



uOttawa

L'Université canadienne  
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES  
ET POSTDOCTORALES



uOttawa

l'Université canadienne  
Canada's university

FACULTY OF GRADUATE AND  
POSTDOCTORAL STUDIES

Karine Circé

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A. (traduction)

GRADE / DEGRÉ

École de traduction et d'interprétation

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Traduction automatique, mémoire de traduction ou traduction humaine? Propositions d'une approche  
pour déterminer la meilleure méthode à adopter, selon le texte

TITRE DE LA THÈSE / TITLE OF THESIS

Lynne Bowker

DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Aline Francoeur

Roda P. Roberts

Gary W. Slater

LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /  
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

**Traduction automatique, mémoire de traduction ou traduction humaine?  
Proposition d'une approche pour déterminer la meilleure méthode à  
adopter, selon le texte**

par

Karine Circé

sous la direction de  
Lynne Bowker

Thèse soumise à la  
Faculté des études supérieures et postdoctorales  
dans le cadre des exigences  
du programme de maîtrise en traduction

École de traduction et d'interprétation  
Faculté des arts  
Université d'Ottawa

© Karine Circé, Ottawa, Canada, 2005



Library and  
Archives Canada

Bibliothèque et  
Archives Canada

Published Heritage  
Branch

Direction du  
Patrimoine de l'édition

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*  
*ISBN: 0-494-11243-3*  
*Our file* *Notre référence*  
*ISBN: 0-494-11243-3*

**NOTICE:**

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

**AVIS:**

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

  
**Canada**

## Table des matières

Remerciements.....	v
Résumé.....	vii
Abstract.....	ix
PARTIE I – Contexte et cadre théorique .....	1
Chapitre 0 – Introduction.....	2
0.1 Contexte.....	2
0.2 Justification/Motivation.....	4
0.3 Objectifs.....	6
0.4 Méthode.....	6
0.5 Portée et limites.....	7
0.6 Plan.....	10
Chapitre 1 – Concepts de base.....	11
1.1 Traduction humaine.....	11
1.1.1 Avantages et inconvénients de la TH.....	11
1.2 Traduction automatique.....	11
1.2.1 Systèmes de TA directe.....	12
1.2.2 Systèmes de TA par transfert.....	14
1.2.3 Systran : système de TA « hybride ».....	16
1.2.4 Avantages et inconvénients de la TA.....	16
1.3 Mémoires de traduction.....	18
1.3.1 Fonctionnement d’une mémoire de traduction « traditionnelle ».....	19
1.3.2 MultiTrans : système de MT.....	21
1.3.3 Avantages et inconvénients des MT.....	24
Chapitre 2 – Typologie des textes et technologies de la traduction.....	27
2.1. Typologie des textes.....	27
2.1.1 Pertinence par rapport aux technologies de la traduction.....	29
2.1.2 Types de textes/caractéristiques des textes et méthodes d’automatisation.....	29
2.1.2.1 Facteurs terminologiques.....	32
2.1.2.2 Facteurs stylistiques/linguistiques.....	32
2.2 Initiatives tentées précédemment en vue de déterminer si certains textes se prêtent à un outil de traduction donné.....	33
2.2.1 TransRouter.....	33
2.2.2 PaTrans – Étude de cas.....	36
2.2.3 Translatability Checker.....	37
2.2.4 FRASSE.....	38
2.2.5 Outils d’analyse de documents.....	39
2.2.6 Comparaison entre la traduction humaine et la traduction automatique.....	41
PARTIE II – Conception d’une approche pour déterminer la meilleure méthode de traduction à adopter.....	43
Chapitre 3 – Travail de préparation.....	44
3.1 Sélection des textes.....	44
3.2 Travail de préparation à l’étape de la traduction.....	45
3.2.1 Création des corpus (TransCorpora) dans MultiTrans.....	47
3.2.2 Création des banques de données terminologiques et des dictionnaires.....	48

3.2.2.1	Création des banques de données terminologiques (TermBase) dans MultiTrans.....	49
3.2.2.2	Création des dictionnaires dans Systran .....	49
3.3	Traduction .....	50
Chapitre 4	– Analyse des données et élaboration des lignes directrices.....	52
4.1	Résultats.....	52
4.2	Analyse des résultats.....	53
4.3	Détermination et analyse des caractéristiques pertinentes.....	55
4.3.1	Caractéristiques pertinentes pour la TA.....	56
4.3.1.1	Problèmes propres à la TA : caractéristiques à éviter dans le TLD.....	56
4.3.1.1.1	Résolution des référents des pronoms.....	57
4.3.1.1.2	Voix passive.....	58
4.3.1.1.3	Interprétation des rapports de coordination entre les composantes de syntagmes.....	58
4.3.1.2	Langage contrôlé (LC) et règles d'écriture.....	59
4.3.2	Caractéristiques pertinentes pour les MT .....	62
4.3.3	Caractéristiques pertinentes pour la TH.....	63
4.3.4	Terminologie.....	64
4.4	Outils.....	64
4.4.1	Statistiques de lisibilité - Microsoft Word.....	65
4.4.2	WordList - WordSmith Tools .....	65
4.4.3	Fonction d'analyse de MultiTrans .....	66
4.5	Première ébauche de la liste de critères .....	66
4.6	Analyse des données.....	68
4.6.1	Analyse des décisions d'appel (DA).....	70
4.6.2	Analyse des énoncés de qualités (EQ).....	72
4.6.3	Analyse des coupures de presse (CP) .....	73
4.6.4	Autres observations.....	74
4.7	Élaboration des lignes directrices .....	75
4.7.1	Lignes directrices pour la TA .....	76
4.7.3	Lignes directrices pour la TH .....	77
PARTIE III	– Essais et évaluation.....	79
Chapitre 5	– Mise à l'essai des lignes directrices .....	80
5.1	Sélection et traduction des textes.....	80
5.2	Essais pilotes - Attribution de valeurs aux éléments des lignes directrices et présentation de celles-ci sous forme de listes de vérification.....	80
5.2.1	Premier essai pilote.....	82
5.2.2	Deuxième essai pilote .....	84
5.2.3	Vérification des prédictions .....	86
5.2.3.1	Premier essai pilote.....	86
5.2.3.2	Deuxième essai pilote .....	87
5.3	Détermination des valeurs et du format définitifs de la liste de vérification.....	88
5.4	Mise à l'essai « officielle » .....	90
5.4.1	Résultats.....	91
Chapitre 6	– Conclusion.....	93

6.1 Commentaires généraux sur la conception et la mise à l'essai de l'approche proposée .....	93
6.2 Recommandations pour les recherches à venir .....	95
6.3 Mot de la fin.....	97
Bibliographie.....	98
Annexe A – Exemples de chacun des trois types de textes ayant servi à l'étape de la conception .....	103
Annexe B – Données relatives à chaque texte ayant servi à l'étape de la conception....	110
Annexe C - Exemples de chacun des deux types de textes ayant servi à l'étape des essais .....	113
Annexe D – Données relatives à chaque texte ayant servi à l'étape des essais .....	126
Annexe E – Trousse remise aux volontaires.....	128
Annexe F – Rétroaction des volontaires .....	133

## **Remerciements**

De nombreuses personnes m'ont soutenue dans cette aventure, mais je tiens d'abord et avant tout à remercier ma directrice de thèse, Lynne Bowker. C'est elle qui m'a fait découvrir la traductique au baccalauréat et qui a fait germer en moi l'idée de faire une maîtrise dans le domaine. Bref, si ce n'avait été de toi, Lynne, rien de tout cela n'aurait été. Je te suis aussi reconnaissante de ta patience et de ta grande disponibilité, et ce, malgré l'ajout d'un membre à ta famille, ainsi que de tes mots d'encouragement lorsque j'avais besoin que quelqu'un me rassure. Tu es franchement l'une des personnes les plus efficaces que je connaisse. Tes nombreuses qualités font vraiment de toi une professeure et une directrice hors pair.

Je savais que la nature du projet que j'entreprenais faisait en sorte que j'allais devoir avoir un bon réseau de collaborateurs. Et j'ai en effet été très bien entourée. Pour leur aide, quelle qu'elle soit, et leur soutien, je remercie les personnes suivantes :

Cathryn Arnold, Manon Boucher (Bureau de la traduction), Michel Caron,  
Monique Charette, Sandra Deschênes, Daniel Gervais (MultiCorpora R&D Inc.),  
Louise Leduc, Céline Marchand, Jocelyne Morin, Véronique Séguin et Kevin Walsh.  
Vous avez tous grandement contribué à la progression (et à la conclusion!) de ce projet.

Finalement, je m'en voudrais de ne pas souligner le soutien que m'ont donné mes camarades de classe, mes collègues, mes amis, mes parents de même que Dean, mon conjoint. Vos mots d'encouragement ont été une source de motivation, surtout dans la

dernière ligne droite. Je sais que je vous ai négligés ces derniers mois, mais sachez que j'ai bien l'intention de me reprendre.

Karine  
Mai 2005

## Résumé

Depuis quelques années, le domaine de la traduction doit faire face à une hausse vertigineuse de la demande, si bien que les traducteurs en poste ont peine à combler les besoins. Cette situation rend d'autant plus attrayants les outils informatisés d'aide à la traduction. Mais une fois ces logiciels en place, comment savoir lequel utiliser? Valent-ils même la peine d'être utilisés? L'objectif de la présente thèse est de fournir une réponse à ces questions en proposant et en mettant à l'essai une approche qui permettra de prédire quelle méthode, entre la traduction automatique, les mémoires de traduction et la traduction humaine, est la meilleure pour un type de textes donné. Dans le présent cas, on considère que la « meilleure » méthode est celle qui permet au traducteur de produire une traduction « livrable » le plus rapidement possible.

Premièrement, à la suite d'un dépouillement des caractéristiques qui, selon divers auteurs, font en sorte qu'un texte se prête mieux à telle ou telle méthode, on analyse puis on traduit trois différents types de textes à l'aide des trois méthodes énoncées dans le but de dégager des corrélations entre les caractéristiques de ces textes et la méthode qui a donné les meilleurs résultats pour chacun. À partir de ces données, on tente de concevoir une liste de vérification contenant des lignes directrices et de fixer les valeurs des différents éléments de celles-ci (p. ex. le nombre idéal de mots par phrase, le pourcentage maximal de phrases conjuguées à la voix passive). Suivent deux essais pilotes pour vérifier la validité de deux séries de valeurs pour prédire la meilleure méthode à adopter pour deux nouveaux types de textes. Une fois ces textes traduits et les prédictions vérifiées, on ajuste les valeurs et on procède à une dernière mise à l'essai. Celle-ci permet de conclure

que l'approche proposée dans cette thèse peut en effet aider les répartiteurs et les traducteurs à choisir la meilleure méthode à adopter.

## **Abstract**

Over the past several years, there has been such an enormous increase in the demand for translation services that translators have been struggling to keep up with it. This situation makes the use of computer translation tools ever more attractive. But once these programs have been installed, how can translators decide which one to use? Or whether it is worth the trouble of using them at all? The objective of this thesis is to answer these questions by proposing and testing an approach to predict which translation method—machine translation, translation memories or human translation—will be the best for a given text type. In this case “best” means the method that allows a translator to produce a “client-ready” translation as quickly as possible.

After a review of the characteristics that, according to various authors, determine that a text is more suitable for a particular translation method, three different text types were analyzed and translated (from English to French) using the three methods listed above in order to establish the correlations between the characteristics of the texts and the method that gave the best results for each text. This data was used to develop a checklist of guidelines and to assign values to their various elements (e.g., the optimal number of words per sentence, the maximum number of sentences in the passive voice). Two pilot tests were then performed to verify the validity of two sets of values for predicting the best method on two new text types. After these texts were translated and the predictions verified, the values were adjusted and a final test was performed. The results of the final test support the conclusion that the approach proposed in this thesis can indeed help text allocators and translators choose the best method to use.

## **PARTIE I – Contexte et cadre théorique**

## **Chapitre 0 – Introduction**

### **0.1 Contexte**

De nombreux auteurs (p. ex. Anobile, 2000; Bowker, 2002a; Allen, 2003) s'entendent pour dire que le phénomène de la mondialisation a fait accroître de façon importante la demande de services de traduction. C'est que des entreprises se rendent compte que, pour atteindre leurs objectifs, elles ne peuvent plus se fier uniquement au marché local ou recourir à une seule langue comme moyen de communication (Allen, 2003, p. 299). Ainsi, pour se tailler une place sur le marché mondial, elles doivent notamment adopter une approche multilingue, ainsi que des stratégies et des technologies langagières qui leur permettront de mettre en œuvre cette approche (Anobile, 2000, p. vii; Allen, 2003, p. 299). Par exemple, les principaux fabricants de logiciels américains ont appris que, plus ils prenaient de temps à lancer un produit traduit sur les marchés étrangers, plus ils perdaient de revenus. C'est pourquoi aujourd'hui ils visent le lancement simultané de toutes les versions d'un même produit (Anobile, 2000, p. vii). Ainsi, il n'est pas rare que des entreprises de traduction aient à traduire de la documentation et des manuels n'ayant jamais été traduits auparavant (Allen, 2003, p. 299-300). L'avènement et la popularisation d'Internet, autre gage de cette mondialisation, ont aussi contribué à augmenter la demande. En effet, les pages Web doivent être mises à jour régulièrement, et si elles sont traduites, leur traduction doit aussi être mise à jour régulièrement et surtout, rapidement, afin qu'elle puisse être affichée en même temps que les pages en langue de départ (LD).

Le fait que plusieurs entreprises aient pris le virage multilingue a fait augmenter considérablement le nombre de demandes de traduction (Allen, 2003, p. 299). Mais il n'y

a pas que les activités commerciales qui contribuent à cette augmentation. Les organismes internationaux et des pays en situation de multilinguisme, comme le Canada, ont aussi beaucoup d'écrits administratifs à faire traduire (Fernández Guerra, 2000, p. 24). En 1996, la Commission européenne prévoyait une croissance de 15 à 18 % par année (Anobile, 2000, p. vii). En 1998, diverses entreprises de traduction connues signalaient une augmentation annuelle de 30 % du nombre de demandes de traduction, tandis que la Commission européenne estimait que cette augmentation était plutôt de 50 % (Allen, 2003, p. 299). En 2002, Allied Business Intelligence Inc. prévoyait que l'ensemble des besoins en traduction dans le monde représenteraient 8,8 milliards de dollars américains en 2003, puis 11,5 milliards de dollars en 2007 (Tunick, 2002, p. 1-8). Malheureusement, malgré une telle croissance, il n'y a pas assez de traducteurs pour suffire à la tâche (Anobile, 2000, p. vii; Fernández Guerra, 2000, p. 24). Le Comité sectoriel de l'industrie canadienne de la traduction (1999, p. 66) observe qu'il y a au Canada, « une pénurie de traducteurs expérimentés face à l'augmentation de la demande ». Ainsi, les traducteurs se retrouvent avec une charge de travail beaucoup plus lourde et des délais plus serrés, mais ne sont pas assez nombreux pour satisfaire à la demande. C'est donc dire que les traducteurs en place doivent traduire plus, plus rapidement, tout en produisant toujours un travail de qualité. Dans ces circonstances, de nombreuses entreprises se demandent comment satisfaire à leurs besoins en traduction de façon abordable. Les technologies langagières pourraient-elles leur servir?

Dans le contexte d'une entreprise de traduction, le sentiment de devoir travailler plus rapidement n'est pas seulement attribuable à une demande accrue, mais aussi à certaines

contraintes financières. Les traducteurs s'attendent à obtenir régulièrement des augmentations de salaire (et y ont droit, quand ils sont syndiqués), mais voilà que les tarifs n'augmentent pas dans une mesure comparable aux hausses de salaire escomptées. Et il ne faut pas se leurrer, dans un contexte de recouvrement des coûts, ce sont les traducteurs qui produisent les revenus qui serviront ensuite à payer leur salaire. Mais une fois qu'on a poussé les tarifs à la limite de ce que le marché est prêt à accepter, comment continuer à accorder des augmentations de salaire tout en assurant la viabilité des entreprises de traduction? Les traducteurs et les dirigeants de telles entreprises devraient peut-être considérer la technologie (traduction automatique, mémoires de traduction) comme une solution possible.

## **0.2 Justification/Motivation**

Lorsque j'ai suivi le cours de traductique à mon dernier semestre du baccalauréat en traduction à l'automne 2001, j'ai découvert toute une gamme d'outils dont j'ignorais presque l'existence, y compris les logiciels de traduction automatique (TA) et les mémoires de traduction (MT). Alors inscrite au programme coop, j'avais déjà une certaine expérience en milieu de travail. Le service de traduction où j'avais fait mes deux derniers stages coop et où on m'avait offert un emploi une fois mon diplôme obtenu, n'avait recours à aucune de ces technologies. Pourquoi ce service, qui faisait tout de même partie d'un des plus importants organismes de traduction au pays, ne bénéficiait-il pas encore des avantages que présentent ces outils? C'est alors que l'idée m'est venue de poursuivre mes études en me concentrant sur ces technologies dans la perspective de mon milieu de travail (Service Criminologie).

Depuis mon entrée sur le marché du travail, j'ai été moi-même témoin, en moins de deux ans et demi, d'une hausse importante des objectifs de production imposés aux traducteurs. Compte tenu de cette situation et du fait conséquent que les traducteurs doivent traduire plus, plus rapidement, tout en produisant toujours un travail de qualité, on est en droit de se demander si les outils de technologie langagière ne pourraient pas aider les traducteurs à composer avec cette pression qui est exercée sur eux.

La traduction automatique (TA) et les mémoires de traduction (MT) peuvent être utiles lorsqu'elles sont utilisées pour certains types de textes. Si elles sont utilisées avec des types de textes qui ne conviennent pas, elles peuvent en revanche faire perdre du temps. Mais, entre la traduction automatique et les mémoires de traduction, voire tout simplement la traduction humaine (TH) – c'est-à-dire sans aucun de ces outils, comment savoir quelle méthode choisir? Selon Sager (1994, p. 194), dans un environnement moderne, les traducteurs doivent, avant de commencer une traduction, choisir de recourir ou non à la technologie et, le cas échéant, à quel type de technologie ils auront recours. Wallace Schwab, un traducteur agréé, a cependant indiqué, dans un article paru en 2003 (p. 30) : « I sometimes pick up a new text to be translated, give it a quick once over to decide which CAT<sup>1</sup> tool will best do the job, then come to the conclusion that none of my fancy tools will really make a difference in getting the job done, so I do the translation simply on my word processor. » Les traducteurs, ou souvent plutôt les personnes qui assument la responsabilité d'assigner le travail (ci-après appelées « répartiteurs »), ont

---

<sup>1</sup> Computer-assisted translation

donc besoin de connaître le « profil » des textes qui se prêtent le mieux à la traduction assistée par l'une ou l'autre des deux technologies en cause ici. Sager (1994, p. 308) fait d'ailleurs justement remarquer que les traducteurs n'ont pas encore appris à catégoriser les textes selon qu'ils conviennent ou non à un système en particulier. Jusqu'à maintenant, les lignes directrices qui existent sur les types de textes à utiliser avec diverses technologies ou méthodes (en l'occurrence la TH, la TA et les MT) ne sont que très vagues et d'ordre assez général. Ce qu'il faut aux répartiteurs, c'est une série de critères précis qu'ils peuvent consulter pour déterminer quelle méthode est la meilleure pour traduire chaque type de textes.

### **0.3 Objectifs**

Le présent projet de thèse vise principalement à concevoir et à mettre à l'essai une approche qui permettra de déterminer quelle méthode de traduction (TH, TA, MT) est la plus efficace pour un type de textes donné. Ainsi, lorsqu'un répartiteur reçoit un texte à traduire, il peut évaluer le texte en fonction des différents critères établis, ce qui lui donnera une bonne idée de la méthode de traduction à adopter (TH, TA, MT) pour traduire le type de textes en question. Le processus de traduction s'en trouvera donc amélioré, puisque les traducteurs ne perdront pas de temps à utiliser un outil qui ne convient pas ou à traduire un texte à partir de zéro lorsque le recours à l'une des deux autres méthodes aurait pu leur faire épargner du temps.

### **0.4 Méthode**

En gros, les étapes que nous suivrons pour créer et mettre à l'essai cette approche sont les suivantes. Nous commencerons par choisir trois types de textes traduits fréquemment au

Service Criminologie<sup>2</sup>. Ensuite, nous traduirons des exemples de chacun de ces types de textes à l'aide de chacune des trois méthodes énoncées précédemment : TH, TA, MT. Deuxièmement, nous analyserons les textes afin de déterminer quelles caractéristiques font en sorte qu'ils sont traduits en moins de temps à l'aide d'une certaine méthode. Troisièmement, compte tenu de cette analyse, nous dresserons trois listes de caractéristiques, soit une pour chaque méthode (TH, TA, MT). Quatrièmement, nous mettrons ces listes à l'essai sur de nouveaux types de textes en comparant ceux-ci aux listes établies afin de déterminer quelle méthode de traduction (TH, TA, MT) sera la meilleure. Cinquièmement, nous traduirons des exemples de chacun de ces nouveaux types de textes à l'aide de chacune des trois méthodes afin de vérifier si nos prédictions étaient bonnes. Finalement, nous ajusterons les listes au besoin compte tenu des résultats de la mise à l'essai.

Bien sûr, nous procéderons d'abord à une recension pour voir notamment ce que d'autres auteurs ont déjà écrit sur les caractéristiques ou les types de textes qui conviennent le mieux à certaines méthodes de traduction, et nous terminerons en faisant des suggestions pour les recherches à venir.

### **0.5 Portée et limites**

Bien que notre approche sera, du moins nous l'espérons, assez générale pour s'appliquer à une vaste gamme d'entreprises de traduction, de traducteurs et de logiciels, les contraintes de temps et le fait que nous avons accès à des ressources limitées font en sorte

---

<sup>2</sup> Service Criminologie, Bureau de la traduction, Travaux publics et Services gouvernementaux Canada

que nous nous en tiendrons à un logiciel de TA, à un logiciel de MT, à des textes provenant du Service Criminologie et à une seule combinaison (unidirectionnelle) de langues.

Plusieurs logiciels de TA sont offerts sur le marché. Une des caractéristiques auxquelles nous tenions absolument pour le logiciel de TA était la possibilité de créer des dictionnaires personnalisés<sup>3</sup>. Les versions de base et les services de TA dans Internet n'offrent habituellement pas cette option. Il fallait aussi pouvoir apporter le logiciel au travail, parce qu'il était très important pour nous que les textes soient traduits dans notre environnement de travail habituel<sup>4</sup>. Il n'était donc pas question de recourir aux ordinateurs de l'Université dotés de tels logiciels. Finalement, le logiciel de TA que nous avons utilisé est Systran Standard 4.0, et ce, pour trois principales raisons. Premièrement, parce qu'il s'agit d'un logiciel de TA très connu et très répandu. Il compte d'ailleurs parmi ses clients le département de la Défense des États-Unis, la Ford Motor Company, la Commission européenne, General Motors du Canada, l'OTAN, Xerox et Autodesk (Hutchins & Somers, 1992, p. 188; *User's Guide for SYSTRAN 4.0*, 2002, p. 3). De plus, ce logiciel n'est pas trop cher et n'exige pas d'ordinateurs trop puissants, alors la plupart des entreprises de traduction ont les moyens de disposer de ce logiciel de TA. Deuxièmement, il offre la possibilité de créer des dictionnaires personnalisés.

---

<sup>3</sup> On peut consigner dans ces dictionnaires la terminologie et la phraséologie propres au domaine et/ou au client et ainsi maximiser les résultats.

<sup>4</sup> Dans notre environnement de travail habituel, nous avons accès à tous les outils dont nous avons besoin. Il était aussi important que tous les textes soient traduits dans les mêmes conditions autant que possible (ergonomie, bruit, réglages).

Troisièmement, la gestionnaire du Service Criminologie a eu l'amabilité de mettre une copie à ma disposition.

Pour ce qui est du logiciel de MT, nous avons choisi MultiTrans, un produit de MultiCorpora R&D Inc., parce que nous avons entendu de bons commentaires de traducteurs qui l'utilisaient dans le cadre de leur travail. Nous avons aussi une collègue qui avait commencé à l'utiliser et qui était satisfaite des résultats. MultiTrans est aussi un produit assez répandu, et il compte parmi ses clients le gouvernement du Canada et certains gouvernements provinciaux, des organismes internationaux, comme l'UNESCO, des sociétés pharmaceutiques, comme Johnson & Johnson, Bayer et Pfizer, ainsi que d'autres grandes entreprises, comme Ernst & Young, Ford, ING, Kraft et Michelin<sup>5</sup>. Finalement, MultiCorpora a bien voulu nous prêter une copie et nous fournir du soutien technique à des fins de recherche.

Quant aux textes utilisés dans le cadre des essais, ils proviendront tous du Service Criminologie et respecteront la combinaison de langues dans laquelle nous travaillons, soit l'anglais et le français, le français étant toujours la langue d'arrivée. Les types de textes sont décrits plus en détail à la section 3.1.

---

<sup>5</sup> [http://www.multicorpora.com/customers\\_f.html](http://www.multicorpora.com/customers_f.html) (consulté le 11 juillet 2004)

## **0.6 Plan**

La présente thèse compte 6 chapitres. Le premier chapitre présente les concepts de base. Il explique notamment le mode de fonctionnement ainsi que les avantages et désavantages de la traduction automatique et des mémoires de traduction. Le deuxième chapitre résume la théorie avancée et les travaux de recherche effectués dans le domaine. Les chapitres 3, 4 et 5 traitent de la conception, de la mise à l'essai et de l'ajustement d'une approche pour déterminer la meilleure méthode de traduction à adopter, selon le texte. Le dernier chapitre contient une évaluation de l'approche élaborée de même que des propositions de recherche pour l'avenir.

## **Chapitre 1 – Concepts de base**

Trois méthodes de traduction différentes sont examinées dans le cadre du présent travail.

Il s'agit de la traduction humaine (TH), de la traduction automatique (TA) et des mémoires de traduction (MT).

### **1.1 Traduction humaine (TH)**

Par traduction humaine, on entend toute situation où un traducteur effectue la traduction proprement dite, mais où il peut se servir d'une gamme d'outils ou de ressources pour l'aider. Voici certains exemples d'outils et de ressources (électroniques ou autres, sauf les mémoires de traduction et la traduction automatique) auxquels les traducteurs professionnels au Canada ont normalement accès : dictionnaires, lexiques, outils de traitement de texte, outils d'aide à la rédaction, archives, Termium<sup>®</sup>, autres bases de données terminologiques, TransSearch (Hansard), Internet et moteurs de recherche, et correcteur orthographique ou grammatical.

#### **1.1.1 Avantages et inconvénients de la TH**

Les textes traduits au moyen de cette méthode sont habituellement de bonne qualité, car le traducteur a le contrôle. Le processus risque cependant d'être parfois un peu long. De plus, quoiqu'il donne un produit de qualité, ce processus peut coûter cher.

### **1.2 Traduction automatique (TA)**

La traduction automatique existe depuis plus de 50 ans. Hutchins et Somers (1992, p. 3) la définissent comme suit : « computerised system responsible for the production of

translations from one natural language into another, with<sup>6</sup> or without human assistance ». Mais selon L'Homme (1999, p. 302), « il est particulièrement difficile, voire impossible, de décrire le fonctionnement des différents logiciels de TA de façon uniforme, car on observe des variantes souvent significatives d'un système à l'autre ». En fait, il existe trois principaux types d'architectures ou stratégies, toutes fondées sur la linguistique. Il y a la méthode directe (la plus ancienne) et deux méthodes indirectes, soit les systèmes par transfert et les systèmes par interlangue (Hutchins et Somers, 1992, p. 71-72; Trujillo, 1999, p. 167-201). Il existe depuis plus récemment une quatrième sorte de systèmes, soit les systèmes non linguistiques, lesquels sont fondés sur des statistiques et des exemples (Trujillo, 1999, p. 203-212; Fernández Guerra, 2000, p. 68-69). Nous nous attarderons ici surtout à la méthode directe et aux systèmes par transfert, dont une brève description figure ci-dessous. Ces deux méthodes sont utilisées par la plupart des systèmes commerciaux, y compris Systran, que nous utilisons aux fins du présent projet (voir la section 1.2.3). En général, les autres méthodes ne sont utilisées que dans le cadre de projets de recherche ou de systèmes de nature prototype.

### 1.2.1 Systèmes de TA directe

Les systèmes de traduction directe ne sont pas vraiment portés sur l'analyse; ils sont plutôt axés sur un dictionnaire bilingue. Ainsi, selon le cas, ils déterminent la catégorie grammaticale, le genre, le nombre et le temps de verbe des mots, puis substituent

---

<sup>6</sup> Il convient de noter que, si un système produit une traduction avec l'aide d'un traducteur, l'humain en question ne participe normalement pas à l'opération traduisante comme telle. S'il y a intervention humaine, c'est la plupart du temps par la *post-édition* (à ne pas confondre avec la *révision*), ce qui signifie que le traducteur corrige les erreurs qui se sont glissées dans la traduction brute produite par le logiciel de TA.

directement des mots de la langue d'arrivée (LA) ayant les mêmes caractéristiques aux mots du texte de la langue de départ (TLD<sup>7</sup>), sans représentation intermédiaire (Hutchins et Somers, 1992, p. 107). Le résultat final est donc presque une traduction mot-à-mot.

LD : The boy ate an apple.  
LA : Le garçon a mangé une pomme.

Parfois, un tel système peut arriver à effectuer quelques modifications simples au niveau de la syntaxe (p. ex., changement de l'ordre des mots) :

LD : the red cars  
LA : les voitures rouges

Cette approche fonctionne bien pour des phrases très simples<sup>8</sup>, mais des problèmes se manifestent avec des structures plus complexes<sup>9</sup> quand la traduction mot-à-mot n'est pas possible.

LD : He walked across the street.  
LA : Il a marché à travers la rue. [*sic*]

Dans l'exemple ci-dessus, la traduction est plus ou moins compréhensible, mais elle n'est pas idiomatique ni de bonne qualité. Une meilleure traduction de cette phrase comprend des changements de catégories grammaticales pour représenter des éléments de manière,

---

<sup>7</sup> Le sigle normalement employé pour désigner ce concept est TD (texte de départ). Nous avons décidé d'utiliser plutôt l'expression « texte de la langue de départ » par souci de cohérence avec l'expression « texte de la langue d'arrivée » (TLA), utilisée au lieu de l'expression « texte d'arrivée » (TA), normalement employée. Dans cette thèse, TA signifie traduction automatique.

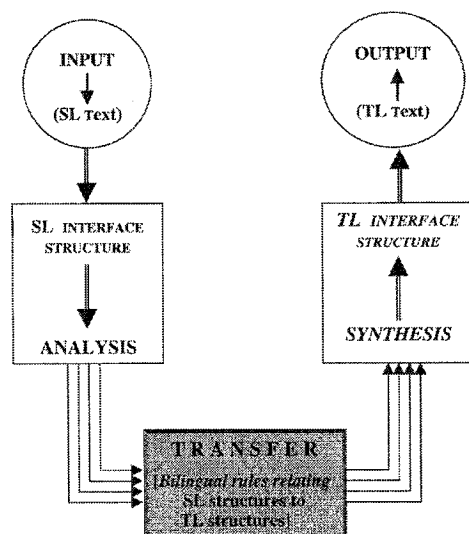
<sup>8</sup> Sujet-verbe-complément, p. ex.

<sup>9</sup> Phrases contenant plusieurs verbes, phrases à la voix passive, p. ex.

de mouvement et de trajectoire : *Il a traversé la rue à pied.* (Littéralement : *He crossed the street on foot.*) Malheureusement, les systèmes de TA directe ne sont pas capables d'effectuer ce type de modification parce qu'ils n'ont pas recours à des règles grammaticales complexes.

### 1.2.2 Systèmes de TA par transfert

Les systèmes de traduction automatique par transfert relèvent des domaines de la linguistique structurale et de l'informatique (Somers, 2003, p. 6). Il s'agit d'ailleurs de l'architecture à laquelle de nombreux systèmes commerciaux ont toujours recours (Fernández Guerra, 2000, p. 57). « Ces logiciels abordent un texte source en suivant *grosso modo* les mêmes étapes de traitement » (L'Homme, 1999, p. 302), qui sont habituellement au nombre de trois : analyse, transfert et synthèse (ou génération) (Arnold, 2003, p. 122) (voir la figure 1).



**Figure 1** : Étapes de la traduction automatique par transfert (Fernández Guerra, p. 60)

Lors de l'analyse syntaxique, les systèmes de TA par transfert convertissent le TLD en une représentation abstraite (décomposée – catégorie grammaticale, genre, nombre, temps, fonctions) de la structure de surface de celui-ci à l'aide de dictionnaires et de grammaires (Hutchins et Somers, 1992, p. 107).

### Exemple

TLD : Girls like white cars.

Représentation abstraite en LD des éléments de la phrase :

- girls = girl (noun) + regular<sup>10</sup> plural
- like = to like (verb) + present tense + 3rd person plural conjugation
- white = white (adjective)
- cars = car (noun) + regular plural

Ensuite, au moyen de dictionnaires et de règles de transfert bilingues, le système procède au transfert lexical (mot-à-mot) et syntaxique de la représentation abstraite du TLD à une représentation abstraite du texte de la langue d'arrivée (TLA<sup>11</sup>).

---

<sup>10</sup> On entend par « pluriel régulier » tout pluriel formé au moyen d'une règle précise. En anglais et en français, le pluriel de la plupart des noms se forme en ajoutant un *s* au singulier (*cow* + *-s* = *cows*; *arbre* + *-s* = *arbres*). Le pluriel de certains autres noms est formé selon d'autres règles. En anglais, le pluriel des noms qui se terminent au singulier par *-s*, *-z*, et *-h*, par exemple, se forme en ajoutant *-es* (*gas* + *-es* = *gases*). En français, les noms qui se terminent par *-al* font *-aux* au pluriel. « Pluriel régulier » est donc utilisé ici par opposition aux « pluriels irréguliers », soit ceux qui constituent des exceptions aux règles, ou qui ne suivent simplement aucune règle (*carnaval* ≠ *carnavaux*, mais bien *carnavals*; *child* ≠ *childs*, mais bien *children*) (de Villers, 1999, p. 1120; Anson et Schwegler, 1997, p. 173).

<sup>11</sup> Le sigle normalement employé pour désigner ce concept est TA (texte d'arrivée). Mais comme il s'agit aussi du sigle de traduction automatique, nous avons décidé d'étoffer et d'utiliser l'expression « texte de la langue d'arrivée » et, par souci de cohérence, de faire de même pour texte de départ (c.-à-d. texte de la langue de départ) ou TLD.

Représentation abstraite en LA des éléments de la phrase :

- fille (nom, féminin) + pluriel régulier
- aimer (verbe) + présent + conjugaison à la 3<sup>e</sup> personne du pluriel
- blanc (adjectif) + féminin + pluriel
- voiture (nom, féminin) + pluriel régulier

Le TLA final est par la suite généré à partir de la représentation du TLA à l'aide de dictionnaires et de grammaires (ces derniers peuvent indiquer, p. ex., s'il est nécessaire d'ajouter des articles, etc., pour produire une phrase bien formée).

TLA : Les filles aiment les voitures blanches.

### **1.2.3 Systran : système de TA « hybride »**

Le système utilisé aux fins du présent projet s'appelle Systran (version Standard 4.0).

Développé en Californie, Systran est le plus ancien système offert sur le marché et celui qui offre le plus grand nombre de combinaisons de langues (Sager, 1994, p. 273). La Commission européenne utilise d'ailleurs beaucoup ce système (Austermühl, 2001, p. 156). Il s'agit d'un système « hybride » en ce sens qu'il combine les méthodes directe et par transfert. Selon Hutchins et Somers (1992, p. 183), Systran est généralement décrit comme étant un système fondé sur l'approche directe, mais il est évident, d'après la description des étapes de la traduction (voir *User's Guide for SYSTRAN 4.0*, p. 2-3), qu'il pourrait presque être considéré comme un système par transfert.

### **1.2.4 Avantages et inconvénients de la TA**

Les logiciels de TA ont l'avantage de produire une traduction brute très rapidement (selon le mode choisi et la longueur du texte, il est souvent question de secondes). Les

résultats ne sont toutefois pas toujours heureux. En effet, ces logiciels ont tendance à donner une traduction qui correspond à la notion de « correspondance formelle »<sup>12</sup> chez Nida et Taber (1969, p. 201), puisqu'elle colle à la forme des phrases du texte de départ, sans en analyser contextuellement le sens, et à la notion de « close translation »<sup>13</sup> chez Sager (1994, p. 155), laquelle favorise le plus haut degré possible d'équivalence formelle. Selon Arnold (2003, p. 128) « this can easily lead to “translationese” where the structure of the source language is wrongly carried over to the target language ». Ainsi, la méthode adoptée par ces systèmes accorde la priorité à la forme du message plutôt qu'à son contenu. Sager (1994, p. 253) explique que le système peut seulement traiter le texte dans la mesure où sa grammaire et ses dictionnaires peuvent reconnaître les formes qu'ils rencontrent dans un texte donné. « Divers facteurs expliquent ces limites. Toutefois, le problème fondamental réside dans le fait que la langue naturelle est extrêmement complexe et difficile à décrire. Aucun modèle formel ne décrit en totalité ne serait-ce qu'une seule langue » (L'Homme, 1999, p. 104).

Il convient de noter que les logiciels de TA qui sont fondés sur des principes linguistiques doivent compter de nouveaux modules, y compris des dictionnaires et des règles de

---

<sup>12</sup> « Quality of a translation in which the features of the form of the source text have been mechanically reproduced in the receptor language. Typically, formal correspondence distorts the grammatical and stylistic patterns of the receptor language, and hence distorts the message, so as to cause the receptor to misunderstand or to labor unduly hard. » (Nida et Taber, 1969, p. 201)

<sup>13</sup> « The *close translation* is concerned with observing the highest possible degree of formal equivalences at the word, phrase, clause and argument level; this is usually associated with the admission of smaller units of translation and a narrower concept of equivalence. The expression 'literal', as synonym to 'close', is here deliberately avoided because of its various and diverse interpretations ranging from word-for-word matches to denotational equivalence only. » (Sager, 1994, p. 155)

grammaire pour chaque combinaison de langues et chaque direction (ang.→fr. et fr.→ang., p. ex.).

### 1.3 Mémoires de traduction (MT)

L'idée des mémoires de traduction est apparue dans les années 70, mais ce n'est que depuis la fin des années 90 que des systèmes de ce genre sont offerts à grande échelle sur le marché (Bowker, 2002b, p. 199). Les mémoires de traduction consistent en une base de données (parfois appelée corpus parallèle ou bitexte) constituée de deux types de textes : des TLD (textes originaux) dans une langue  $x$  et des TLA correspondants (leur traduction) en langue  $y$ . Les deux types de textes sont alignés (c.-à-d. liés l'un à l'autre), normalement<sup>14</sup> au niveau de la phrase (ou d'éléments s'apparentant à une phrase<sup>15</sup>). Le but des mémoires de traduction? Permettre la réutilisation de documents ou de passages déjà traduits.

Il convient de noter que les traductions qui sont stockées dans une MT ont été effectuées par des traducteurs. Une MT est donc un système qui permet la traduction assistée par ordinateur (TAO), mais il ne s'agit pas d'un système de TA, car le système de MT ne fait aucune tâche de traduction proprement dite. Le rôle de la MT est de chercher dans la base de données et d'extraire les passages (traduits antérieurement par un traducteur) qui,

---

<sup>14</sup> Selon le logiciel utilisé, l'alignement peut aussi se faire au niveau du paragraphe.

<sup>15</sup> Ex. : titres, éléments de listes, cellules de tableaux (Bowker, 2002a, p. 94).

selon le logiciel et les paramètres établis, sont plus ou moins semblables aux passages dans le nouveau TLD et que le traducteur peut alors « recycler ».

### 1.3.1 Fonctionnement d'une mémoire de traduction « traditionnelle »<sup>16</sup>

D'abord, le système divise tous les textes consignés dans le corpus de la mémoire de traduction en segments (qui correspondent normalement aux phrases, mais qui peuvent être des titres, des sous-titres, ou même des paragraphes). Ensuite, les segments en langue  $x$  sont alignés aux segments correspondants en langue  $y$ . On appelle *unité de traduction* une paire de segments alignés. Lorsqu'un traducteur reçoit un nouveau texte en langue  $x$  à traduire en langue  $y$ , le système compare ce texte au contenu de la base de donnée en vue de trouver des segments déjà traduits. Le cas échéant, le système extrait l'unité de traduction en question et la présente au traducteur (Bowker, 2002b, p. 199). Ce dernier peut ensuite réutiliser les passages trouvés (s'il y a lieu) ou du moins, s'en inspirer.

Il y a deux façons de créer une base de données alignée. On peut soit commencer avec une mémoire vide, à laquelle on intègre des textes à mesure qu'on les traduit, soit verser des textes déjà traduits dans la base de données (Bédard, 1998, p. 25; Bowker, 2002a, p. 108-111). La deuxième façon est plus rapide, mais comporte un plus grand risque d'erreurs, car l'alignement est effectué automatiquement par le système<sup>17</sup> (Iverson, 2003,

---

<sup>16</sup> Qui effectuent l'alignement et les correspondances au niveau de la phrase.

<sup>17</sup> Dans le premier cas, l'alignement se fait au fur et à mesure (Bowker, 2002a, p. 109).

p. 35). Pour garantir que l’alignement est bien fait, il faut le vérifier manuellement – un processus qui peut prendre beaucoup de temps.

Selon les logiciels et leurs algorithmes de recherche, différents types de correspondances peuvent être présentées au traducteur, dont des correspondances parfaites ou totales (le segment extrait de la base de données est identique au nouveau segment à traduire), des correspondances partielles (la MT permet des divergences au niveau des nombres, des dates, des devises, des noms propres, etc.) (Bowker, 2002a, p. 98) et des correspondances floues (le segment extrait de la base de données est semblable au nouveau segment à traduire, mais sa traduction doit être révisée) (Iverson, 2003, p. 35) (voir la figure 2).

Correspondance parfaite ou totale	The filename is not a valid name. The filename is not a valid name.
Correspondance partielle	The 6.0 version of the software will be issued in 1999. The 5.0 version of the software will be issued in 1997.
Correspondance floue	The specified file is not valid. The specified file is not a valid file.

**Figure 2** : Types de correspondances  
(L’Homme, 1999, p. 215; Bowker, 2002a, p. 95, 100)

Pour ce qui est des correspondances floues, le niveau de correspondance ou de similarité est normalement exprimé en pourcentage (L’Homme, 1999, p. 216; Bowker, 2002a, p. 99). C’est au traducteur qu’il revient de fixer le seuil de similarité qui doit être atteint pour qu’une correspondance lui soit présentée. Ce seuil doit cependant être soigneusement fixé pour tirer le maximum de cet outil. Si le seuil est trop élevé, le système risque d’omettre d’extraire des correspondances utiles. Si le seuil est trop bas, des segments non pertinents risquent d’être relevés et de faire perdre du temps au traducteur (Bowker, 2002b, p. 200).

La plupart des mémoires de traduction fonctionnent parallèlement avec une banque de données terminologiques. Le traducteur doit d'abord verser des termes et leurs équivalents dans la banque; le logiciel de mémoire de traduction peut ensuite vérifier automatiquement si les termes qui figurent dans un nouveau TLD se trouvent dans la banque de données terminologiques. Si on le désire, ceux-ci peuvent d'ailleurs être automatiquement remplacés par leur équivalent dans la LA. De cette façon, même s'il n'y a pas de correspondance parfaite, partielle ou floue dans le corpus, le traducteur peut quand même obtenir un peu d'aide grâce à la banque de données terminologiques (Bowker, 2003, p. 56-57).

### **1.3.2 MultiTrans : système de MT**

Le logiciel de mémoire de traduction utilisé aux fins du présent projet s'appelle MultiTrans. Il s'agit d'un produit de MultiCorpora R&D Inc., dont les bureaux se trouvent ici-même dans la région, à Gatineau (Québec). Contrairement aux systèmes traditionnels de mémoire de traduction, qui sont fondés uniquement sur l'alignement au niveau de la phrase<sup>18</sup>, la technologie de MultiTrans repose sur la reconnaissance de chaînes de caractères de toute longueur présentes dans le contexte pertinent (*Guide de l'utilisateur – MultiTrans Light 3.6*, p. 1).

---

<sup>18</sup> Voir note à ce sujet à la section 1.3.

MultiTrans compte plusieurs modules, dont trois nous intéressent ici. Il s'agit des modules TransCorpora, TermBase et TransTerm. Le module TransCorpora est un outil qui permet la création et la gestion de corpus (fonction TransCorpora Builder), de même que la recherche dans ces corpus (fonction TransCorpora Search) afin de récupérer des mots, des expressions, des phrases ou des paragraphes. Lorsqu'un corpus est créé, les documents qu'il contient sont automatiquement alignés. Ainsi, les résultats des recherches sont présentés en contexte, dans des fenêtres parallèles (voir la figure 3), ce qui permet de corriger tout alignement erroné. Le module TermBase (voir la figure 4) est un outil de gestion de la terminologie. Il permet la création et la gestion de banques de données terminologiques, y compris l'ajout, la modification et la suppression d'entrées. Il permet aussi la recherche dans ces banques. Le module TransTerm est intégré à la barre de menu du logiciel de traitement de texte (menu TransTerm). Ainsi, il permet la comparaison d'un document à traduire ouvert dans Microsoft Word, par exemple, aux documents contenus dans le corpus créé dans TransCorpora, de même que le remplacement automatique des expressions d'un document par leurs expressions équivalentes respectives repérées dans la banque de données terminologiques créée dans TermBase (les expressions remplacées sont affichées en couleur). Lorsqu'une expression de deux mots ou plus est repérée dans le corpus, elle est surlignée dans le document à traduire. On peut consulter les résultats en sélectionnant l'expression en question et en sélectionnant la fonction Extraire de TransCorpora à partir du menu TransTerm. L'écran de MultiTrans s'affiche. Il ne suffit plus que de trouver la traduction dans la fenêtre Cible et de l'insérer dans le document à traduire. Le module TransTerm comporte aussi une fonction de Traduction interactive, laquelle cherche simultanément dans la banque

TermBase et le corpus TransCorpora, mais cette fonction n'a pas été utilisée dans le cadre de ce projet, par souci d'uniformité. On accède au module TransTerm à partir du logiciel de traitement de texte, puis on ouvre une session TermBase, une session TransCorpora, ou les deux. (*Guide de l'utilisateur – MultiTrans Light 3.6*, p. 1, 3-5, 11-12, 107-108, 112, 114, 116). La figure 3 présente un écran typique dans MultiTrans.

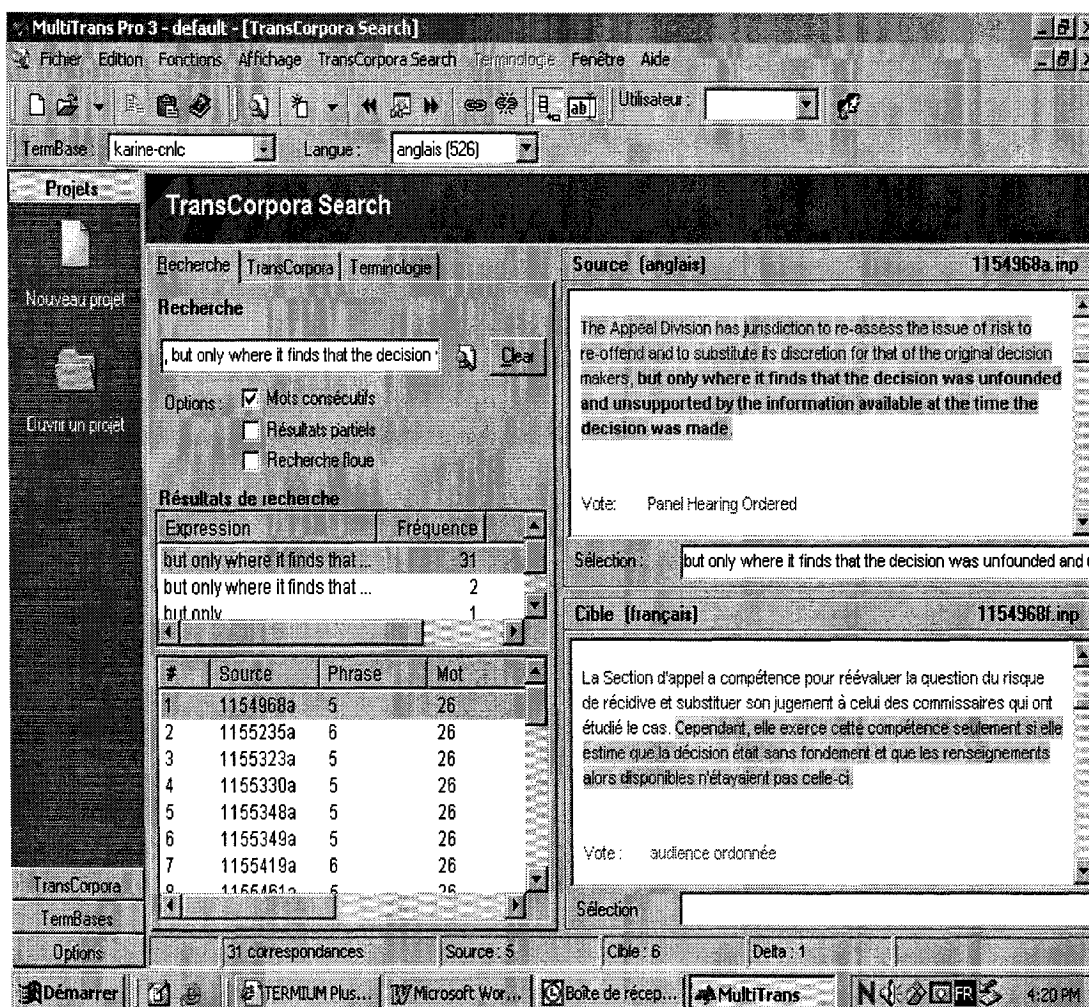


Figure 3 : Écran typique dans MultiTrans

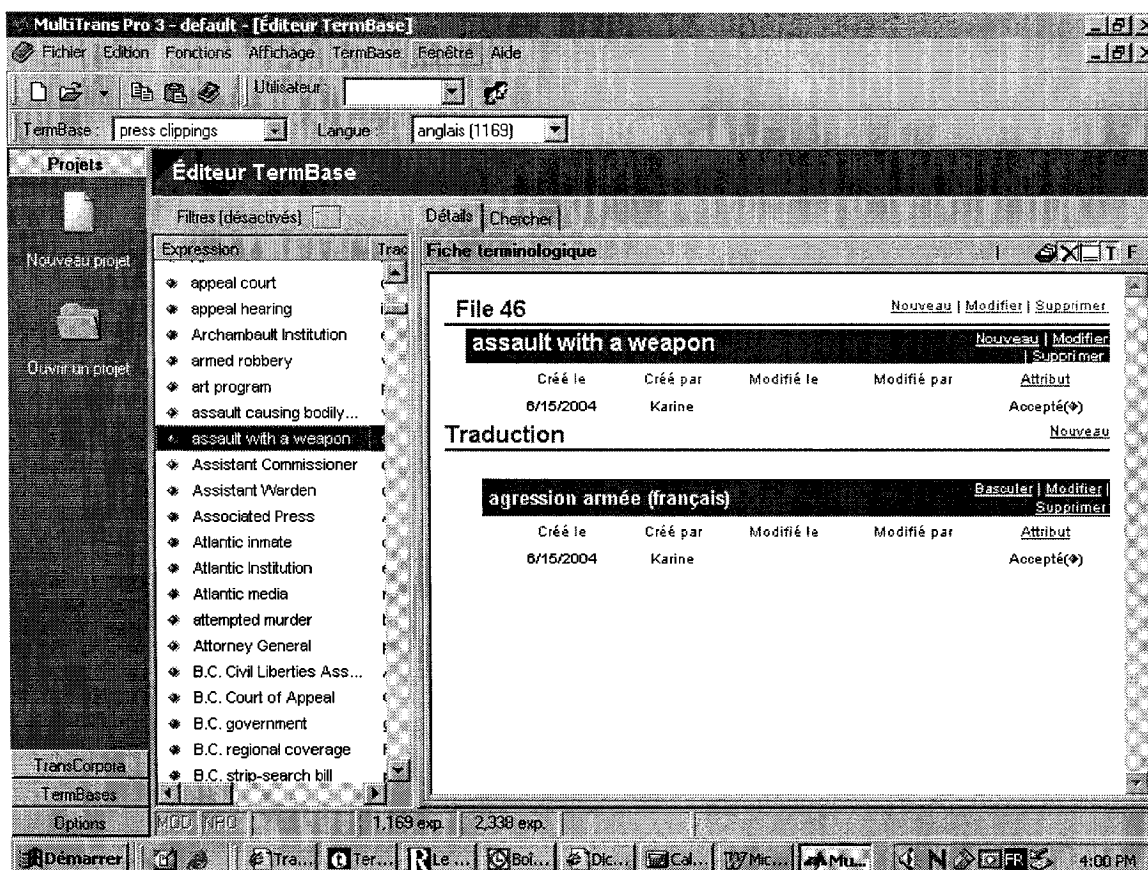


Figure 4 : Écran typique dans TermBase

### 1.3.3 Avantages et inconvénients des MT

Les MT assurent une certaine uniformité, ce qui améliore la qualité des textes, et permettent de travailler plus rapidement, ce qui entraîne une augmentation de la productivité (*Guide de l'utilisateur – MultiTrans Light 3.6*, p. 1). En effet, si on trouve des passages déjà traduits, il en reste moins à traduire, ce qui implique une économie de temps et d'argent (Iverson, 2003, p. 35). Les MT ont aussi l'avantage de ne pas être conçues pour une combinaison de langues en particulier. Comme elles ne sont fondées sur aucune connaissance linguistique, on peut y verser toutes sortes de textes, peu importe la langue (*Guide de l'utilisateur – MultiTrans Light 3.6*, p. 1).

Les MT ne conviennent cependant pas à tous. Pour qu'elles soient utiles, il faut traduire souvent les mêmes types de textes, des textes sur un même sujet, des textes modifiés<sup>19</sup>, des textes répétitifs, etc. L'efficacité des MT est normalement proportionnelle à la taille de la base de données et à la qualité des textes qu'elle contient. Ainsi, elles exigent un investissement important au début, question d'alimenter la base de données. Bien sûr, plus de traducteurs contribuent à la base de données, plus rapidement celle-ci sera alimentée. Il peut cependant aussi y avoir des conséquences négatives si plusieurs traducteurs avec des styles différents alimentent une même base de données. Si on met ensemble différentes phrases provenant de différentes traductions, on risque de se retrouver avec une « salade de phrases » plutôt qu'avec un texte cohérent (Bédard, 2000, p. 45).

Une fois l'étape de l'alimentation de la base de données franchie, le principal désavantage des MT est que, contrairement à la TA, qui donne presque toujours une traduction, peu importe sa qualité, elles ne répondent pas aux besoins inédits. Ces passages inédits et leur traduction peuvent toutefois être versés à la base de données pour être possiblement récupérés plus tard. On aurait cependant remarqué que les traducteurs qui ont recours aux MT font des traductions « phrase-à-phrase » (même ordre et même nombre de phrases que le TLD) pour favoriser la « recyclabilité » de la traduction et améliorer leurs chances de trouver des segments correspondants dans la MT dans le cadre de traductions ultérieures. Cette « symétrie quantitative », qui a pour but de faciliter la

---

<sup>19</sup> Textes mis à jour, ou qui ont fait l'objet de révisions, p. ex. (Bowker, 2002a, p. 112).

réutilisation des traductions, peut nuire au caractère naturel, au style et à la qualité des traductions et entraîner une « déformation professionnelle » et une part de démotivation chez les traducteurs (Bédard, 2000, p. 44-45; Bowker, 2002a, p. 117).

Aussi, il est entendu que toute traduction effectuée à l'aide d'un logiciel de MT doit faire l'objet d'une relecture finale attentive. Sans quoi, plusieurs genres d'erreurs risquent de se produire : erreurs d'ordre terminologique (un mot ne veut pas nécessairement toujours dire la même chose), erreurs de déictique (pronoms, etc.), reproduction à grande échelle ou à long terme de traductions douteuses ou erronées, etc. (Bédard, 2000, p. 45-46).

Maintenant que nous avons présenté un aperçu des trois méthodes de traduction, y compris une explication du fonctionnement des outils et des avantages et désavantages de chacun, nous allons passer au chapitre 3, dans lequel est présentée une analyse de recherches menées sur les types de textes qui semblent se prêter le mieux à chaque méthode.

## **Chapitre 2 – Typologie des textes et technologies de la traduction**

Avant d'entreprendre tout essai pratique, nous avons consulté la documentation publiée pour orienter notre démarche. Nous nous sommes surtout penchée sur deux sujets : 1) la typologie des textes, surtout par rapport à l'utilisation de la technologie de la traduction, et 2) les initiatives tentées précédemment en vue de déterminer dans quelle mesure un texte donné se prête à un outil de traduction donné.

### **2.1. Typologie des textes**

La notion de typologie des textes renvoie à l'idée selon laquelle il est possible de classifier des textes selon différentes catégories fondées sur diverses caractéristiques. Il n'existe pas de classification absolue. En fait, il existe plusieurs classifications de textes, fondées sur différents critères : public visé, intention, fonction, contenu, domaine, structure, etc. (Trosborg, 1997a, p. viii; L'Homme, 1999, p. 344; García Izquierdo, 2000, p. 284). Le but du présent projet n'est pas de passer ces classifications en revue.

Soulignons cependant que García Izquierdo (2000, p. 284), pour illustrer l'ampleur du phénomène, se contente de renvoyer aux typologies proposées par Sandig (1972), House (1981), Reiß (1982), Wilss (1984) et Emery (1991) en ajoutant que « all these typologies diverge with regard to the criteria on which the classification itself is based (text-function, macrofunction, domain to which text belongs, etc.). » De plus, mentionnons aussi que Biber (1989, p. 4), qui a lui-même proposé des classifications (celles-ci, fondées sur des critères linguistiques), renvoie quant à lui à celles présentées par Longacre (1976) et Chafe (1982). Pour leur part, Hatim et Mason (1990, p. 198) ont présenté une typologie basée sur les fonctions des textes. En outre, il convient de noter que les limites entre les divers types de textes consignés dans un même système de

classification sont loin d'être nettes (García Izquierdo, 2000, p. 284). Il est donc souvent plus réaliste d'imaginer un continuum (Snell-Hornby, 1995, p. 35; Biber, 1989, p. 4 et 16) entre deux extrêmes [sous-langages et textes littéraires, p. ex. (Melby, 1992, p. 148)].

La détermination de types de textes est dans une certaine mesure liée à la reconnaissance de caractéristiques textuelles (Neubert et Shreve, 1992, p. 126). Trosborg (1997b, p. 16) renvoie d'ailleurs à la classification de Biber (1989), fondée sur des critères strictement linguistiques<sup>20</sup>. La différence entre l'approche de Biber et la nôtre est que Biber voulait établir un cadre théorique pour d'autres études (1989, p. 39) tandis que notre approche est plus pragmatique, plus pratique. Ainsi, elle sera fondée strictement sur des caractéristiques de surface, surtout au niveau du lexique et de la phrase, parce que c'est seulement au niveau des mots et des phrases que les outils d'aide à la traduction sont en mesure de traiter un texte (généralement, ils ne peuvent même pas faire de liens entre les phrases). Donc, dans le contexte de la traduction assistée par ordinateur, la typologie des textes doit être fondée, entre autres, sur des caractéristiques terminologiques (stabilité, cohérence, étendue de la terminologie) et stylistiques/linguistiques (Gordon, 1990, p. 3-1, 3-3) (voir les sections 2.1.2.1 et 2.1.2.2 ci-dessous), puisque ce n'est pas ce qui est dit (c.-à-d. le contenu) ou la fonction de ce qui est dit qui compte, mais bien la façon de le dire (formulation, syntaxe) et le vocabulaire employé pour le dire (c.-à-d. la forme), qui peuvent avoir une influence sur les résultats.

### **2.1.1 Pertinence par rapport aux technologies de la traduction**

Lorsqu'il est question de traduction automatique ou de mémoires de traduction, un des critères à prendre en considération est sans contredit la nature des textes à traduire. En effet, « les propriétés des textes déterminent le type d'automatisation auquel ils peuvent se prêter » (L'Homme, 1999, p. 343).

Dans le contexte de la traduction assistée par ordinateur (TAO), « une typologie des textes permet de déterminer la catégorie à laquelle ils appartiennent et leur degré de "taosabilité" » (Gordon, 1990, p. 3-1). Ainsi, il est possible, à partir d'un examen du profil des textes à traduire, de faire un tri et de déterminer si un texte se prête à ce genre de traduction. Il convient de noter que parfois, l'utilisation de lignes de mention d'objet, de titres ou d'en-têtes peut faciliter cette tâche (Smith, 1985, p. 233; García Izquierdo, 2000, p. 293).

### **2.1.2 Types de textes/caractéristiques des textes et méthodes d'automatisation**

Comme il a été mentionné précédemment, il existe différentes façons de classifier les textes, chacune fondée sur des caractéristiques différentes. Ainsi, les types de textes peuvent être définis en fonction de catégories cognitives ou en fonction de critères strictement linguistiques (similarité des caractéristiques linguistiques) (Trosborg, 1997b, p. 16). Certains ouvrages traitant des technologies de la traduction conseillent ou déconseillent le traitement informatique de textes appartenant à certains domaines assez

---

<sup>20</sup> Le modèle de Biber tient compte de 67 caractéristiques linguistiques, lesquelles relèvent de seize catégories grammaticales, dont : pronoms et pro-verbès; questions; syntagmes prépositionnels, adjectifs et

vagues. Par exemple, bien que les auteurs de ces ouvrages donnent tout de même aussi d'autres critères plus précis, ils indiquent que les textes techniques, scientifiques, administratifs, informatifs et commerciaux sont idéals pour la TA (Gordon, 1990; Fernández Guerra, 2001), mais que les textes littéraires sont à éviter (Gordon, 1990; Melby, 1995). Ils ne s'entendent pas non plus à tous égards : Fernández Guerra dit que les textes juridiques se prêtent à la TA, tandis que Gordon prétend qu'ils sont à éviter. Du côté des MT, les documents techniques, scientifiques et juridiques sont ceux qui donneraient normalement les meilleurs résultats (Austermühl, 2001; Bowker, 2002a), tandis que les textes littéraires et publicitaires sont ceux qui ont tendance à donner de moins bons résultats (Bowker, 2002a). L'Homme (1999) est d'autant plus vague lorsqu'elle précise que le traducteur peut recourir aux mémoires de traduction et aux logiciels de traduction automatique pour traduire des textes de natures semblables, c'est-à-dire, par exemple, « des textes liés à un domaine de spécialité unique ou à des domaines connexes » (p. 351).

Certains auteurs parlent aussi de circonstances ou de caractéristiques favorables ou défavorables à la TA ou aux MT. Comme celles-ci sont beaucoup trop nombreuses pour les nommer toutes et que de le faire deviendrait énormément fastidieux, nous nous contenterons ici de présenter celles qui sont les plus simples et/ou les plus pertinentes à nos yeux.

---

adverbes; coordination; négation.

L'Homme conseille le recours aux MT et à la TA pour traduire des textes fournis par un même client (même terminologie) et des textes de même forme (formulations semblables, passages répétés)<sup>21</sup> (1999, p. 351-352). L'Homme rappelle aussi que plus un texte est prévisible (tendance générale propre à certains types de textes), plus il se prête à l'informatisation (1999, p. 356).

*Grosso modo*, pour la TA, il faut, dans le TLD, éviter les ambiguïtés (Lavroff, 1990; L'Homme, 1999), les phrases trop longues et complexes (Lavroff, 1990; Gordon, 1990; Hutchins et Somers, 1992; L'Homme, 1999; Auster Mühl, 2001), les ellipses (Gordon, 1990; Hutchins et Somers, 1992; Fernández Guerra, 2001), les empilages et cascades de mots (Gordon, 1990; Hutchins et Somers, 1992; Fernández Guerra, 2001) et les omissions (Gordon, 1990; Sager, 1994; Auster Mühl, 2001). En revanche, il faut rechercher les textes dont la terminologie est stable et limitée (Gordon, 1990; L'Homme, 1999).

Pour ce qui est des MT, les textes comportant des passages répétés (répétitions à l'intérieur d'un même texte ou entre différents textes), les textes mis à jour ou modifiés et les textes de nature semblable sont idéals (L'Homme, 1999; Auster Mühl, 2001; Bowker, 2002a).

---

<sup>21</sup> Ex. : rapports, guides d'utilisation ou d'entretien, procédures, descriptifs des caractéristiques d'un produit. (p. 351)

Compte tenu de tout ce qui précède, il semble que, pour obtenir les meilleurs résultats possible dans le cadre du présent projet, il faudra, pour chaque type de texte, tenir compte de la terminologie du sujet traité et des constructions linguistiques qu'il contient.

#### **2.1.2.1 Facteurs terminologiques**

Pour qu'un système donne de bons résultats, il faut souvent créer, en plus du dictionnaire de base fourni avec le système, un dictionnaire personnalisé contenant le vocabulaire propre à un domaine ou à un client en particulier. Ce travail peut être de plus ou moins longue haleine selon l'étendue du vocabulaire utilisé et du domaine en question (Gordon, 1990, p. 3-2). Une stratégie souvent proposée pour assurer la stabilité et la cohérence de la terminologie est l'adoption d'un langage contrôlé (voir la section 4.3.1.2). Cependant, les traducteurs qui ont recours à la technologie n'ont pas toujours le contrôle sur la nature ou la qualité des textes qu'ils doivent traduire (Underwood et Jongejan, 2001, p. 368), et c'est le cas des textes utilisés aux fins du présent projet. En revanche, certains types de textes sont habituellement écrits par la même personne ou par le même groupe de personnes, ce qui garantit une certaine cohérence de style.

#### **2.1.2.2 Facteurs stylistiques/linguistiques**

Les caractéristiques stylistiques/linguistiques sont beaucoup plus nombreuses et difficiles à cerner et à inventorier. Ainsi, Gordon (1990, p. 3-4, 3-5) donne en exemple quelques caractéristiques faciles à reconnaître ayant un effet négatif sur les résultats de la TAO : erreurs d'orthographe (y compris les coquilles), de grammaire ou d'usage, fautes d'accord (créent des ambiguïtés), ellipses, omissions d'articles, empilage de

compléments, d'attributs, de noms ou d'adjectifs, utilisation d'abréviations et d'acronymes.

## **2.2 Initiatives tentées précédemment en vue de déterminer si certains textes se prêtent à un outil de traduction donné**

Tel qu'il a déjà été mentionné, de plus en plus d'outils de traduction sont offerts sur le marché. Il y a plusieurs types d'outils (TA et MT, p. ex.) de même que plusieurs marques concurrentes. Un grand nombre de chercheurs ont mené des projets dans le cadre desquels ils ont tenté de déterminer dans quelle mesure certains textes se prêtaient à certains outils. Voici un résumé des travaux réalisés par certains de ces chercheurs et, le cas échéant, une explication de la façon dont le présent projet prend appui sur ces travaux ou diffère de ceux-ci.

### **2.2.1 TransRouter**

Le projet TransRouter, financé par la Commission européenne, s'est déroulé de janvier 1998 à avril 2000, soit pendant une période de 28 mois<sup>22</sup>. Il s'agit du projet qui a poussé le plus à fond l'idée d'établir quels types de textes se prêtent mieux à certains types d'outils. Il visait en effet la conception d'un outil entièrement automatisé qui aiderait les répartiteurs à déterminer la meilleure façon de traiter un projet de traduction (traduction humaine, mémoire de traduction, traduction automatique, entre autres, ou même une combinaison de méthodes) compte tenu de la qualité exigée, du temps et du budget accordés, et, bien évidemment, des méthodes, des applications et des ressources

---

<sup>22</sup> <http://www.hltcentral.org/projects/print.php?acronym=TRANSROUTER> (consulté le 13 octobre 2004)

linguistiques disponibles. En ce moment, ce genre de décision, s'il y a lieu, se prend en fonction de l'expérience et de l'intuition plutôt qu'à partir de données précises<sup>23</sup>. Reste qu'avant de décider du meilleur parcours de traduction à suivre, il faut avoir autant d'information que possible sur la nature du texte à traduire. L'établissement du profil des textes à traduire est donc un élément essentiel de ce processus. Ainsi, on s'est d'abord demandé quels facteurs pouvaient influencer de telles décisions, dans le but d'ensuite fournir des outils qui examineraient ces différents facteurs pour un projet de traduction donné. On compte donc parmi les facteurs de détermination du meilleur parcours de traduction les facteurs relatifs au projet de traduction en soi (langues source et cible, échéance, coût maximum), les facteurs concernant la nature (ou permettant l'établissement du profil) du texte à traduire (longueur du texte, longueur moyenne des phrases, degré de simplicité des phrases, degré de répétition interne), les facteurs à savoir si le texte a déjà été traduit et les facteurs en ce qui a trait à l'existence de ressources linguistiques. TransRouter appuie le processus décisionnel en offrant une suite d'outils informatiques (qui comptent les mots, détectent les répétitions, évaluent la longueur des phrases, vérifient la simplicité des phrases, comparent des versions, détectent les termes inconnus, estiment la couverture de mémoires de traduction) qui permettent l'analyse du

---

<sup>23</sup> Il convient ici de noter que les répartitrices du Service Criminologie effectuent déjà une certaine analyse du texte à traduire, en ce sens qu'elles doivent déterminer, aux fins de la facturation, de statistiques, de l'attribution du travail au sein du service (compte tenu des compétences et des goûts de chacun des traducteurs) ou de l'exportation du travail vers un autre service (selon le domaine de spécialité de chacun), si celui-ci est général, spécialisé ou ultra-spécialisé de même que le sujet dont il traite (informatique, administration, sociologie, droit, etc.). En ce moment, le recours à la technologie – ou l'évaluation d'un texte pour savoir s'il se prête à certains outils – n'est pas un facteur qui est pris en considération. Toutefois, comme le Bureau de la traduction a démontré qu'il était disposé à recourir aux technologies langagières (il a p. ex. récemment mis sur pied un service d'« Infolangagerie »), il est possible qu'à l'avenir, les répartiteurs aient à prendre des décisions sur les types d'outils à utiliser pour la traduction des textes qu'ils traitent, et nous espérons que les résultats du présent projet seront utiles à cette fin.

texte à traduire. Il convient de noter que les outils en question peuvent être utilisés de façon indépendante, c'est-à-dire à l'extérieur de l'environnement TransRouter. Ensuite, en se fondant sur toutes ces données (certaines sont compilées manuellement, d'autres, par les outils informatiques), un logiciel d'évaluation calcule les parcours de traduction possibles et présente les avantages et les désavantages de chacun. Il estime aussi les coûts associés à chaque parcours de même que le temps requis et la qualité prévue. La décision définitive revient à l'utilisateur. (TransRouter Consortium, 1999; Underwood et Jongejan, 1999; Cleary et Schäler, 2000)

La dernière mise à jour de la page Web de TransRouter remonte à février 2000<sup>24</sup>, et le dernier article trouvé sur le projet (Cleary et Schäler) est paru la même année. Dans le rapport annuel du projet pour 1999<sup>25</sup>, il est indiqué que le second prototype est prêt et à la phase d'essai, et qu'un modèle d'essai sera disponible à des fins de démonstration en janvier 2000. Enfin, le rapport parle d'examiner la possibilité de développer une application commerciale et d'offrir gratuitement une version de TransRouter dans Internet, mais une recherche rapide dans Internet permet de conclure qu'il n'en est rien.

Le présent projet de thèse est beaucoup plus modeste. Il vise premièrement à déterminer les caractéristiques des textes qui se prêtent le mieux à la TH, à la TA et à la MT, caractéristiques que les répartiteurs pourront chercher « manuellement » ou à l'aide d'outils automatisés (en utilisant certaines fonctions dans MS Word ou la fonction

---

<sup>24</sup> <http://irc.csis.ul.ie/research/projects/TransRouter/index.html> (consulté le 14 octobre 2004)

d'analyse de MultiTrans, p. ex.) dans les textes à traduire afin de décider de la meilleure méthode de traduction à adopter. De plus, tel qu'il a été mentionné à la section 0.4, les textes en question sont limités à ceux ordinairement traduits au Service Criminologie.

### **2.2.2 PaTrans – Étude de cas**

PaTrans est un système de TA anglais-danois élaboré par le Center for Sprogteknologi (CST) au Danemark et conçu à l'origine pour traduire des documents traitant de la pétrochimie et des brevets dans le domaine de la mécanique. En 1996, le CST et le cabinet de traduction pour lequel il a conçu PaTrans ont lancé un projet dont le but était de déterminer s'il serait profitable d'adapter le système à de nouveaux types de textes.

Cette étude de cas rejoint le projet TransRouter en ce sens qu'on a automatisé le plus possible l'analyse des nouveaux types de textes afin de déterminer s'ils se prêtaient à la TA avec PaTrans. Une suite de programmes a été élaborée, chacun examinant un certain critère mesurable et attribuant une note à chaque texte en fonction du nombre d'occurrences du phénomène en question. Povlsen *et al* (1998) donnent une liste de phénomènes causant des problèmes dans les traductions, mais comme ils sont propres à la traduction anglais-danois, la plupart ne s'appliquent pas nécessairement à la combinaison de langues traitée ici.

La méthode adoptée pour le présent projet ressemble à certains égards à celle adoptée pour cette étude de cas. En effet, le seul critère permettant de juger de l'utilité et de

---

<sup>25</sup> [http://www.hltcentral.org/usr\\_docs/project-source/transrouter/ar-99.html](http://www.hltcentral.org/usr_docs/project-source/transrouter/ar-99.html) (consulté le 15 octobre 2004)

l'efficacité des différentes méthodes sera exprimé en fonction des efforts (temps) de postédition requis. Il est en effet raisonnable de présumer que plus le texte brut est de qualité, moins les efforts de postédition sont importants. (Povlsen *et al*, 1998)

### **2.2.3 Translatability Checker**

Le Translatability Checker est un outil conçu pour évaluer la mesure dans laquelle des textes anglais se prêtent à la traduction automatique, et ce, dans le but d'assurer l'utilisation la plus efficace et la plus efficiente de la technologie. Cet outil a été élaboré dans le cadre du projet TQPro (Translation Quality for Professionals), appuyé par la Commission européenne.

Le Translatability Checker procède à une analyse rapide et en surface du texte de départ, puis non seulement attribue une note au texte, mais présente aussi l'analyse de chaque phrase du texte. La note est calculée selon l'occurrence d'« indicateurs » considérés comme ayant un effet négatif sur la qualité des traductions automatiques. Parmi les phénomènes causant des difficultés aux systèmes de TA, on donne notamment en exemple l'ambiguïté structurelle, les mots composés comportant trois éléments ou plus (ou « empilage », pour reprendre l'expression de Gordon utilisée plus haut), les « phrases » sans verbe, l'ambiguïté lexicale, les phrases soit très longues, soit très courtes, et les homographes (nom-verbe, adjectif-verbe, adjectif-nom-verbe). (Underwood et Jongejan, 2001)

La liste d'indicateurs utilisée par le Translatability Checker pourrait servir de point de départ pour la liste de critères qui sera dressée dans le cadre du présent projet. Mais

comme le Translatability Checker ne traite que de la traduction automatique, il faudra ajouter d'autres critères à la liste afin que celle-ci tienne également compte des mémoires de traduction et de la traduction humaine.

#### **2.2.4 FRASSE**

FRASSE est un outil élaboré dans le cadre d'un projet de l'Université de Linköping (Suède) intitulé *Linguistic Engineering for Generation and Translation of Documentation*. Ahrenberg et Merkel (1996) expliquent que ce projet a notamment pour objet de voir comment l'information contenue dans des textes de départ et leur traduction peut être exploitée à différentes étapes d'un projet de traduction (p. 184), d'étudier les avantages et les désavantages des systèmes fondés sur une mémoire de même que d'examiner et d'élaborer des outils analytiques pouvant prédire l'effet de tels systèmes sur un ensemble de textes donné (p. 187). Parmi les caractéristiques pouvant être exploitées, Ahrenberg et Merkel (1996, p. 185) mentionnent notamment la répétition de syntagmes ou de phrases. Divers outils sont examinés dans le cadre de ce projet, dont des outils de diagnostic qui caractérisent les textes et les types de textes en fonction de paramètres [longueur des mots et des phrases, occurrence des prépositions et des pronoms, etc. (p. 195)] ayant un effet direct sur le rendement et l'application de diverses méthodes de traduction assistée par ordinateur (Ahrenberg et Merkel, 1996, p. 185 et 191). Le profil ainsi établi indique quel type d'outil devrait être utilisé pour la traduction. FRASSE est l'un des outils de cette catégorie. Il sert à relever les phrases et les syntagmes récurrents dans un texte et calcule également les taux de répétition interne et externe (p. 192).

En lisant cet article, il devient clair qu'Ahrenberg et Merkel décrivent un précurseur des outils d'analyse de documents que l'on retrouve actuellement dans la plupart des progiciels de mémoire de traduction (voir la section 2.2.5), lesquels peuvent être utilisés pour relever les chaînes de texte récurrentes.

### **2.2.5 Outils d'analyse de documents**

Un outil d'analyse de documents analyse de nouveaux textes de la langue de départ (TLD) afin de relever les passages qui sont répétés à l'intérieur du texte ou dans d'autres textes à traduire, ou contenus dans différentes ressources (banques de données terminologiques, corpus de mémoires de traduction, p. ex.). Pour y arriver, l'outil divise le TLD en segments et compare chaque segment avec le reste du TLD (pour trouver les répétitions internes) ou avec le contenu d'autres documents ou ressources (pour trouver les répétitions externes). Une fois le processus de comparaison terminé, la plupart des outils d'analyse de documents génèrent un genre de rapport statistique détaillé permettant de faire la distinction entre les nouveaux passages et les passages répétés (segments que le traducteur peut « recycler »). De tels outils peuvent aider les répartiteurs et les traducteurs à déterminer, par exemple, le temps nécessaire à la traduction ou le coût de celle-ci.

Un extrait d'une analyse de document générée par la fonction d'analyse du logiciel MultiTrans est présenté à la figure 5. L'analyse donne d'abord certains renseignements généraux (langues source et cible, identification du document source, nombre de mots et de phrases contenus dans le document source, p. ex.), puis elle fournit des détails sur le nombre de mots et de phrases du document source pour lesquels le logiciel a trouvé une

correspondance (parfaite ou floue) dans la TermBase ou la mémoire de traduction  
(TransCorpora).

### Renseignements généraux

Langues de: anglais à: français  
Document(s) source : C:\@Data\karine\1823457a.doc

### Statistiques relatives au(x) fichier(s) traité(s)

#### Total

Fichier(s) : 1  
Phrases : 29  
Mots : 221  
Temps écoulé : 7 sec.

### TermBase - Rendement

TermBase utilisé : karine-desc-tache  
Expressions dans TermBase : 295  
Mots remplacés : 33  
% Remplacement : 15%

Phrases avec 100% remplacement : 4  
# Mots: 9  
% Mots: 4%

Phrases avec 75 - 84% remplacement : 0  
# Mots: 0  
% Mots: 0%

### TransCorpora - Rendement

Corpora Utilisé : karine-desc-tache  
# phrase source : 681  
# mots source : 1,426  
# total mots trouvés : 166  
% Mots: 75%

Préférences du TransCorpora Search  
2 mots consécutifs ou plus

Phrases  
# Mots  
% Mots

Phrases avec 100% remplacement  
12  
33  
15%

Phrases avec 85 - 99% remplacement  
1  
58  
26%

Phrases avec 75 - 84% remplacement  
1

36  
16%

Total dans phrases 75-100% pareil:  
14  
127  
58%

*Figure 5 : Extrait de résultats d'analyse – MultiTrans*

Il convient de noter qu'habituellement, les outils d'analyse de documents ne consistent pas en des outils indépendants ou autonomes; ils sont plutôt normalement intégrés à d'autres outils, comme des mémoires de traduction. Par exemple, MultiTrans et Trados sont tous deux dotés d'un module d'analyse. Quoique chaque outil d'analyse de documents soit propre à un certain produit, il pourrait être utile pour les répartiteurs d'utiliser la fonction d'analyse de documents de n'importe quel produit à leur disposition pour analyser les nouveaux textes à traduire en fonction de la liste de vérification.

Dans le cadre du présent projet, la fonction d'analyse de MultiTrans est l'un des outils que nous utiliserons pour cerner les caractéristiques des textes et élaborer la liste de vérification.

### **2.2.6 Comparaison entre la traduction humaine et la traduction automatique**

La comparaison qu'Oleg Lavroff (1990) décrit dans son article ressemble en certains points (objectif, méthode, outil) au présent projet. Dans le cadre de son expérience, Lavroff a notamment calculé le temps qu'a pris un traducteur pour postéditer des traductions brutes produites à l'aide d'un logiciel de TA, en l'occurrence, Systran. Il présente aussi un tableau des temps et des coûts de traduction selon la stratégie adoptée, de même que les gains réalisés par rapport à la traduction humaine. Le présent projet de

thèse tient compte d'une troisième option, soit les MT, et cherche non seulement à établir quelle option, entre la TH, la TA et les MT, est la plus efficace pour certains textes, mais aussi à proposer des critères pour choisir dès le départ la meilleure option possible à l'avenir.

Cette première partie comportait l'introduction, une explication des concepts de base (traduction humaine, traduction humaine et mémoires de traduction) et un résumé de la théorie avancée et des initiatives tentées précédemment dans le domaine. La deuxième partie décrit le travail de préparation en vue de l'élaboration des lignes directrices qui permettront de déterminer la meilleure méthode de traduction à adopter, selon le texte.

**PARTIE II – Conception d’une approche pour déterminer la meilleure  
méthode de traduction à adopter**

## **Chapitre 3 – Travail de préparation**

Comme il a été énoncé à la section 0.3, le présent projet vise à trouver une façon de déterminer quelle méthode de traduction (TH, TA, MT) est la plus efficace pour un type de textes donné. Le présent chapitre décrit le travail de préparation nécessaire pour arriver à une approche pour déterminer la meilleure méthode de traduction à adopter.

### **3.1 Sélection des textes**

Comme nous menons ce projet dans la perspective du Service Criminologie, nous avons commencé par choisir trois types de textes traduits fréquemment dans ce service : des décisions de la Section d'appel de la Commission nationale des libérations conditionnelles (CNLC), des énoncés de qualités et des coupures de presse. Comme on peut le constater en lisant les descriptions qui suivent, ces types de textes ont chacun une fonction et une forme distincte. Un exemple de chacun de ces types de textes est présenté à l'annexe A.

La Section d'appel est un tribunal administratif auquel s'adressent les détenus qui estiment avoir fait l'objet d'une décision inéquitable de la part de la CNLC au terme de leur audience de libération conditionnelle. Les décisions d'appel comportent normalement les rubriques suivantes : Le rôle de la Section d'appel, Sommaire de la décision de la Section d'appel, Analyse des motifs et Conclusion. La première rubrique, Le rôle de la Section d'appel, compte trois paragraphes standards qui demeurent inchangés d'un texte à l'autre. La section suivante, Sommaire de la décision de la Section d'appel, comporte habituellement les mêmes formulations; il n'y a pratiquement que les dates, le type de décision et les motifs d'appel énumérés qui changent d'une fois à l'autre.

La Conclusion est moins rigide dans ses formulations, mais il reste qu'essentiellement, elle contient plus ou moins toujours les mêmes choses. La rubrique Analyse des motifs est normalement la plus longue. C'est là que le commissaire examine point par point chacun des motifs d'appel du délinquant et les réfute, selon le cas. Comme on peut dire que cette rubrique est plus « personnalisée », son contenu est forcément plus inusité et exige donc un peu plus d'effort de la part du traducteur. Comme ces textes sont protégés, tous les renseignements confidentiels (noms et numéro d'identification des détenus) ont été supprimés des textes.

Quant aux énoncés de qualités, ils indiquent notamment les études, le profil linguistique, l'expérience, les capacités, les connaissances et les qualités personnelles exigés des candidats à un poste en particulier.

Pour ce qui est des coupures de presse, il s'agit de résumés d'articles publiés ou de reportages diffusés dans les médias partout au pays concernant dans une certaine mesure le Service correctionnel du Canada (SCC). Ces coupures de presse paraissent chaque jour ouvrable sur le site intranet du SCC.

### **3.2 Travail de préparation à l'étape de la traduction**

L'étape suivante consistait à traduire trois exemples de chacun de ces types de textes à l'aide de chacune des trois méthodes énoncées précédemment : TH, TA, MT (donc 27 textes en tout). Il était important ici de choisir trois exemples *différents* de chaque type de textes, parce que si le *même* texte avait été traduit trois fois, toutes les recherches auraient été effectuées et tous les problèmes de traduction auraient été résolus à la

première traduction, ce qui aurait fait en sorte d'accélérer, voire de rendre routinier le travail de traduction avec la deuxième et la troisième méthode, et ainsi, de fausser les résultats. En outre, nous avons décidé d'utiliser plusieurs exemples<sup>26</sup> de chaque type de textes afin d'avoir un aperçu le plus représentatif possible des résultats que donnait chaque méthode pour chaque type de textes. Si nous n'avions utilisé qu'un seul exemple de chaque type de texte, il aurait été difficile de déterminer si les problèmes de traduction, le cas échéant, étaient propres au texte en question ou au type de textes. Nous avons tenu compte des différences de longueur et de niveau de difficulté des textes en calculant la moyenne du nombre de mots traduits à l'heure pour chaque méthode pour chaque type de textes (voir la section 4.1 pour plus de détails).

Cependant, la TA et la MT exigeaient d'abord un certain travail de préparation. En effet, il nous fallait premièrement créer un corpus et une banque de données terminologiques dans MultiTrans pour chaque type de textes. Pour garantir une certaine « égalité des chances » entre les logiciels de TA et de MT, il allait de soi qu'il fallait du même coup créer des dictionnaires équivalents pour chaque type de textes dans Systran. Il convient de noter que les banques de données terminologiques créées dans MultiTrans et les dictionnaires créés dans Systran sont sensiblement les mêmes, la plupart des différences étant attribuables aux différentes règles de création des deux systèmes. Par exemple, comme MultiTrans n'est fondé sur aucune connaissance linguistique, il a fallu,

---

<sup>26</sup> Nous reconnaissons évidemment qu'un échantillon de trois exemples de chaque type de textes pour chaque méthode constitue un très petit échantillon; toutefois, trois valent mieux qu'un, et compte tenu du temps et des ressources dont nous disposions pour ce projet de recherche, nous estimons qu'il s'agit d'un nombre raisonnable.

notamment, créer des entrées pour les noms au singulier, et d'autres pour les noms au pluriel. Dans Systran, une entrée pour la forme canonique était censée suffire.

### **3.2.1 Création des corpus (TransCorpora) dans MultiTrans**

Ainsi, pour les corpus des différents types de textes, il a d'abord fallu décider de la taille du corpus et choisir les textes qui le composeraient. Le nombre de textes pour chaque corpus a été fixé à 30, car, pour la portée du présent projet, il semblait suffisant pour former des échantillons assez représentatifs<sup>27</sup>.

Pour les décisions de la Section d'appel, nous avons donc créé un corpus tenant compte de différentes décisions pouvant être rendues : libération conditionnelle totale refusée/révoquée, semi-liberté refusée/révoquée, libération d'office révoquée, etc. Certaines des décisions contenues dans le corpus confirmaient la décision de la CNLC, tandis que d'autres ordonnaient la tenue d'une nouvelle audience.

Pour ce qui est des énoncés de qualités, nous voulions choisir 30 énoncés représentant différents groupes professionnels et niveaux de poste. Nous avons donc décidé de composer ce corpus d'énoncés de qualités relevant de dix groupes professionnels importants représentés au SCC : AS (Services administratifs), CR (Commis aux écritures et règlements), CS (Gestion des systèmes d'ordinateurs), ES (Économique, sociologie et statistique), EX (Direction), FI (Gestion des finances), IS (Services d'information), PE

---

<sup>27</sup> Francie Gow (2003, p. 64), qui a rédigé une thèse sur l'évaluation des mémoires de traduction, avait aussi créé un corpus composé de 30 textes pour son projet, ce qu'elle avait estimé suffisant.

(Gestion du personnel), PM (Administration des programmes) et WP (Programmes de bien-être social). Pour chaque groupe, nous avons choisi trois énoncés de niveau différent.

Quant aux coupures de presse, comme certaines nouvelles qui font les manchettes reviennent souvent dans les coupures de presse pendant des jours, voire des semaines, nous avons décidé de créer un corpus avec les coupures des 30 jours ouvrables précédant l'essai des trois méthodes de traduction sur celles-ci.

Le processus de création de corpus dans MultiTrans est simple, rapide et convivial. Le *Guide de l'utilisateur* (p. 13-21) explique clairement les différentes étapes à suivre, imprimés d'écran à l'appui. On compte notamment les étapes suivantes : sélection des langues, sélection des documents source et cible, sélection du fichier de sortie (création d'un nouveau corpus ou ajout à un corpus existant) et auto-alignement.

### **3.2.2 Création des banques de données terminologiques et des dictionnaires**

Il ne restait donc plus qu'à créer les banques de données terminologiques et les dictionnaires. Pour décider du contenu de ceux-ci, nous avons dépouillé les 90 textes des corpus pour en dégager la terminologie et la phraséologie propres aux types de textes et aux domaines en question. Il convient de noter que, comme une partie de la terminologie est propre aux clients (p. ex. *offender* = délinquant, et non contrevenant, bien que cet équivalent ne soit pas faux en soi), certains termes qui pourraient être perçus comme « faciles » ou comme ne justifiant pas la création d'une entrée ont été retenus.

### **3.2.2.1 Création des banques de données terminologiques (TermBase) dans MultiTrans**

Le processus de création de banques de données terminologiques dans MultiTrans est également simple et convivial. Il devrait toutefois normalement être moins rapide que le processus de création de corpus, puisque dans ce cas-ci, il ne suffit pas de sélectionner des fichiers, mais bien de taper les différentes expressions et leur traduction. Le temps requis dépend donc du nombre d'entrées et de la longueur de celles-ci. Il convient de noter que le module TermBase permet l'importation de listes balisées, ce que nous avons fait pour une des banques de données (nous avons commencé le dépouillement des textes avant d'obtenir le logiciel et nous avons consigné les entrées dans un tableau en Word). C'est en tentant cette opération que nous avons eu notre seule et unique difficulté avec ce logiciel. Nous avons communiqué sans problème avec MultiCorpora R&D Inc.; on était déjà au courant qu'il y avait un bogue puis on nous a indiqué la marche à suivre pour contourner le problème.

### **3.2.2.2 Création des dictionnaires dans Systran**

Nous devons tout d'abord mentionner que le guide de l'utilisateur de ce logiciel nous a paru très incomplet et que nous nous sommes lancée dans la création des dictionnaires un peu à l'aveuglette. Évidemment, le temps nécessaire à la création d'un dictionnaire est tributaire du nombre d'entrées et de la longueur de ces entrées. Nous estimons qu'en général, si tout se passe bien, on peut consigner environ 100 entrées à l'heure. Mais voilà que tout ne s'est pas toujours bien passé lors de la création des dictionnaires dans Systran. Au moment de créer le troisième dictionnaire (celui pour les décisions d'appel), nous avons été estomaquée de remarquer que les entrées consignées et enregistrées

disparaissaient lorsque nous rouvrons le fichier. Nous avons communiqué avec le soutien technique de Systran et, à notre grand désarroi, on n'avait jamais entendu parler de ce problème. En utilisant nos entrées, le technicien a réussi à reproduire notre problème, mais n'avait pas de solution concrète à nous proposer. Après plusieurs appels et bien des embêtements, on en a déduit que nos entrées comportaient trop de caractères et que nous devions les séparer. C'est donc ce que nous avons fait, à contrecœur, ce qui semble avoir « fonctionné ». Un autre irritant a été le fait que le guide de l'utilisateur ne précise pas le nombre de dictionnaires pouvant être sélectionnés pour une traduction donnée. Nous en avons donc profité pour poser la question au technicien; il nous a répondu qu'avec notre version du logiciel, il était possible de sélectionner jusqu'à dix dictionnaires à la fois. Comme nous craignons de perdre à nouveau toutes les entrées du dictionnaire, nous avons décidé par précaution de créer plusieurs dictionnaires afin de limiter le nombre d'entrées dans chacun (et ainsi minimiser les pertes, au cas où); nous avons donc fini par créer huit dictionnaires pour les décisions d'appel. Mais lorsque le moment de traduire les décisions d'appel à l'aide de ce logiciel est venu, le système ne nous a pas permis de choisir plus de cinq dictionnaires. Nous avons donc dû passer quelques heures à combiner certains dictionnaires à l'aide de la fonction copier-coller.

### **3.3 Traduction**

Une fois les textes choisis et les corpus, banques de données terminologiques et dictionnaires créés, nous avons traduit, sans interruption et en nous chronométrant, trois exemples différents de chaque type de textes à l'aide de chacune des méthodes prévues : TH, TA, MT. Donc, pour les décisions de la Section d'appel, nous avons traduit trois textes à l'aide du logiciel de TA, trois textes à l'aide du logiciel de MT et trois textes sans

l'aide de ces deux logiciels (pour un total de neuf textes différents). Il en était de même pour les énoncés de qualités et les coupures de presse (donc pour un total global de 27 textes). Il convient de noter que le temps chronométré correspond au temps de traitement à l'aide des différents logiciels (le cas échéant), au temps de traduction et/ou de révision ainsi qu'au temps de correction, le produit final étant considéré comme un texte « livrable » selon les normes de qualité du Bureau de la traduction ou, à tout le moins, prêt pour l'étape de la révision ou de la relecture, le cas échéant.

## Chapitre 4 – Analyse des données et élaboration des lignes directrices

L'étape de traduction menée à bien, l'étape suivante consiste à analyser les données dans le but de déterminer quelle méthode a donné les meilleurs résultats pour chaque type de textes. Les textes seront par la suite examinés afin d'en tirer les caractéristiques pertinentes, lesquelles serviront à établir une première série de lignes directrices à l'intention des répartiteurs. Ces étapes sont décrites plus en détail dans le présent chapitre.

### 4.1 Résultats

Voici donc, pour chacun des 27 textes traduits, le nombre total de mots, le temps de traduction chronométré et le nombre de mots traduits à l'heure. Pour chaque type de texte, on donne le nombre moyen de mots traduits à l'heure et l'intervalle de variation<sup>28</sup> pour chaque méthode de traduction.

#### Énoncés de qualités

Méthode	N <sup>bre</sup> de mots	Temps requis	N <sup>bre</sup> de mots/heure	Moyenne de mots/heure	Intervalle de variation
MT	231	26 min.	533,08	490,82	73,70
MT	245	32 min.	459,38		
MT	240	30 min.	480,00		
TA	186	25 min.	446,40	475,90	74,31
TA	261	34 min.	460,59		
TA	243	28 min.	520,71		
TH	227	46 min.	296,09	289,15	47,51
TH	227	52 min.	261,92		
TH	361	70 min.	309,43		

---

<sup>28</sup> Différence entre les deux valeurs extrêmes d'une série statistique.

## Coupages de presse

Méthode	N <sup>bre</sup> de mots	Temps requis	N <sup>bre</sup> de mots/heure	Moyenne de mots/heure	Intervalle de variation
MT	677	85 min.	477,88	465,54	50,95
MT	800	99 min.	484,85		
MT	1092	151 min.	433,90		
TA	740	93 min.	477,42	485,49	26,61
TA	881	111 min.	476,22		
TA	771	92 min.	502,83		
TH	846	121 min.	419,50	461,96	84,20
TH	680	81 min.	503,70		
TH	802	104 min.	462,69		

## Décisions d'appel

Méthode	N <sup>bre</sup> de mots	Temps requis	N <sup>bre</sup> de mots/heure	Moyenne de mots/heure	Intervalle de variation
MT	773	51 min.	909,41	817,88	139,95
MT	949	74 min.	769,46		
MT	1188	92 min.	774,78		
TA	730	49 min.	893,88	913,83	273,33
TA	760	43 min.	1060,47		
TA	1102	84 min.	787,14		
TH	1234	125 min.	592,32	723,63	231,32
TH	755	55 min.	823,64		
TH	843	67 min.	754,93		

## 4.2 Analyse des résultats

On remarque que, pour les énoncés de qualités, la mémoire de traduction obtient la moyenne de mots traduits à l'heure la plus élevée. La traduction automatique suit de très près (moins de 15 mots à l'heure de moins), tandis que la traduction humaine arrive bonne troisième (différence de 201,67 mots à l'heure avec la MT). Lorsqu'on examine les résultats pour chacun des textes, on constate que la méthode qui est arrivée au deuxième rang, la TA, compte tout de même deux des quatre meilleurs résultats (le 2<sup>e</sup>, soit 520,71 mots traduits à l'heure, et le 4<sup>e</sup>, soit 460,59 mots traduits à l'heure).

L'intervalle de variation pour ces deux méthodes (MT et TA) est d'ailleurs sensiblement

le même (73,70 et 74,31 respectivement). On peut donc en conclure que la MT a donné les meilleurs résultats, mais que la TA n'est pas une mauvaise méthode pour autant.

C'est avec les coupures de presse que les résultats sont les plus serrés et, par conséquent, les moins concluants. Reste que, puisqu'il faut déterminer une méthode « gagnante », la traduction automatique est celle qui a donné les meilleurs résultats (moyenne de 485,49 mots traduits à l'heure), suivie pas très loin derrière de la mémoire de traduction (moyenne de 465,54 mots traduits à l'heure, soit moins de 20 mots à l'heure de moins) et encore moins loin ensuite par la traduction humaine (moyenne de 461,96 mots traduits à l'heure, ce qui donne une différence de 3,58 mots à l'heure entre la MT et la TH). En fait, 23,53 mots à l'heure séparent la « meilleure » et la « pire » méthode. Fait intéressant et gage probable de la quasi-égalité entre les méthodes, la traduction humaine (3<sup>e</sup> place) a obtenu le meilleur et le pire temps (503,70 et 419,50 mots traduits à l'heure). La TA (1<sup>re</sup> place) est la méthode avec l'intervalle de variation le plus faible (26,61), ce qui explique peut-être que sa moyenne de mots traduits à l'heure est supérieure. Il convient de noter que ce type de textes était probablement celui au contenu le plus varié. Puisqu'il pouvait traiter d'une vaste gamme de sujets différents, l'étendue du vocabulaire utilisé était probablement plus grande, et il y avait aussi probablement moins de répétition, ce qui pourrait expliquer pourquoi la TA et la MT n'ont pas donné des résultats de beaucoup supérieurs à ceux de la TH.

Pour ce qui est des décisions d'appel, il s'agit du type de textes pour lequel l'écart entre les résultats des différentes méthodes est le plus net et précis, mais aussi pour lequel les

intervalles de variation sont les plus grands pour une même méthode. Ainsi, la traduction automatique arrive en première place avec une moyenne de 913,83 mots traduits à l'heure, suivie de la MT (différence de 95,95 mots à l'heure entre les deux). La TH ferme la marche avec un écart de 94,25 mots à l'heure avec la MT. Quant aux intervalles de variation, ils s'expliquent peut-être par la variation de la longueur des textes. En effet, il semble y avoir, pour chaque méthode de traduction, un rapport plus ou moins inversement proportionnel entre le nombre de mots des textes et le nombre de mots traduits à l'heure : moins le texte est long, plus le nombre de mots traduits à l'heure augmente. Ce phénomène est peut-être attribuable au fait que le nombre de passages répétitifs des décisions d'appel (retrouvés normalement dans l'introduction et la conclusion, tel que noté à la section 3.1) demeure à peu près le même peu importe la longueur du texte. Ainsi, plus le texte est long, plus il comprend des passages nouveaux ou inusités.

Il y a une chose qui est commune à chaque type de textes : la TH obtient la pire moyenne dans chaque cas.

### **4.3 Détermination et analyse des caractéristiques pertinentes**

Maintenant que nous savons quelle méthode de traduction fonctionne le mieux pour chaque type de textes (la MT pour les énoncés de qualités, la TA pour les coupures de presse et les décisions d'appel), il faut tenter de déterminer quelles caractéristiques de ces textes font en sorte qu'ils se prêtent mieux à une méthode plutôt qu'à une autre. Le chapitre 2 donne un bref aperçu des caractéristiques qui, selon plusieurs auteurs, favorisent ou non l'automatisation du processus de traduction. La présente section vise à

déterminer quelles caractéristiques pertinentes (et précises et mesurables à l'aide des outils à notre disposition) figureront sur la liste de vérification qui permettra aux répartiteurs de décider de la meilleure méthode de traduction possible pour un type de textes donné.

#### **4.3.1 Caractéristiques pertinentes pour la TA**

À la section 2.1.2.1, nous avons brièvement parlé de l'adoption d'un langage contrôlé comme stratégie pour assurer la stabilité et la cohérence de la terminologie. Même si, comme nous l'avons alors expliqué, nous ne pouvons pas recourir à cette stratégie, les règles de grammaire qui la sous-tendent pourront peut-être nous aider à cerner les caractéristiques des textes qui se prêtent mieux (ou moins) à la TA. Après tout, si un texte ne semble pas suivre ces règles, il n'est probablement pas un bon candidat pour la TA (et peut-être même pas non plus pour la MT). En revanche, s'il semble suivre ces règles, il est possible qu'il s'y prête mieux. Nous présenterons d'abord les principaux problèmes propres à la TA au niveau syntaxique de même que la source de ces problèmes afin de savoir quelles caractéristiques sont à éviter dans le TLD.

##### **4.3.1.1 Problèmes propres à la TA : caractéristiques à éviter dans le TLD**

En traduction automatique, certaines erreurs découlent de problèmes structurels, c'est-à-dire lorsque la structure fait en sorte qu'il y a plus d'une signification possible ou lorsque, dans une LA donnée, il n'existe pas de correspondance directe pour une tournure syntaxique donnée dans la LD. Il n'est donc pas étonnant que les erreurs syntaxiques constituent l'une des principales catégories d'erreurs. Voici par conséquent un bref survol

de certains problèmes de traduction que les règles d'écriture énumérées plus loin (voir la section 4.3.1.2) pourraient permettre d'éviter.

#### 4.3.1.1.1 Résolution des référents des pronoms

Parmi les problèmes propres à la TA, on compte la résolution des référents des pronoms.

Ce problème survient lorsqu'un antécédent n'est pas clair pour le système (et parfois même pour les humains, tel que le démontre l'exemple qui suit; toutefois, les humains peuvent habituellement se reporter à l'ensemble du contexte pour les aider à trouver la bonne interprétation, tandis que les ordinateurs, qui généralement n'ont aucune forme d'intelligence, ne peuvent se servir du contexte aussi facilement).

Exemple<sup>29</sup>

The cat tries to catch the swallow. It looks tired.

Il y a ici deux possibilités : *it* renvoie soit à *cat*, soit à *swallow*, ce qui donne les deux options de traduction suivantes :

Le chat poursuit l'hirondelle. Il est visiblement fatigué.  
Le chat poursuit l'hirondelle. Elle est visiblement fatiguée.

En limitant le recours aux pronoms par la répétition du nom ou la reformulation de la phrase, on évite ce genre de problème.

---

<sup>29</sup> Exemple inspiré de L'Homme, 1999, p. 57.

#### 4.3.1.1.2 Voix passive

La voix passive donne du fil à retordre aux logiciels de TA anglais-français. En anglais, cette tournure est formée de l’auxiliaire être suivi du participe passé. En français, cette construction est parfois acceptable, mais pas toujours.

Exemples<sup>30</sup>

This building was built in 1923.  
Ce bâtiment fut construit en 1923.

Mary was given a book.  
Mary fut donné un livre. [*sic*]

This bed has been slept in.  
Ce lit a été dormi dans. [*sic*]

En limitant la voix passive par la reformulation de la phrase, on évite ce genre d’erreur.

#### 4.3.1.1.3 Interprétation des rapports de coordination entre les composantes de syntagmes

Certaines suites de mots, bien qu’elles obéissent aux règles syntaxiques, peuvent être interprétées de plus d’une façon, ce qui peut causer des difficultés aux logiciels de TA (et, encore une fois, aux humains aussi). Ainsi, lorsque des conjonctions comme *and* et *or* figurent dans une phrase, il peut arriver que le logiciel de TA juge mal leur portée.

Exemple<sup>31</sup>

Smog and pollution control

---

<sup>30</sup> Exemples inspirés/tirés de Hutchins et Somers, 1992, p. 103.

Il y a deux interprétations possibles. Soit *control* est modifié par *pollution* seulement, soit il est modifié à la fois par *smog* et *pollution*, ce qui donnerait en français :

Le smog et la lutte contre la pollution  
ou  
La lutte contre le smog et la pollution

On peut remédier à ce problème en reformulant la phrase en anglais de l'une des deux façons suivantes, selon l'interprétation voulue :

Pollution control and smog  
Smog control and pollution control

#### 4.3.1.2 Langage contrôlé (LC) et règles d'écriture

Nyberg *et al* (2003, p. 245) expliquent la notion de langage contrôlé comme suit : « A CL is an explicitly defined restriction of a natural language that specifies constraints on lexicon, grammar and style. » Autrement dit, il s'agit de « a form of language usage restricted by grammar and vocabulary rules » (Arnold *et al*, 1994, p. 156). En traduction automatique, ces règles tiennent compte des difficultés de traitement et visent à éliminer les ambiguïtés afin d'éviter les problèmes d'analyse (L'Homme, 1999, p. 324). Les textes de départ sont donc plus faciles à lire et à comprendre, ce qui améliore la qualité des traductions produites à l'aide d'un logiciel de TA (Nyberg *et al*, 2003, p. 245). La traduction automatique donne donc de meilleurs résultats avec des textes rédigés en fonction d'un langage contrôlé qu'avec ceux produits sans l'aide d'un langage contrôlé

---

<sup>31</sup> Exemple tiré de Hutchins et Somers, 1992, p. 182.

(Arnold *et al*, 1994, p. 156; Mogensen, 2000, p. 29; Austermühl, 2001, p. 165; Ehgoetz, 2004, p. 86).

Comme nous l'avons mentionné plus haut, nous n'examinerons que l'aspect des langages contrôlés qui traite des règles d'écriture. Pour ce projet, le choix des règles dépendra de notre capacité à vérifier le respect de celles-ci de façon automatisée (voir Nyberg *et al*, 2003, p. 252). Voici donc quelques exemples de règles relevées dans certains ouvrages qui pourraient servir à l'établissement de la liste de vérification, dont certaines sont tirées du PACE<sup>32</sup> et du AECMA<sup>33</sup> Simplified English :

- Limiter la longueur des phrases (20 mots, p. ex.) (Arnold *et al*, 1994, p. 157; Nyberg *et al*, 2003, p. 247 et 255; Ehgoetz, 2004, p. 29)
- Ne pas utiliser la voix passive (Nyberg *et al*, 2003, p. 247; Ehgoetz, 2004, p. 110 et 114)
- Limiter le recours aux pronoms (Nyberg *et al*, 2003, p. 247)
- Éviter les ellipses (Arnold *et al*, 1994, p. 157; Nyberg *et al*, 2003, p. 255)
- Ne pas omettre les conjonctions ou les pronoms relatifs (Arnold *et al*, 1994, p. 157; Nyberg *et al*, 2003, p. 255; Ehgoetz, 2004, p. 114)

---

<sup>32</sup> Perkins Approved Clear English

<sup>33</sup> Association européenne des constructeurs de matériel aérospace (European Association of Aerospace Industries)

- Limiter le recours aux gérondifs, aux adjectifs et aux participes présents en *-ing* (Arnold *et al*, 1994, p. 157; Nyberg *et al*, 2003, p. 255; Ehgoetz, 2004, p. 109 et 114)

Un peu dans la même veine que les règles des langages contrôlés, de nombreux auteurs ont publié dans Internet<sup>34</sup> et dans certaines publications (Hine, 2003, p. 20; Kramasz, 2003, p. 16; Harkus, 2003, p. 23) des règles d'écriture pour les textes destinés à la traduction automatique ou à la traduction en général. Sans toutefois s'inscrire nécessairement dans le cadre de langages contrôlés, ces règles d'écriture peuvent également servir à l'établissement de la liste de vérification. En voici un compte rendu :

- Ne pas omettre les pronoms relatifs, même s'ils peuvent l'être (ex. : *that, whom, which*)
- Éviter le recours aux verbes en *-ing*
- Ne pas utiliser de pronoms
- Rédiger des phrases simples
- Utiliser la voix active plutôt que la voix passive
- Limiter la longueur des phrases (20 mots, c'est bien; 40, c'est trop)
- Ne pas utiliser de contractions (*can't, don't, won't*)

---

<sup>34</sup> <http://www.sdl.com/products-translation/products/products-ref-center/products-ref-center-white-papers/products-ref-center-white-papers-mt/products-ref-center-white-papers-mt-writing-for.htm>  
<http://www.multilingualwebmaster.com/library/writing-TR.html>  
<http://www.etc.nl/fortrans.htm>  
<http://translationsoftware4u.com/translation-tips.htm>  
[http://www.izar.com/resources/knowledge\\_base5.html](http://www.izar.com/resources/knowledge_base5.html)  
 (consultés le 24 février 2005)

### 4.3.2 Caractéristiques pertinentes pour les MT

Nous n'avons pas trouvé beaucoup de documentation en ce qui a trait à des règles d'écriture précises pouvant améliorer le rendement des mémoires de traduction, mais, dans un autre ordre d'idées, il semble qu'au moins une entreprise se sert désormais d'un système de MT comme d'un système de rédaction (donc dans une langue, plutôt que deux) afin d'assurer l'uniformité de ses textes et ainsi simplifier le travail de traduction (Clark, 2002).

Pour revenir aux langages contrôlés, bien que ceux-ci soient normalement associés à la TA, Nyberg *et al* (2003, p. 248) sont d'avis qu'ils peuvent aussi améliorer l'efficacité des MT puisqu'en améliorant l'uniformité des textes, ils les rendent plus facilement réutilisables. Il est en effet logique de croire que, si le contenu similaire était normalisé, il pourrait être réutilisé plus efficacement (Brown, 2003, p. 8; Smart, 2003, p. 20). Dans le même ordre d'idées, Mogensen (2000, p. 29-30) fait remarquer que, pour maximiser la recyclabilité et ainsi l'utilité des MT, les textes doivent être écrits dans un style « simplifié », c'est-à-dire avec des phrases courtes et un vocabulaire stable. Les auteurs ne sont pourtant pas tous du même avis. Selon Heyn (1998, p. 135), il n'y a pas lieu de recourir à un langage contrôlé; il reconnaît pourtant que certains traducteurs ont tendance

à éviter d'utiliser des références anaphoriques ou cataphoriques afin de maximiser la réutilisabilité des unités de traduction qui figurent dans le corpus<sup>35</sup>.

Autrement, il semble que nous devrions nous en tenir aux résultats fournis par la fonction d'analyse du logiciel MultiTrans pour le volet « mémoire de traduction » de la liste de vérification (voir la section 2.2.5).

### **4.3.3 Caractéristiques pertinentes pour la TH**

Un traducteur peut, en principe, traduire n'importe quel type de textes. Plutôt que de préciser les caractéristiques qui font en sorte qu'un texte se prête à la TH, nous proposerons que celui-ci soit confié à un traducteur lorsqu'un texte ne se prête *pas* à la TA ou aux MT. Autrement dit, les caractéristiques qui feront en sorte qu'un texte se prête mieux à la TH seront opposées à celles qui feront en sorte qu'un texte se prête à la TA ou aux MT. Par exemple, si un texte contient de longues phrases, des constructions passives, de multiples pronoms, etc., il devrait être confié à un traducteur.

---

<sup>35</sup> Ces remarques sont intéressantes dans le contexte, bien que les questions relatives à l'utilisation des MT comme systèmes de rédaction, d'une part, et aux avantages que comportent les LC pour les MT, d'autre part, dépassent le cadre du présent projet.

#### 4.3.4 Terminologie

Comme il a été mentionné au chapitre 2, dans le contexte de la traduction assistée par ordinateur, il faut rechercher les textes dont la terminologie est stable et limitée. Les textes qui contiennent beaucoup de synonymes ou qui décrivent un même concept de plusieurs façons différentes se prêtent probablement moins à la TA ou aux MT.

Par exemple, dans le domaine du SRAS, le terme *SARS-associated coronavirus* compte plusieurs variantes, dont *SARS-CoV*, *SARS CoV*, *SCoV*, *SARS coronavirus*, *severe acute respiratory syndrome coronavirus* et *SARS virus* (Kweon, 2005). Un texte qui contiendrait toutes ces variantes se prêterait moins bien à la traduction assistée par ordinateur parce que les systèmes de TA et de MT donnent de meilleurs résultats lorsque le vocabulaire est stable<sup>36</sup>.

On peut avoir une idée de l'étendue du vocabulaire en calculant le rapport entre le nombre de mots différents et le nombre total de mots (type/token ratio). Plus le type/token ratio est bas, plus le contenu lexical est restreint, et vice versa. Un des outils d'analyse utilisés donne ce ratio (voir la section 4.4.2).

#### 4.4 Outils

Afin d'analyser les caractéristiques des différents textes, nous avons utilisé des outils que nous avons déjà à notre disposition, soit les statistiques de lisibilité dans Microsoft

---

<sup>36</sup> On peut bien ajouter les variantes au dictionnaire, mais le travail de dépouillement et d'entrée s'en trouve alors multiplié dans la même mesure. Dans le cas des mémoires de traduction, les correspondances seront

Word, WordList dans WordSmith Tools et la fonction d'analyse de MultiTrans. Voici une brève description de chacun.

#### **4.4.1 Statistiques de lisibilité - Microsoft Word**

Les statistiques de lisibilité dans Microsoft Word<sup>37</sup> présentent, entre autres, les données suivantes : nombre moyen de mots par phrase, degré de lisibilité Flesch<sup>38</sup>, niveau scolaire<sup>39</sup> et pourcentage de verbes conjugués à la voix passive.

#### **4.4.2 WordList - WordSmith Tools**

Cet outil comporte de nombreuses composantes, dont WordList, que nous utiliserons pour obtenir des données sur les textes que nous analyserons. WordList donne, pour un texte (ou un groupe de textes) donné, la liste des mots qu'il contient, de même que certaines statistiques, dont le rapport entre le nombre de mots différents et le nombre total de mots. Les listes de mots peuvent être présentées soit en ordre alphabétique, soit en ordre de fréquence.

---

de moins bonne qualité (c.-à-d. moins de 100 %) s'il contiennent des synonymes au lieu d'un terme constant pour désigner un même concept.

<sup>37</sup> Pour obtenir les statistiques de lisibilité dans Microsoft Word, cliquer Outils>Options>onglet Grammaire et orthographe, cocher Vérifier la grammaire et l'orthographe et Afficher les statistiques de lisibilité, cliquer OK>Outils>Grammaire et orthographe; lorsque Word termine de vérifier l'orthographe et la grammaire, il affiche les statistiques.

<sup>38</sup>  $206,835 - (1,015 \times \text{nombre moyen de mots par phrase}) - (84,6 \times \text{nombre moyen de syllabes par mot})$

<sup>39</sup>  $(0,39 \times \text{nombre moyen de mots par phrase}) + (11,8 \times \text{nombre moyen de syllabes par mot}) - 15,59$

#### 4.4.3 Fonction d'analyse de MultiTrans

Un extrait d'une analyse de document générée par la fonction d'analyse du logiciel MultiTrans est présenté à la section 2.2.5. Nous devons sélectionner, parmi les nombreuses données présentées, celles qui figureront sur notre liste de vérification. Il devra s'agir de données fournissant une bonne idée de l'ensemble des résultats, par exemple le nombre de mots et/ou de phrases du document source pour lesquels le logiciel a trouvé une correspondance (parfaite et/ou floue ou toutes confondues) dans la TermBase ou la mémoire de traduction (TransCorpora).

#### 4.5 Première ébauche de la liste de critères

La liste qui suit a été élaborée en tenant compte des règles des langages contrôlés et des règles d'écriture dont il a été question à la section 4.3.1.2, des caractéristiques qui, selon plusieurs auteurs, favorisent ou non l'automatisation du processus de traduction (voir le chapitre 2) et, bien sûr, des outils à notre disposition (voir la section 4.4).

- Nombre de mots par phrase
- Degré de lisibilité (100=très facile)
- Niveau scolaire (1<sup>re</sup> à la 12<sup>e</sup> année)
- Voix passives (% de verbes conjugués)
- Nombre d'occurrences – pron. pers. *them, they, it*<sup>40</sup>
- Nombre d'occurrences – pron. poss. *mine, yours, hers, his, its, ours, theirs*

---

<sup>40</sup> Les pronoms personnels qui nous intéressent ici sont ceux qui ont plusieurs équivalents possibles (masculin/féminin, p. ex.) en français (p. ex., *they = ils ou elles*).

- Nombre d'occurrences – pron. rel. *who, whom, which, that, whose*
- Nombre de mots en *-ing*<sup>41</sup>
- Nombre de contractions – *can't, don't, won't, they're, that's*
- Nombre de conjonctions de coordination – *and, but, or, nor*
- Rapport entre le nombre de mots différents et le nombre total de mots (type/token ratio)
- TermBase – % de mots remplacés
- TermBase – Nombre de phrases avec 100 % remplacement
- TransCorpora – % de mots trouvés
- TransCorpora – Nombre de phrases avec 75-99 % remplacement
- TransCorpora – Nombre de phrases avec 100 % remplacement

Il convient de noter que, comme les problèmes de traduction ne sont pas les mêmes pour toutes les langues, l'ensemble de cette liste ne s'applique qu'aux textes traduits de l'anglais vers le français.

---

<sup>41</sup> Nous savons que tous les mots en *-ing* ne sont pas nécessairement des participes présents (*bring, thing, sing, wing, etc.*). Nous croyons toutefois que la fréquence de ces mots est assez négligeable pour ne pas fausser les données. Si nous avions voulu comptabiliser seulement les participes présents, il aurait fallu déterminer la catégorie grammaticale de chaque mot de chaque texte, et les logiciels qui peuvent servir à cette fin ne sont pas infaillibles. Quoi qu'il en soit, il n'est pas réaliste d'imposer cette étape supplémentaire aux répartiteurs. On peut obtenir une estimation raisonnable du nombre de participes présents en utilisant WordSmith pour chercher les mots en *-ing*.

#### 4.6 Analyse des données

Pour chaque élément sur la liste, nous avons inscrit les données pour chaque type de textes. Ces données sont résumées au tableau I (pour les données relatives à chaque texte, voir l'annexe B) et sont analysées plus en détail un peu plus loin. Il convient de noter que, pour faciliter la comparaison entre les différents types de textes (qui étaient de différentes longueurs), nous avons pondéré les données en affichant soit des moyennes, soit le nombre d'occurrences pour 1000 mots.

**Tableau I – Données relatives aux DA, aux EQ et aux CP**

	DA <sup>42</sup>	EQ <sup>43</sup>	CP <sup>44</sup>
Mots par phrase (moyenne)	25,6	12	21,9
Degré de lisibilité (100=très facile) (moyenne)	38,9	5,1	33,0
Niveau scolaire (moyenne)	11,8	12,0	12,0
Voix passives (% de verbes conjugués)	28	1	17
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	7 (7,2)	0 (1,8)	4 (4,3)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/1000 mots)	3 (2,9)	0 (1,3)	6 (7,2)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/1000 mots)	31 (34,3)	0 (0,5)	8 (10,7)
Nombre de mots en <i>-ing</i> (occurrences/1000 mots)	*18 (19,4)	13 (49,6)	27 (33,5)
Nombre de contractions – <i>can't, don't, won't, they're, that's</i> (occurrences/1000 mots)	0 (0)	0 (0)	1 (1)
Nombre de conjonctions de coordination – <i>and, but, or, nor</i> (occurrences/1000 mots)	35 (37,4)	21 (88)	16 (19,3)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	34,95	56,92	49,42
TermBase – % de mots remplacés	23	15	21
TermBase – Nombre de phrases avec 100 % remplacement (%)	10 (14)	4 (13)	4 (7)
TransCorpora – % de mots trouvés	76	75	61

<sup>42</sup> Décisions d'appel

<sup>43</sup> Énoncés de qualités

<sup>44</sup> Coupures de presse

TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)	13 (19)	3 (11)	0 (0)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	25 (36)	11 (34)	12 (22)

\*en excluant les occurrences du nom « hearing » (dans le sens juridique de « audience »), qui revient souvent dans ces textes

À titre de rappel, nous avons aussi résumé les principales constatations des sections 4.1 et 4.2 dans le tableau II, ci-dessous.

**Tableau II – Résumé des résultats**

	« Meilleure » méthode	Méthode de second rang	Méthode la moins efficace	Observations
DA	TA (913,83 mots/heure)	MT (817,88 mots/heure)	TH (723,63 mots/heure)	La TA l'a nettement emporté sur les autres méthodes
EQ	MT (490,82 mots/heure)	TA (475,90 mots/heure)	TH (289,15 mots/heure)	Les résultats de la MT et de la TA étaient très proches
CP	TA (485,49 mots/heure)	MT (465,54 mots/heure)	TH (461,96 mots/heure)	Les résultats des 3 méthodes étaient proches

Comme le but de cet exercice est d'établir des lignes directrices qui tiennent compte des caractéristiques qui font en sorte qu'un texte se prête bien à une méthode de traduction donnée, nous avons analysé chacun des types de textes afin de cerner leurs caractéristiques.

#### 4.6.1 Analyse des décisions d'appel (DA)

La TA l'a nettement emporté dans le cas des DA. Nous avons donc commencé par l'analyse des caractéristiques de ces textes afin de rédiger des lignes directrices pour cerner les textes qui semblent mieux se prêter à la traduction à l'aide de la TA. Certaines des caractéristiques vont dans le sens des points soulevés dans la documentation, d'autres non.

Parmi les trois types de textes examinés, les DA affichent le plus haut degré de lisibilité, ce qui semble logique compte tenu du fait que les logiciels de TA ont tendance à donner de meilleurs résultats si les textes sont clairs et sans ambiguïté. Dans le même ordre d'idées, le fait que les DA comptent un grand nombre de pronoms relatifs, qui aident aussi à désambiguïser un texte (p. ex. en éliminant les ellipses), semble aussi faire en sorte qu'ils se prêtent bien à la TA. Dans la même veine, les DA comptent aussi le moins de mots en *-ing*. Puisqu'il est reconnu que ces constructions donnent du fil à retordre aux logiciels de TA, il semble logique qu'un texte contenant peu de mots en *-ing* se prête bien à la TA. Finalement, les DA affichent le type/token ratio le plus bas, ce qui veut dire que leur contenu lexical est le plus restreint (c.-à-d. que la terminologie est uniforme plutôt que variée).

Certaines de nos constatations semblent toutefois moins logiques. Par exemple, parmi les trois types de textes, les DA affichent le plus haut nombre moyen de mots par phrase (25,6). Tel qu'il a été mentionné précédemment, les logiciels de TA ont tendance à donner de meilleurs résultats lorsque les phrases sont courtes, et on recommande souvent de limiter la longueur des phrases à environ 20 mots. Bien que les DA contiennent les

phrases les plus longues, elles ne sont toutefois pas exceptionnellement longues; il semble donc que ces textes se prêtent tout de même bien à la TA.

Comme la résolution des référents des pronoms cause souvent des difficultés aux logiciels de TA, il est aussi surprenant que, de tous les types de textes, les DA contiennent le plus grand nombre de pronoms personnels. Il convient cependant de noter que, dans l'ensemble, le nombre de pronoms contenu dans chacun des types de textes est très faible – moins de une occurrence pour 100 mots, et ce, pour tous les types de textes. Ainsi, le fait qu'il y ait relativement peu de ce type de pronoms dans les DA donne à penser qu'elles semblent tout de même bien se prêter à la TA.

Le résultat le plus surprenant a été que les DA contiennent une proportion beaucoup plus importante de constructions passives que les deux autres types de textes. Comme les constructions passives peuvent causer des difficultés aux logiciels de TA, nous ne nous attendions pas à ce résultat. Tel qu'il est indiqué à la section 4.3.1.1.2, parfois le transfert linguistique de certaines constructions passives peut se faire facilement, mais pas toujours. Il est possible que les constructions passives contenues dans ces textes soient de celles qui sont faciles à traduire; toutefois, il faudrait procéder à une analyse plus approfondie pour confirmer cette hypothèse. Quoi qu'il en soit, puisque le nombre de constructions passives n'est que l'une des caractéristiques examinées, le fait que les DA contiennent un grand nombre de constructions passives n'empêche pas ce type de textes de bien se prêter à la TA dans l'ensemble.

#### 4.6.2 Analyse des énoncés de qualités (EQ)

La MT a été la méthode de traduction la plus efficace pour les EQ, bien que celle-ci ait été suivie d'assez près par la TA. Nous analyserons donc les EQ en vue de cerner les caractéristiques qui font en sorte qu'un texte se prête à la traduction à l'aide d'une MT.

Les EQ comptent le moins de mots par phrase (12 en moyenne). Comme les MT « recyclent » des textes déjà traduits, une phrase courte est plus susceptible d'être répétée qu'une phrase longue. Plus une phrase est longue, plus il y a de risques d'introduire des variantes et ainsi de réduire les chances de trouver des correspondances.

Les EQ contiennent aussi le moins de pronoms personnels et possessifs. Tel qu'il est noté à la section 4.3.2, il est plus probable de trouver une correspondance utile dans une MT pour une phrase qui ne contient pas de pronoms.

Nous nous rappellerons que le fonctionnement des MT repose strictement sur la reconnaissance de chaînes de caractères (voir la section 1.3). Ainsi, des facteurs comme la lisibilité, le recours aux pronoms relatifs, ainsi que le recours limité aux mots en *-ing* et à la voix passive, lesquels peuvent aider les logiciels de TA à désambiguïser des textes, n'ont pas vraiment d'incidence sur les MT. Parmi les trois types de textes examinés ici, les EQ affichent le plus petit nombre de pronoms relatifs et le degré de lisibilité le plus faible, de même que le plus grand nombre de mots en *-ing*. Toutefois, puisque ces facteurs n'ont pas d'incidence sur le fonctionnement des MT, on peut considérer que ces textes se prêtent quand même bien aux MT.

Les EQ affichent aussi de hauts taux de correspondances dans TermBase et TransCorpora (MultiTrans). Bien qu'il s'agisse des taux de second rang et non des plus hauts taux, ces données ne sont dans plusieurs cas qu'à un ou deux points de pourcentage du meilleur taux enregistré.

La constatation qui est probablement la plus surprenante dans cette analyse est le fait que le type/token ratio des EQ est le plus élevé parmi les trois types de textes. Un type/token ratio élevé indique que le vocabulaire contenu dans le texte est varié plutôt que stable, et on s'attendrait à ce qu'un texte dont le vocabulaire est stable se prête mieux aux MT. Ce ratio élevé s'explique peut-être par le fait que, tel qu'indiqué à la section 3.2.1, nous avons choisi 30 énoncés représentant dix groupes professionnels importants œuvrant au SCC. Idéalement, dans une situation « réelle », il serait probablement préférable de créer une base de données différente pour chacun de ces dix groupes. Le contenu terminologique de chaque base de données serait ainsi plus normalisé. Pourtant, malgré le fait que le type/token ratio est élevé, il ne s'agit que de l'une des nombreuses caractéristiques examinées, et cette surprise n'empêche pas les EQ de bien se prêter aux MT dans l'ensemble.

#### **4.6.3 Analyse des coupures de presse (CP)**

Lorsqu'on examine le nombre de mots traduits à l'heure à l'aide de chacune des trois méthodes, on constate qu'il n'y a pas de gagnant évident pour les CP. Les trois méthodes ont donné des résultats très similaires; moins de 25 mots à l'heure séparent la méthode la plus efficace (TA) et la méthode la moins efficace (TH).

Dans les analyses précédentes (des DA, qui semblent très bien se prêter à la TA, et des EQ, qui semblent surtout se prêter aux MT), nous avons constaté que parfois, il est avantageux d'obtenir des scores très élevés dans certaines catégories (p. ex. le plus haut degré de lisibilité), tandis que d'autres fois, il est préférable d'obtenir des scores très faibles dans d'autres catégories (p. ex. le plus petit nombre de mots par phrase). Dans le cas des CP toutefois, on remarque que les scores ne sont ni les plus élevés ni les plus faibles dans presque toutes les catégories. Les scores des CP ont plutôt tendance à se situer entre les deux, ce qui semble indiquer qu'il n'est pas vraiment avantageux de recourir à la technologie pour traduire ces textes.

La TH semble donc concurrentielle pour les coupures de presses, surtout si on considère qu'il faut du temps pour construire des corpus et/ou des dictionnaires pour travailler avec les systèmes de MT ou de TA.

#### **4.6.4 Autres observations**

Un certain nombre d'éléments qui figuraient sur notre liste initiale se sont révélés inutiles; nous les avons donc supprimés de notre liste révisée. Par exemple, il n'y a pas de différence significative entre le niveau scolaire attribué à chaque type de textes. Ce n'est pas vraiment surprenant puisque ces textes sont tous rédigés à l'intention d'adultes. Ce critère a donc été supprimé de la liste parce qu'il ne nous aide pas à déterminer quelle méthode de traduction adopter pour chaque type de textes.

Dans le même ordre d'idées, il n'y a aucune différence significative entre le nombre de contractions figurant dans chaque type de textes – une seule contraction enregistrée dans

les 27 textes. Ce critère a donc été supprimé de la liste parce qu'il ne nous aide pas à déterminer quelle méthode de traduction adopter pour chaque type de textes.

Finalement, les données relatives au nombre de conjonctions sont très bruitées.

Autrement dit, bien que les conjonctions puissent parfois faire en sorte qu'il soit difficile d'interpréter les rapports de coordination (voir la section 4.3.1.1.3), les conjonctions peuvent aussi être utilisées dans de nombreux autres contextes qui ne présentent aucune ambiguïté. Parce qu'il faudrait beaucoup de temps pour examiner chaque occurrence d'une conjonction pour voir si oui ou non elle pourrait causer un problème d'ambiguïté, nous avons décidé d'éliminer cet élément de notre liste révisée. Il ne serait d'ailleurs pas réaliste de s'attendre à ce que les répartiteurs se prêtent à cet exercice. Les éléments qui feront partie de nos lignes directrices définitives doivent pouvoir être calculés automatiquement.

#### **4.7 Élaboration des lignes directrices**

En nous fondant sur notre analyse à la section 4.6, nous avons déterminé les caractéristiques qui semblaient les plus pertinentes pour chacune des trois méthodes de traduction; ces caractéristiques font donc partie de notre liste initiale de lignes directrices. Pour l'instant, les valeurs des différents éléments des lignes directrices ont été laissées en blanc (représentées par X), tout comme la valeur qui correspond au nombre de critères à remplir pour prédire que telle ou telle méthode est la meilleure à adopter. Les essais pilotes décrits au chapitre 5 nous aideront peut-être à déterminer les meilleures valeurs possible pour chaque élément des lignes directrices.

Il est important de comprendre que les lignes directrices ne sont pas absolues. Un texte donné peut ne pas répondre à tous les critères établis. Toutefois, s'il remplit la majorité des critères relatifs à une méthode, il se prête probablement bien à cette méthode. Nous tenterons de vérifier la validité de ces lignes directrices au chapitre 5.

#### **4.7.1 Lignes directrices pour la TA**

- Nombre de mots par phrase = X ou moins
- Degré de lisibilité = au moins X
- % de verbes conjugués à la voix passive = moins de X
- Nombre de pronoms personnels = moins de X pour 1000 mots
- Nombre de pronoms possessifs = moins de X pour 1000 mots
- Nombre de pronoms relatifs = au moins X pour 1000 mots
- Nombre de mots en *-ing* = moins de X pour 1000 mots
- Type/token ratio = moins de X

Si le texte répond à au moins X de ces critères, il se prête probablement bien à la TA.

#### **4.7.2 Lignes directrices pour les MT**

- Nombre de mots par phrase = X ou moins
- Nombre de pronoms personnels = moins de X pour 1000 mots
- Nombre de pronoms possessifs = moins de X pour 1000 mots
- Type/token ratio = moins de X
- TermBase - % de mots remplacés = au moins X
- TermBase - % de phrases avec 100 % remplacement = au moins X

- TransCorpora - % de mots trouvés = au moins X
- TransCorpora - % de phrases avec 75-99 % remplacement = au moins X
- TransCorpora - % de phrases avec 100 % remplacement = au moins X

Si le texte répond à au moins X de ces critères, il se prête probablement bien aux MT.

#### 4.7.3 Lignes directrices pour la TH

- Nombre de mots par phrase = au moins X
- Degré de lisibilité = moins de X
- % de verbes conjugués à la voix passive = au moins X
- Nombre de pronoms personnels = au moins X pour 1000 mots
- Nombre de pronoms possessifs = au moins X pour 1000 mots
- Nombre de pronoms relatifs = moins de X pour 1000 mots
- Nombre de mots en *-ing* = au moins X pour 1000 mots
- Type/token ratio = au moins X
- TermBase - % de mots remplacés = moins de X
- TermBase - % de phrases avec 100 % remplacement = moins de X
- TransCorpora - % de mots trouvés = moins de X
- TransCorpora - % de phrases avec 75-99 % remplacement = moins de X
- TransCorpora - % de phrases avec 100 % remplacement = moins de X

Si le texte répond à au moins X de ces critères, il se prête probablement mieux à la TH.

Comme les caractéristiques qui font en sorte qu'un texte se prêterait probablement mieux à la TH sont opposées à celles qui font en sorte qu'un texte se prêterait à la TA ou aux

MT, il serait peut-être plus facile de sauter l'analyse des critères relatifs à la TH et de dire simplement que, par déduction, si le texte ne se prête pas à la TA ni aux MT, il se prête alors probablement mieux à la TH. Les deux scénarios seront examinés à l'étape des essais pilotes.

Cette deuxième partie traitait du travail de préparation en vue de l'élaboration des lignes directrices qui permettront de déterminer la meilleure méthode de traduction à adopter, selon le texte. La troisième partie portera sur la mise à l'essai et l'ajustement de cette approche, et le dernier chapitre contient une évaluation de l'approche élaborée de même que des propositions de recherche pour l'avenir.

## **PARTIE III – Essais et évaluation**

## **Chapitre 5 – Mise à l’essai des lignes directrices**

### **5.1 Sélection et traduction des textes**

Pour pouvoir mettre les lignes directrices à l’essai, nous avons besoin de nouveaux types de textes<sup>45</sup>. Nous en avons choisi deux (qui sont aussi traduits régulièrement au Service Criminologie), soit des comptes rendus de réunions hebdomadaires d’un comité (CR) et des résumés de l’état de plaintes relatives aux droits de la personne (PDP). Comme les seconds textes sont relativement longs et que nous en avons en moins grand nombre à notre disposition (explicable du fait qu’ils ne sont rédigés que quatre fois par année), nous avons séparé un texte en six extraits de longueur semblable. Un exemple de chacun de ces types de textes est présenté à l’annexe C. Cette fois, nous avons traduit deux exemples de chacun de ces types de textes à l’aide de chacune des trois méthodes (TH, TA, MT) (donc 12 textes en tout). Encore une fois, il a d’abord fallu créer un corpus et une banque de données terminologiques dans MultiTrans de même qu’un dictionnaire équivalent dans Systran, et ce, pour chaque type de textes. Toutes ces tâches ont été réalisées selon la méthode décrite pour les trois premiers types de textes (voir le chapitre 3).

### **5.2 Essais pilotes - Attribution de valeurs aux éléments des lignes directrices et présentation de celles-ci sous forme de listes de vérification**

Bien sûr, nous avons décidé des valeurs en tenant compte des résultats de l’analyse des données relatives aux trois premiers types de textes, mais il reste qu’il s’agit tout de

---

<sup>45</sup> Les données servant à l’élaboration doivent toujours être différentes des données servant aux essais. Pour voir si les lignes directrices étaient vraiment efficaces, il fallait les mettre à l’essai sur des textes différents de ceux utilisés à l’étape de l’élaboration.

même d'un exercice aléatoire dans une certaine mesure. Par exemple, pour attribuer les valeurs aux éléments des lignes directrices pour la TA, nous avons tenu compte du fait que cette méthode avait donné les meilleurs résultats pour les décisions d'appel (DA) et avons à partir de là attribué aux éléments pertinents deux séries de valeurs très proches de celles des DA, en les arrondissant au multiple de 5 le plus près (voir le tableau I).

Afin d'augmenter les chances d'efficacité de la liste de vérification qui sera utilisée pour la mise à l'essai « officielle » avec volontaires, nous avons décidé de procéder d'abord à deux essais pilotes, avec les deux séries de valeurs présentées selon deux approches différentes. Par souci d'économie d'espace, les listes de vérification sont présentées avec les résultats de l'analyse des données relatives aux deux derniers types de textes (pour les données relatives à chaque texte, voir l'annexe D). Il convient de noter que les valeurs ont été déterminées avant que nous prenions connaissance de ces données, que voici :

**Tableau III – Données relatives aux CR et aux PDP**

	PDP	CR
Mots par phrase	17,7	15,4
Degré de lisibilité (100=très facile)	39,1	43,9
Voix passives (%/verbes conjugués)	13	15
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	2 (3,3)	5 (6,8)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/1000 mots)	12 (16,6)	0 (0,6)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/1000 mots)	27 (37,4)	22 (29,1)
Nombre de mots en <i>-ing</i> (occurrences/1000 mots)	17 (23,5)	16 (21,6)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	31,11	41,61
TermBase – % de mots remplacés	36	23
TermBase – Nombre de phrases avec 100 % remplacement (%)	19 (30)	12 (16)

TransCorpora – % de mots trouvés	94	76
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)	11 (17)	5 (6)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	34 (55)	25 (35)

### 5.2.1 Premier essai pilote

PDP	TA		MT		TH	
Mots par phrase (moyenne)	≤ 25	✓	≤ 20	✓	≥ 20	
Degré de lisibilité (100=très facile) (moyenne)	≥ 35	✓			< 35	
Voix passives (% de verbes conjugués)	< 20	✓			≥ 15	
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	< 10	✓	< 30	✓	≥ 30	
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/1000 mots)	< 50	✓	< 30	✓	≥ 50	
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/1000 mots)	≥ 20	✓			< 20	
Nombre de mots en <i>-ing</i> (occurrences/1000 mots)	< 30	✓			≥ 30	
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	< 45	✓	< 45	✓	≥ 45	
TermBase – % de mots remplacés			≥ 15	✓	< 15	
TermBase – Nombre de phrases avec 100 % remplacement (%)			≥ 10	✓	< 10	
TransCorpora – % de mots trouvés			≥ 65	✓	< 65	
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)			≥ 10	✓	< 10	
TransCorpora – Nombre de phrases avec 100 % remplacement (%)			≥ 30	✓	< 30	

Si le texte répond à au moins 5 des critères de la TA, il se prête probablement bien à la TA.

Résultat : le texte répond à 8 critères sur 8.

Si le texte répond à au moins 6 des critères des MT, il se prête probablement bien aux MT.

Résultat : le texte répond à 9 critères sur 9.

Si le texte répond à au moins 8 des critères de la TH, il se prête probablement mieux à la TH.

Résultat : le texte répond à 0 critère sur 12.

Prédiction : La MT et la TA donneront toutes deux de bons résultats.

CR	TA		MT		TH	
Mots par phrase (moyenne)	≤ 25	✓	≤ 20	✓	≥ 20	
Degré de lisibilité (100=très facile) (moyenne)	≥ 35	✓			< 35	
Voix passives (% de verbes conjugués)	< 20	✓			≥ 15	✓
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	< 10	✓	< 30	✓	≥ 30	
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/1000 mots)	< 50	✓	< 30	✓	≥ 50	
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/1000 mots)	≥ 20	✓			< 20	
Nombre de mots en <i>-ing</i> (occurrences/1000 mots)	< 30	✓			≥ 30	
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	< 45	✓	< 45	✓	≥ 45	
TermBase – % de mots remplacés			≥ 15	✓	< 15	
TermBase – Nombre de phrases avec 100 % remplacement (%)			≥ 10	✓	< 10	
TransCorpora – % de mots trouvés			≥ 65	✓	< 65	
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)			≥ 10	✓	< 10	✓
TransCorpora – Nombre de phrases avec 100 % remplacement (%)			≥ 30	✓	< 30	

Si le texte répond à au moins 5 des critères de la TA, il se prête probablement bien à la TA.

Résultat : le texte répond à 8 critères sur 8.

Si le texte répond à au moins 6 des critères des MT, il se prête probablement bien aux MT.

Le texte répond à 9 critères sur 9.

Si le texte répond à au moins 8 des critères de la TH, il se prête probablement mieux à la TH.

Le texte répond à 2 critères sur 12.

Prédiction : La MT et la TA donneront toutes deux de bons résultats.

## 5.2.2 Deuxième essai pilote

<i>PDP – Liste de vérification pour la traduction automatique</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il moins de 20 mots par phrase?	✓	
Le degré de lisibilité est-il supérieur à 40?		✓*
Le nombre de verbes conjugués à la voix passive est-il inférieur à 20 %?	✓	
Le texte compte-t-il moins de 5 pronoms personnels pour 1000 mots?	✓	
Le texte compte-t-il moins de 5 pronoms possessifs pour 1000 mots?		✓
Le texte compte-t-il plus de 30 pronoms relatifs pour 1000 mots?	✓	
Le texte compte-t-il moins de 20 mots en <i>-ing</i> pour 1000 mots		✓*
Le type/token ratio est-il inférieur à 35 %?	✓	

\* nombre/taux proche du point de démarcation

Si vous avez répondu OUI à au moins 6 de ces questions, votre texte se prête probablement bien à la TA.

Résultat : 5 OUI évidents, 2 très proche de OUI, 1 NON

Ce type de textes semble donc se prêter assez bien à la traduction automatique.

<i>PDP - Liste de vérification pour les mémoires de traduction</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il moins de 20 mots par phrase?	✓	
Le texte compte-t-il moins de 10 pronoms personnels pour 1000 mots?	✓	
Le texte compte-t-il moins de 10 pronoms possessifs pour 1000 mots?		✓
Le type/token ratio est-il inférieur à 40 %?	✓	
Le taux de mots remplacés (TermBase) est-il supérieur à 25 %?	✓	
Le taux de phrases avec 100 % remplacement (TermBase) est-il supérieur à 25 %?	✓	
Le taux de mots trouvés dans TransCorpora est-il supérieur à 75 %?	✓+	
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) a été trouvée dans TransCorpora est-il supérieur à 20 %?		✓*
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans TransCorpora est-il supérieur à 30 %?	✓+	

\*nombre/taux proche du point de démarcation

+valeur beaucoup plus élevée que le point de démarcation

Si vous avez répondu OUI à au moins 6 de ces questions, votre texte se prête probablement bien aux MT.

Résultat : 7 OUI évidents, y compris 2+, 1 proche de OUI, 1 NON

Ce type de textes semble donc se prêter très bien aux mémoires de traduction.

Prédiction : La mémoire de traduction donnera les meilleurs résultats, mais la traduction automatique sera aussi une option viable.

<b>CR – Liste de vérification pour la traduction automatique</b>	<b>OUI</b>	<b>NON</b>
Le texte compte-t-il moins de 20 mots par phrase?	✓	
Le degré de lisibilité est-il supérieur à 40?	✓	
Le nombre de verbes conjugués à la voix passive est-il inférieur à 20 %?	✓	
Le texte compte-t-il moins de 5 pronoms personnels pour 1000 mots?		✓*
Le texte compte-t-il moins de 5 pronoms possessifs pour 1000 mots?	✓	
Le texte compte-t-il plus de 30 pronoms relatifs pour 1000 mots?		✓*
Le texte compte-t-il moins de 20 mots en <i>-ing</i> pour 1000 mots?		✓*
Le type/token ratio est-il inférieur à 35 %?		✓

\* nombre/taux proche du point de démarcation

Si vous avez répondu OUI à au moins 6 de ces questions, votre texte se prête probablement bien à la TA.

Résultat : 4 OUI évidents, 3 proche de OUI, 1 NON

Ce type de textes semble donc se prêter assez bien à la traduction automatique.

<b>CR – Liste de vérification pour les mémoires de traduction</b>	<b>OUI</b>	<b>NON</b>
Le texte compte-t-il moins de 20 mots par phrase?	✓	
Le texte compte-t-il moins de 10 pronoms personnels pour 1000 mots?	✓	
Le texte compte-t-il moins de 10 pronoms possessifs pour 1000 mots?	✓+	
Le type/token ratio est-il inférieur à 40 %?		✓*
Le taux de mots remplacés (TermBase) est-il supérieur à 25 %?		✓*
Le taux de phrases avec 100 % remplacement (TermBase) est-il supérieur à 25 %?		✓
Le taux de mots trouvés dans TransCorpora est-il supérieur à 75 %?	✓	
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) a été trouvée dans TransCorpora est-il supérieur à 20 %?		✓
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans TransCorpora est-il supérieur à 30 %?	✓	

\* nombre/taux proche du point de démarcation

+ valeur beaucoup plus élevée que le point de démarcation

Si vous avez répondu OUI à au moins 6 de ces questions, votre texte se prête probablement bien aux MT.

Résultat : 5 OUI évidents, y compris 1+, 2 proches de OUI, 2 NON

Ce type de textes semble donc se prêter assez bien aux mémoires de traduction.

Prédiction : La traduction automatique donnera les meilleurs résultats, mais la mémoire de traduction sera aussi une option viable.

### 5.2.3 Vérification des prédictions

Pour vérifier la validité de ces valeurs et des prédictions, voici la comparaison des temps de traduction de ces types de textes, et ce, pour chacune des méthodes :

Comptes rendus

Méthode	N <sup>bre</sup> de mots	Temps requis	N <sup>bre</sup> de mots/heure	Moyenne de mots/heure
TH	696	115 min.	363,13	410,11
TH	678	89 min.	457,08	
TA	579	53 min.	655,47	632,50
TA	640	63 min.	609,52	
MT	1092	94 min.	697,02	756,85
MT	735	54 min.	816,67	

Plaintes relatives aux droits de la personne

Méthode	N <sup>bre</sup> de mots	Temps requis	N <sup>bre</sup> de mots/heure	Moyenne de mots/heure
TH	662	47 min.	845,11	868,36
TH	743	50 min.	891,60	
TA	728	40 min.	1092,00	1202,25
TA	700	32 min.	1312,50	
MT	695	26 min.	1603,84	1490,05
MT	734	32 min.	1376,25	

#### 5.2.3.1 Premier essai pilote

Pour les PDP, compte tenu des valeurs attribuées aux éléments des lignes directrices, les résultats indiqués sur la liste de vérification ont permis de prédire que la MT et la TA donneraient toutes deux de bons résultats. Si on examine les résultats ci-dessus, on constate que la MT et la TA ont en effet donné toutes les deux de bons résultats. On remarque cependant que la MT a donné de meilleurs résultats que la TA; il aurait donc été bien que les résultats de la liste de vérification permettent de prédire que l'une des

deux méthodes serait meilleure que l'autre. Nous tenterons donc d'ajuster les valeurs attribuées à certains éléments des lignes directrices afin que cette différence puisse être prévisible.

Quant aux CR, compte tenu des valeurs attribuées aux éléments des lignes directrices, les résultats de la liste de vérification ont permis de déterminer que la MT et la TA donneraient toutes deux de bons résultats. Si on examine les résultats ci-dessus, on remarque que la MT a donné de meilleurs résultats que la TA. Nous tenterons donc d'ajuster les valeurs attribuées à certains éléments des lignes directrices en conséquence.

Pour ce qui est du format de la liste utilisée dans le premier essai pilote, on constate qu'il semble un peu trop « encombrant » de présenter les données dans un seul tableau. Aussi, les symboles mathématiques font en sorte que le tableau n'est pas très facile d'utilisation.

### **5.2.3.2 Deuxième essai pilote**

Pour les PDP, compte tenu des valeurs attribuées aux éléments des lignes directrices dans le cadre du deuxième essai pilote, les résultats de la liste de vérification ont permis de déterminer que la mémoire de traduction donnerait les meilleurs résultats, mais que la traduction automatique serait aussi une option viable. Ces prédictions étaient exactes.

Quant aux CR, compte tenu des valeurs attribuées aux éléments des lignes directrices dans le cadre du deuxième essai pilote, les résultats de la liste de vérification ont permis de déterminer que la traduction automatique donnerait les meilleurs résultats, mais que la mémoire de traduction serait aussi une option viable. Dans la réalité, c'était plutôt le

contraire. Afin d'améliorer les prédictions, nous tenterons donc d'ajuster les valeurs attribuées à certains éléments des lignes directrices, en tentant bien sûr de préserver les valeurs qui ont contribué à la prédiction exacte dans le cas des PDP.

Pour ce qui est de la présentation des lignes directrices dans le second essai pilote, il semble en effet plus simple de faire deux listes (une pour la TA et une pour les MT) et de ne pas avoir de liste pour la TH, si on prend soin de noter que si un texte ne se prête pas à la TA ni aux MT, alors la TH est probablement la meilleure méthode. De plus, il semble plus convivial de présenter les lignes directrices sous forme de questions simples que de recourir à des symboles mathématiques. L'utilisation de symboles comme \* et + complique d'ailleurs peut-être aussi un peu trop les choses.

### **5.3 Détermination des valeurs et du format définitifs de la liste de vérification**

En comparant les résultats des deux essais pilotes pour déterminer quelles valeurs ont produit les prédictions les plus justes, nous avons tenté d'ajuster les valeurs attribuées aux éléments des lignes directrices afin de rendre les prédictions encore plus justes. Il convient de répéter qu'il s'agit d'un exercice plus ou moins aléatoire. Comme de nombreuses formes de recherche scientifique ayant recours à une approche par essais et erreurs, nous avons eu recours à un processus itératif. Idéalement, il aurait été souhaitable de procéder à d'autres essais afin d'ajuster davantage les valeurs; toutefois, dans le contexte de la présente thèse de maîtrise, où le temps à notre disposition et la portée du projet sont limités, nous nous en tiendrons aux deux essais pilotes décrits ci-dessus.

Pour ce qui est du format de la liste de vérification, nous avons opté pour celui du deuxième essai pilote, car il nous semblait plus convivial. Nous espérons ainsi simplifier la tâche de nos trois volontaires et des utilisateurs à l'avenir et ainsi, optimiser la validité des résultats. Voici donc la liste de vérification définitive, telle que présentée aux volontaires dans le cadre de l'essai « officiel ». Nous avons opté pour le format utilisé dans le cadre du deuxième essai pilote, lequel présente des listes séparées pour la TA et les MT et pose des questions simples. Nous avons cependant décidé de laisser tomber l'utilisation des symboles (\*, +, etc.) afin de simplifier les choses autant que possible.

<i>Liste de vérification pour la traduction automatique</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il <b>25 mots par phrase ou moins</b> ?		
Le <i>degré de lisibilité</i> est-il <b>égal ou supérieur à 35</b> ?		
Le nombre de <i>verbes conjugués à la voix passive</i> est-il <b>inférieur à 20 %</b> ?		
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 5 pronoms possessifs</b> pour 1000 mots?		
Le texte compte-t-il <b>au moins 20 pronoms relatifs</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 20 mots en -ing</b> pour 1000 mots?		
Le <i>type/token ratio</i> est-il <b>inférieur à 40 %</b> ?		

Si vous avez répondu OUI à au moins 6 de ces questions, votre texte se prête probablement bien à la TA.

<i>Liste de vérification pour les mémoires de traduction</i>	<b>OUI</b>	<b>NON</b>
Le texte compte-t-il <b>20 mots par phrase ou moins</b> ?		
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 20 pronoms possessifs</b> pour 1000 mots?		
Le <i>type/token ratio</i> est-il <b>inférieur à 45 %</b> ?		
Le taux de <i>mots remplacés (TermBase)</i> est-il <b>égal ou supérieur à 20 %</b> ?		
Le taux de <i>phrases avec 100 % remplacement (TermBase)</i> est-il <b>égal ou supérieur à 15 %</b> ?		
Le taux de <i>mots trouvés dans TransCorpora</i> est-il <b>égal ou supérieur à 75 %</b> ?		
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) a été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 10 %</b> ?		
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 30 %</b> ?		

Si vous avez répondu OUI à au moins 6 de ces questions, votre texte se prête probablement bien aux MT.

Si vous avez répondu NON à la majorité des questions des deux listes, ce type de textes se prête probablement mieux à la traduction humaine (TH).

#### **5.4 Mise à l'essai « officielle »**

Aux fins de cette mise à l'essai, nous avons remis à trois de collègues (deux répartitrices en titre et une répartitrice suppléante) une trousse de cinq pages (voir l'annexe E) comprenant des instructions quant à la procédure à suivre, un tableau de données, deux listes de vérification (une pour chaque type de texte) et une fiche de rétroaction. *Grosso modo*, nous leur avons demandé de répondre aux questions des deux listes (TA et MT) en se référant au tableau de données (PDP et CR) et de prédire, en fonction de leurs réponses et des instructions données à la fin de chaque liste de questions, la meilleure approche pour chaque type de textes. Nous leur avons aussi demandé si elles trouvaient que les listes de questions et les instructions données étaient claires et faciles à consulter/suivre. Cette mise à l'essai visait non seulement à valider les valeurs définitives attribuées aux

éléments des listes de vérification, mais aussi à vérifier la facilité d'utilisation des listes de vérification.

#### 5.4.1 Résultats

Les répartitrices ont toutes les trois coché exactement les mêmes réponses aux quatre listes de questions (deux listes, une pour la TA, une pour les MT, par type de texte) :

<b><i>PDP - Liste de vérification pour la traduction automatique</i></b>	<b><i>OUI</i></b>	<b><i>NON</i></b>
Le texte compte-t-il <b>25 mots par phrase ou moins</b> ?	✓	
Le <i>degré de lisibilité</i> est-il <b>égal ou supérieur à 35</b> ?	✓	
Le nombre de <i>verbes conjugués à la voix passive</i> est-il <b>inférieur à 20 %</b> ?	✓	
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>moins de 5 pronoms possessifs</b> pour 1000 mots?		✓
Le texte compte-t-il <b>au moins 20 pronoms relatifs</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>moins de 20 mots en -ing</b> pour 1000 mots?		✓
Le <i>type/token ratio</i> est-il <b>inférieur à 40 %</b> ?	✓	

<b><i>PDP - Liste de vérification pour les mémoires de traduction</i></b>	<b><i>OUI</i></b>	<b><i>NON</i></b>
Le texte compte-t-il <b>20 mots par phrase ou moins</b> ?	✓	
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>moins de 20 pronoms possessifs</b> pour 1000 mots?	✓	
Le <i>type/token ratio</i> est-il <b>inférieur à 45 %</b> ?	✓	
Le taux de <i>mots remplacés (TermBase)</i> est-il <b>égal ou supérieur à 20 %</b> ?	✓	
Le taux de <i>phrases avec 100 % remplacement (TermBase)</i> est-il <b>égal ou supérieur à 15 %</b> ?	✓	
Le taux de <i>mots trouvés dans TransCorpora</i> est-il <b>égal ou supérieur à 75 %</b> ?	✓	
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 10 %</b> ?	✓	
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 30 %</b> ?	✓	

<b><i>CR - Liste de vérification pour la traduction automatique</i></b>	<b><i>OUI</i></b>	<b><i>NON</i></b>
Le texte compte-t-il <b>25 mots par phrase ou moins</b> ?	✓	
Le <i>degré de lisibilité</i> est-il <b>égal ou supérieur à 35</b> ?	✓	
Le nombre de <i>verbes conjugués à la voix passive</i> est-il <b>inférieur à 20 %</b> ?	✓	
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>moins de 5 pronoms possessifs</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>au moins 20 pronoms relatifs</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>moins de 20 mots en -ing</b> pour 1000 mots?		✓

Le <i>type/token ratio</i> est-il <b>inférieur à 40 %</b> ?		✓
---	--	---

<b>CR - Liste de vérification pour les mémoires de traduction</b>	<b>OUI</b>	<b>NON</b>
Le texte compte-t-il <b>20 mots par phrase ou moins</b> ?	✓	
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?	✓	
Le texte compte-t-il <b>moins de 20 pronoms possessifs</b> pour 1000 mots?	✓	
Le <i>type/token ratio</i> est-il <b>inférieur à 45 %</b> ?	✓	
Le taux de <i>mots remplacés (TermBase)</i> est-il <b>égal ou supérieur à 20 %</b> ?	✓	
Le taux de <i>phrases avec 100 % remplacement (TermBase)</i> est-il <b>égal ou supérieur à 15 %</b> ?	✓	
Le taux de <i>mots trouvés dans TransCorpora</i> est-il <b>égal ou supérieur à 75 %</b> ?	✓	
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 10 %</b> ?		✓
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 30 %</b> ?	✓	

Conformément aux instructions qui leur avaient été données, les trois répartitrices ont prédit que la MT donnerait les meilleurs résultats pour les deux types de textes, ce qui était effectivement le cas.

Ces résultats font deux choses : d'une part, ils valident les nouvelles valeurs attribuées aux éléments des listes de vérification et, d'autre part, ils prouvent que les listes de vérification et les instructions connexes sont faciles d'utilisation. Les commentaires des trois volontaires abondent d'ailleurs aussi dans ce sens (voir l'annexe F).

## **Chapitre 6 – Conclusion**

### **6.1 Commentaires généraux sur la conception et la mise à l'essai de l'approche proposée**

L'objectif de cette thèse était de proposer et de mettre à l'essai une approche qui permettrait de déterminer quelle méthode, entre la traduction automatique, les mémoires de traduction et la traduction humaine, était la meilleure pour un type de textes donné.

Nous avons effectivement réussi à concevoir et à mettre à l'essai une approche qui, selon les résultats des essais, permet de déterminer la meilleure méthode pour un type de textes donné et qui, de surcroît, est facile d'utilisation.

Indépendamment des résultats, il convient de noter que certains facteurs pourraient avoir influencé la conception de l'approche proposée et les résultats.

Aux étapes de la traduction, il y a peut-être eu un certain manque d'impartialité du fait que la personne qui a dépouillé les textes et qui a créé les bases de données terminologiques et les dictionnaires est celle qui a ensuite aussi traduit les textes. Cette personne connaissait déjà la terminologie relative à la plupart des types de textes utilisés, mais pas tous. Or, l'étape du dépouillement des textes et la création des bases de données terminologiques et des dictionnaires ont permis à cette personne de se familiariser avec la terminologie des textes en question, ce qui pourrait avoir indirectement influé sur les résultats pour la traduction humaine. On peut donc présumer que les résultats auraient été différents (différence encore plus importante entre les résultats de la TH et ceux de la TA et de la MT, p. ex.) si les textes avaient été traduits par un traducteur débutant dans le domaine.

Dans la même veine, puisque nous espérons que cette approche puisse servir au plus grand nombre de traducteurs possible, il aurait peut-être justement été préférable que des traducteurs de divers niveaux d'expérience (débutants à expérimentés) fassent les traductions et qu'on calcule la moyenne du temps requis par les divers traducteurs pour traduire les différents types de textes à l'aide des trois méthodes de traduction.

L'un des principaux désavantages de cette approche est qu'elle demande beaucoup de temps. Pour procéder à l'analyse des textes à traduire dans MultiTrans, il faut déjà avoir en place un corpus dans TransCorpora et une base de données terminologiques dans TermBase pour le type de textes donné. Aussi, pour analyser les textes à traduire et obtenir les données nécessaires pour répondre aux questions de la liste de vérification, il faut recourir à trois logiciels : MultiTrans (fonction d'analyse), WordSmith Tools (WordList) et Word (statistiques de lisibilité). Il convient de noter que, pour l'essai « officiel », les données ont été fournies aux volontaires. Dans une situation « réelle », les répartitrices auraient eu à obtenir ces données elles-mêmes, ce qui veut dire qu'elles auraient eu à apprendre à se servir de deux nouveaux logiciels, soit WordSmith Tools et MultiTrans.

Finalement, si nous avions eu plus de temps, il y a deux choses que nous aurions faites. Idéalement, nous aurions utilisé des types de textes différents pour l'essai « officiel », et non ceux qui avaient déjà été utilisés pour les essais pilotes. Aussi, comme nous l'avons mentionné à la section 5.3, nous avons eu recours à un processus itératif. Idéalement, il

aurait été souhaitable de procéder à plus d'essais afin d'ajuster davantage les valeurs; toutefois, dans le contexte de la présente thèse de maîtrise, nous nous sommes limitée à deux essais pilotes. Comme nous n'avons pas procédé à de nombreux ajustements et avons tout de même réussi à obtenir des résultats positifs, nous croyons que d'autres ajustements pourraient améliorer la liste de vérification encore davantage.

Dans l'ensemble, l'approche proposée semble être valide et permettre effectivement de déterminer la meilleure méthode pour un type de textes donné. Il faudrait cependant procéder à d'autres essais pour le confirmer.

## **6.2 Recommandations pour les recherches à venir**

Si des chercheurs menaient des études à plus grande échelle dans ce domaine, nous leur proposerions d'abord d'utiliser plus de types de textes, tant à l'étape de la conception qu'à celle des essais et de l'ajustement des valeurs, et d'incorporer d'autres outils ou méthodes de traduction, comme la prétraduction automatique (voir Bédard, 1992; Bédard, 2004), qui, à nos yeux, vaudrait la peine d'être explorée. « Le terme "prétraduction" désigne le remplacement global, dans un document à traduire, de mots simples ou de chaînes de mots provenant d'un dictionnaire bilingue » (Bédard, 2004, p. 30). Ce processus produit ainsi une traduction partielle du texte à traduire ou, si on préfère, un texte en langue de départ ponctué d'éléments en langue d'arrivée. D'un côté, si les éléments prétraduits sont valides, le traducteur gagnera fort probablement du temps. D'un autre côté, il se peut qu'un mot ou une chaîne de mots valable dans un contexte donné ne convienne pas dans d'autres contextes. Donc, si la prétraduction contient des

éléments qui ne conviennent pas, le traducteur doit réviser le texte et risque de perdre du temps. Il pourrait donc être utile, par exemple, de procéder à des essais comparatifs pour déterminer les circonstances dans lesquelles le recours à la prétraduction automatique vaut la peine.

En outre, la méthode utilisée pour concevoir cette approche pourrait être appliquée à d'autres combinaisons de langues compte tenu des caractéristiques linguistiques et terminologiques de ces langues. Aussi, la conception de l'approche et l'approche comme telle pourraient donner des résultats différents avec d'autres outils que Systran et MultiTrans. On pourrait ainsi vérifier si l'approche est valide pour tous les logiciels d'une même catégorie (TA ou MT) ou seulement pour les logiciels utilisés dans le cadre du projet. De plus, tel qu'il a été mentionné plus haut, l'étape de la conception pourrait être menée à bien avec des traducteurs de divers niveaux d'expérience. Il se pourrait que certains outils de traduction assistée par ordinateur soient plus utiles pour des traducteurs débutants que pour des traducteurs chevronnés, par exemple.

Tel qu'il a été démontré au chapitre 2, les chercheurs connaissent déjà certaines caractéristiques des textes qui se prêtent bien ou mal aux outils de traduction assistée par ordinateur. Peut-être que l'application de notre approche ou d'une version modifiée de celle-ci (critères linguistiques et terminologiques différents, p. ex.) à d'autres types de textes pourrait contribuer à approfondir leurs connaissances à cet égard.

Comme la technologie évolue rapidement, il serait souhaitable que notre approche ou toute autre approche connexe soit mise à jour compte tenu d'autres fonctions des logiciels de TA et de MT (fonction de traduction interactive de MultiTrans mentionnée à la section 1.3.2, p. ex.) et des nouvelles fonctions ajoutées à ceux-ci, de même que des nouveaux outils d'analyse. N'oublions pas que nous avons dû recourir à trois logiciels pour obtenir les données nécessaires pour répondre aux questions de la liste de vérification. Il serait bien si ce nombre était réduit à deux, voire à un seul.

### **6.3 Mot de la fin**

Avec la hausse de la demande et la pénurie de traducteurs que le domaine de la traduction connaît aujourd'hui, les traducteurs risquent de ne pas avoir le choix de se tourner vers la traduction assistée par ordinateur s'ils veulent augmenter leur production et répondre à cette demande, qui excède largement l'offre. Toutefois, il ne suffit pas de se procurer des outils, encore faut-il savoir, une fois ces logiciels en place, comment éviter les pertes de temps et choisir l'outil qui se prête le mieux à chaque type de textes. L'approche proposée dans la présente thèse devrait faciliter ce processus décisionnel.

## Bibliographie

AHRENBERG, Lars, et Magnus MERKEL. « On translation corpora and translation support tools : A project report », dans Karin Aijmer, Bengt Altenberg et Mats Johansson, *Languages in Contrast*, Lund (Suède), Lund University Press, 1996, p. 183-200.

ALLEN, Jeffrey. « Post-editing », dans Harold Somers, *Computers and Translation: A translator's guide*, Amsterdam, John Benjamins, 2003, p. 297-317.

ANOBILO, Michael. « Foreword », dans R.C. Sprung, *Translating into Success: Cutting-edge strategies for going multilingual in a global age*, Amsterdam, John Benjamins, 2000, p. vii.

ANSON, Chris M., et Robert A. SCHWEGLER. *The Longman Handbook for Writers and Readers*, New York, Addison-Wesley Educational Publishers Inc., 1997.

ARNOLD, Doug, Lorna BALKAN, Siety MEIJER, R. Lee HUMPHREYS, et Louisa SADLER. *Machine Translation: An Introductory Guide*, Oxford, Blackwell Publishers, 1994.

ARNOLD, Doug. « Why Translation is Difficult for Computers », dans Harold Somers, *Computers and Translation: A Translator's Guide*, Amsterdam, John Benjamins, 2003, p. 119-142.

AUSTERMÜHL, Frank. *Electronic Tools for Translators*, Manchester, St. Jerome Publishing, 2001.

BÉDARD, Claude. « La prétraduction automatique, outil de productivité et d'évolution professionnelle », *META*, vol. 37, n° 4 (décembre 1992), p. 738-760.

BÉDARD, Claude. « Ce qu'il faut savoir sur les mémoires de traduction », *Circuit*, n° 60 (été 1998), p. 25-26.

BÉDARD, Claude. « Mémoire de traduction cherche traducteur de phrases... », *Revue TRADUIRE*, n° 186 (2000), p. 41-49.

BÉDARD, Claude, et Marie-Pierre HÉTU. « La prétraduction automatique, vous connaissez? », *Circuit*, n° 84 (été 2004), p. 30.

BIBER, Douglas. « A typology of English texts », *Linguistics*, vol. 27, n° 1 (1989), p. 3-43.

BOWKER, Lynne. *Computer-Aided Translation Technology: A Practical Introduction*, Ottawa, University of Ottawa Press, 2002a.

- BOWKER, Lynne. « Information Retrieval in Translation Memory Systems: Assessment of Current Limitations and Possibilities for Future Development », *Knowledge Organization*, vol. 29, n° 3/n° 4 (2002b), p. 198-203.
- BOWKER, Lynne. « Terminology Tools for Translators », dans Harold Somers, *Computers and Translation: A Translator's Guide*, Amsterdam, John Benjamins, 2003, p. 49-65.
- BROWN, M. Katherine. « Trends in Writing for Translation », *MultiLingual Computing & Technology #59 Supplement* (octobre/novembre 2003), p. 20-21.
- CLARK, Bob. « E-term + TM = AM », *Language International*, vol. 14, n° 4 (août 2002), p. 21-25.
- CLEARY, Róisín, et Reinhard SCHÄLER. « R&D opens doors to Translation Portals », *Localisation Ireland*, vol. 4, n° 1 (2000), p. 9.
- COMITÉ SECTORIEL DE L'INDUSTRIE CANADIENNE DE LA TRADUCTION. *L'Industrie canadienne de la traduction : Stratégie de développement des ressources humaines et d'exportation* (en ligne), 1999 (consulté le 11 juillet 2004). Sur Internet : <URL : <http://www.uottawa.ca/associations/csict/princi-f.htm>>.
- EHGOETZ, Melissa. *Designing and Implementing a Small Scale Controlled Language for Multilingual Machine Translation*, 2004. Thèse de maîtrise en traduction, Université d'Ottawa.
- FERNANDEZ GUERRA, Ana. *Machine Translation: Capabilities and Limitations*, València, Universitat de València, 2000.
- GARCIA IZQUIERDO, Isabel. « The Concept of Text Type and Its Relevance to Translator Training », *Target*, vol. 12, n° 2 (2000), p. 283-295.
- GORDON, Jean, « Essai de typologie de textes source, dans le cadre de la traduction assistée par ordinateur », dans *L'intérêt de la traduction assistée par ordinateur pour les responsables de centres d'information et pour les utilisateurs finaux*, France, AGARD Lecture Series No. 171, 1990, p. 3-1 - 3-8.
- GOW, Francie. *Metrics for Evaluating Translation Memory Software*, 2003. Thèse de maîtrise en traduction, Université d'Ottawa.
- HARKUS, Susan. « A Writing Checklist », *MultiLingual Computing & Technology #59 Supplement* (octobre/novembre 2003), p. 23.
- HATIM, Basil et Ian MASON. *Discourse and the Translator*, Londres, Longman, 1990.

HINE, Jonathan T. Jr. « Writing for Translation », *MultiLingual Computing & Technology* #53 Supplement (janvier/février 2003), p. 20-21.

HEYN, Matthias « Translation Memories: Insights and Prospects », dans L. Bowker, M. Cronin, D. Kenny and J. Pearson, *Unity in Diversity? Current Trends in Translation Studies*, Manchester, St. Jerome Publishing, 1998, p. 123-136.

HUTCHINS, W. John et Harold L. SOMERS. *An Introduction to Machine Translation*, Londres, Academic Press, 1992.

IVERSON, Steve. « Working With Translation Memory: When and how to use TM for a successful translation project », *MultiLingual Computing & Technology*, vol. 14, n° 7 (2003), p. 35-37.

KWEON, Sangmi. *A Commented Bilingual Terminology File of SARS-related Terms*, 2005. Maîtrise en traduction, Université d'Ottawa.

KRAMASZ, Debra. « Writing and Translation as Parallel Processes », *MultiLingual Computing & Technology* #59 Supplement (octobre/novembre 2003), p. 14-18.

LAVROFF, Oleg, « Comparaison entre les traductions humaines et les traductions automatiques (qualités, coûts et délais) », dans *L'intérêt de la traduction assistée par ordinateur pour les responsables de centres d'information et pour les utilisateurs finaux*, France, AGARD Lecture Series No. 171, 1990, p. 10-1 - 10-8.

L'HOMME, Marie-Claude. *Initiation à la traductique*, Brossard (Québec), Linguatéc éditeur inc., 1999.

MELBY, Alan. « The translator workstation », dans John Newton, *Computers in Translation: A practical appraisal*, New York, Routledge, 1992, p. 147-165.

MOGENSEN, Else. « Orwellian Linguistics: How using computer-aided translation tools impacts the target language », *Language International*, vol. 12, n° 5 (octobre 2000), p. 28-31.

MULTICORPORA R&D INC. *Guide de l'utilisateur – MultiTrans Light 3.6*, Gatineau, MultiCorpora R&D Inc., s.d.

NEUBERT, Albrecht et Gregory R. SHREVE. *Translation as Text*, Kent (Ohio) et Londres, The Kent State University Press, 1992.

NIDA, Eugene A. et Charles R. TABER. *The Theory and Practice of Translation*, Leiden, E.J. Brill, 1969.

NYBERG, Eric, Teruko MITAMURA et Willem-Olaf HUIJSEN. « Controlled Language for Authoring and Translation », dans Harold Somers, *Computers and Translation: A translator's guide*, Amsterdam, John Benjamins, 2003, p. 245-281.

POVLSEN, Claus, Nancy UNDERWOOD, Bradley MUSIC et Anne NEVILLE. « Evaluating Text-Type Suitability for Machine Translation: a Case Study on an English-Danish MT System », dans A. Rubio, *et al.*, *Proceedings of the First International Conference on Language Resources and Evaluation*, Grenade, Espagne, 28-30 mai 1998, vol. 1, p. 27-31.

SAGER, Juan C. *Language Engineering and Translation: Consequences of automation*, Philadelphie, John Benjamins, 1994.

SCHWAB, Wallace. « Translating through the Looking Glass », *Circuit*, n° 78 (hiver 2003), p. 30.

SMART, John M. « Controlled English for Global Business », *MultiLingual Computing & Technology #59 Supplement* (octobre/novembre 2003), p. 19-21.

SNELL-HORNBY, Mary. *Translation Studies: An integrated approach*, Philadelphie, John Benjamins, 1995.

SMITH, Edward L. « Text type and discourse framework », *Text*, vol. 5, n° 3 (1985), p. 229-247.

SOMERS, Harold. « Introduction », dans Harold Somers, *Computers and Translation: A Translator's Guide*, Amsterdam, John Benjamins, 2003, p. 1-11.

SYSTRAN SOFTWARE, INC. *User's Guide for SYSTRAN 4.0*, San Diego, Systran Software, Inc., 2002.

TRANSROUTER CONSORTIUM. « TransRouter: a decision support tool for translation managers » (en ligne), dans *Proceedings of Machine Translation Summit VII (MT Summit '99) "MT in the Great Translation Era"*, Kent Ridge Digital Labs, Singapour, 13-17 sept. 1999 (consulté le 11 août 2004). Sur Internet : <URL : <http://www.localisation.ie/research/projects/TransRouter/MTSummit.htm>>.

TROSBORG, Anna. « Introduction », dans Anna Trosborg, *Text Typology and Translation*, Philadelphie, John Benjamins, 1997a, p. vii-xvi.

TROSBORG, Anna. « Text Typology: Register, Genre and Text Type », dans Anna Trosborg, *Text Typology and Translation*, Philadelphie, John Benjamins, 1997b, p. 3-23.

TRUJILLO, Arturo. *Translation Engines: Techniques for Machine Translation*, Londres, Springer, 1999.

TUNICK, Laraine. *Language Translation, Localization, and Globalization: World Market Forecasts, Industry Drivers, and eSolutions*, New York, Allied Business Intelligence Inc., 2002.

UNDERWOOD, Nancy L. et Bart JONGEJAN. « Profiling Translation Projects: An Essential Part of Routing Translation », dans *Proceedings of the 8<sup>th</sup> International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-99)*, University College, Chester, Angleterre, 1999.

UNDERWOOD, Nancy L. et Bart JONGEJAN. « Translatability Checker: A Tool to Help Decide Whether to Use MT », dans Bente Maegaard, *Proceedings of MT Summit VIII*, Santiago de Compostela, Espagne, 18-22 septembre 2001, p. 363-368.

VILLERS, Marie-Éva de. *Multidictionnaire de la langue française*, Montréal, Éditions Québec/Amérique, 1999.

**NATIONAL PAROLE BOARD  
APPEAL DIVISION  
DECISION**

---

**Name:** DOE, John  
**F.P.S.:** 123456A  
**Institution:** Dorchester Penitentiary  
**Decision(s) Appealed:** Statutory Release – Change Condition

---

**The Role of the Appeal Division**

The role of the Appeal Division is to ensure that the law and the Board policies are respected, and that the rules of fundamental justice are adhered to and that the Board's decisions are based upon relevant and reliable information.

The Appeal Division reviews the decision-making process to confirm that it was fair and that the procedural safeguards were respected.

The Appeal Division has jurisdiction to re-assess the issue of risk to re-offend and to substitute its discretion for that of the original decision makers, but only where it finds that the decision was unfounded and unsupported by the information available at the time the decision was made.

**Vote:** Decision Affirmed

***Summary of Appeal Decision:***

You have appealed the National Parole Board's decision of May 20, 2004, to impose the special condition to abstain from intoxicants on your statutory release.

The Appeal Division has carefully reviewed your file. We have also considered your written submissions dated May 25, 2004.

In arriving at our decision to deny the appeal and affirm the Board's decision, we have summarized and analyzed your grounds under the following headings:

- Reasonableness of the Decision
- Information Issues: Erroneous Information

## ***Analysis of Grounds:***

- **Reasonableness of the Decision**
- **Information Issues: Erroneous Information**

*You submit that the “alcohol restriction” imposed on your statutory release is unwarranted and based on “untrue facts”.*

*You claim that alcohol is not one of your criminogenic risk factors and you have never been in trouble due to drinking. You admit that drugs played a “big part” in your return to crime and argue that the Board should only have imposed a special condition to abstain from drugs.*

*You note that on your last parole release, the Board did not impose a condition to abstain from alcohol because you do not have an alcohol problem. You have never been in trouble in the institution due to drinking. Your Parole Officer also agrees that you do not have an alcohol problem. You are not a violent offender and, therefore, there is “little risk” if the alcohol restriction was removed.*

Mr. Doe, the Appeal Division finds that there was sufficient relevant and reliable information to support the Board’s decision to impose the special condition to abstain from intoxicants (alcohol and drugs) on your statutory release.

Contrary to what you submit, C.S.C. did recommend the imposition of the condition to abstain from all intoxicants, which included both drugs and alcohol. To support the alcohol prohibition, C.S.C. in its Assessment for Decision (A.F.D.) dated April 13, 2004, indicated that you had a lengthy history of becoming aggressive when you drank. C.S.C. referred to your half-sister’s comments that “*alcohol and drug use*” could easily lead you back to prison. As well, reference was made to Psychologist Dewhurst who was of the view that your feelings of confusion and anger resulted in your inability to generate positive choices and “*often alcohol becomes involved*”.

While it is true that your last day parole release in 2001 did not include a condition to abstain from alcohol, your previous day parole release in 2000 did include such a condition (see A.F.D. dated July 13, 2000 and N.P.B. decision dated August 16, 2000, which imposed the special condition to abstain from intoxicants). In fact, your previous day parole release was subsequently revoked because you had violated your special condition related to alcohol within two weeks of your release, absconded from the halfway house and were suspected of committing property offences (see N.P.B. decision dated December 14, 2000).

Furthermore, information in your file reveals that you were known to be argumentative and aggressive when under the influence of alcohol. While on day parole (provincial) in 1999, concern was expressed that you could possibly reoffend in a violent manner if you did not stop drinking. Psychological opinion at the time noted that you had significant problems coping with stress, frustration and anger and you abused alcohol (C.P.P.R. dated July 13, 2000 and Correctional Plan dated March 23, 2000). File information also indicated that you were convicted in 1999 for Providing Alcohol to a Minor while on day parole. At the time of the offence, you were found to be drinking and suspension warrants were issued when you failed to return to the halfway house following an A.A. meeting. Your day parole was subsequently revoked (Correctional Plan dated March 23, 2000).

Family members described you as having an “*addictive personality*” (Community Assessment dated February 1, 2000) and your mother was willing to support you provided you stayed away from alcohol, drugs and your criminal associates (C.P.P.R. dated September 13, 2001). You also prepared a Relapse Prevention/Maintenance Plan during the O.S.A.P.P. Program, where you identified having liquor in the house as a high-risk situation for you (O.S.A.P.P. Report dated October 14, 2000).

Mr. Doe, in light of the above information, the Appeal Division is satisfied that the Board’s decision to prohibit you from using alcohol (in addition to drugs) is reasonable and supported by sufficient relevant and reliable information. Substance abuse is an identified risk factor in your case and is linked to your aggressive and criminal behaviour. The Board’s decision to impose the special condition to abstain from intoxicants is justified and is necessary to protect society and to facilitate your successful reintegration into the community.

**Conclusion:**

For the reasons set out above, the Appeal Division denies the appeal and affirms the Board’s decision of May 20, 2004, to impose the special condition to abstain from intoxicants on your statutory release.

M. Charbonneau, Q.C.  
Board Member

\sr

**DECISION DATE:** July 14, 2004

**Statement of Qualifications  
Policy Analyst PM-03  
Position number: SGPOL-0142T  
Hemispheric Cooperation**

**Education**

Graduation with an acceptable degree from a recognized university, with specialization in criminology or a related discipline or an acceptable combination of experience, training and/or education.

**Language Proficiency**

English Essential

**Experience**

- Experience in analyzing Cross-border issues and providing related policy advice for the use of senior officials and Ministers.
- Experience in preparing ministerial correspondence, briefing notes and memoranda to summarize research and policy materials.
- Experience in consulting and liaising with a variety of other government departments, the US, as well as non-governmental associations on Cross-border issues.

**Knowledge**

- A knowledge of the organization and responsibilities of the Department of Public Safety & Emergency Preparedness Canada
- A knowledge of the federal policy development and Cabinet decision making process as well as legislative processes.
- Knowledge of the policing and law enforcement environment in Canada.
- Knowledge of the issues relating to Canada – US relations and the Department.

**Abilities**

- Ability to provide advice to managers on policy and program implications.
- Ability to work independently and efficiently and to manage competing priorities.
- Ability to communicate effectively both orally and in writing.
- Ability to maintain effective working relationships with other federal departments and agencies in the policy development process.

**Personal Suitability**

Strong communication and interpersonal skills

Dependability

Initiative

Flexibility

**Condition of Employment**

Secret security clearance

## Annexe B – Données relatives à chaque texte ayant servi à l'étape de la conception

Coupures de presse	MT1	MT2	MT3	TA1	TA2	TA3	TH1	TH2	TH3	Moy.
Mots par phrase	18,1	18,4	22,6	27,3	20,3	18,6	22,6	20,5	29,0	21,9
Degré de lisibilité (100=très facile)	35,5	38,2	32,9	25,6	38,0	36,6	26,4	33,0	30,5	33,0
Niveau scolaire	12,0	12,0	12,0	12,0	12,0	12,0	12,0	12,0	12,0	12,0
Voix passives (%/verbes conjugués)	30	30	22	19	17	6	6	3	22	17
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	2 (3,0)	2 (2,5)	5 (4,6)	1 (1,4)	7 (7,9)	1 (1,3)	5 (5,9)	6 (8,8)	3 (3,7)	4 (4,3)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/1000 mots)	6 (8,9)	9 (11,3)	4 (3,7)	4 (5,4)	2 (2,3)	8 (10,4)	2 (2,4)	6 (8,8)	9 (11,2)	6 (7,2)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/1000 mots)	8 (11,8)	6 (7,5)	8 (7,3)	15 (20,3)	4 (4,5)	9 (11,7)	8 (9,5)	4 (5,9)	14 (17,5)	8 (10,7)
Nombre de mots en <i>-ing</i> (occurrences/1000 mots)	17 (25,1)	36 (45)	38 (34,8)	26 (35,1)	24 (27,2)	24 (31,1)	27 (31,9)	23 (33,8)	30 (37,4)	27 (33,5)
Nombre de contractions – <i>can't, don't, won't, they're, that's</i> (occurrences/1000 mots)	0 (0)	2 (2,5)	1 (0,9)	0 (0)	0 (0)	1 (1,3)	1 (1,2)	2 (2,9)	0 (0)	1 (1)
Nombre de conjonctions de coordination – <i>and, but, or, nor</i> (occurrences/1000 mots)	7 (10,3)	16 (20)	31 (28,4)	10 (13,5)	18 (20,4)	10 (13,0)	18 (21,3)	15 (22,1)	20 (24,9)	16 (19,3)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	53,63	47,31	46,58	47,82	48,80	47,88	49,95	54,20	48,57	49,42
TermBase – % de mots remplacés	21	24	19	20	19	23	21	20	18	21
TermBase – Nombre de phrases avec 100 % remplacement (%)	4 (8)	4 (7)	4 (6)	4 (9)	4 (6)	4 (7)	4 (8)	4 (8)	4 (8)	4 (7)
TransCorpora – % de mots trouvés	61	66	59	62	59	63	62	61	60	61
TransCorpora – Nombre de phrases avec 75-99% remplacement (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	12 (23)	12 (21)	12 (17)	11 (24)	13 (21)	12 (20)	13 (25)	14 (28)	11 (23)	12 (22)

Décisions d'appel	MT1	MT2	MT3	TA1	TA2	TA3	TH1	TH2	TH3	Moy.
Mots par phrase	26,8	26,9	24,0	23,2	34,1	23,3	25,4	22,1	25,2	25,6
Degré de lisibilité (100=très facile)	46,3	28,3	34,6	43,5	24,8	39,6	35,2	56,6	40,9	38,9
Niveau scolaire	12,0	12,0	12,0	12,0	12,0	12,0	12,0	10,3	12,0	11,8
Voix passives (%/verbes conjugués)	30	37	20	25	20	30	24	19	46	28
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	6 (7,8)	3 (3,2)	10 (8,4)	2 (2,7)	5 (6,6)	10 (9,1)	8 (6,5)	8 (10,6)	8 (9,5)	7 (7,2)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/1000 mots)	2 (2,6)	2 (2,1)	2 (1,7)	2 (2,7)	1 (1,3)	4 (3,6)	4 (3,2)	3 (4,0)	4 (4,7)	3 (2,9)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/1000 mots)	31 (40,1)	26 (27,4)	30 (25,3)	27 (37,0)	24 (31,6)	35 (31,8)	43 (34,8)	33 (43,7)	31 (36,8)	31 (34,3)
Nombre de mots en <i>-ing</i> (occurrences/1000 mots)	*13 (16,8)	*18 (19)	*19 (16)	*13 (17,8)	*25 (32,9)	*15 (13,6)	*25 (20,3)	*16 (21,2)	*14 (16,6)	*18 (19,4)
Nombre de contractions – <i>can't, don't, won't, they're, that's</i> (occurrences/1000 mots)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Nombre de conjonctions de coordination – <i>and, but, or, nor</i> (occurrences/1000 mots)	24 (31,0)	36 (37,9)	44 (37,0)	28 (38,4)	29 (38,2)	40 (36,3)	50 (40,5)	29 (38,4)	33 (39,1)	35 (37,4)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	33,79	33,63	35,19	36,08	37,19	33,30	32,14	38,50	34,74	34,95
TermBase – % de mots remplacés	24	21	21	32	25	18	20	22	23	23
TermBase – Nombre de phrases avec 100 % remplacement (%)	6 (11)	9 (12)	10 (11)	9 (15)	10 (17)	10 (13)	12 (14)	10 (16)	10 (16)	10 (14)
TransCorpora – % de mots trouvés	83	75	74	79	73	75	73	75	77	76
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)	22 (40)	12 (16)	12 (13)	16 (27)	6 (10)	10 (13)	11 (13)	7 (11)	20 (32)	13 (19)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	20 (36)	23 (30)	22 (24)	22 (37)	22 (37)	30 (38)	26 (30)	24 (39)	32 (52)	25 (36)

\*en excluant les occurrences du nom « hearing », qui revient souvent dans ces textes

Énoncés de qualités	MT1	MT2	MT3	TA1	TA2	TA3	TH1	TH2	TH3	Moy.
Mots par phrase	15,3	13,4	8,9	11,7	11,9	8,4	10,9	10,4	17,2	12
Degré de lisibilité (100=très facile)	3,0	0,0	9,6	5,7	0,0	0,0	13,8	3,1	10,5	5,1
Niveau scolaire	12,0	12,0	12,0	12,0	12,0	12,0	12,0	12,0	12,0	12,0
Voix passives (%/verbes conjugués)	0	6	0	0	0	0	0	0	0	1
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrences/1000 mots)	0 (0)	4 (16,3)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (1,8)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrences/s)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4,4)	1 (4,4)	1 (2,8)	0 (1,3)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrences/s)	0 (0)	1 (4,1)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0,5)
Nombre de mots en <i>-ing</i> (occurrences/s)	13 (56,3)	8 (32,7)	11 (45,8)	11 (59,1)	15 (57,5)	15 (61,7)	7 (30,8)	7 (30,8)	26 (72)	13 (49,6)
Nombre de contractions – <i>can't, don't, won't, they're, that's</i> (occurrences/s)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Nombre de conjonctions de coordinacoordination <i>ut, or, nor</i> (occurrences/s)	19 (82,3)	18 (73,5)	19 (79,2)	20 (107,5)	20 (76,6)	24 (98,8)	18 (79,3)	20 (98,1)	35 (97)	21 (88)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	57,33	60,08	56,39	58,95	58,74	52,65	57,49	58,85	51,84	56,92
TermBase – % de mots remplacés	15	15	6	26	16	11	17	17	8	15
TermBase – Nombre de phrases avec 100 % remplacement (%)	4 (14)	6 (21)	2 (6)	4 (15)	5 (14)	2 (7)	6 (18)	5 (20)	2 (5)	4 (13)
TransCorpora – % de mots trouvés	75	76	64	81	94	72	80	70	66	75
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)	2 (7)	5 (17)	4 (13)	3 (11)	3 (9)	2 (7)	6 (18)	0 (0)	5 (13)	3 (11)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	12 (41)	17 (59)	6 (19)	11 (41)	13 (37)	3 (10)	14 (41)	5 (20)	14 (37)	11 (34)

## Annexe C – Exemple de chacun des deux types de textes ayant servi à l'étape des essais

### QUARTERLY STATUS REPORT - OFFENDER HUMAN RIGHTS COMPLAINTS

During this latest reporting period (June 11 to September 3, 2004), the Human Rights Division has managed fifty complaints filed by offenders and visitors with the Canadian Human Rights Commission (CHRC). Twelve files have been officially settled/closed since the last report and sixteen new complaints have been received in the same time period.

Table 1 provides a breakdown of CHRC offender complaints by Region and stage in the complaint process. Table 2 provides a brief summary of CHRC complaints. The major issues identified in CHRC complaints filed by offenders are outlined in Annex A.

**TABLE 1 (a): STAGES OF COMPLAINTS BY REGION**

REGIONS	Intake	Mediation	Investig.	Pre-Tribunal	Tribunal	Settled/ Closed	Total
Atlantic	1	2	2	0	0	2	7
Quebec	10	0	2	0	1	4	17
Ontario	4	0	5	0	0	1	10
Prairies	4	0	1	0	0	3	8
Pacific	5	0	1	0	0	2	8
<b>Total</b>	<b>24</b>	<b>2</b>	<b>11</b>	<b>0</b>	<b>1</b>	<b>12</b>	<b>50</b>

**TABLE 1 (b): COMPARISON TABLE FROM PREVIOUS REPORT (March 3, to June 11, 2004).**

REGIONS	Intake	Mediation	Investig.	Pre-Tribunal	Tribunal	Settled/ Closed	Total
Atlantic	2	1	2	0	0	0	5
Quebec	6	0	2	1	1	5	15
Ontario	2	2	3	0	0	3	10
Prairies	3	0	4	0	0	1	8
Pacific	4	0	1	0	0	1	6
<b>Total</b>	<b>17</b>	<b>3</b>	<b>12</b>	<b>1</b>	<b>1</b>	<b>10</b>	<b>44</b>

**TABLE 2: SUMMARY OF CHRC COMPLAINTS**

(\*) Denotes new complaint received since last Quarterly Report.

Facility	Ground of discrimination/ Stage	Status
Nova Institution for Women	Sexual Orientation, Race, Colour and National / Ethnic Origin	The offender alleges that she was moved from House 2 to House 4 because she was intimately involved with a black inmate in House 2. She also alleges that CSC has delayed her parole because of her involvement in a same-sex, inter-racial relationship. <b>Mediation was</b>

	Investigation	<b>unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. A CHRC investigator concluded that Nova Institution does not treat same-sex relationships consistently and recommended that the Commission appoint a conciliator to bring about a resolution to the complaint. CSC is preparing a response to that recommendation.</b>
Nova Institution for Women	Sexual Orientation, Race, Colour and National / Ethnic Origin  Investigation	The offender alleges that she was separated from her partner because the two were involved in a same-sex, inter-racial relationship. In addition, she claims that CSC has falsely accused her of being a sexual predator because she presents herself as a man. <b>Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. A CHRC investigator concluded that Nova Institution does not treat same-sex relationships consistently and recommended that the Commission appoint a conciliator to bring about a resolution to the complaint. CSC is preparing a response to that recommendation.</b>
Nova Institution for Women*	Retaliation  Mediation	The offender alleges that CSC is discriminating against her by treating her in an adverse differential manner in retaliation for lodging her initial CHRC complaint (under section 14.1 of the <i>Canadian Human Rights Act</i> ). <b>CSC will attempt to resolve this complaint through mediation.</b>
Nova Institution for Women	Religion  CLOSED	The offender alleges that, she has been denied access to sweats and other events related to her Aboriginal spirituality because she is in the maximum security unit. She also claims that the institutional Elder does not visit her. CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>
Dorchester	Disability  Mediation	The offender alleges that CSC has discriminated against him by treating him in an adverse differential manner based on his disability (asthma and spinal nerve damage). He claims that he is exposed to second-hand smoke, which affects his asthma. In addition, he alleges that CSC has failed to provide him with medication to manage his back pain. <b>CSC is attempting to resolve this complaint through mediation.</b>
Dorchester*	Disability  Intake	The offender alleges that CSC has treated him in an adverse differential manner based on his disability. He claims that CSC's discriminatory policies result in disabled offenders being paid less than able-bodied inmates. In addition, the offender alleges that able-bodied offenders are provided with more work opportunities. CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>
Springhill	Race	The offender alleges that CSC is treating him in an adverse differential

	<b>CLOSED</b>	manner based on his race. He claims that he has been placed in segregation and denied access to telephone calls and programs because he is Aboriginal. CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>
Sainte-Anne des Plaines (RRC)	Disability <b>Investigation</b>	The offender alleges that CSC has discriminated against him based on his history of mental illness. He claims that decisions related to his penitentiary placement were based on CSC's misperception that he was a depressive individual with a high suicide risk. He adds that a staff member made inappropriate comments about his mental state. <b>CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. However, the Commission decided that CSC's internal redress procedure will not adequately address the complaint and decided to initiate an investigation. CSC has asked the Federal Court of Canada to review the Commission's decision.</b>
Sainte-Anne des Plaines (SHU)	Disability <b>Tribunal</b>	The offender alleges that his mental disability (ADHD) was not accommodated because he was not given appropriate prescription medication that may have assisted him in obtaining a lower security classification. The file moved to investigation and the investigator found that CSC had failed to accommodate the offender's disability. <b>As such, the file has been forwarded to the Canadian Human Rights Tribunal (CHRT). It is expected that the CHRT will examine this complaint in the fall of 2004.</b>
Sainte-Anne des Plaines (SHU)	Disability <b>CLOSED</b>	The offender alleges that CSC has failed to provide him with equal access to health services (prosthetic dental work) because of his HIV status. Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. <b>A CHRC investigator recently recommended that the Commission dismiss the complaint because the offender's complaint has been redressed (received the dental work). The Commission concurred with the investigator. As such, this file is now closed.</b>  NOTE: The complainant also names Atlantic Institution (where he was incarcerated prior to the SHU) in his complaint.
Port Cartier	Disability <b>CLOSED</b>	The offender alleges that staff did not accommodate his mental disability, which includes being illiterate, by insisting that he make requests for services, etc. in writing. The offender also alleges that he has been harassed as a result of complaints he has made regarding his disability. An investigator found that the evidence did not support the complainant's allegations and recommended that the CHRC dismiss the complaint. However, the Commission disagreed with the investigator's findings and sent the complaint to the Canadian Human Rights Tribunal. <b>Pre-tribunal mediation was successful in bringing about a resolution to this complaint and the Commission approved the minutes of</b>

		settlement. As such, this file is now closed.
Port Cartier*	Retaliation  <b>Intake</b>	The offender (referred to in complaint # 8 above) alleges that CSC is discriminating against him by treating him in an adverse differential manner in retaliation for lodging his initial CHRC complaint (under section 14.1 of the <i>Canadian Human Rights Act</i> ). <b>CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. CSC awaits the Commission's decision.</b>
Port Cartier	Colour  <b>Investigation</b>	The offender alleges CSC has discriminated against him by treating him in an adverse differential manner and by failing to provide him with a service free of harassment based on his colour (black). This is a complex complaint, involving 2 regions (QUE and ATL) and 5 institutions (Atlantic, Drummond, Archambault, Donnacona and Port Cartier--PCI being the institution currently housing the offender). <b>CSC Legal Services is co-ordinating this file. The CHRC may not proceed with this investigation as the offender is not a Canadian citizen.</b>
Port Cartier	Religion  <b>Intake</b>	The offender alleges that CSC's programs are discriminatory. He refuses to participate in CSC's Western, Judeo-Christian programs because they go against his religious beliefs. More specifically, he claims that his religion (Scientology) considers psychology and psychological programs to be detrimental to him. As a result, he claims that he has been denied transfers and a lower security classification and pay raises. <b>It has since been determined that there is a deportation order in place for the offender. As such, the Commission may not deal with the complaint.</b>
Port Cartier*  (6 individual complaints)	Religion  <b>Intake</b>	Six offenders filed identical complaints alleging that CSC is discriminating against them based on their religion (Scientology). The offenders claim that they are not allowed to participate in religious meetings or view Scientology videos. <b>CSC has suggested that the CHRC encourage the offenders to address this issue through the Inmate Complaint and Grievance System. CSC awaits the Commission's decision.</b>
Drummond	Colour and Race  <b>CLOSED</b>	The offender alleges that he has been subjected to racial slurs and other forms of harassment based on his race and colour. He also alleges that staff have threatened him and used retaliation tactics to deter him from filing complaints. He adds that when he has filed institutional complaints, he has not received a response. CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>

Drummond	Retaliation (for submission of earlier complaint)  <b>CLOSED</b>	The offender alleges that CSC is discriminating against him by treating him in an adverse differential manner in retaliation for lodging his first CHRC complaint (under section 14.1 of the <i>Canadian Human Rights Act</i> ). CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>
Drummond	Colour and Race  <b>Intake</b>	The offender alleges that he has been subjected to discrimination and harassment based on his race and colour. He claims he has been subjected to racial slurs from more than one staff member. He also alleges that staff have threatened to "make his time hard" if he files complaints. He adds that his complaints at the institutional level are ignored. CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>
Drummond	Retaliation (for submission of earlier complaint)  <b>Intake</b>	The offender alleges that CSC is discriminating against him by treating him in an adverse differential manner in retaliation for lodging his first CHRC complaint (under section 14.1 of the <i>Canadian Human Rights Act</i> ). CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>
Kingston Penitentiary	National/Ethnic Origin and Disability  <b>Investigation</b>	The offender alleges that contrary to CD 702: 1) Several spiritual items have been confiscated from his cell and never been returned. 2) He and other Aboriginal inmates at KP have no access to Aboriginal programs. 3) The offender has been denied access to spiritual medicines. In addition, he alleges that KP refuses to accommodate his disability by providing him with a device to improve circulation in his hand.. <b>Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. A CHRC investigator has recently recommended that the Commission dismiss the complaint as there is no evidence that CSC discriminated against the offender. CSC awaits the Commission's decision in this matter.</b>
Frontenac	Race  <b>Investigation</b>	The offender alleges that CSC has discriminated against him and other black inmates by not providing a permanent funding base for the Black Inmates' Friends Assembly (BIFA). <b>The offender decided not to participate in mediation. As such, this file has now proceeded to the investigation stage.</b>
Bath	Disability	The offender alleges that CSC is denying him medication that is

	<b>Intake</b>	necessary to control his disability (agoraphobia). <b>CSC has informed the CHRC that the offender is also pursuing this matter in Federal Court and through the Inmate Complaint and Grievance System. A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>
Bath*	Disability <b>Intake</b>	The offender alleges that CSC has failed to accommodate his disability (allergy to cigarette smoke) by not providing him smoke-free living conditions. <b>CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. The CHRC is currently deciding whether or not to pursue this complaint.</b>
Collins Bay (VISITOR)	Race, Colour, National/Ethnic Origin <b>Investigation</b>	Visitor alleges that she and her three children have been harassed and intimidated by CSC staff while visiting her husband. She alleges that staff centre her out and "badger" her in front of other visitors because she is black. She also claims that she has been subject to racial slurs by CSC staff. <b>Mediation was unsuccessful in bringing about a resolution to this complaint. As such, this file has now proceeded to the investigation stage.</b>
Warkworth	Colour and Race <b>Investigation</b>	The offender alleges that he has been denied a transfer to a minimum security facility on two separate occasions despite having completed his correctional plan. He claims that this denial is based on his colour and race. In addition, the offender notes that it has been two years since Warkworth has granted an ETA to a black inmate. CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>However, the Commission decided that CSC's internal redress procedure will not adequately address the complaint. As a result, the complaint has moved to the investigation stage.</b>
Warkworth	Religion and National/Ethnic Origin <b>CLOSED</b>	The offender alleges that the IPSO is harassing him based on his religion and national/ethnic origin. He claims that he is referred to as "Tony Soprano" because he is Italian. In addition, he claims that he has been subjected to derogatory comments about Catholicism. <b>CSC has informed the CHRC that the offender is also pursuing this matter through the Inmate Complaint and Grievance System. The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>
Warkworth*	Disability and Sex <b>Intake</b>	Offender alleges that CSC is failing to accommodate his disability (Borderline Personality Disorder-BPD) by not providing him with the most current treatment techniques (Dialectical Behaviour Therapy-DBT). He also claims that he is being discriminated against based on his sex because female inmates with the BPD have access to DBT. <b>CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. CSC</b>

		<b>awaits the Commission's decision in this matter.</b>
Joyceville and Frontenac	Race, Colour and National/Ethnic Origin and Religion  <b>Investigation (settlement in progress)</b>	The offender alleges that CSC has discriminated against him based on race, colour and national/ethnic origin and religion. The offender initially refused mediation and an investigation was initiated by the CHRC. However, both parties then agreed to mediation. <b>Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved back to the investigation stage. CSC and the offender have since reached an informal resolution and await the Commission's approval of the minutes of settlement.</b>
Grand Valley Institution for Women*	Race, National / Ethnic Origin, Disability  <b>Intake</b>	The offender alleges that she has been denied a voluntary transfer to Okimaw Ohci Healing Lodge for two reasons: First, she claims that the Healing Lodge refuses to accept her because she is an Aboriginal from Eastern Canada (Mi'kmaq) as opposed to being from Prairies (Cree). Second, she claims that the denial of her transfer is an excuse to avoid accommodating her disability (cancer treatment). <b>The CHRC may not pursue this complaint as the alleged discrimination took place more than a year before she filed her complaint. CSC awaits the Commission's decision.</b>  NOTE: The offender also names Springhill Institution (where she was incarcerated prior to Grand Valley) in her complaint.
Saskatchewan Penitentiary*	Religion and National / Ethnic Origin  <b>Intake</b>	The offender alleges that CSC has discriminated against him because he is a Muslim and a Pakistani national. He claims that his placement in involuntary segregation is a direct result of his religion and nationality. In addition, he alleges that CSC has placed false information on his file regarding his and his visitors' connections to organized crime. <b>The CHRC may not pursue this complaint as the offender is not a Canadian citizen.</b>
Saskatchewan Penitentiary	Religion  <b>Intake</b>	The offender alleges that CSC has discriminated against him by diagnosing him with a mental illness (schizophrenia) because of his religious experiences and beliefs. He also claims that CSC has transferred him to RPC Prairies and forced him to take medication against his will. <b>CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. The CHRC is currently deciding whether or not to pursue this complaint.</b>
Saskatchewan Penitentiary	Religion  <b>Intake</b>	The offender alleges that CSC treated him in an adverse differential manner based on his Aboriginal spirituality. He claims that CSC did not accommodate his religious needs by providing the services of an Aboriginal Elder to search his medicine bundle before leaving the institution for medical ETA. He also alleges that CSC has a discriminatory policy that allows Correctional Officers to search Aboriginal offenders' medicine bundles. <b>CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>

Edmonton	Race and Colour <b>CLOSED</b>	The offender alleges that correctional staff have made several racial slurs towards him and have sexually assaulted him during a search. <b>However, a CHRC investigator recommended that the Commission not proceed with this complaint as the offender is not a Canadian citizen. The Commission concurred with the investigator's recommendation. As such, this file is now closed.</b>
Edmonton	Religion <b>Intake</b>	The offender alleges that his religious needs are not being met. He claims that he was deprived of access to his religious books and other articles. The offender also claims that some of his religious effects were confiscated and thrown out during a search. In addition, he alleges that his religious diet be cancelled. <b>CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>
Bowden	Sexual Orientation <b>Investigation</b>	The offender alleges that CSC discriminated against him based on his sexual orientation, making erroneous statements about his orientation in reports and on his file, removing his PFVs, and not supporting him for release. <b>Mediation was unsuccessful in bringing a resolution to this matter. This complaint has now moved to the investigation stage.</b>
Bowden	Race, Colour, and National / Ethnic Origin <b>CLOSED</b>	The offender alleges that CSC has discriminated against him on the basis of his race, national or ethnic origin and colour. Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. <b>A CHRC investigator recommended that the Commission dismiss the complaint as there is no evidence that CSC discriminated against the offender. The Commission concurred with the investigator. As such, this file is now closed.</b>
Bowden	Retaliation for submission of earlier complaint <b>CLOSED</b>	The offender alleges that CSC is discriminating against him by treating him in an adverse differential manner in retaliation for lodging his first CHRC complaint (under section 14.1 of the <i>Canadian Human Rights Act</i> ). Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. <b>A CHRC investigator has recently recommended that the Commission dismiss the complaint as there is no evidence that CSC discriminated against the offender. The Commission concurred with the investigator. As such, this file is now closed.</b>
Kent	Disability <b>Intake</b>	The offender alleges that CSC has not accommodated his mental disability (ADHD) because he has not been provided the appropriate prescription medication to control his impulsive behaviour. As a result, he claims that he has been unable to cascade to a lower security level. <b>CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>

Kent*	Sex  <b>Intake</b>	The offender alleges that CSC has discriminated against him by not accommodating his needs as a transsexual. The offender claims that CSC has not taken his needs into account when performing strip-searches and urinalysis testing. In addition the offender alleges that CSC has deprived him access to personal effects such as appropriate undergarments and make-up. <b>CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. The CHRC is currently deciding whether or not to pursue this complaint.</b>
Mission	Religion  <b>CLOSED</b>	Offender alleges that Mission Institution is not providing him full access to his Buddhist practices and needs. He notes that in the past, other CSC institutions have accommodated his religious needs. He also alleges that he has been threatened with placement in segregation or a transfer to a higher security level if he pursues his complaint. CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. <b>The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>
Mission	Religion  <b>CLOSED</b>	Offender alleges that CSC's programs are discriminatory. He refuses to participate in CSC's Western, Judeo-Christian programming because they go against his Buddhist beliefs. As a result, he claims that he has been denied ETAs, transfers and a lower security classification. <b>CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. The CHRC decided that they should not deal with this complaint because the offender has not exhausted CSC's internal redress system. As such, the file is now closed. However, the Commission reserves the right to revisit this complaint within a year of the offender exhausting CSC's grievance process.</b>
Mission	Religion, Disability, National / Ethnic Origin  <b>Intake</b>	The offender has made numerous allegations of mistreatment based on his age, national/ethnic origin, and religion. He claims that during a cell search, many of his religious effects were stolen and/or destroyed by CSC officers. In addition, he claims that despite specific requests, his Spinal Trauma Injury Support System (STISS) was removed from his cell and not returned. <b>CSC has informed the CHRC that the offender is pursuing this issue through the Inmate Complaint and Grievance System. The CHRC is currently deciding whether or not to pursue this complaint.</b>
Mission*	Disability  <b>Intake</b>	The offender alleges that CSC has failed to accommodate his disability (visual impairment) in a number of ways. He claims that CSC refuses to let him have software and peripherals necessary for someone with a visual impairment to operate a personal computer. The offender also claims that CSC's refusal to accommodate his needs has contributed to the accelerated deterioration of his vision as well as a number of other adverse health effects (stress, insomnia, weight loss, backaches etc...).

		<b>CSC suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. A CHRC investigator has recently recommended that the Commission not pursue the complaint at this time as the offender has not exhausted CSC's internal redress procedure. CSC awaits the Commission's decision.</b>
Mountain	Race and Colour  <b>Investigation</b>	The offender alleges that the CSC has discriminated against him by treating him in an adverse manner and by failing to provide him with service free of harassment based on his race. He indicates that he has been subjected to racial slurs as well as repeated cell searches and unjust property seizure because he is a black inmate. <b>Mediation was unsuccessful in bringing a resolution to this matter and the complaint moved to the investigation stage. However, the CHRC may not proceed with this investigation as the offender is not a Canadian citizen.</b>
Matsqui*	Disability and Sex  <b>Intake</b>	Offender alleges that CSC is failing to accommodate his disability (Borderline Personality Disorder-BPD) by not providing him with the most current treatment techniques (Dialectical Behaviour Therapy-DBT). He also claims that he is being discriminated against based on his sex because female inmates with the BPD have access to DBT. <b>CSC has suggested that the CHRC encourage the offender to address this issue through the Inmate Complaint and Grievance System. CSC awaits the Commission's decision in this matter.</b>

**NOTE:**

- Complaints 2 and 3 filed by one offender**
- Complaints 5 and 6 filed by one offender**
- Complaints 8 and 12 filed by one offender**
- Complaints 21 and 22 filed by one offender**
- Complaints 23 and 24 filed by one offender**
- Complaints 41 and 42 filed by one offender**

## **Annex A: Major Issues in CHRC Complaints**

The following is a summary of the three major issues identified in CHRC complaints filed by offenders:

**Adverse treatment based on race:** Thirteen of the thirty-eight (38) active complaints allege adverse treatment based on race, colour or national / ethnic origin. Currently, there is at least one complaint of this nature in every Region of the country. The major issues identified in these complaints are: being subjected to harassment and racial slurs by CSC staff; unwarranted cell searches; unjustified seizures of personal property; being denied access to programs and transfers; being held at a higher security level than required; .and having adverse (but incorrect) information placed on file.

**Failure to accommodate disability:** Overall, thirteen of the thirty-eight (38) active complaints allege that CSC has failed to accommodate offenders' disabilities. The complaints refer to a diverse number of issues such as: failing to provide required medication; failing to provide equal access to mental health services; failing to provide assistive devices; and, failing to addressing offenders' environmental sensitivities.

**Failure to accommodate religion:** Currently, twelve (12) of the thirty-eight (38) active complaints are related to religious accommodation. They involve the following key issues: the availability and preparation of religious diets, access to religious items and services and respect for religious observances.

## Minutes from the last weekly IM/IT Service Management Meeting

March 16<sup>th</sup>, 2005  
11:15am

### Compte rendu de la dernière réunion hebdomadaire de gestion de services de la GI/TI

**Chairperson / Président** : Robert Hitsman

**Regional Participants /**  
**Participants des régions:**

Atlantic  
Eric Leger

Québec  
Marc Favreau

Ontario  
Tom Burns

Prairies  
No representation.

Pacific  
Garry Louden  
Bob Dobos

**NHQ Participants /**  
**Participants de l'AC:**

Trevor Mann (Service Management)  
Patrick Buchanan (Service Desk)  
Scott Leibbrandt (NPB Desktop)  
Jackie Mitchell (NPB)  
Charles Lacaille (VMS)  
Susan Poapst (Tech Services)  
Cameron MacDonald (QA)  
Doug Smith (Messaging)  
Eric Moisan (NAT-OPS)  
Gerry Mulcahey (CJIL)  
Keith Murray (OMSR)

### **Action Items**

No action items reported at this time.

### **Emergency**

No emergency items reported at this time.

### **Problems**

Patrick Buchanan stated that one of the Unix Server's have a hardware issue (memory). This affects CJIL, VIP, IMRS & other small applications. The server is not available at the current moment & is being worked on.

P223444 – Logon Scripts – Marc Favreau was wondering the status of this CR. Eric Moisan stated that Paul Harvey is looking into this CR at the moment.

## Changes

C6880 - Performance patch for IFMMS – No representation.

C6981 - IFMMS Production database upgrade – No representation.

C7147 - RCMP access to OWA: firewall changes – No representation.

C7227 - Upgrade Antigen on Exchange to build 1494 – Doug Smith stated that they are waiting for security to complete their tasks.

C7274 - Change 3 UNIX parameter settings & 4GB memory – No new information.

C7224 - New Offender Reports – Update – No new information.

C5371 - Community Residential Facilities (CRF) – Eric Moisan stated that once the Unix server is back up, they will start installing; there was not enough memory prior to the memory upgrade.

C7196 - Nfix040.exe - Peoplesoft DR DNS Alias - CEDV2 – Robert Hitsman stated that this CR is currently in Service Management. Robert is currently working with Paul Harvey to do a pilot in the Pacific replication.

C7075 - IMRS for CORCAN – Gerry Mulcahey stated that this CR should be done by its target date, if not, he will update the Ticket's date.

C7230 - Nfix124 - IAS Update for CEDv1/PTCEv1 – Robert Hitsman stated that this CR is in Service Management & is out for P.A.T. testing. The NFix is due Friday March 18, 2005 @ 12:00.

C7223 - Nfix123.exe - Remove Peoplesoft 7 icons - CEDv1 – Robert Hitsman stated that he spoke with Robert Crane this morning about the two NFix's that are reliant on the Peoplesoft release. Robert Crane said that the Peoplesoft is going to be delayed for another 2-3 weeks from its previous release date.

C5414 - MS RDP vs. MS SMS Remote Tools – Robert Hitsman stated that this CR was in Service Management, but he changed it to NAT – OPS.

C5719 - Upgrade Tru64 5.1 Servers to 5.1b & Update Firmwar – No representation.

C6056 - NFix024 - Client requirements for OMSR, OMSM – Robert Hitsman stated that this CR was sent out for replication on Monday March 14, 2005 & ran on Tuesday March 15, 2005. The Pacific was the last Region to replicate & they reported that they finished it this morning.

## New Business

-Tom Burns asked Eric Moisan if there were any updates about the DC's, like when they might be coming out. Eric Moisan stated that he is not sure, but they will be going out as 2000 Servers.

-Charles Lacaille stated that the Ontario Upgrade CR6787, is going ahead this weekend. Everything is on schedule & everyone has been notified.

**Annexe D – Données relatives à chaque texte ayant servi à l'étape des essais**

PDP	MT1	MT2	TA1	TA2	TH1	TH2	Moy.
Mots par phrase	17,6	18,5	17,5	16,9	18,0	17,5	17,7
Degré de lisibilité (100=très facile)	41,2	41,4	35,3	35,5	38,3	42,8	39,1
Voix passives (%/verbes conjugués)	13	15	12	10	17	12	13
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrence/1000 mots)	2 (2,9)	3 (4,1)	5 (6,9)	0 (0)	2 (3,0)	2 (2,7)	2 (3,3)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrence/1000 mots)	10 (1,44)	16 (2,18)	12 (1,65)	17 (2,43)	5 (0,76)	11 (1,48)	12 (1,66)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrence/100 mots)	25 (36)	26 (35,4)	27 (37,1)	28 (40)	28 (42,3)	25 (33,6)	27 (37,4)
Nombre de mots en <i>-ing</i> (occurrence/1000 mots)	15 (21,6)	17 (23,2)	18 (24,7)	16 (22,9)	16 (24,2)	18 (24,2)	17 (23,5)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	26,49	32,62	35,34	28,67	30,13	33,42	31,11
TermBase – % de mots remplacés	48	29	35	45	31	26	36
TermBase – Nombre de phrases avec 100 % remplacement (%)	20 (34)	13 (22)	25 (37)	23 (37)	16 (29)	14 (22)	19 (30)
TransCorpora – % de mots trouvés	99	88	95	100	92	87	94
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)	6 (10)	7 (12)	15 (22)	11 (17)	5 (9)	19 (29)	11 (17)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	31 (53)	28 (47)	44 (65)	42 (67)	23 (42)	36 (55)	34 (55)

CR	MT1	MT2	TA1	TA2	TH1	TH2	Moy.
Mots par phrase	18,7	14,8	14,9	13,0	12,9	17,8	15,4
Degré de lisibilité (100=très facile)	50,6	46,4	44,9	35,5	39,0	47,0	43,9
Voix passives (%/verbes conjugués)	22	17	16	12	18	6	15
Nombre d'occurrences – pron. pers. <i>them, they, it</i> (occurrence/1000 mots)	8 (7,3)	0 (0)	9 (15,5)	2 (3,1)	3 (4,3)	7 (10,3)	5 (6,8)
Nombre d'occurrences – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i> (occurrence/1000 mots)	0 (0)	0 (0)	2 (3,5)	0 (0)	0 (0)	0 (0)	0 (0,6)
Nombre d'occurrences – pron. rel. <i>who, whom, which, that, whose</i> (occurrence/1000 mots)	39 (35,7)	23 (31,3)	15 (25,9)	15 (23,4)	19 (27,3)	21 (31,0)	22 (29,1)
Nombre de mots en <i>-ing</i> (occurrence/1000 mots)	24 (22,0)	15 (20,4)	16 (27,6)	9 (14,1)	15 (21,6)	16 (23,6)	16 (21,6)
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	37,46	40,93	45,63	45,10	39,52	41,04	41,61
TermBase – % de mots remplacés	18	26	22	27	26	18	23
TermBase – Nombre de phrases avec 100 % remplacement (%)	14 (15)	10 (14)	15 (22)	9 (14)	10 (14)	14 (19)	12 (16)
TransCorpora – % de mots trouvés	74	82	68	74	81	76	76
TransCorpora – Nombre de phrases avec 75-99 % remplacement (%)	4 (4)	3 (4)	2 (3)	11 (17)	4 (5)	3 (4)	5 (6)
TransCorpora – Nombre de phrases avec 100 % remplacement (%)	31 (34)	27 (39)	25 (36)	21 (32)	23 (32)	25 (35)	25 (35)

## **Annexe E – Trousse remise aux volontaires**

Chères collègues,

Vous trouverez ci-joint un tableau contenant différentes données sur différents types de textes (nommés PDP et CR). Vous devrez vous référer à ce tableau pour répondre deux fois (1 fois pour chaque type de texte) à deux listes de questions auxquelles vous devrez répondre oui ou non.

Selon vos réponses et les instructions données à la fin de chaque liste de questions, vous devrez ensuite prédire quelle méthode de traduction (entre la traduction humaine, la traduction automatique et la mémoire de traduction) est la meilleure pour chaque type de textes N'oubliez pas d'indiquer vos prédictions au bas de chaque page.

Il n'y a pas de limite de temps; prenez tout le temps dont vous avez besoin.

Je vous demanderais toutefois de ne pas vous consulter. Si vous avez des questions, n'hésitez pas à me les poser. Vous pouvez m'appeler à la maison si jamais vous avez besoin d'aide en soirée.

Finalement, à la dernière page, je vous demande de me donner brièvement vos impressions.

Tableau de données

	PDP	CR
Mots par phrase	17,7	15,4
Degré de lisibilité (100=très facile)	39,1	43,9
Voix passives (%/verbes conjugués)	13	15
Nombre d'occurrences/1000 mots – pron. pers. <i>them, they, it</i>	3,3	6,8
Nombre d'occurrences/1000 mots – pron. poss. <i>mine, yours, hers, his, its, ours, theirs</i>	16,6	0,6
Nombre d'occurrences/1000 mots – pron. rel. <i>who, whom, which, that, whose</i>	37,4	29,1
Nombre d'occurrences/1000 mots - mots en <i>-ing</i>	23,5	21,6
Rapport entre le nombre de mots différents et le nombre total de mots (Type/token ratio)	31,11	41,61
TermBase – % de mots remplacés	36	23
TermBase – Taux de phrases avec 100 % remplacement	30	16
TransCorpora – % de mots trouvés	94	76
TransCorpora – Taux de phrases avec 75-99 % remplacement	17	6
TransCorpora – Taux de phrases avec 100 % remplacement	55	35

## Type de textes: PDP

**\*Vous remarquerez qu'il n'y a pas de liste de vérification pour la traduction humaine. C'est parce que nous avons adopté le principe de déduction selon lequel si une traduction ne se prête ni à la traduction automatique ni aux mémoires de traduction, elle se prête alors probablement mieux à la traduction humaine.**

<i>Liste de vérification pour la traduction automatique</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il <b>25 mots par phrase ou moins</b> ?		
Le <i>degré de lisibilité</i> est-il <b>égal ou supérieur à 35</b> ?		
Le nombre de <i>verbes conjugués à la voix passive</i> est-il <b>inférieur à 20 %</b> ?		
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 5 pronoms possessifs</b> pour 1000 mots?		
Le texte compte-t-il <b>au moins 20 pronoms relatifs</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 20 mots en -ing</b> pour 1000 mots?		
Le <i>type/token ratio</i> est-il <b>inférieur à 40 %</b> ?		

Si vous avez répondu OUI à au moins 6 de ces questions, ce type de textes se prête probablement bien à la traduction automatique (TA).

<i>Liste de vérification pour les mémoires de traduction</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il <b>20 mots par phrase ou moins</b> ?		
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 20 pronoms possessifs</b> pour 1000 mots?		
Le <i>type/token ratio</i> est-il <b>inférieur à 45 %</b> ?		
Le taux de <i>mots remplacés (TermBase)</i> est-il <b>égal ou supérieur à 20 %</b> ?		
Le taux de <i>phrases avec 100 % remplacement (TermBase)</i> est-il <b>égal ou supérieur à 15 %</b> ?		
Le taux de <i>mots trouvés dans TransCorpora</i> est-il <b>égal ou supérieur à 75 %</b> ?		
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 10 %</b> ?		
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 30 %</b> ?		

Si vous avez répondu OUI à au moins 6 de ces questions, ce type de textes se prête probablement bien aux mémoires de traduction (MT).

Si vous avez répondu NON à la majorité des questions des deux listes, ce type de textes se prête probablement mieux à la traduction humaine (TH).

**Selon vous, quelle est la meilleure méthode de traduction (TA, MT ou TH) pour ce type de texte?**

**Prédiction :** \_\_\_\_\_

## Type de textes : CR

**\*Vous remarquerez qu'il n'y a pas de liste de vérification pour la traduction humaine. C'est parce que nous avons adopté le principe de déduction selon lequel si une traduction ne se prête ni à la traduction automatique ni aux mémoires de traduction, elle se prête alors probablement mieux à la traduction humaine.**

<i>Liste de vérification pour la traduction automatique</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il <b>25 mots par phrase ou moins</b> ?		
Le <i>degré de lisibilité</i> est-il <b>égal ou supérieur à 35</b> ?		
Le nombre de <i>verbes conjugués à la voix passive</i> est-il <b>inférieur à 20 %</b> ?		
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 5 pronoms possessifs</b> pour 1000 mots?		
Le texte compte-t-il <b>au moins 20 pronoms relatifs</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 20 mots en -ing</b> pour 1000 mots?		
Le <i>type/token ratio</i> est-il <b>inférieur à 40 %</b> ?		

Si vous avez répondu OUI à au moins 6 de ces questions, ce type de textes se prête probablement bien à la traduction automatique (TA).

<i>Liste de vérification pour les mémoires de traduction</i>	<i>OUI</i>	<i>NON</i>
Le texte compte-t-il <b>20 mots par phrase ou moins</b> ?		
Le texte compte-t-il <b>moins de 10 pronoms personnels</b> pour 1000 mots?		
Le texte compte-t-il <b>moins de 20 pronoms possessifs</b> pour 1000 mots?		
Le <i>type/token ratio</i> est-il <b>inférieur à 45 %</b> ?		
Le taux de <i>mots remplacés (TermBase)</i> est-il <b>égal ou supérieur à 20 %</b> ?		
Le taux de <i>phrases avec 100 % remplacement (TermBase)</i> est-il <b>égal ou supérieur à 15 %</b> ?		
Le taux de <i>mots trouvés dans TransCorpora</i> est-il <b>égal ou supérieur à 75 %</b> ?		
Le taux de phrases pour lesquelles une correspondance floue (75-99 %) été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 10 %</b> ?		
Le taux de phrases pour lesquelles une correspondance parfaite (100 %) a été trouvée dans <i>TransCorpora</i> est-il <b>égal ou supérieur à 30 %</b> ?		

Si vous avez répondu OUI à au moins 6 de ces questions, ce type de textes se prête probablement bien aux mémoires de traduction (MT).

Si vous avez répondu NON à la majorité des questions des deux listes, ce type de textes se prête probablement mieux à la traduction humaine (TH).

**Selon vous, quelle est la meilleure méthode de traduction (TA, MT ou TH) pour ce type de texte?**

**Prédiction :** \_\_\_\_\_

## **Rétroaction**

**D'après vous, les listes de questions et les instructions données à la fin de chaque liste étaient-elles claires et faciles à consulter/suivre?**

**Sinon, avez-vous des suggestions pour les améliorer?**

**MERCI BEAUCOUP D'AVOIR ACCEPTÉ DE PARTICIPER À MON PROJET! ☺**

## **Annexe F – Rétroaction des volontaires**

**D'après vous, les listes de questions et les instructions données à la fin de chaque liste étaient-elles claires et faciles à consulter/suivre?**

V1 – Oui, assez faciles à suivre.

V2 – Oui.

V3 – Elles étaient claires, faciles à consulter et très bien détaillées pour chacun des tableaux.

**Sinon, avez-vous des suggestions pour les améliorer?**

Aucune suggestion