la sécurité des patients<sup>3</sup>.

---

2

<http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM095266.pdf>

3

Pour plus d'information :

<http://www.fda.gov/MedicalDevices/Safety/ReportaProblem/default.htm>

Les rapports d'interventions sont enregistrés dans la base de données de GE Healthcare et l'étape de préparation des fichiers pour la traduction peut commencer.

Nous pouvons résumer le contexte :

- ✓ Les ingénieurs de service de GE Healthcare effectuent des rapports en cas de problème, dans la langue locale ;
- ✓ GE Healthcare dispose d'un service de traduction qui aide à traduire ces rapports ;
- ✓ Les rapports traduits sont utilisés par les techniciens pour résoudre le problème ;
- ✓ Ils sont également envoyés la FDA dans le cas de problèmes liés aux dispositifs

La plupart des agences et des entreprises pré-traitent (formatage du document, correction orthographique, segmentation de phrases, etc.) les fichiers pour faciliter la traduction avant de les envoyer. Les agences les post-traitent ensuite pour intégrer ces traductions dans leur système. Cette standardisation des processus et des fonctions est directement liée à l'industrialisation de métier de traduction suite aux demandes des grands entreprises pour lesquelles la traduction est une partie intégrante du développement de leurs produits, et impose la mise en place de processus documentaires très précis, implémentés à leur mode de production.

La qualité orthographique d'un texte peut influencer le traitement automatique de la langue. La normalisation est une tâche qui a pour but de proposer, pour chaque mot considéré comme incorrect, sa forme correcte.

Une partie du travail de ce mémoire se concentre sur la normalisation des fautes lexicales, qui correspondent à des mots inconnus de la mémoire de traduction (TM), en laissant de côté les fautes grammaticales. Ce travail vise à optimiser le temps du prétraitement des fichiers.

Nous voulons normaliser un texte et le corriger ensuite. Traiter les fautes grammaticales n'est pas pertinent dans ce cas. Bien que la majorité des formes

(que nous considérons comme fausses) correspondent à des fautes lexicales, on peut trouver dans les rapports d'intervention une forme involontairement réduite.

De manière générale, si nous voulons normaliser une forme incorrecte, deux possibilités s'offrent à nous. Nous pouvons soit normaliser la forme comme le ferait un outil de correction automatique, en proposant sa forme correcte (ex : *examin* - *examen*), soit la remplacer par une forme qui partage le même lemme que sa forme attendue (ex : *demarez* - *démarrer*). Ce dernier provoquera aucun bruit puisque nous cherchons à améliorer un texte pour qu'il puisse être analysé ultérieurement qui s'appuie non pas sur les tokens mais sur les lemmes.

## Objectifs

L'objectif de cette recherche est d'analyser le processus d'amélioration de la qualité de traduction et de la productivité.

Pour atteindre ce but plusieurs **méthodes** ont été utilisées :

- ✓ Création des macros/des scripts qui permettent de préparer les fichiers pour une traduction optimisée ;
- ✓ Optimisation de l'utilisation de l'outil TMS (abréviations, concaténation de phrases, création de glossaire) ;
- ✓ Proposition d'améliorations des outils/plateformes IT ;
- ✓ Étude statistique (data management) ;
- ✓ Rédaction de guides de bonnes pratiques à l'usage des ingénieurs ;
- ✓ Nettoyage des mémoires de traduction

## Structure

Ce mémoire comprend Introduction, État de l'art, Matériel et Méthode, Résultats, Conclusion et Bibliographie.

Dans le Chapitre État de l'art nous présentons plusieurs articles sur le sujet d'optimisation des données, de la traduction.



Dans le Chapitre Matériel et Méthode nous présentons les données du projet et les méthodes et des pratiques qui ont été utilisées pour améliorer le processus du traitement des fichiers, et de la traduction suivis des résultats obtenus, ainsi que des coûts de traduction, le rôle de la traduction automatique statistique ont été analysés.

Dans le Chapitre Perspectives nous présentons les perspectives d'amélioration du processus de la traduction selon des recherches et des tests pratiques.

Dans le Chapitre Conclusion nous présentons un bilan sur la problématique en cours et donne les pistes pour le futur développement du processus de la traduction.

## **Projet CHU Translation Productivity**

Ce travail s'inscrit dans le cadre du projet *CHU (Complaint Handling Unit) Translation Productivity* au sein de GEHC à Buc, France, sur lequel j'ai travaillé durant la période de stage du juin 2014 au janvier 2015. Le travail a été effectué afin de pouvoir optimiser l'extraction, traitement des rapports d'intervention générés dans les systèmes de gestion de services, ainsi que leur traduction à la demande de la FDA (Food and Drug Administration). Le projet a été réalisé entre l'équipe de Traduction, de Global Services et de Sourcing, en coopération avec l'équipe de SDL. SDL s'occupe de la gestion de l'expérience client à l'échelle internationale (leur système de gestion de projets de traduction (TMS) gère le flux du contenu global à chaque étape du processus de localisation, notamment la traduction), et de GENPACT, leader de la transformation et de la gestion des processus et activités d'entreprises.

## I. ÉTAT DE L'ART

Bien que la traduction n'occupe pas la place principale dans le monde de la médecine, elle joue certainement un rôle important dans le partage des résultats de la recherche médicale, la publication de nouveaux résultats dans la communauté scientifique internationale et la commercialisation de nouveaux médicaments produits et dispositifs médicaux sont des éléments clés [Andriesen 2006; Montalt-Resurrecció et Shuttleworth 2012].

Par exemple, des livres, des articles et des présentations, les textes médicaux traduits comprennent principalement documents d'enregistrement, tels que les dossiers de demande d'enregistrement de nouveaux médicaments et des dispositifs médicaux, ainsi que des manuels d'instruction pour équipement et instruments médicaux, et les documents pour les essais cliniques.

La qualité de traduction est une question complexe parce que, en dehors de la précision et de l'utilisation correcte de la langue, elle implique des facteurs tels que la satisfaction du client et le respect des exigences contractuelles, et elle est largement déterminée par le type de texte, la fonction et les attentes de la communauté liée à la traduction [Gouadec 2010]. L'assurance de la qualité de la traduction médicale consiste à concevoir des méthodes de contrôle efficaces pour la détection des erreurs, les tests de lisibilité et de la mise en service des professionnels suffisamment qualifiés pour effectuer la traduction médicale.

Par conséquent, l'une des questions essentielles est de déterminer quelles mesures peuvent être prises afin d'assurer la qualité adéquate de la traduction.

Comme indiqué plus haut, la partie de recherche effectuée dans ce mémoire se base sur la correction des fautes lexicales. Les travaux sur la correction lexicale ont commencé dans les années 60 [Blair, 1960 ; Damerau, 1964]. Les premiers systèmes proposés se limitaient au mot à corriger sans prendre en compte les mots proches. Ils fonctionnaient principalement à base de règles typographiques et de distance d'édition [Damerau, 1964 ; Kernighan et al. 1990], ou avec un système de vérification de lexique plus complet et compréhensif [Oflazer, 1996]. Puis, le contexte a commencé à être pris en

compte. Pour ce faire, certains travaux se sont appuyés sur des modèles de langue *n-grammes* de mots [Brill et Moore, 2000 ; Park et Levy, 2011], phonétiques (Toutanova et Moore, 2002), ou encore sur des modèles de langue *n-grammes* qui combinaient ces deux caractéristiques [Boyd, 2009]. Ces *n-grammes* sont souvent associés à d'autres paramètres tels que la catégorie grammaticale, donc la transcription phonétique ou encore la longueur du mot à corriger.

Les méthodes utilisées dépendent toujours de l'objectif visé par le correcteur. Un correcteur qui cherche à corriger des fautes grammaticales en plus des fautes lexicales (comme ceux proposés par Yvon [2011]) doit être automatiquement plus robuste, la détection d'erreurs étant plus complexe à réaliser et la génération d'erreurs pour la correction devient plus risquée.

Dans ce mémoire, nous voulons corriger des mots sans prendre en compte le contexte dans lequel ils apparaissent. De tels systèmes se basent sur la distance d'édition introduite par Damerau [1964] et Levenshtein [1966].

Cette notion comprend quatre types d'opérations (l'insertion et la suppression d'une lettre, la substitution d'une lettre par une autre, l'inversion de deux lettres consécutives). Le but est de s'appuyer sur ces types, que l'on peut donc considérer comme des règles de correction, pour passer d'un mot mal orthographié à sa forme attendue. Cette méthode a été reprise et améliorée par de nombreux chercheurs. Des règles spécifiques pour les fautes de proximité clavier sont proposées [Sagot et Boullier, 2008]. Des connaissances linguistiques ont également été intégrées. Par exemple, Véronis (1988) et Sagot et Boullier (2008) traitent les fautes dites phonétiques (ex : remplacement de *o* par *eau*). Le contexte dans lequel s'applique la règle peut être pris en compte (Kernighan et al. 1990). Par ailleurs, la pondération de ces règles ne se fait plus systématiquement en fonction du nombre d'opérations effectuées sur le mot. Le choix du poids d'une correction se fait, manuellement ou non, en fonction de l'opération effectuée et des lettres concernées par cette opération (Véronis, 1988 ; Mitton, 1996 ; Sagot et Boullier, 2008). Ces règles s'appuient majoritairement soit sur des règles simples que l'on tente d'appliquer sur toutes les lettres d'un mot inconnu, soit sur des règles plus spécifiques (proximité clavier, phonétique) souvent listées

manuellement. L'objectif était d'acquérir automatiquement ces règles et de les pondérer en fonction de la fréquence de la faute correspondante. Pour cela, ils utilisent les techniques d'apprentissage par analogie pour induire des règles de correction.

La plupart des articles sur l'extraction d'information ne proposent pas des méthodes pour pouvoir obtenir des résultats pertinents à partir de corpus non normalisés car il n'existe pas d'outils de fouille de textes résistants à un bruit trop important. Des approches essaient de se passer de certaines parties de la normalisation [Dias et al. 2000], qui se base uniquement sur des statistiques de mots sans lemmatisation ni groupement des locutions.

Pour relever certaines ambiguïtés ou réduire le nombre d'opérations effectuées, il est souvent nécessaire d'orienter les traitements en fonction de l'objectif final qui peut être différent de la traduction et de l'indexation de textes. Deux principaux prétraitements sont présents dans la littérature du domaine de la fouille de textes. D'une part, les méthodes de lemmatisation des formes fléchies des mots et, d'autre part, les méthodes de segmentation d'un texte, notamment en phrases.

Les méthodes de segmentation des textes par apprentissage des frontières des phrases [Zhang et al. 2003] ou par utilisation d'automates et de listes élaborés manuellement [Friburger et al. 2000] font aussi l'objet d'un nombre significatif d'articles bien que chaque méthode reste très spécifique au corpus traité.

Pour le reste des prétraitements, il n'existe que peu d'articles [Grefenstette et Tapanainen, 1994 ; Habert et al. 1998 ; Mikheev, 2000 ; Sagot et Boullier, 2005] traitant explicitement du prétraitement des données textuelles.

Afin de faciliter la communication avec les patients étrangers ou d'immigrés avec maîtrise de la langue limitée, et de fournir des versions traduites de documents médicaux (documents réglementaires, articles scientifiques, des fiches de patients), il est nécessaire de faire appel à des traducteurs médicaux professionnels. Que ce soit pour employer des linguistes médicaux compétents ou des professionnels linguistiquement compétents c'est encore un sujet de controverse [Langlais, 2009], mais il semble que la solution la plus raisonnable

serait de concevoir une traduction et une vérification de système dans lequel les textes seraient traduits par un traducteur expérimenté qui se spécialise dans le domaine médical et ensuite vérifiés par un expert, et enfin subir une relecture par une tierce personne [Bowker, 2005]. Un tel processus change la situation actuelle dans laquelle le traducteur est souvent la seule personne responsable du processus de la traduction.

Dans les entreprises multinationales les départements de documentation technique publient des manuels pour leurs produits. Depuis plusieurs années, les rédacteurs techniques travaillent avec les équipes informaticiens et de la traduction pour la gestion de la traduction et pour la vérification des documents traduits.

Pour la gestion de la traduction, la langue source du système de gestion de contenu doit être vérifiée, ensuite les données sont envoyées au fournisseur de services de traduction et, après la traduction, il faut analyser l'exactitude formelle et terminologique des traductions qui comprend aussi une vérification complète et précise de la qualité.

Les traducteurs travaillant dans le secteur de la localisation sont confrontés à la tâche de produire des traductions de haute qualité dans un délai d'exécution très court. Une façon d'essayer d'équilibrer ces objectifs est d'utiliser la technologie de traduction comme des outils de TAO (traduction assistée par ordinateur). Un des outils de technologie de traduction les plus populaires disponibles est la mémoire de traduction (MT). Les MT sont souvent promues/présentées comme des outils qui peuvent aider les traducteurs à améliorer leur productivité, mais on en sait moins sur l'impact que l'utilisation de ces outils peut avoir sur la qualité de la traduction.

Les systèmes de gestion de mémoires de traduction stockent dans une base de données appelée une mémoire de traduction des segments de textes qui ont été traduits, ainsi que leurs traductions. (Ces segments correspondants, parfois appelés unités de traduction, peuvent être créés automatiquement lorsqu'un utilisateur traduit, ou peuvent être assemblés à partir d'un texte source et sa traduction existante en utilisant un aligneur bilingue.) La plupart des systèmes de

gestion de mémoires de traduction s'intègrent à ou intègrent des éditeurs de textes, afin d'en permettre l'usage pendant un processus de traduction. [O'Brien, 1998]. À l'aide d'un tel outil, un traducteur peut traduire un plus grand volume d'information dans une période de temps donnée. [Bowker, 2005]

Suite au développement du processus de la traduction, l'activité de post-édition devient une activité plus fréquente dans la localisation, comme opposition à la traduction complète de nouveaux textes. Dans une industrie qui se déplace si rapidement, il est mieux de se concentrer sur la finalisation de projets que sur le processus lui-même. Par conséquent, ces aides à la traduction sont utilisés dans la localisation flux de travail avec peu de données pour quantifier l'effort de traduction réelle et la qualité finale après la post-édition. Depuis, la productivité et la qualité ont un impact direct sur les prix, il est d'une importance capitale pour explorer cette relation en termes de productivité et de qualité de la post-édition des textes, provenant de systèmes de traduction-mémoire et traduit par les outils de TAO, par rapport à la traduction de textes sans aucune aide.

Une base de données est créée pour la combinaison de langues, par exemple russe - français et elle est alimentée au cours de la traduction. Le texte source est divisé automatiquement en segments (unités sémantiques). À chaque segment dans la langue source correspond un segment dans la langue cible. Ces segments sont enregistrés dans la MT.

Lors d'une nouvelle traduction, le texte à traduire est comparé avec le contenu de la base de données. Les segments qui se trouvent tels quels dans la base de données (*100 % match*) peuvent être prétraduits automatiquement. Lorsqu'un segment proche se trouve dans la base de données, on parle de *Fuzzy Match* (75 à 99 %). Lorsque cette valeur est inférieure à 75 %, on ne tient pas compte de la correspondance approximative car son utilisation n'est pas rentable.

Le processus de traduction et de localisation est un domaine où les multiples pratiques, systèmes et fournisseurs à travers l'entreprise avait un problème pour la vitesse, la qualité et le coût.

Dans ce contexte, il semble logique de penser que si les prix, la qualité et les temps sont déjà établis pour les MT selon le niveau différent des *Fuzzy Match*

alors nous avons seulement besoin de comparer des segments de MT avec des segments de MT, plutôt que de comparer la sortie de MT par la traduction humaine. Par conséquent, une fois que la corrélation est établie, le même ensemble de normes pour le temps, la qualité et prix peut être utilisé pour les types d'aide à la traduction. [Guerberof Arenas, 2008].

Des tâches à viser ont été les suivantes :

- ✓ correction automatique des caractères spéciaux
- ✓ revue d'un processus de chargement chez SDL
- ✓ création de dictionnaire pour les abréviations

Pour continuer à travailler sur le projet d'amélioration du processus, il me fallait de revoir toute l'histoire de changements du début pour trouver une solution d'avancer vers la simplification de traitement de fichier. Dans le Chapitre Matériel et méthodes inclut le matériel, ainsi que méthodes utilisées et aussi une petite modification à une méthode déjà appliquée.

## II. MATÉRIEL ET MÉTHODE

Dans le mémoire présent nous avons analysé les documents internes de GE tels que procédures qualité, les dossiers de l'historique de conception technique (qui sont considérés hors du domaine).

Les rapports d'intervention du CHU (Complaint Handling Unit-Service Report) visent à améliorer la qualité des produits et de soins de santé. Ils sont écrits dans un style narratif avec des phrases techniques standards, qui décrivent l'état des dispositifs médicaux et ils doivent être traduits en anglais pour la FDA, parce que la langue de service est bien l'anglais.

Selon la demande de la FDA les fabricants (manufacturers) sont tenus de signaler les problèmes de sécurité et d'analyser les problèmes de qualité de tous les événements de service non planifiés.

L'objectif de cette procédure consiste à définir les exigences et le procédé utilisés dans la reconnaissance, l'évaluation et la détermination du rapport sur les dispositifs médicaux (MDR - Medical Device Reporting), du signalement d'événements à l'autorité compétente ou une autre agence de réglementation, ainsi que pour la détermination du besoin d'investigations supplémentaires et la résolution des réclamations liées à des dispositifs médicaux distribués [GE Healthcare Global Quality Procedure GEHC\_GQP\_14.01, octobre 2014].

Quant au rapport sur les dispositifs médicaux, son règlement fournit un mécanisme qui permet à la FDA, ainsi qu'aux fabricants, d'identifier et de surveiller les événements indésirables (décès, blessures graves et des défaillances) impliquant les dispositifs médicaux. Le but est de détecter et de corriger les problèmes au plus vite. Les exigences de la réglementation du MDR sont appliquées sous l'autorité de la loi FFDCa (*Federal Food, Drug and Cosmetic Act*, principale loi fédérale qui régit les résidus de produits antiparasitaires et les autres additifs aux produits alimentaires qui sont potentiellement dangereux<sup>4</sup>). Les

---

<sup>4</sup> FD&C Act : Chapitre V Drugs and Devices  
[http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAAct/FDCAActChapterVDrugsandDevices/default.htm#Part\\_A](http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAAct/FDCAActChapterVDrugsandDevices/default.htm#Part_A)



mécanismes d'application comprennent saisie, l'injonction, les pénalités monétaires civiles, et de poursuites pénales.

Le règlement prévoit aussi des exigences de MDR, des effets indésirables de rapports d'événement et de tenue de dossiers pour les installations de dispositifs médicaux de l'utilisateur (par exemple, les hôpitaux et les maisons de soins infirmiers) et les importateurs de dispositifs médicaux, ainsi que les exigences de tenue de registres pour les distributeurs de dispositifs médicaux.

Cette procédure s'applique à tous les sites, modalités et fonctions où les dispositifs médicaux sont utilisés et qui fonctionne sous le système de management de la qualité de GE Healthcare (GEHC). Cela inclut les produits fabriqués, importés ou distribués par GEHC qui ont été mis sur le marché.

Les rapports RFS (**R**equest **F**or **S**ervice) sont collectés par téléphone (call center) et par écrit [par mail, ou via application spéciale *SnF-Store & Forward* (une application offline utilisée par les techniciens de maintenance pour faire des comptes rendus des interventions sur site)] rapportant des déficiences en rapport avec l'identité, la qualité, la durabilité, la fiabilité, l'efficacité, la sécurité, la performance ou la conformité réglementaire d'un produit. Tous les rapports ont format d'application pour les RFS qui s'appelle *MUST* (**M**ulti **U**ser **S**ervice **T**ool). Le processus de traitement des réclamations comporte les étapes suivantes :

- ✓ Examen
- ✓ Évaluation
- ✓ Clôture après intervention

Le CHU lit toutes les informations documentées par différents contributeurs afin de comprendre l'historique de l'événement. Cet historique doit donner suffisamment d'informations pour évaluer le risque de la plainte, un rapport.

Chaque RFS contient l'information sur le risque de danger. Le système traite les rapports avec mention SAFETY ISSUE en priorité, même si le rapport n'a pas été terminé. Les rapports rédigés en anglais sont immédiatement chargés dans le système TrackWise et sont visibles par le CHU. Les rapports rédigés dans

les autres langues sont traduits vers l'anglais au cours de 2 jours et chargés ensuite sur le système TrackWise. Les rapports avec mention NON-SAFETY sont extraits une fois il est terminé.

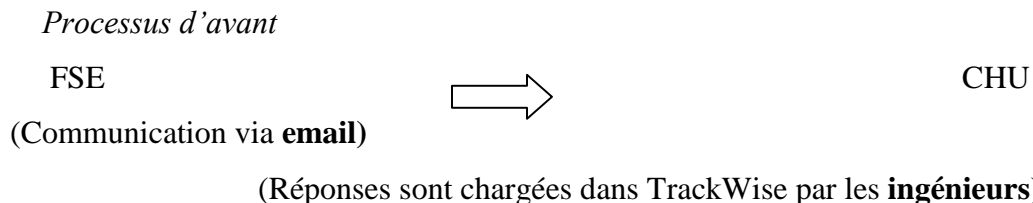
## Les étapes de traitement de fichiers

La première étape de traitement de fichiers consiste à charger les fichiers de rapports sur le serveur FTP (**F**ile **T**ransfer **P**rotocol). Ensuite, les fichiers sont téléchargés du FTP pour pouvoir les formater dans Excel et charger sur la plateforme TMS (**T**ranslation **M**anagement **S**ystem), application de gestion globale de l'information qui permet de gérer les traductions et de s'occuper de localisation et apporter le flux de contrôle, d'intégration et de processus requis pour préparer un contenu propre (<http://www.sdl.com/fr/>), pour traduire. La troisième et quatrième étape sont gérées par SDL qui comprend l'utilisation de Mémoire de Traduction (Translation Memory) et l'intervention des traducteurs freelance.

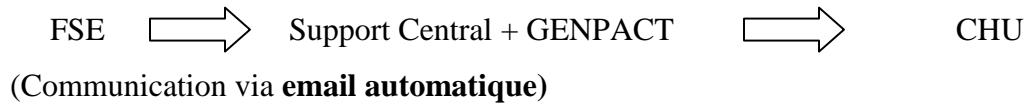
C'est au moment du transfert de fichiers que nous rencontrons les problèmes de caractères corrompus pour certaines langues, notamment les langues avec les accents et caractères spéciaux, ainsi que les langues qui utilisent l'alphabet cyrillique, les lettres non-accentuées, un manque d'espace insécable qui posent un problème pour la traduction, en faisant du bruit et augmente les coûts de traduction.

## CHU Translation Process et le fichier MUST

L'équipe de service de GE Healthcare a lancé un projet de simplification CHU en 2013 pour améliorer le processus de réception de réponses des techniciens de maintenance (réclamations) :



*Nouveau processus*



(Réponses sont chargées dans TrackWise par **GENPACT**)

Cette mise à jour permet de traiter les rapports d'intervention à la façon plus rapide, de créer la base de données de connaissances pour analyser les défauts et surveiller le travail de techniciens. Suite à la formation préalable, les rapports ont plus de qualité (ils sont mieux rédigés).

En janvier 2014 l'équipe de service a créé l'initiative de réduction de coût de traduction de rapports d'intervention par segmentation de phrases. Cela permet de diminuer le pourcentage de mots payés par intégrer la segmentation de phrases pour augmenter 100 % match en vue :

- ✓ Améliorer la qualité de la traduction de rapports d'intervention;
- ✓ Une exécution plus rapide ;
- ✓ Une plus grande visibilité;
- ✓ Réduire les coûts.

Le principe de segmentation de phrases est de diviser les phrases longues en phrases courtes afin de créer la base avec les phrases prédéfinies, ainsi que les dates et enrichir la mémoire de traduction.

En février 2014 plusieurs tests ont été effectués par SDL pour analyser les impacts sur les traductions de rapports et, selon les résultats, la mémoire de traduction (MT) a été modifiée (mars 2014). Les résultats de cette expérimentation ont été satisfaisants :

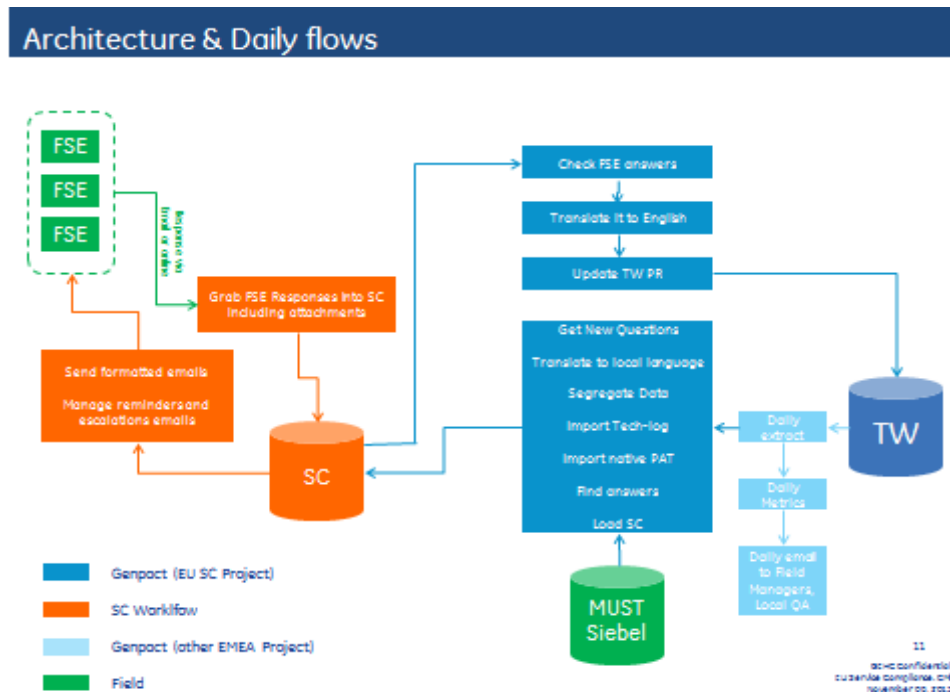
- 30% de la réduction de coûts (=360K€) pour les 4 langues principales qui sont le français (FR), l'espagnol (ES), l'allemand (DE) et l'italien (ITA) de rapports d'intervention (85 % de rapports sont rédigés en ces 4 langues)

- Exécution plus rapide

**Pour atteindre le but nous avons désigné un plan de développement suivant :**

- Renforcer la communication et le suivi avec le fournisseur SDL
- Optimiser l'extraction de données et de préparer des fichiers pour la traduction
- Renforcer l'outil de TMS : des abréviations, des phrases concaténation, la création de glossaire
- Mise en place de rapports réguliers sur l'utilisation de la plateforme TMS et proposer des moyens d'optimisation. Insérez rapports statistiques.
- Améliorer la communication entre l'équipe de traduction et d'opérations (mises à jour de SnF)
- Formation des ingénieurs de terrain sur les bonnes pratiques en fonction des améliorations futurs de TMS.

L'image ci-dessus montre l'architecture globale du processus du traitement des réclamations : le fichier MUST doit être traduit de la langue locale vers l'anglais.



Chaque réclamation comprend 3 parties importantes pour décrire un problème rencontré lors de l'intervention, qui s'appellent PAT (**P**roblem Statement, **A**ction Taken, **T**est) :

P // Décrire problème

A// Indiquer ce qu'a été fait pour réparer ; faire un diagnostic, pas de recommandations à faire sur place

T// Vérification des données et résultats des tests (ex : *Équipement passé tous les tests nécessaires et est prêt à l'emploi – 01*).



Pour information, le Centre Service collecte des informations de l'ingénieur (quels problèmes ont été détectés), il signale *Actual Safety* ou *Potential Safety*. Il doit également demander si le patient avait été impliqué, les dossiers des patients, les informations de l'expérience.

Le fichier MUST<sup>5</sup> (application Mainframe utilisé par CSC<sup>6</sup> et BOS<sup>7</sup>) est généré à partir des réclamations techniciens de maintenance sous forme de texte qui est ensuite traité sous format Excel avec les colonnes spécifiques à l'aide d'une macro créée par GENPACT (Voir annexe D).

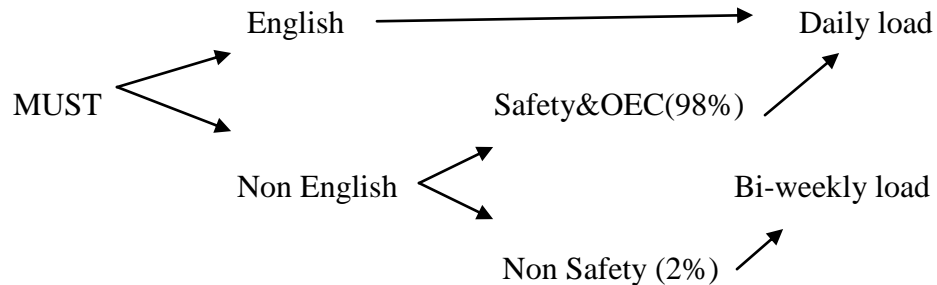
Avant de charger le fichier Excel de MUST sur TMS il y a un processus géré par GENPACT qui comprend les étapes suivantes :

---

<sup>5</sup> **M**ulti **U**ser **S**ervice **T**ool  
<sup>6</sup> **C**ustomer **S**ervice **C**enter  
<sup>7</sup> **B**usiness **O**perations **S**pecialist

## Load Timing

Le nombre de fichiers dépend des degrés de risque de danger de la santé du patient et de la langue du compte-rendu :



**Modality split** : Modalités d'imagerie médicale les plus utilisées

DI (Vascular+LCS+U/S+HCIT) ~70%

OEC~10%

CS~10%

HCIT~10%

**Field Load by Language Split** : Il comprend 21 langues

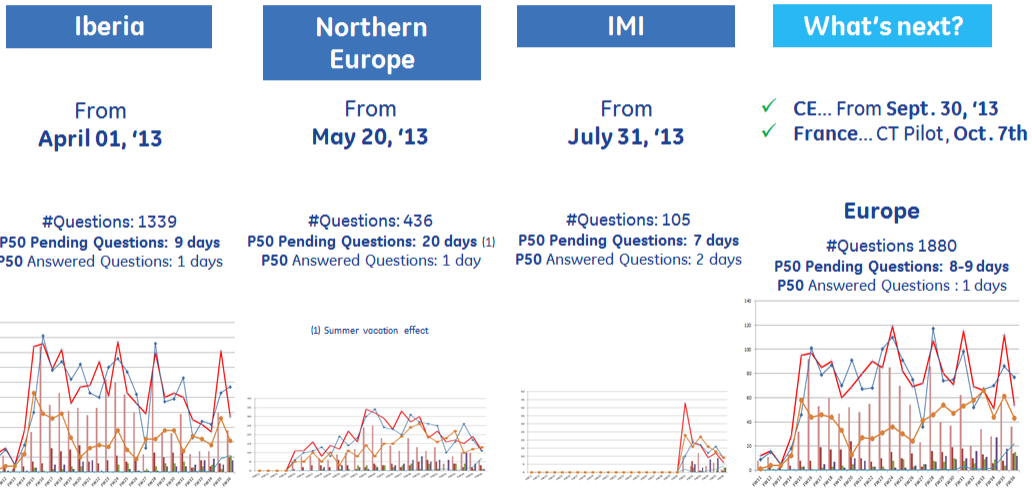
All languages : 3 fields translated —→ PAT

(CUST\_INCDNT\_ST ; FE INCDNT ST; ACTION), sauf Bulgarian, Greek:

multiple fields translated —→PAT+CUST name, address, etc.

Ci-dessous on peut voir un planning de déploiement, qui montre l'évaluation des nombres de questions/demandes, ainsi que évaluation du temps passé pour le traitement des questions selon la région et la période :

# Progressive Rollout @ fw36



Après la génération des fichiers à partir d'un fichier globale avec des réclamations pour chaque pays et par modalité nous obtenons le fichier MUST:

COUNTRY	CT	CODELCT	NAM	ZONE	CUST	NAI	CONT	PH	MODALITY	PRODUCT	PRODUCT	DEVICE	II	ASSOCIA	DEVICE	L	DEATH	IN	DEATH	IN	CUS	INCL	FE	INCDN	ACTION	T	COMPON	FAILURE	REPAIR	C	TEST	
FRANCE	00000000	FRANCE	LI		MR	BILLAF	06 64 90 3	DXR	SENO	ESSI	SENO	I1919220		N	N	N					#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		MME	BAS	03 20 48 3	DXR	SENO	ESSI	SENO	ESS. M4471522		N	N	N					#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		MATHILDE	032199308	DXR		DEFINIUM	DEFINIUM	M4167160		N	N	N						#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		MR	MARTI	03 21 23 0	CT	CT	LIGHTS	ECT	SELEC	M1870683		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		MR	DRUEL	03 22 26 8	DXR	PRESTIGE	PRESTIGE	B1529607		Y	N	N						#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		ME	ADELA	03 22 26 8	MR	3.0T	HDX	3.0T	HDX	M4160077		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		VEROVE	J1	032022505	MR	3.0T	MR	73-0T	MR75	M9481853		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MELANIE	03 21 31 5	DXR		SENO	ESSI	SENO	ESSI	X1520492C		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	
FRANCE	00000000	FRANCE	LI		MME	WA	032724328	DXR	SENO	ESSI	SENO	ESSI	M4481018		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		emelyne	032121187	MR		1.5T	MR	41	1.5T	MR45	M4162221		N	N	N			#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MR	DEHEZ		MR	1.5T	HDX	1.5T	HDX	M4164337		NOT	REPC	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MME	LEG	05 56 12 8	MR	1.5T	MR	41	1.5T	MR45	X2181501		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		PIERRE	M		MR	1.5T	HDX	1.5T	HDX	M4164331		Y	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MARIANN	04 92 03 1	CT		CT	OPTIM	OPTIMA	C	M4134633		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		SILVIE	04 92 03 7	MR		1.5T	HDX	1.5T	LX	TO	M4194447		N	N	N			#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MANIP	00000000	CT		CT	LIGHTS	LS	RT	16	B	M4134623		N	N	N		#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		SYLVIE	049203771	MR		1.5T	HDX	1.5T	LX	TO	M4194447		N	N	N			#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MME	ANN	04 93 26 5	DXR	PRESTILIX	PRESTILIX	M6004403			N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####
FRANCE	00000000	FRANCE	LI		MME	AIEL	06 72 76 4	DXR	SENO	ESSI	SENO	ESSI	I9237511		N	N	N				#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####	#####

W34      QUANDO SI RICHIEDE LA STAMPA GIORNALIERA SI BLOC CA IL PC - PATIENT EXPERIENCE 29-JAN-2015: NESSUN PAZIENTE COINVOLTO.

I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA					
ITALY	00000000	AREZZO CENTER	DSSA	BIAP	05755680	HCIT	CENTRICIT	PACS	COR	A5803803	RA1000	N	N	N	PC	REFERT	25-FEB-20	25-FEB-20	KB	EC			
ITALY	00000000	AREZZO CENTER	SEGRETER	05593866	HCIT	CENTRICIT	CENTRICIT	A5368605	WORKSTA	N	N	N	N	UNA	POST	26-FEB-20	26-FEB-20	JM	EC				
ITALY	00000000	ITALY NORTH EAST	SIG.A	MO	05442851	HCIT	IDI	Works	IDI	WORKI	A5139640	IDI	4.7	N	N	N	HA	BUFFA	03-DEC-20	03-DEC-20	B5	S1	
ITALY	00000000	ITALY NORTH EAST	SIG.RA	MC	05442851	HCIT	IDI	Works	IDI	WORKI	A5139640	IDI	4.7	N	N	N	IL	CLT	SEG	09-FEB-20	09-FEB-20	B5	S1
ITALY	00000000	ITALY NORTH EAST	MONTALT	3.35E+09	HCIT	IDI	Works	IDI	WORKI	A5139640	IDI	4.7	N	N	N	NON	FUNC	02-MAR-20	02-MAR-20	B5	S1		
ITALY	00000000	ITALY NORTH WEST	DR	DANES	02695169	HCIT	CENTRICIT	CENTRICIT	A5121808	RIS	N	N	N	N	NON	SI	RH	01-DEC-20	01-DEC-20	WS	VI		
ITALY	00000000	ITALY NORTH WEST	REPARTO	03825921	HCIT	CENTRICIT	CENTRICIT	F5119170	PACS	IV	N	N	N	N	PROBLEM	07-JAN-20	07-JAN-20	B5	S1				
ITALY	00000000	ITALY CENTER	EGIDIO	LA	3.45E+09	HCIT	CRITICAL	(	B5104395	Anaesthe	NOT	REPC	N	N	SU	RICHIE	03-MAR-20	03-MAR-20	AD	XA			
ITALY	00000000	ITALY NORTH WEST	CAPOSAL	01343933	HCIT	CENTRICIT	A5604847	CCC	N	N	N	N	N	NON	FUN	28-OCT-20	28-OCT-20	N5	SF				
ITALY	00000000	ITALY NORTH WEST	STEFANO	011	31316	HCIT	CENTRICIT	C5177244	CA	N	N	N	N	EMAIL	CL	06-NOV-20	06-NOV-20	B5	S1				
ITALY	00000000	ITALY NORTH WEST	SIG.RA	ME	011313591	HCIT	CENTRICIT	C5177244	CA	N	N	N	N	NON	SI	CC	02-DEC-20	02-DEC-20	N5	TE			
ITALY	00000000	ITALY NORTH WEST	DSA	AGN	3.92E+09	HCIT	CENTRICIT	CENTRICIT	B5177230	N.A	N	N	N	N	DUE	POST	16-FEB-20	16-FEB-20	B5	S1			
ITALY	00000000	ITALY NORTH WEST	GIADA	03526750	HCIT	CENTRICIT	A5360413	CCC	N	N	N	N	N	QUANDO	19-JAN-20	19-JAN-20	B5	S1					
ITALY	00000000	ITALY NORTH WEST	ALESSAND	03526750	HCIT	CENTRICIT	A5360413	CCC	N	N	N	N	N	ALCUNE	E	02-MAR-20	02-MAR-20	SA	DM				
ITALY	00000000	ITALY NORTH WEST	SIG.	ALESS	03526750	HCIT	CENTRICIT	A5360413	CCC	N	N	N	N	UN	UTENT	03-MAR-20	03-MAR-20	B5	S1				
ITALY	00000000	ITALY NORTH WEST	ALESSAND	03526750	HCIT	CENTRICIT	A5360413	CCC	N	N	N	N	N	RICHIE	03-MAR-20	03-MAR-20	B5	S1					
ITALY	00000000	ITALY NORTH EAST	SIG.	CARR	0423.7318	HCIT	CARDOAS	A5105110	CARDOAS	N	N	N	N	SI	RICHIE	04-NOV-20	04-NOV-20	B5	S1				
ITALY	00000000	ITALY NORTH EAST	SIG.	GRAZI	042373241	HCIT	CENTRICIT	CENTRICIT	A5105147	PACS	EE	N	N	NON	RIES	20-JAN-20	20-JAN-20	B5	S1				
ITALY	00000000	ITALY NORTH EAST	SIG.	GRAZI	042373241	HCIT	CENTRICIT	CENTRICIT	A5105196	PACS	EE	N	N	IN	FASE	D	20-JAN-20	20-JAN-20	B5	S1			

En prenant en compte le transfert de données, il y a toujours une forte possibilité de trouver des informations manquantes et des caractères spéciaux pour les langues accentuées comme le français, l'italien, l'allemand et l'espagnol qui peuvent empêcher de traduire les fichiers correctement et rapidement. Après avoir analysé les fichiers FR, ITA et DE, nous avons vite remarqué les caractères corrompus, le manque des accents, par exemple dans le fichier FR :

L ANTENNE RACHIS SE METS EN PAUSE EN COURS D EXAMEN - PATIENT EXPERIENCE 20-JUN-2014: Examen interrompu/repris.

Les caractères corrompus ont été corrigés par l'équipe de IT Services, et pour la correction des accents et des mots mal tapés, avec des fautes d'orthographe, ainsi que des abréviations qui peuvent être non-traduits nous avons créé un glossaire pour 4 langues principales (le français, l'italien, l'allemand et l'espagnol), suivi par macro afin d'impliquer sur le fichier Excel. Pour créer un glossaire nous avons analysé plusieurs fichiers MUST déjà compilés et en cours de traduction pour trouver les erreurs répétées et les corriger afin d'optimiser



le processus et augmenter l'effet de levier (leverage)<sup>8</sup> de MT (mémoire de la traduction). Pour certains cas c'est essentiel parce que si le mot est écrit correctement, ça peut faire partie d'une phrase standard. En dessous, on peut voir l'extrait d'un glossaire pour la langue française (Voir annexe II) :

acquisition	d'acquisition	d'acquisition	l'acquisition	l'acquisition	acquisition	acquisition	pb	problème		C'est	C'est	n a pas	n'a pas	s'est	s'est
affaire	d'affaire	d'affaire	l'affaire	l'affaire	ecran	écran	err	erreur		c'était	c'était	n est pas	n'est pas	s'arrete	s'arrête
affiche	d'affiche	d'affiche	l'affiche	l'affiche	arriere	arrière	mess	message		c'était	c'était	n ya	n'y a	s'arrete	s'arrête
alarme	d'alarme	d'alarme	l'alarme	l'alarme	controle	contrôle	msg	message				n arrive	n'arrive	s'arrete	s'arrête
allumage	d'allumage	d'allumage	l'allumage	l'allumage	cliches	clichés	qd	quand				n a que	n'à que	s'endench	s'endench
ampoule	d'ampoule	d'ampoule	l'ampoule	l'ampoule	tete	tête	tech	technicien				n ayant	n'ayant	s'allume	s'allume
anode	d'anode	d'anode	l'anode	l'anode	consequence	conséquence	tec	technicien						s'il	s'il
antenne	d'antenne	d'antenne	l'antenne	l'antenne	succes	succès	pr	pour						s'ouvrir	s'ouvrir
apercevoir	d'apercevoir	d'apercevoir	l'apercevoir	l'apercevoir	egalement	également	bcp	beaucoup						s'eteint	s'êteint
appareil	d'appareil	d'appareil	l'appareil	l'appareil	examin	examen	apm	après-midi						s'eteint	s'êteint

Nous avons intégré le glossaire pour chaque langue dans une macro VBA de GENPACT afin de faciliter les corrections automatique des mots mal orthographiés et saisis (Voir annexe III).

L'équipe de Service a également créé un fichier avec des phrases standards pour chaque langue afin d'améliorer la TM. Nous intégrons les phrases standards dans la mémoire de traduction. Le résultat est suivant :

<sup>8</sup> L'effet de levier explique le taux de rentabilité comptable des capitaux propres en fonction du taux de rentabilité après impôt de l'actif économique (rentabilité économique) et du coût de la dette. Si ce scénario prévu par les associés se réalise, le taux de rentabilité économique sera supérieur au coût de l'emprunt. Sous cette condition, une augmentation relative des dettes se traduira par une augmentation du taux de rentabilité financière des associés.

Le taux de rentabilité financière sera d'autant plus grand que le montant du résultat courant devra rémunérer un montant faible de capitaux propres, ce qui est le cas si le projet a été financé par un endettement important. On dit que les associés bénéficient d'un effet de levier.

<http://www.iutbayonne.univ-pau.fr/~berterre/FormationsGestion/diagnosticfina/rentabilite/effelevier.html>

1	Sheet1	1	Sheet1
2	SYSTEM WZORAZ ZERWAL BADANIE GLOWY, DZIS POJAWIL SIE KOMUNIKAT SCANNER HAR DWARE STOPPED SCAN , RESTART KONSOLI NIE POMOGL, CALKOWITY TEZ NIE, TERAZ NIE R EAGUJE NA PRZYCISK CONFIRMED- PATIENT NOT INVOLVED, SYSTEM DOWN	2	YESTERDAY THE SYSTEM ABORTED THE HEAD EXAMINATION, TODAY THERE WAS A MESSAGE DISPLAYED: SCANNER HARDWARE STOPPED SCAN, THE RESTART OF THE CONSOLE DID NOT HELP, FULL RESTART DID NOT HELP EITHER, NOW IT DOES NOT REACT ON CONFIRMED BUTTON - PATIENT WAS NOT INVOLVED, SYSTEM DOWN
3	PATIENT EXPERIENCE	3	PATIENT EXPERIENCE
4	04-JUL-2014:	4	04-JUL-2014:
5	NO PATIENT IMPACT	5	NO IMPACT ON PATIENTS
6	16-DEC-2013:	6	16-DEC-2013:
7	system przerwal badanie i nie skanuje - badania nie mozna bylo dokonczyc	7	the system has aborted examinations and it is not scanning now - examinations could not be completed
8	CAUSE:	8	CAUSE:
9	corrupted database	9	corrupted database
10	16-DEC-2013:	10	16-DEC-2013:
11	zrekonstruowano baze danych surowych (raw data)	11	the raw database was reconstructed
12	16-DEC-2013:	12	16-DEC-2013:
13	wykonano badania testowe	13	tests exams were performed

Nous avons analysés une grande partie des rapports d'intervention de techniciens de maintenance pour pouvoir constituer une liste de phrases fréquentes lors d'un bilan sur intervention. Le petit script de Python pour compter la fréquence des phrases a été utilisé (Voir annexe IV)

Nous pouvons donc résumer toutes les modifications qui ont été impliquées pour le traitement automatique de réclamations :

#### 1/Modification de concaténation à l'aide de l'outil

- Phrases pré-traduites : la TM est mise à jour avec la liste des phrases pré-traduites. Cela devrait améliorer l'effet de levier (leverage) de la TM.
- Traduction automatique des dates
- Segmentation //

#### Le compte rendu typique dans TMS :

- ✓ 18-JAN-2014 : *I have found this problem.* – CAUSE : *This was because of this* // 19-JAN-2014: *I have more details now* // 20-JAN-2014: *I have done this repair*

**Est modifié à la façon suivante:**

- 18-JAN-2014 (ignored field)
- ✓ *I have found this problem.*
- – CAUSE: (ignored field)
- ✓ *This was because of this*
- // (ignored field)
- 19-JAN-2014:
- ✓ *I have more details now*
- // (ignored field) (ignored field)
- 20-JAN-2014: (ignored field)
- ✓ *I have done this repair*

**HCIT\_Weekly\_MUST\_FW4\_FRANCE.xls avec la segmentation:**

**Source excel file**

**Old segmentation**

119	13-DEC-2013: LES EXAMENS N'ARRIVENT PLUS DANS L'ARCHIVE. PAS DE PERTE DE DONNES PATIENT. - CAUSE: INCONNUE. // 20-JAN-2014: PROBLEME D ARCHIVAGE VERS L EA DU DC - CAUSE: MISE EN CONFIG DE LA PURGE	119	13-DEC-2013: LES EXAMENS N'ARRIVENT PLUS DANS L'ARCHIVE. PAS DE PERTE DE DONNES PATIENT. - CAUSE: INCONNUE. // 20-JAN-2014: PROBLEME D ARCHIVAGE VERS L EA DU DC - CAUSE: MISE EN CONFIG DE LA PURGE
120	13-DEC-2013: JOURNAL D'ERREUR CONSULT, PAS DE CONCLUSION // 20-JAN-2014: DEMANDE DE CLOTURE PAR ORANGE	120	13-DEC-2013: JOURNAL D'ERREUR CONSULT, PAS DE CONCLUSION, // 20-JAN-2014: DEMANDE DE CLOTURE PAR ORANGE
121	13-DEC-2013: SANS OBJET. - TEST RESULT: FIX OR REPAIR NOT COMPLETE, FOLLOW-UP ACTION NEEDED // 20-JAN-2014: TICKET FERM - TEST RESULIT: FOIIPMENT PASSED ALI REFOIIRFD TESTS AND IS	121	DEC-02-2013: NOT APPLICABLE. - TEST RESULT: FIX OR REPAIR NOT COMPLETE, FOLLOW-UP ACTION NEEDED // DEC-04-2013: check 20 cm OA - TEST RESULIT: FOIIPMENT PASSED ALI REFOIIRFD TESTS AND IS

**New segmentation**

201	//	201	//
202	20-JAN-2014: PROBLEME D ARCHIVAGE VERS L EA DU DC - CAUSE: MISE EN CONFIG DE LA PURGE	202	20-JAN-2014: PROBLEME D ARCHIVAGE VERS L EA DU DC - CAUSE: MISE EN CONFIG DE LA PURGE
203	13-DEC-2013: JOURNAL D'ERREUR CONSULT, PAS DE CONCLUSION	203	DEC-13-2013: CHECKED ERROR LOG, NO CONCLUSION
204	//	204	//
205	20-JAN-2014: DEMANDE DE CLOTURE PAR ORANGE	205	20-JAN-2014: DEMANDE DE CLOTURE PAR ORANGE
206	13-DEC-2013: SANS OBJET. - TEST RESULT: FIX OR REPAIR NOT COMPLETE, FOLLOW-UP ACTION NEEDED	206	SEP-23-2013: NOT APPLICABLE. - TEST RESULT: FIX OR REPAIR NOT COMPLETE, FOLLOW-UP ACTION NEEDED

2/ Les résultats attendus pour la segmentation //

	File	New	Fuzzy	Reps	100 %	Leverage	Costs
<b>Old segmentation</b>	SPAIN FW04	5,804	888	124	269	3.80%	\$999.68
<b>New segmentation</b>		5,232	979	268	608	8.58%	\$915.44
<b>Difference</b>		-572	91	144	339	4.78%	-\$84.24

En utilisant la nouvelle règle de segmentation, nous obtenons:

- ✓ 572 nouveaux mots
- ✓ +339 100% matchs
- ✓ Augmentation de l'effet de levier de 4,78%
- ✓ 84.24 \$ d'économie

	File	New	Fuzzy	Reps	100 %	Leverage	Costs
<b>Old segmentation</b>	FRANCE FW04	4,226	1,166	23	28	0.51%	\$769.44
<b>New segmentation</b>		3,681	1,599	36	129	2.37%	\$716.88
<b>Difference</b>		-545	433	13	101	1.85%	-\$52.56

En utilisant la nouvelle règle de segmentation, nous obtenons:

- ✓ -545 Nouveaux mots
- ✓ +101 100% matchs
- ✓ Augmentation de l'effet de levier de 1,85%
- ✓ 52.56 \$ d'économie

## Old segmentation

927-NOV-2013: NINGUNA, EL INGENIERO SE DESPLAZAR AL CENTRO -  
 TEST RESULT: PROBLEM NOT CORRECTED. COMMUNICATED FINDINGS  
 TO CUSTOMER // 07-FEB-2014: No se hacen verificaciones debido a la no  
 aceptacion del presupuesto. TEST RESULT: NO SERVICE PERFORMED

927-NOV-2013: NINGUNA, EL INGENIERO SE DESPLAZAR AL CENTRO -  
 TEST RESULT: PROBLEM NOT CORRECTED. COMMUNICATED FINDINGS  
 TO CUSTOMER // 07-FEB-2014: No se hacen verificaciones debido a la no  
 aceptacion del presupuesto. TEST RESULT: NO SERVICE PERFORMED

- Text split into 1197 segments
- New words: 24,807
- 100% matches: 1,127

## New segmentation

33 27-NOV-2013:  
 34 NINGUNA, EL INGENIERO SE DESPLAZAR AL CENTRO.  
 35 TEST RESULT: PROBLEM NOT CORRECTED. COMMUNICATED FINDINGS  
 TO CUSTOMER  
 36 //  
 37 07-FEB-2014:  
 38 No se hacen verificaciones debido a la no aceptacion del presupuesto.  
 39 TEST RESULT: NO SERVICE PERFORMED

33 NOV-27-2013:  
 34 NINGUNA EL INGENIERO SE DESPLAZAR AL CENTRO  
 35 na - TEST RESULT: PROBLEM NOT CORRECTED. COMMUNICATED  
 FINDINGS TO CUSTOMER  
 36 //  
 37 FEB-07-2014: HUB  
 38 No se hacen verificaciones debido a la no aceptacion del presupuesto.  
 39 TEST RESULT: NO SERVICE PERFORMED

- Text split into 5647 segments
- New words: 17,741
- 100% matches: 7,530

Lors de l'analyse des fichiers et les mémoires de traduction nous avons détecté l'apparition de l'anglais dans la mémoire de traduction des autres langues (mélange de langues) (voir les images ci-dessus), ainsi que la traduction de l'anglais vers l'anglais qui empêchait de gagner en productivité. Si le technicien rédige la partie de sa réclamation en anglais, il fallait mettre en fonction la règle pour la mémoire de traduction qui dit de ne pas traduire l'anglais, mais seulement la langue locale. C'est un problème pour l'entreprise de devoir corriger des documents mal rédigés en anglais, et que les prétraitements peuvent servir à améliorer et faciliter le processus de traduction.

## Exemple MUST\_DI\_Weekly\_Poland\_10 07

MUST_DI_Weekly_POLAND0_11_07.xls:itd - SDL Edit	
File Edit Format Format Paint View Tools Translation Memory TermBase Help	
1 Sheet1	1 Sheet1
2 SYSTEM W CZORAJ ZERWAŁ BADANIE GŁOWY, DZIS POJAWIŁ SIĘ KOMUNIKAT SCANNER HARDWARE STOPPED SCAN, RESTART KONSOLI NIE POMOGŁ, CAŁKOWITY TEŻ NIE, TERAZ NIE REAGUJE NA PRZYCIŚNĄCIE POTWIERDZONE - PACJENTA NIE WOLNO WOLNO, SYSTEM DŁUGO	2 YESTERDAY THE SYSTEM ABORTED THE HEAD EXAMINATION, TODAY THERE WAS A MESSAGE DISPLAYED: SCANNER HARDWARE STOPPED SCAN, THE RESTART OF THE CONSOLE DID NOT HELP, FULL RESTART DID NOT HELP EITHER, NOW IT DOES NOT REACT ON CONFIRMED BUTTON - PATIENT WAS NOT INVOLVED, SYSTEM DOWN
3 PATIENT EXPERIENCE	3 PATIENT EXPERIENCE
4 04-JUL-2014:	4 04-JUL-2014:
5 NO PATIENT IMPACT	5 NO IMPACT ON PATIENTS
6 16-DEC-2013:	6 16-DEC-2013:
7 system przerwał badanie i nie skanuje - badania nie można było dokończyć	7 the system has aborted examinations and it is not scanning now - examinations could not be completed
8 CAUSE:	8 CAUSE:
9 corrupted database	9 corrupted database
10 16-DEC-2013:	10 16-DEC-2013:
11 zrekonstruowano bazy danych surowych (raw data)	11 the raw database was reconstructed
12 16-DEC-2013:	12 16-DEC-2013:
13 wykonano badania testowe	13 tests exams were performed

## Exemple DI\_Weekly\_MUST\_Belgium\_FW26\_FRA

DI_Weekly_MUST_BELGIUM_FW26_FRA.xls:itd - SDL Edit	
File Edit Format Format Paint View Tools Translation Memory TermBase Help	
34 DISQUE DEFECT	34 DISK DEFECT
35 20-JUN-2014:	35 JUN-20-2014:
36 REMPLACEMENT DISQUE DUR FORMATAGE	36 REPLACED HARD DRIVE FORMATTING
37 20-JUN-2014:	37 JUN-20-2014:
38 Une acquisition de test a été réalisée. Le test a réussi.	38 A test acquisition was performed; test was successful.
39 TEST RESULT: EQUIPMENT PASSED ALL REQUIRED TEST AND IS READY FOR USE	39 TEST RESULT: EQUIPMENT PASSED ALL REQUIRED TESTS AND IS READY FOR USE
40 WHEN SYSTEM TURNED ON CUST GETS THE MESSAGE THAT THE CALIBRATION SOLUTIONS WILL EXPIRE WITHIN 3 DAYS AND STRANGE NUMBERS APPEAR SUCH AS 1, 5 DT50 210	40 WHEN SYSTEM TURNED ON CUST GETS THE MESSAGE THAT THE CALIBRATION SOLUTIONS WILL EXPIRE WITHIN 3 DAYS AND STRANGE NUMBERS APPEAR SUCH AS 1, 5 DT50 210
41 PATIENT EXPERIENCE	41 PATIENT EXPERIENCE
42 22-JUN-2014:	42 JUN-22-2014:
43 no patient involved	43 No patient involved
44 20-JUN-2014:	44 JUN-20-2014:
45 pop up calibration detector needed sur aero DR	45 Pop up calibration detector needed on aero DR
46 CAUSE:	46 Cause:
47 Calibrations nécessaires tous les 3 mois	47 Calibrations needed every 3 months
48 20-JUN-2014:	48 JUN-20-2014:

### III. PERSPECTIVES

La recherche et le travail effectués lors de projet CHU Translation Productivity avaient pour but d'améliorer le système de traitement des fichiers avant de les envoyer en traduction. Les corrections des fichiers source nous ont permis d'améliorer le système de prétraitement des fichiers de réclamations de CHU, ainsi que d'avoir un certain niveau de productivité de la traduction.

Bien évidemment, il reste pas mal de choses à améliorer au niveau de traitement des rapports d'intervention, qui s'adapte en fonction du développement des nouveaux outils et plateformes pour le traitement des fichiers.

Récemment, après avoir changé la plateforme principale pour traitements de réclamations, l'équipe de service a découvert un problème concernant l'apparence de phrases dupliquées dans les fichiers pour le français de manière un peu étrange : dans le fichier il y avait une phrase contenant des caractères corrompus suivis par une phrase déjà corrigée. Il est apparu sur la plate-forme PQM suite à l'intégration du processus *IRSR* (*IRSR* = Initial Review of Service Records) dans le PQM. Cela peut avoir un impact significatif sur le volume de la traduction, ainsi que sur les coûts ou on perd l'effet de levier.

Un autre problème que nous avons détecté lors des analyses de fichiers est la façon du traitement des fichiers en Cyrillique, notamment le russe (les autres langues qui utilisent l'alphabet cyrillique, comme bulgare, serbe, macédonien, sont traités correctement dans le système). Généralement, les rapports des interventions dans les hôpitaux russes sont écrits en anglais, mais il y a éventuellement des phrases en russe qui le système ne traite pas correctement, en laissant les caractères corrompus, non lisible, qu'ensuite empêchent de traduire les fichiers proprement et laisse les informations non claires.

Il y a deux étapes de résoudre ce problème :

- ✓ Effectuer une extraction de test en cyrillique et vérifier que les deux caractères cyrilliques et latins ne sont pas corrompus (cela sera

possible à faire sur un environnement de test avec la collaboration de l'équipe IT) ;

- ✓ Mettre en œuvre un changement de conception pour extraire le cyrillique avec le bulgare et le russe. Nécessité d'évaluer l'impact de validation de documentation et contrôler si le changement a eu un impact sur les plateformes de GENPACT (SOP [Standard Operating Procedure Software] et éventuellement sur les logiciels).



## CONCLUSION

L'objectif de cette recherche était d'analyser le processus d'amélioration de la qualité de traduction et de la productivité.

Pour atteindre ce but plusieurs **méthodes** ont été utilisées :

- ✓ Création des macros/des scripts qui permettent de préparer les fichiers pour une traduction optimisée ;
- ✓ Optimisation de l'utilisation de l'outil TMS (abréviations, concaténation de phrases, création de glossaire) ;
- ✓ Proposition d'améliorations des outils/plateformes IT ;
- ✓ Étude statistique (data management) ;
- ✓ Rédaction de guides de bonnes pratiques à l'usage des ingénieurs ;
- ✓ Nettoyage des mémoires de traduction

Comme résultats globaux, avec les améliorations appliquées nous avons obtenu :

- ✓ La lisibilité d'informations a devenue meilleure et plus pertinente ;
- ✓ Le meilleur suivi ;
- ✓ Productivité ;
- ✓ Des coûts partagés.

L'autre service de GE Healthcare HCIT s'est intéressé au projet d'optimisation et de la traduction et ils nous ont demandé de pouvoir appliquer la même méthode sur ses ressources à eux.

## BIBLIOGRAPHIE

Guerberof Arenas, Ana Productivity and quality in the post-editing of outputs from translation memories and machine translation, 2008

Bowker L. (2005) Productivity vs Quality? A pilot study on the impact of translation memory systems, *Localisation Focus* 4(1):13-20.

Pecina P., Dušek O., Goeuriotb L., Hajic J. Adaptation of machine translation for multilingual information retrieval in the medical domain, 2013

Andriesen, Simon (2006). 'Quality Control in Medical Translations', *Innovations in Pharmaceutical Technology*

Mobaraki, Mohsen and Aminzadeh, Sirwan (2012). 'A Study on Different Translation Evaluation Strategies to Introduce an Eclectic Method'a. *International Journal of English Linguistics*

Gouadec, Daniel (2010). 'Quality in translation' Yves Gambier and Luc van Doorslaer (eds.) *Handbook of Translation Studies: Volume 1*, 270–275.

Bulletin de l'Organisation mondiale de la Santé, Vol.93, Numéro 6, juin 2015, p.361- 436

Fontanet (2005) "La traduction technique : le texte sous l'empire de l'extratextuel". *Mots, termes et contextes*, 309, <http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Fontanet.pdf>

Rouleau M. La langue médicale : une langue de spécialité à emprunter le temps d'une traduction. *Traduction, Terminologie et Rédaction (TTR)* 1995; 8 (2): 29-49

Rouleau M. (2003) *La terminologie médicale et ses problèmes*, Panacea, vol IV

Max A., Névéol A., Yvon F., Zweigenbaum P., Ravaud P. Traduction automatique en français des revues Cochrane : Statistical machine translation of Cochrane reviews into French, In: Session francophone de MEDINFO, 2013

Eck M., Vogel S., Waibel A. (2004) Improving Statistical Machine Translation in the Medical Domain using the Unified Medical Language System, In: Proc of Coling.

Sagot, B. et Boullier P. (2005). From raw corpus to word lattices : robust pre-parsing processing [d'un corpus brut à des treillis de mots : traitement pré-syntaxique robuste]. In Proceedings of L&T 05, Poznan, Pologne : 348-351

Dias, G., S. Guilloire, et J. P. Lopes (2000). Extraction automatique d'associations textuelles à partir de corpora non traités. In M. Rajman and J. Chapelier (Eds.), Actes des 5es Journées Internationales d'Analyse Statistique des Données Textuelles (JADT 2000), Vol. 1 : 213-221.

Korenus, T., J. Laurikkala, K. Jarvelin, et M. Juhola (2004). Stemming and lemmatization in the clustering of finnish text documents. In CIKM '04 : Proceedings of the thirteenth ACM conference on Information and knowledge management, New York, NY, USA, ACM Press : 625-633.

Mikheev, A. (2000). Document centered approach for text normalization. In Proceedings of the 23<sup>rd</sup> ACM SIGIR Conference on Research and Development in Information Retrieval, Athens, Greece :136-143.

Zhang, T., F. Damerou, et D. Johnson (2003). Updating an nlp system to fit new domains : an empirical study on the sentence segmentation problem. In W. Daelemans et M. Osborne (Eds.), Proceedings of CoNLL-2003, Edmonton, Canada : 56-62.

Heitz T. Modélisation du prétraitement des textes, JADT 2006

Baranes M., Sagot B. Normalisation de textes par analogie: le cas des mots inconnus, 21ème Traitement Automatique des Langues Naturelles, Marseille, 2014

Levenstein, V. (1966). Binary Codes Capable of Correcting Deletions, Insertions and Reversals

Li, M., Zhang, Y., Zhu, M. et Zhou, M. (2006). Exploring distributional similarity based models for query spelling correction. In Proceedings of ACL-CoLing'06, pages 1025–1032

Sagot, B. (2010). The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In Proceedings of LREC'10, La Valette, Malta.

Sagot, B. et Boullier, P. (2008). SxPipe 2: architecture pour le traitement pré-syntaxique de corpus bruts. TAL, 49 (2), pp.155-188.

Sagot, B., Nouvel, D., Moulleron, V., Baranes, M. (2013). Extension dynamique de lexiques morphologiques pour le français à partir d'un flux textuel. In Actes de TALN'13.

## ANNEXE

### ANNEXE I

```
Sub upload_data_MUST()
Dim upload_date As Date

'actworkbook = ActiveWorkbook.Name
'

'textpos = InStr(1, actworkbook, "MUST", vbTextCompare)
'

'If textpos = 0 Then
'    MsgBox "File Name not contain MUST. Please check. Operation Abort",
vbInformation + vbOKOnly, "Admin"
'    Exit Sub
'End If

ct = Application.WorksheetFunction.CountA(Rows("1:1"))

If ct <> 47 Then
    MsgBox "Column Count <> 47. Operation Abort", vbInformation +
vbOKOnly, "Admin"
    Exit Sub
End If

If IfCheck_CorrectLoad (44, "MUST") Then Exit Sub
If IfCheck_CorrectLoad (45) Then Exit Sub
If IfCheck_CorrectLoad (9) Then Exit Sub

With ActiveSheet
    LRow = .Cells(.Rows.Count, "A").End(xlUp).Row
End With

For x = 2 To LRow
    If Cells(x, 46) <> "English" And Cells(x, 46) <> "Non-English" Then
        Cells(x, 46).Select
        MsgBox "Comments=English OR Non-English", vbInformation +
vbOKOnly, "Admin"
        Exit Sub
    End If
Next

If upload_old_data = True Then
```

```

upload_date = InputBox("Enter Date in DD MMM YYYY format. For Ex." &
Format(Now(), "dd mmm yyyy"), "Admin")
    If Not IsDate(upload_date) Then
        MsgBox "Incorrect Date Enter. Operation Abort", vbInformation +
vbOKOnly, "Admin"
        Exit Sub
    End If
End If
DoEvents
i = 2
'Columns("M:M").Select
' Selection.Replace What:=",", Replacement:="", LookAt:=xlPart, _
' SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, _
' ReplaceFormat:=False

Call open_database
strr = "select * from [filename] where [File Name]='" & Range("au2") & "'"
Call open_rs(strr)

If rs.RecordCount >= 1 Then
    ans = MsgBox("File already in the database." & vbCrLf & vbCrLf & "Are you
sure you want to replace the existing file?", vbQuestion + vbYesNo, "Admin")
    If ans = vbYes Then
        con.Execute "delete * from mustdata where [file name]='" & Cells(2, 47) & "'"
        con.Execute "delete * from filename where [file name]='" & Cells(2, 47) & "'"

        ElseIf ans = vbNo Then
            Exit Sub
        End If
    End If
End If

```

## ANNEXE II

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
	D		L			Orthographe			Abbreviations				C		N		S	
acquisition	d acquisition	d'acquisition	l'acquisition	l'acquisition		aquisition	acquisition		pb	problème			Cest	C'est	n a pas	n'a pas	s est	s'est
affaire	d affaire	d'affaire	l'affaire	l'affaire		ecran	écran		err	erreur			c'était	c'était	n est pas	n'est pas	s'arrete	s'arrête
affiche	d affiche	d'affiche	l'affiche	l'affiche		arriere	arrière		mess	message			c'etait	c'était	n y a	n'y a	s'arrete	s'arrête
alarme	d alarme	d'alarme	l'alarme	l'alarme		controle	contrôle		msg	message					n arrive	n'arrive	s'arrete	s'arrête
allumage	d allumage	d'allumage	l'allumage	l'allumage		cliches	clichés		qd	quand					n a que	n'a que	s'endend	s'endend
ampoule	d ampoule	d'ampoule	l'ampoule	l'ampoule		tete	tête		tech	technicien					n ayant	n'ayant	s'allume	s'allume
anode	d anode	d'anode	l'anode	l'anode		consequence	conséquence		tec	technicien							s'il	s'il
antenne	d antenne	d'antenne	l'antenne	l'antenne		succes	succès		pr	pour							s'ouvrir	s'ouvrir
apercevoir	d apercevo	d'apercevo	l'apercevoir	l'apercevoir		egalement	également		bcp	beaucoup							s'eteint	s'éteint
appareil	d appareil	d'appareil	l'appareil	l'appareil		examin	examen		apm	après-midi							s'eteint	s'éteint
aquisition	d aquisition	d'acquisition	l'acquisition	l'acquisition		creer	créer		tjs	toujours							s'affiche	s'affiche
arriere	d arriere	d'arriere	l'arriere	l'arriere		demiere	dernière		chgt	changement								
arriere	d arriere	d'arriere	l'arriere	l'arriere		meme	même		qqun	quelqu'un								
autre	d autre	d'autre	l'autre	l'autre		precedent	précédent											
autres	d autres	d'autres	l'autres	l'autres		depart	départ											
avance	d avance	d'avance	l'avance	l'avance		arrivee	arrivée											
eau	d eau	d'eau	l'eau	l'eau		impliquee	impliquée											
ecran	d ecran	d'ecran	l'ecran	l'ecran		apres midi	après-midi											
écran	d écran	d'écran	l'écran	l'écran		apres-midi	après-midi											
envoi	d envoi	d'envoi	l'envoi	l'envoi		apres midi	après-midi											
erreur	d erreur	d'erreur	l'erreur	l'erreur		parametrage	paramétrage											
espace	d espace	d'espace	l'espace	l'espace		maniere	manière											
espace	d espace	d'espace	l'espace	l'espace		donnees	données											

## ANNEXE III

Sub Correct\_Errors\_SR()

,

' Select Folder where all Records are saved

,

Dim intResult As Integer

Dim strPath, Filename As String

Dim wb As Workbook

'FRENCH CORRECTIONS

'This table contains the pairs of words (wrong word vs corrected word)

tableau = Array("d activité", "d'activité", "d accès", "d'accès", "d activer", "d'activer", "d activité", "d'activité", "d affaire", "d'affaire", "d affiche", "d'affiche", "d aggravation", "d'aggravation", "d alarme", "d'alarme", "d alimentation", "d'alimentation", "d allumage", "d'allumage", "d allumer", "d'allumer", "d ampoule", "d'ampoule", "d analyser", "d'analyser", "d anode", "d'anode", "d apercevoir", "d'apercevoir", "d appareil", "d'appareil", "d appel", "d'appel", "d application", "d'application", "d application", "d'application", "d après", "d'après", "d acquisition", "d'acquisition", "d arrêt", "d'arrêt", "d arrêter", "d'arrêter", "d arriere", "d'arriere", "d arrière", "d'arrière", "d arrivée", "d'arrivée", "d attention", "d'attention", "d autre", "d'autre", "d autres", "d'autres", "d avance", "d'avance", "d avoir", "d'avoir", "d eau", "d'eau", "d échec", "d'échec", "d ecran", "d'ecran", "d écran", "d'écran", "d electricité", "d'electricité", "d enregistrement", "d'enregistrement")

\*\*\*\*\*

'GERMAN CORRECTIONS

'This table contains the pairs of german words (wrong word vs corrected word)

tableau2 = Array("bzgl", " bezüglich ", " bzw ", " beziehungsweise ", " z.B ", " zum Beispiel ", " ggf ", " gegebenenfalls ", \_  
" MfG ", " Mit freundlichen Grüßen ", " Probledm ", " Problem ", " upgedatet ", " aktualisiert ", " Nr ", " Nummer ", " Corrupte ", \_  
" defekte ", " d.h ", " das heißt ", " dringede ", " dringende ", " happend ", " happened ", " imoact ", " impact ", " initparameter ", " init parameter ", \_  
" kontrolliert ", " kontrolliert ", " nessessary ", " necessary ", " ordnungsmässig ", " ordnungsmäßig ", " pateint ", " patient ", " patinet ", " patient ", \_  
" problemmen ", " Problemen ", " reponse ", " response ", " UniversalViewer ", " Univeral Viewer ", " Blokiert ", " blockiert ", " checke ", " checked ", \_  
" configuration ", " configuration ", " datas ", " data ", " diesses ", " dieses ", " druchgeführt ", " durchgeführt ", " erfogreich ", " erfolgreich ", \_  
" immernoch ", " immer noch ", " itseld ", " itself ", " Log-Files ", " logfiles ", " LogFiles ", " logfiles ", " mammo ", " mammo ", " mfd ", " Mit freundlichen Grüßen ", \_  
" necessary ", " necessary ", " o.ä ", " oder ähnlich ", " Offline-Schalten ", " offline schalten ", " resendet ", " neu gesendet ", " RIS-Version ", " RIS Version ", \_  
", \_



"Send-Queue", "send queue", "Sende-Queue", "send queue",  
"StorageControllers", "Storage Controller", "Studien-Status", "Studienstatus", "\_  
" systme ", " System ", " telefonsich ", " telefonisch ", " thsi ", " this ", " usw ", "  
und so weiter", " Context-Adaption ", " context adaptation ", \_  
" Contextmatching ", " context matching ", " Custome ", " Customer ", " databse  
", " database ")

\*\*\*\*\*

### SPANISH CORRECTIONS

This table contains the pairs of german words (wrong word vs corrected word)

```
tableau3 = Array(" algun ", " algún ", " alimentacion ", " alimentación ", "  
amplifcador ", "amplificador ", " automatico ", " automático ", _  
" averia ", " avería ", " bateria ", " batería ", " boton ", " botón ", " caido ", " caído  
", " calibracion ", " calibración ", _  
" ccomprobar ", " comprobar ", " colimdor ", " colimador ", " comunicacion ", "  
comunicación ", " comprobacion ", " comprobación ", _  
" conectividdad ", " conectividad ", " conexion ", " conexión ", "  
correspondiemntes ", " correspondientes ", _  
" defectuso ", " defectuoso ", " desnganche ", " desenganche ", " desnganche ", "  
desenganche ", " desocnocida ", " desconocida ", " despues ", " después ", "  
deteccion ", " detección ", _  
" deteccion ", " detección ", " dia ", " día ", " diagnostica ", " diagnóstica ", "  
diagnostico ", " diagnóstico ")
```

\*\*\*\*\*

### ITALIAN CORRECTIONS

This table contains the pairs of german words (wrong word vs corrected word)

```
tableau4 = Array(" errore ", " errore ", " cancewllata ", " cancellata ", "  
couinvolto ", "coinvolto", " piu ", "più", "PIU", "più", " positvo ", " positivo ", _  
" Nessu ", " nessun ", " Nesusn ", " nessun ", " CONNSESSION ", "  
CONNESSIONE ", " CONNSESSIONE ", " CONNESSIONE ", " parient ", "  
"patient", _  
" NECCSARIA ", "NECESSARIA ", " lciente ", " cliente ", " dlla ", " della",  
"d'integrazione", "di integrazione", " DIFFICOLTA' ", "DIFFICOLTÀ", _  
" funzionalita' ", "funzionalità", "funzioante", "funzionate", "installaizone", "  
installazione", " irtaliano", "italiano", "occiasionalmente", " occasionalmente",  
"rprecedenti", "precedenti", "QUALITA", "QUALITÀ", "regolamrnete",  
"regolarmente", "salc", "sala", "\", """)
```

\*\*\*\*\*

### Folder dialog displayed to the user

```
intResult = Application.FileDialog(msoFileDialogFolderPicker).Show
```

Checks if user has canceled the dialog or selected a folder

```
If intResult <> 0 Then
```

```
strPath = Application.FileDialog(msoFileDialogFolderPicker).SelectedItems(1) &  
"\"
```

'Avoid displaying workbooks

With Application

.ScreenUpdating = False

.DisplayAlerts = False

'Loop for Excel files in folder

Filename = Dir(strPath & "\*.xls")

Do While Filename <> ""

Set wb = Workbooks.Open(strPath & Filename)

**FRENCH, GERMAN, ITALIAN AND SPANISH CORRECTIONS**

'This loop finds matching countries, and then replace the wrong words with the correct ones

If Cells(2, 9) = "FRANCE" Or Cells(2, 9) = "ALGERIA" Or Cells(2, 9) = "LUXEMBOURG" Then

For p = 0 To UBound(tableau, 1) Step 2

Range("W:Y", "AC:AC").Replace What:=tableau(p), Replacement:=tableau(p + 1), LookAt:=xlPart, \_

SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, \_

ReplaceFormat:=False

Next p

ElseIf Cells(2, 9) = "GERMANY" Or Cells(2, 9) = "AUSTRIA" Or Cells(2, 9) = "SWITZERLAND" Then

For p = 0 To UBound(tableau2, 1) Step 2

Range("W:Y", "AC:AC").Replace What:=tableau2(p), Replacement:=tableau2(p + 1), LookAt:=xlPart, \_

SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, \_

ReplaceFormat:=False

Next p

ElseIf Cells(2, 9) = "ITALY" Then

For p = 0 To UBound(tableau4, 1) Step 2

Range("W:Y", "AC:AC").Replace What:=tableau4(p), Replacement:=tableau4(p + 1), LookAt:=xlPart, \_

SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, \_

ReplaceFormat:=False

Next p

ElseIf Cells(2, 9) = "SPAIN" Then

For p = 0 To UBound(tableau3, 1) Step 2

Range("W:Y", "AC:AC").Replace What:=tableau3(p), Replacement:=tableau3(p + 1), LookAt:=xlPart, \_

SearchOrder:=xlByRows, MatchCase:=False, SearchFormat:=False, \_

ReplaceFormat:=False

Next p  
'MsgBox ("ok")  
End If

\*\*\*\*

## ANNEXE IV

```
1 g1=open('franch.txt').read()
2 >>> g2=open('phrases.txt').read().split('\n')
3 >>> count={}
4 >>> for phrase in g2:
5     if phrase in g1:
6         compte = {phrase: g1.count(phrase) for phrase in g2}
7
8
9 >>> print compte
```

## **ABBREVIATIONS**

<b>CHU</b>	Complaint Handling Unit
<b>CRC</b>	Customer Remote Communication
<b>CSC</b>	Customer Service Center
<b>CTC</b>	Customer Technical center
<b>FE</b>	Field Engineer
<b>FSE</b>	Field Service Engineer
<b>MUST</b>	Multi User Service Tool (Mainframe application managing RFS activities)
<b>RFS</b>	Request For Service
<b>SE</b>	Service Engineer (includes FSE and OLSE)
<b>SnF</b>	Store & Forward, (offline application used by FSE for debrief of site interventions)
<b>SR</b>	Service Report (result of debrief)
<b>TMS</b>	Translation Management System