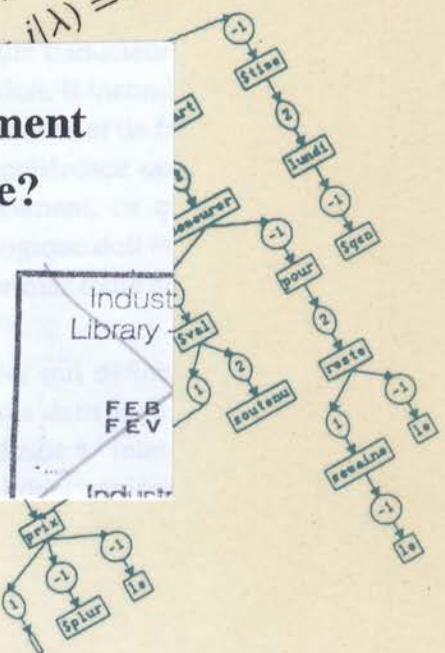


**Peut-on vérifier automatiquement
la cohérence terminologique?**

Elliott Macklovitch



CITI

Centre d'innovation
en technologies de l'information

Centre for Information
Technology Innovation

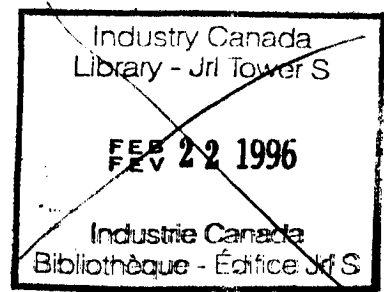
QUEEN
P
308
.M314
1995

Canada

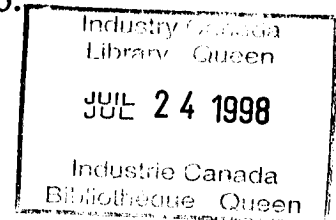
Industrie Canada
Centre d'innovation en technologies de l'information

**Peut-on vérifier automatiquement
la cohérence terminologique?**

Elliott Macklovitch



À paraître dans les Actes des IV^{es} Journées scientifiques,
Lexicommatique et Dictionnairiques, organisées par AUPELF-UREF,
Lyon, France, le 28-30 septembre 1995.



Laval
décembre 1995

Ce document fait état de travaux de recherche réalisés dans le cadre des activités du Centre d'innovation en technologies de l'information (CITI), Industrie Canada. Les opinions exprimées dans ce document n'engagent que l'auteur.

Also available in English.

© Copyright Industrie Canada 1995

No. catalogue Co28-1/125-1995F

ISBN No. 0-662-80729-4

P
308

. M314
1995

Peut-on vérifier automatiquement la cohérence terminologique?

Elliott Macklovitch¹

1. Introduction

Il arrive souvent dans les services de traduction que l'on soit obligé de morceler les textes volumineux, c'est-à-dire de les diviser et de les répartir entre plusieurs traducteurs, y compris des pigistes qui travaillent à l'extérieur du service. Dans une telle situation, il incombe généralement au réviseur de fusionner les parties traduites par ces différentes personnes et de faire en sorte que le tout soit cohérent. Le réviseur doit notamment s'assurer de la cohérence ou de l'uniformité terminologique, ce qui peut représenter une lourde tâche. Intuitivement, ce qu'on entend par cohérence terminologique est assez clair ici : chaque unité terminologique doit être traduite de la même façon partout dans le texte, dans le but, par exemple, de supprimer toute ambiguïté pour le lecteur.

La cohérence terminologique est généralement l'un des critères qui définissent une bonne traduction et, bien sûr, la situation décrite ci-dessus n'est pas la seule dans laquelle on doit s'en préoccuper. Au CITI, nous sommes en train de développer un outil d'aide à l'intention du réviseur (qui peut être le traducteur lui-même) dont le but est justement de valider certaines propriétés d'un texte traduit. Pour le moment, ce système, que l'on a baptisé TransCheck, est en mesure de déceler certaines des erreurs de traduction les plus courantes, comme les faux amis, les calques, les emprunts et certaines impropriétés. Il est important de noter que la plupart de ces erreurs sont généralement hors de portée des outils de vérification monolingue comme les correcteurs d'orthographe ou de grammaire, précisément parce que les erreurs de traduction sont d'ordre *bilingue* et dépendent des relations qui existent entre deux textes de langues différentes. Si TransCheck est en mesure de les déceler, par contre, c'est parce qu'il a été conçu expressément pour traiter ce genre de textes parallèles. Plus précisément, le système cherche à reconstituer le travail du traducteur humain en effectuant ce qu'on appelle un *alignement automatique* des deux textes – une opération qui consiste à rétablir explicitement les liens entre les segments d'un texte en LD et les segments correspondants dans sa traduction en LA. Pour une description détaillée du premier prototype de TransCheck, voir (Macklovitch 1994); mais voici en gros comment on envisage l'emploi d'un tel vérificateur de traductions.

Avant que TransCheck puisse vérifier une propriété d'une traduction, le texte source et le texte cible doivent être soumis au système pour que celui-ci en fasse l'alignement. Nous n'allons pas décrire ici les algorithmes dont se sert le programme pour calculer automatiquement les correspondances entre les deux textes. Mentionnons seulement que lorsque la résolution des alignements ne va pas au-delà de la phrase, le programme est d'une grande précision²; en outre, le programme peut traiter les cas où une phrase dans une langue est traduite par deux ou même trois

1. Voir les remerciements à la page 15.

2. Pour plus de détails sur l'algorithme d'alignement utilisé actuellement au CITI, voir (Simard et al. 1992).

phrases dans l'autre langue, et vice-versa. Suivant la terminologie utilisée par (Harris 1988), nous appellerons le produit d'un tel programme d'alignement un «bi-texte». Supposons maintenant qu'un réviseur désire valider une ébauche de traduction avant de la livrer au client, pour s'assurer qu'elle soit exempte d'interférence de la langue source, car c'est là une autre qualité qu'on associe généralement à une bonne traduction. Il pourrait faire appel à TransCheck pour l'aider dans cette tâche, de la même façon qu'un rédacteur monolingue utilise un correcteur d'orthographe pour s'assurer que ses textes sont exempts de fautes d'orthographe. Le premier prototype du système TransCheck du CITI incorpore une base de données d'environ 2 800 paires de mots qui ne doivent pas être pris pour des équivalents, dont un grand nombre des exemples classiques de faux amis comme «library//librairie» et «deception//déception». Ce que fait le système, concrètement, c'est de prendre tour à tour les entrées de cette base de données et, pour chacune, d'examiner le bi-texte produit par le programme d'alignement. Si, par exemple, il trouve un segment en langue source contenant «library» qui est aligné avec un segment en langue cible contenant «librairie», il marque cette paire alignée pour la soumettre à l'attention du réviseur. Au cours de la séance d'édition qui suit, le réviseur examine tous les segments marqués et apporte toutes les corrections nécessaires au texte cible, y compris des corrections que le système propose lui-même d'après l'information de chaque entrée de la base de données. (Une recopie d'écran d'une séance d'édition TransCheck est présentée à la page suivante).

Voilà essentiellement comment fonctionne TransCheck pour repérer les cas d'interférence de la langue source dans une ébauche de traduction. Nous travaillons sur plusieurs extensions du premier prototype de TransCheck qui, nous l'espérons, permettront au système de détecter automatiquement l'omission d'unités importantes du texte et de vérifier la transposition de divers types d'expressions numériques. Une autre application importante que nous explorons actuellement, c'est la possibilité d'utiliser TransCheck comme outil de vérification de la cohérence terminologique d'une ébauche de traduction. Nous présentons, dans ce rapport, les résultats de ces premières expériences sur la terminologie.

Encore ici, l'idée de base est très simple et, en fait, très voisine de l'approche utilisée pour détecter les cas d'interférence de la langue source. Supposons que nous prenions, comme point de départ, la définition naïve de cohérence terminologique donnée plus haut, à savoir que chaque occurrence d'un terme désigné du texte source doit être traduite exactement par le même terme dans la langue cible. Supposons aussi que le réviseur soit en mesure de dresser la liste des équivalences à vérifier avant de commencer la révision de l'ébauche de traduction. Il pourrait s'agir en fait d'un glossaire propre au texte traduit, dont chaque entrée serait un énoncé d'équivalence terminologique simple – quelque chose comme «e-terme1 = f-terme1». Comme précédemment, un bi-texte serait produit à partir du texte source et de l'ébauche de traduction. Ensuite, le système convertirait les entrées du glossaire d'équivalences terminologiques du réviseur en une série de requêtes TransCheck qu'il appliquerait tour à tour au bi-texte. Les segments alignés pour lesquels le système trouverait le terme source du glossaire, MAIS NON le terme cible correspondant, seraient, comme plus haut, marqués en vue d'être portés à l'attention du réviseur.

Nous avons décidé, pour tester cette idée, de réaliser une étude de faisabilité à petite échelle, sans nous faire d'illusion sur ce schéma qui nous semblait nettement trop simpliste pour mener à

TransCheck : une recopie d'écran

```

▽ *TransCheck Text*
Buffers File Edit Help
That is all I ask. Je n'en demande pas plus.

If you have a marketing and sales guy
to sell Dash-8s, where does he get
his direction right now? Le représentant des ventes et de la
commercialisation des Dash-8, de qui
reçoit-il ses directives en ce
moment?

The history of the boundary, the A-B
line and this body of water which is
sovereign and internal to Canada, is
one of a trail of bullying, broken
promises, treachery, deception, lies
and treason, and it carries on to
this very day. L'histoire de l'établissement de la
frontière, de la ligne AB et de ce
plan d'eau, qui fait partie des eaux
intérieures canadiennes, est marquée
par l'intimidation, les promesses
brisées, la trahison, la déception
ainsi que le mensonge, et ça se
continue jusqu'à ce jour.

As well, the 1987 CPP changes Les modifications apportées au RPC en
included the payment of two benefits 1987 permettaient, en outre, le

--*-Emacs: *TransCheck Text* (TransCheck)--55%-----
<Couple id=357>
<Error no=357.1>
  Compiled query: f(déception+/NomC\*) e(deception+/NC\*)
  Français incorrect: déception+/NomC
  Anglais: deception+/NC
  Français correct: tromperie ; duperie
  Type d'erreur: faux ami complet
  Source: Van Roey
</Error>

<Text>
CoupleInfo: Couple ID = 357,
            Document = "text",
            Source = English
tokens[L1] = {13567}
tokens[L2] = {15913}
The: DT L': Déte-ddéf-fémi-sing
history: NC histoire: NomC-fémi-sing
of: PP de: Prép
the: DT l': Déte-ddéf-masc-sing
boundary: NC établissement: NomC-masc-sing
,: PC de: Prép
the: DT la: Déte-ddéf-fémi-sing
A: ZZ frontière: NomC-fémi-sing
-: PC ,: Punc-pccm

--*-Emacs: *TransCheck Info* (TransCheck Narrow)--Top-----
Garbage collecting...done

```

un système opérationnel de vérification terminologique. Mais quelles seraient exactement ses faiblesses? Quels problèmes seraient soulevés et quelle serait leur importance relative? Lesquels de ces problèmes pourraient être résolus rapidement, et quels autres devraient attendre les résultats d'une longue recherche? Pour nous aider à répondre à ces questions, cruciales si nous voulons réellement mettre au point un vérificateur de cohérence terminologique opérationnel, nous avons décidé d'incorporer une version rudimentaire du schéma décrit ci-dessus au prototype actuel de TransCheck, et de l'appliquer à plusieurs traductions véritables.

2. L'étude de faisabilité

2.1 Méthodologie

Pour notre étude de faisabilité, nous avons essayé d'obtenir un certain nombre de textes de domaines variés, comportant chacun deux versions de la traduction en langue cible : une version préliminaire ou ébauche, et une version finale révisée. (Rappelons que nous voulons, grâce à TransCheck, signaler les erreurs possibles dans une *ébauche* de traduction.) Cependant, il s'est révélé plus difficile que prévu d'obtenir d'authentiques ébauches de traduction, peu importe la définition qu'on en donne. Quand on leur demande des textes pour mener des expériences sur la détection d'erreurs, les traducteurs et les services de traduction se montrent tout naturellement réticents à fournir des produits non finis, même si on les assure que ces textes serviront uniquement à la recherche et que les noms des personnes ne seront pas révélés. Il nous a fallu, par conséquent, faire certains compromis dans notre méthodologie, notamment en ce qui touche le premier texte que nous avons analysé.

Pour chacun des quatre textes sources que nous avons obtenus, nous avons choisi vingt-cinq des termes les plus fréquents ou les plus évidents, avec l'aide d'un programme-maison d'extraction de termes candidats appelé F-TERM. F-TERM se fonde sur les idées initialement proposées dans (Justeson et Katz 1993). Appliqué à un texte où la catégorie grammaticale de chaque mot a déjà été assignée automatiquement, il extrait des séquences de mots qui correspondent à une définition syntaxique de terme complexe (ou terme à mots multiples) : en anglais, il s'agit essentiellement d'un groupe nominal dépourvu de son déterminant et formé d'une chaîne de noms et/ou d'adjectifs se terminant par un nom, laquelle chaîne peut être suivie ou non d'un groupe prépositionnel. Encore une fois, précisons que le programme produit une liste de termes **candidats**, qui sont triés selon leur nombre d'occurrences dans le texte. En fait, la plupart de ces termes, à tout le moins au sommet de la liste, se révèlent des termes valides. En revanche, on ne retrouve pas dans la liste tous les termes d'un texte. D'une part, F-TERM ne relève pas les termes à mot unique, et d'autre part, les imperfections de l'étiquetage automatique peuvent entraîner l'inclusion de séquences qui ne sont pas des groupes nominaux bien formés et l'omission de séquences qui en sont. De surcroît, F-TERM n'a aucun critère pour distinguer un groupe nominal non lexical (ou descriptif) d'un terme en bonne et due forme, si ce n'est la répétition littérale de l'expression dans le texte³.

3. Voir (Justeson et Katz 1993) pour une analyse intéressante de la répétition comme critère permettant de déterminer ce qui constitue un «terme», et (Hannan 1995) pour une analyse critique des limites des approches fondées sur la syntaxe pour l'extraction automatique de termes.

Nous avons repéré la traduction de chaque terme source de la liste dans la version finale du texte correspondant en langue cible. Lorsque le texte en langue cible contenait des équivalents différents du même terme source, nous avons choisi le terme cible le plus fréquent, en faisant appel de temps à autre à TERMIUM, la banque de données linguistiques bien connue du gouvernement du Canada, pour nous aider dans notre choix. Ces traductions reflètent une partie des décisions du réviseur quant à la terminologie qui convenait au texte, et nous les avons formalisées sous forme d'énoncés d'équivalences terminologiques simples, puis converties en requêtes TransCheck, de la manière décrite ci-dessus. TransCheck a alors pu parcourir le bi-texte produit à partir de chaque texte source et de son ébauche de traduction, et signaler tous les cas dans lesquels un terme source n'a pas été rendu **exactement** par le terme cible indiqué, ou une variante flexionnelle du terme cible⁴. Enfin, par une analyse des résultats du système, c'est-à-dire en comparant les incohérences signalées par TransCheck et la terminologie de la version finale, nous espérons nous faire une meilleure idée des principales difficultés de cette façon de procéder⁵.

2.2 Les textes

Compte tenu de la difficulté que nous avons eue à obtenir des ébauches de traduction, nous avons décidé d'effectuer notre premier test de vérification de la cohérence terminologique sur un manuel des Forces armées de 80 000 mots portant sur la formation et le déploiement des tireurs d'élite, pour lequel nous avons l'original anglais et une version française finale, mais pas d'ébauche de traduction. Puisque le texte avait été publié dans les deux langues officielles, nous pouvions supposer que la cohérence de la terminologie de la version française finale avait déjà été vérifiée. Ces lacunes pouvaient sembler à première vue des obstacles majeurs; toutefois, à ce stade, nous étions plus intéressés à relever les types de bruit⁶ que TransCheck générerait dans sa vérification de la cohérence terminologique qu'à découvrir toute erreur isolée que le réviseur de ce texte aurait pu oublier.

La liste des vingt-cinq termes que nous avons choisis pour le texte «Tireurs d'élite» forme la première colonne du tableau 1, qui se trouve à la fin de ce rapport. La troisième colonne du tableau donne le terme cible (TC) qui correspond à chaque terme source dans la version française finale. La deuxième colonne indique le nombre total d'occurrences de chaque terme source (TS), et la quatrième colonne, le nombre de fois que ce terme source n'est PAS traduit exactement par le terme cible. Si nous prenons la première entrée du tableau, par exemple, nous voyons que «sniper» apparaît au total 1 277 fois dans le texte, et que de ce nombre, 105 occurrences ne sont pas traduites par «tireur d'élite». Les cinq autres colonnes du tableau 1 présentent une ventilation des cas où TS \neq TC. «Tête seulement» désigne les occurrences où un terme cible complexe est tronqué, c'est-

4. Par variance flexionnelle, on entend ici la forme plurielle des noms, la forme plurielle et la forme féminine des adjectifs, etc.

5. Même si nous savions, d'entrée de jeu, que les ébauches de traduction n'avaient pas fait l'objet d'une révision systématique, et donc que les versions finales pouvaient fort bien contenir des incohérences terminologiques qui n'avaient pas été corrigées.

6. Le mot «bruit» est employé ici dans un sens analogue à celui qui est le sien en recherche documentaire pour désigner des résultats qui ne sont pas pertinents, dans ce cas-ci parce qu'ils ne correspondent pas à des erreurs de traduction véritables.

à-dire que seulement la tête est utilisée plutôt que le terme complet; voir des exemples en (1) ci-dessous. La colonne suivante, «Pronom», comprend les cas où le texte français utilise un pronom ou un autre type d'anaphore plutôt que le groupe nominal cible complet; voir des exemples en (2) ci-dessous. La colonne intitulée «TS non traduit» désigne les cas où le terme source n'est pas réellement traduit dans le texte cible : ou bien il est entièrement omis, ou bien, dans certains cas, la traduction contient une périphrase au lieu d'un terme équivalent; voir des exemples de chacune de ces possibilités en (3) ci-dessous. La colonne «Autre TC» désigne les cas où le texte français utilise un autre terme que celui indiqué dans la troisième colonne. Quelques cas de ce genre, qui sont peut-être de véritables incohérences terminologiques, ont été relevés dans le texte Tireurs d'élite, mais ils ne sont pas très nombreux, ce qui est normal puisqu'il s'agit d'un texte publié; des exemples possibles sont donnés en (4) ci-dessous. La dernière colonne, «Autres» vise tous les cas qui ne sont pas compris dans les catégories précédentes. Cela comprend des cas de coordination et d'autres constructions grammaticales qui ont pour effet de fragmenter le terme cible (voir des exemples en (5) ci-dessous), des cas de bruit du système attribuables aux problèmes d'étiquetage, et diverses coquilles, comme des fautes d'orthographe ou de dactylographie, qui sont détectées par TransCheck mais ne constituent pas vraiment des incohérences *terminologiques*.

- (1a) ... the **sniper** moves his head back and forth =>
... le **tireur** déplace la tête vers l'avant ou vers...
- (1b) The Unertl **telescopic sight** is a fixed 10 power... =>
La **lunette** Unertl grossit 10 fois et...
- (2a) The **sniper** places his hand to his chest... =>
Il place la main contre sa poitrine...
- (2b) **Snipers** will infiltrate enemy areas... =>
Ces derniers infiltrèrent les secteurs de l'ennemi...
- (3a) IDENTIFYING **SNIPER TARGETS** =>
IDENTIFICATION DES CIBLES
- (3b) Adjustment for **eye relief** should be made by... =>
L'ajustement de la distance entre l'oeil et la lunette se fait en...
- (4a) The sniper can insert a pad on the **ghillie suit**... =>
Le tireur d'élite peut insérer un coussin dans sa **tenue de camouflage**...
- (4b) ... which is supported from underneath by the top of the **trigger**... =>
... qui est supportée du dessous par la tête du **percutateur**...
- (5a) direct or **indirect fire** => **tirs** direct ou **indirect**
- (5b) Determining correct **eye relief** => Déterminer le **dégagement** correct de l'**oeil**

Le deuxième texte examiné, que nous avons baptisé «Lait», porte sur les produits laitiers. Il s'agit d'une analyse économique du ministère de l'Agriculture sur la compétitivité de l'industrie canadienne du lait et des produits laitiers. La traduction avait été confiée en sous-traitance à un cabinet, mais avant d'être retourné au ministère client, le texte a fait l'objet d'une révision sommaire appelée assurance de la qualité, faite par une traductrice principale du Bureau de la traduction du

gouvernement. Bien que le cabinet n'ait pas précisé combien de traducteurs avaient participé à ce travail, tout porte à croire qu'au moins deux personnes étaient intervenues : il y avait un changement de police de caractères au milieu du fichier livré par le cabinet, lequel coïncidait avec un changement de la terminologie du texte cible pour plusieurs termes sources. L'une des préoccupations du réviseur du gouvernement, par conséquent, était de s'assurer de la cohérence de la terminologie du début à la fin de la version française finale. Le tableau 2, à la fin de cet article, indique les vingt-cinq termes sources que nous avons choisis pour ce texte, leurs équivalents en langue cible et la ventilation de tous les cas de non-correspondance.

Le troisième texte, que nous appelons «PPCI», fait partie d'un guide du ministère des Affaires extérieures décrivant les politiques et pratiques du ministère pour l'organisation de promotions commerciales dans le cadre du Programme de promotion du commerce international (PPCI) du gouvernement. L'original anglais, comprenant environ 16 000 mots, se divise en deux parties distinctes : une description bien structurée, en style administratif, du programme et de son application, puis une série d'annexes disparates regroupant des formulaires du gouvernement, des questionnaires à choix multiples, des contrats types, etc. Dans ce cas, nous savions que les deux parties avaient été attribuées à des traducteurs différents du Bureau de la traduction du gouvernement. Les termes sources choisis pour ce texte, ainsi que leurs équivalents en langue cible, figurent au tableau 3.

Le dernier texte analysé, que nous appelons «Grappes», ne comptait que 4 235 mots, mais avait quand même été attribué à deux traducteurs en raison du délai très court dont on disposait. Il s'agissait d'un appel de propositions du ministère de l'Industrie, des Sciences et de la Technologie relatif à un type particulier d'étude économique devant être réalisé dans l'Ouest canadien et faisant appel au concept d'analyse des grappes. Les parties de l'appel de propositions décrivant le travail à accomplir (produits à livrer, dates, critères de sélection, etc.) sont destinées au grand public et ne sont pas très techniques; en revanche, le texte présente de l'information de référence sur l'analyse des grappes et porte donc indirectement sur un domaine relativement spécialisé des sciences économiques. Le tableau 4 présente les termes sources que nous avons retenus dans le texte Grappes, ainsi que leurs équivalents en langue cible et la ventilation des non-correspondances.

Avant de passer à l'analyse détaillée des résultats produits par TransCheck pour ces quatre textes, il y a une importante question de définition que nous aimerions soulever concernant l'utilisation du mot «terme». Au moment d'amorcer cette étude, nous avons supposé que les unités qu'un réviseur voudrait vérifier dans une ébauche de traduction seraient les termes en bonne et due forme, au sens technique de l'expression, c'est-à-dire les éléments d'un vocabulaire spécialisé décrivant les concepts et relations fondamentaux d'un domaine spécialisé. Cette hypothèse, cependant, n'était pas tout à fait justifiée. En effet, l'une des choses que nous avons découvertes dans nos consultations avec le réviseur ayant effectué l'assurance de la qualité de trois de nos textes, c'est que la distinction terme/non-terme n'est pas critique dans le travail de vérification de la cohérence d'une traduction. Certes, une grande partie des unités dont le réviseur avait tenu à vérifier la cohérence dans ces ébauches de traduction étaient des termes (avec fiches TERMIUM), mais certaines ne l'étaient pas et, dans d'autres cas, il était difficile de déterminer si l'on avait affaire à un terme ou non. En fait, la distinction entre termes spécialisés et mots du vocabulaire général

dans un texte donné n'est pas toujours évidente, même pour un terminologue⁷. Le lecteur devrait garder ce fait à l'esprit en examinant les listes de termes sources des tableaux 1 à 4, car certains de ces termes – les acronymes, par exemples, qui abondent dans le texte PPCI – peuvent, à première vue, ne pas sembler des termes. Nous ne prétendons pas que les 100 unités que nous avons choisies pour notre étude de faisabilité constituent effectivement des termes au sens propre du mot. De même que Martin Kay a défini la traduction comme étant *ce que les traducteurs font*, nous voulons que TransCheck soit un outil d'aide pour les réviseurs dans ce qu'ils font. En ce qui touche la vérification de la cohérence terminologique, le réviseur nous a signalé que l'objectif primordial était d'éviter la confusion chez le lecteur. Que les unités devant être rendues uniformément pour atteindre cet objectif soient des termes en bonne et due forme ou non, c'est une question plus ou moins accessoire. En l'absence d'une solution de rechange commode, toutefois, nous continuerons d'utiliser le mot «terme» dans cet article, mais en nous affranchissant, si possible, de sa définition technique.

3. Résultats

Le tableau 5 à la page suivante présente une synthèse des résultats fournis séparément aux tableaux 1 à 4. Les chiffres de ce tableau sommaire correspondent aux totaux des colonnes des tableaux 1 à 4, qui résultent de l'analyse des vingt-cinq paires de termes choisies dans chacun des textes (Tireurs d'élite, Lait, PPCI et Grappes).

Examinons, pour commencer, les chiffres de la troisième colonne du tableau 5, qui donnent les taux globaux de non-correspondance pour chaque texte, c'est-à-dire la proportion des occurrences de termes sources non traduites exactement par les termes cibles figurant dans les glossaires d'équivalences des textes. Notons que le taux de non-correspondance du texte Tireurs d'élite, à 8,5 %, est sensiblement moins élevé que ceux des trois autres textes. Rappelons, toutefois, que la version française du texte Tireurs d'élite soumise à TransCheck était une version publiée qui, peut-on supposer, avait été vérifiée sous l'angle de la cohérence terminologique; dans le cas des trois autres textes, en revanche, la version soumise à Transcheck était une traduction préliminaire. Par conséquent, cette disparité des taux de non-correspondance n'est pas vraiment une anomalie; en fait, si notre hypothèse sur la façon dont les textes ont été traduits est exacte, c'est un résultat qui apparaît normal. Notons, par ailleurs, que les taux de non-correspondance des trois autres textes sont relativement voisins : 30 % pour les textes Lait et Grappes, et 24 % pour le texte PPCI.

Le tableau 5 révèle une autre différence importante entre les résultats obtenus pour le texte Tireurs d'élite et ceux obtenus pour les trois autres traductions : le nombre de cas où un terme cible est remplacé par un pronom ou un autre type d'anaphore est beaucoup plus élevé dans le texte Tireurs d'élite (35 occurrences sous «Pronom» contre 11 au total pour les trois autres textes); il en est de même, bien qu'à un moindre degré, pour le nombre d'occurrences sous «Tête seulement», c'est-à-dire de cas où le terme cible est tronqué (29 % de non-correspondances dans Tireurs d'élite, comparativement à 19 %, 18 % et 0 % pour les textes Lait, PPCI et Grappes respectivement).

7. Toutefois, on peut présumer que la distinction serait claire pour un expert dans le domaine de spécialité du texte. Voir (Ahmad et al. 1994) pour une analyse intéressante de la difficulté de définir ce qui constitue un terme dans le contexte de l'extraction automatique de termes.

Pourquoi il en est ainsi n'est pas évident. C'est peut-être que le terme «sniper» est si fréquent dans le premier texte que le recours à ce genre d'anaphore évite la répétition encombrante de «tireur d'élite» et permet d'alléger et de diversifier le style du texte cible⁸. L'explication tient peut-être encore plus à la réalité extra-linguistique : il n'y a qu'un seul référent possible pour «sniper» dans ce texte (le stagiaire type à qui s'adresse ce manuel), tandis que les entités auxquelles renvoient les termes tronqués dans les autres textes sont peut-être moins évidentes⁹.

Tableau 5 : Synthèse des résultats pour les quatre textes:

Texte / nombre de mots	Nbre. d'occ. de TS	TS =/= TC Nbre / %	Tête seul. Nbre / %	Pron. Nbre / %	Autre TC Nbre / %	TS non trad. Nbre / %	Autres Nbre / %
Tireurs d'élite / 80 000	1 799	153 / 8,5%	44 / 29%	35 / 23%	10 / 6,5%	54 / 35%	10 / 6,5%
Lait / 16 500	547	165 / 30%	31 / 19%	4 / 2%	36 / 22%	86 / 52%	8 / 4,8%
PPCI / 16 00	367	87 / 24%	16 / 18%	5 / 5,7%	46 / 53%	10 / 11%	10 / 11%
Grappes / 4 235	211	64 / 30%	0	2 / 3%	35 / 55%	17 / 27%	10 / 16%
Total / 116 735	2 924	469 / 16%	91 / 19%	46 / 9,8%	127 / 27%	167 / 36%	38 / 8%
Total sauf Tireurs d'élite	1 125	316 / 28%	47 / 15%	11 / 3,5%	117 / 37%	113 / 36%	28 / 8,8%

Si nous passons maintenant aux autres catégories du tableau 5, nous remarquons que la proportion de non-correspondances attribuables à l'emploi d'autres termes (Autre TC) est beaucoup plus élevée dans les textes PPCI et Grappes que dans les textes Tireurs d'élite et Lait. Il se trouve que les textes PPCI et Grappes sont les moins techniques des quatre textes que nous avons analysés; c'est du moins l'impression distincte qui ressort d'un survol de ces textes sources. En fait,

8. Il faut se garder de conclure qu'il y aura moins d'occurrences de «tireur d'élite» que de «sniper»; en fait, on observe le contraire dans ce texte. Les structures anaphore-antécédent permises relèvent des propriétés intrinsèques d'une langue et, fréquemment, ne sont pas transposées directement de la langue source à la langue cible. Les répétitions encombrantes auxquelles nous faisons allusion sont celles que le traducteur voudrait éviter pour répondre aux exigences stylistiques du texte cible.

9. Cette dernière hypothèse a été proposée par Pierre Isabelle.

nous pouvons peut-être confirmer cette impression au moyen des données produites par notre programme F-TERM. Rappelons que F-TERM repère les séquences de mots qui correspondent à une définition syntaxique de terme complexe. À l'instar de (Justeson et Katz 1993), nous supposons pour les fins de cet argument que la plupart des séquences complexes qui se répètent intégralement au moins deux fois dans un texte sont en fait des unités terminologiques. Dans le texte Grappes, par exemple, F-TERM détecte 377 termes candidats complexes, dont seulement 49 se répètent plus d'une fois dans le texte. Le nombre total d'occurrences correspondant à ces 49 termes candidats est de 125; si nous ajoutons les 43 occurrences du terme simple «cluster», qui, comme le mot «sniper», est le terme le plus fréquent du texte, nous obtenons un total de 168 occurrences de termes dans un texte de 4 235 mots. En divisant le nombre total de mots du texte par le nombre d'occurrences de termes candidats, nous devrions obtenir un corrélat très approximatif de la densité terminologique : dans le cas du texte Grappes, on obtient 25,21 mots par terme. En comparaison, le même calcul appliqué au texte Lait, qui est peut-être le plus technique des textes analysés, donne un ratio beaucoup moins élevé de mots par terme : 9,42. Le ratio du texte Tireurs d'élite est de 19,14, et celui du texte PPCI, qui est encore moins technique que Grappes, de 32,60. Certes, comme nous le disions plus haut à la section 2, les résultats produits par F-TERM ne sont pas entièrement fiables, mais la seule hypothèse à faire ici, c'est que F-TERM est constant dans ses faiblesses d'un texte à l'autre. Ces chiffres sont alors sans signification en valeur absolue, mais peuvent refléter la densité terminologique *relative* des quatre textes que nous avons analysés. Or, la densité terminologique est sûrement un élément qui contribue à la perception de *technicité* d'un texte, de sorte que nous pourrions être tentés de postuler une corrélation inverse entre elle et la tendance à permettre une plus grande variabilité terminologique dans un texte. Toutefois, la densité terminologique globale n'est certainement pas le facteur crucial qui détermine dans quels cas un réviseur décidera qu'un terme cible substitut est acceptable dans le contexte d'une traduction donnée. Notre analyse semble indiquer que les termes les plus susceptibles d'être traduits par de tels substituts sont les termes généraux et non techniques, dont le référent sera normalement évident pour tous les lecteurs. Par exemple, «la province» est un substitut acceptable de «gouvernement provincial» dans la plupart des contextes et, en fait, a été accepté par le réviseur dans le texte PPCI. Encore une fois, s'il n'y a aucun risque d'ambiguïté pour le lecteur, ce genre de variante peut contribuer à rendre le texte cible plus lisible. Ailleurs dans l'ébauche de traduction du texte PPCI, toutefois, les termes «project manager» et «reporting officer» sont tous deux rendus, dans certains cas, par «agent responsable»; ici, le réviseur a jugé bon de corriger l'équivalent substitut et de maintenir trois termes distincts, car ces diverses fonctions pourraient bien être assumées par des personnes différentes.

Un autre résultat à signaler, dans le tableau 5, est la forte proportion de cas, dans le texte Lait, où le terme source n'est pas traduit, mais plutôt omis ou remplacé par une périphrase dans le texte cible : dans ce texte, la catégorie «TS non traduit» représente 52 % des non-correspondances. En analysant les segments signalés par TransCheck dans l'ébauche de traduction et en les comparant à la version française finale, nous avons relevé plusieurs omissions flagrantes. Un exemple est donné en (6) ci-dessous (la flèche simple indique la traduction préliminaire et la flèche double, la traduction finale).

(6) ...Canadian consumption of **milk fat and milk proteins**... is relatively similar to that of the U.S. ...

-> ... la consommation canadienne est comparable à celle de...

=> ... la consommation individuelle canadienne de **matière grasse et de protéines du lait** est comparable à celle...

Dans de nombreux autres cas, cependant, le problème n'est pas tant attribuable à une omission du traducteur qu'à la verbosité du texte source. La version anglaise du texte Lait se distingue par une surabondance de longs composés nominaux pour lesquels, dans bien des cas, les éléments qui précèdent la tête sont superflus ou encore peuvent être déduits facilement du contexte, et ne sont donc pas essentiels à la compréhension de l'expression. Il se peut, dans de tels cas, que le traducteur décide d'omettre une partie du syntagme nominal complexe, afin de produire un texte cible à la fois plus allégé et plus clair. Un exemple simple est donné en (7) ci-dessous. Le terme cible pour «industrial milk» est «lait de transformation», qui est défini comme du lait cru qui sera transformé en fromage, en yogourt, etc.; l'inclusion de l'adjectif «raw» est donc superflue dans (7). TransCheck, évidemment, n'a aucun moyen de le savoir, et dès qu'il trouve une occurrence de «raw milk» non rendue par «lait cru», il signale une incohérence terminologique possible – à tort, dans ce cas.

(7) ... for increased shares of **industrial raw milk supplies** =>

... accroître leur part du marché de l'approvisionnement en lait de transformation...

Cet exemple soulève une importante question : comment savoir lesquelles des non-correspondances signalées par TransCheck dans les traductions préliminaires représentent de réelles incohérences terminologiques, et lesquelles correspondent à de «faux positifs»¹⁰ attribuables à une déficience de notre approche? Au départ, nous espérions pouvoir répondre à cette question en faisant la comparaison entre les segments signalés et la terminologie des traductions finales; toutefois, comme nous l'avons déjà indiqué, les traductions préliminaires n'ont pas fait l'objet d'une révision systématique, mais seulement d'un processus d'assurance de la qualité visant à corriger les erreurs les plus flagrantes. La personne ayant effectué l'assurance de la qualité a été très franche à ce sujet : elle a admis que les traductions finales pouvaient encore contenir certaines incohérences; malheureusement, elle n'avait disposé ni du temps ni des outils pour faire un travail plus complet.

Néanmoins, il y a des conclusions que nous pouvons tirer sur le rendement de TransCheck à partir de nos analyses détaillées de ces traductions, dont les résultats figurent au tableau 5. Par exemple, il est manifeste que la quasi-totalité des non-correspondances des catégories «Tête seulement» et «Pronom» sont attribuables à de faux positifs, c'est-à-dire à un bruit du système. Pratiquement aucune d'elles n'a été corrigée dans les versions françaises finales. Ensemble, ces deux catégories regroupent près de 30 % de tous les segments signalés par TransCheck, et, si l'on ajoute les cas de termes coordonnés classés dans la catégorie «Autres», nous pouvons sans crainte

10. «Faux positif» est un terme qui vient du domaine des méthodes de diagnostic; il désigne un individu qui, d'après les données ou les résultats de laboratoire, est classé dans un groupe de référence ou une catégorie de diagnostic auquel il n'appartient pas, notamment en raison d'une déficience méthodologique.

affirmer que ces non-correspondances comptent pour environ un cas sur trois¹¹. Par contre, une proportion beaucoup plus grande des non-correspondances des catégories «Autre TC» et «TS non traduit» ont effectivement été corrigées par le réviseur, sans doute parce qu'elles reflétaient de véritables incohérences terminologiques. Ces deux dernières catégories représentent globalement entre 63 % et 73 % de toutes les erreurs potentielles signalées, soit environ deux cas sur trois. Notons par ailleurs qu'il existe une différence assez fondamentale entre les non-correspondances de ces deux regroupements des catégories de notre taxonomie. Les non-correspondances du premier groupe proviennent toutes de l'emploi de variantes de la **forme** du terme désigné : simplification d'un terme complexe pour n'en garder que la tête, combinaison avec un autre terme dans une coordination et, cas moins évident, remplacement par un pronom ou un autre type d'anaphore. Par contre, les incohérences possibles signalées par le système en raison de l'omission du terme, ou de son remplacement par une périphrase ou un autre terme, ne concernent pas seulement la forme superficielle du terme; elles sont liées à sa redondance dans le texte, ou à la synonymie du terme avec une périphrase ou un substitut proposé – des facteurs qui touchent davantage le **sens** du terme. Nous allons revenir sur cette distinction dans la conclusion qui suit.

4. Conclusion

Dans (Bédard 1986, notamment au chapitre 2), l'auteur critique ce qu'il appelle «l'obsession des équivalents», c'est-à-dire la tendance à reproduire mécaniquement dans le texte cible les équivalents de tous les termes trouvés dans le texte source. Il en résulte invariablement, affirme l'auteur, des traductions de qualité inférieure. Le traducteur n'est pas lié par le mot-à-mot du texte source, mais par le sens du texte, et pour rendre ce dernier convenablement, il ne doit pas hésiter à se servir de son jugement professionnel. Pour ce qui est de la terminologie technique, le traducteur ne doit pas se sentir obligé d'utiliser «les équivalences directes ou toutes faites». Afin de produire un texte compréhensible et transparent dans la langue cible, il doit parfois modifier ou abrégé certains termes, omettre d'autres termes qui figurent dans le texte source, et même créer de nouveaux termes.

«Une réalité donnée peut souvent être décrite de plus d'une façon, et le discours technique n'échappe pas à cette vérité générale. Les termes employés par le locuteur, si techniques et si justes qu'ils soient, ne sont pas nécessairement les seuls dont il aurait pu se servir. Corrolaire : dans la même mesure, le traducteur n'est pas condamné à utiliser les équivalents directs pour faire passer le message.» (Bédard 1986, p.31)

Les arguments de Bédard semblent aller directement à l'encontre de l'approche automatisée que nous présentons ici pour la vérification de la cohérence terminologique. En effet, notre application de TransCheck se fonde sur une transposition littérale des termes du texte source dans le texte cible – du moins pour les termes indiqués par le réviseur dans le glossaire d'équivalences terminologiques. Il importe de rappeler, toutefois, que Transcheck est un **outil d'aide à la**

11. Nous pourrions inclure d'autres constructions syntaxiques qui, comme la coordination, ont pour effet de fragmenter une expression complexe que TransCheck cherche à reconnaître comme une chaîne continue de caractères. Quand un terme anglais est précédé d'un adjectif, comme dans (5b) ci-dessus, l'adjectif correspondant peut, en français, être placé entre la tête et son complément. Dans nos textes, ces cas étaient peu fréquents, toutefois.

traduction, non un système de traduction automatique; il n'a pas pour fonction d'imposer des termes au réviseur, mais seulement de l'aider à valider une traduction préliminaire. C'est le réviseur, et non le système, qui a le dernier mot. C'est au réviseur de décider s'il doit intervenir ou non pour chaque incohérence terminologique possible signalée par le système et, le cas échéant, quelle correction apporter. En revanche, l'approche visée par cette application de TransCheck suppose que, pour certains types de textes du moins, un degré élevé d'uniformité terminologique est souhaitable. Si ce n'était pas le cas, c'est-à-dire si les utilisateurs passaient outre systématiquement aux erreurs possibles signalées par TransCheck, ils auraient tôt fait d'abandonner le système.

Dans notre examen des résultats à la section précédente, nous avons vu que pour les 100 termes choisis dans les quatre textes analysés, TransCheck a produit un taux de non-correspondance global se situant entre 16 % et 28 % (selon qu'on inclut ou non dans le calcul le texte Tireurs d'élite). Nous avons également observé que, bien que notre méthodologie ne nous ait pas permis d'établir précisément dans quelle proportion ces non-correspondances étaient de réelles incohérences, environ le tiers des segments signalés par le système pouvaient être considérés comme de faux positifs attribuables à l'incapacité de TransCheck de reconnaître certaines variantes formelles des termes cibles. À cela s'ajoute un nombre indéterminé d'omissions ou de termes substitués que le réviseur peut décider, pour diverses raisons, de laisser tels quels. En bref, si ce prototype de TransCheck était mis à la disposition des utilisateurs dans sa forme actuelle, dans au moins un cas sur trois le système demanderait à l'utilisateur de vérifier des incohérences possibles qui, en réalité, n'exigeraient pas d'intervention. Est-ce un niveau de bruit que les utilisateurs sont prêts à tolérer? C'est difficile à dire; cela dépendrait du temps que le système permettrait de gagner par rapport aux autres erreurs signalées et des erreurs gênantes qu'il permettrait d'éviter¹². Mais une chose est sûre : la vérification terminologique automatisée serait acceptée beaucoup plus facilement si nous trouvions le moyen de réduire l'incidence du bruit actuellement engendré par le système. Cela pourrait bien être possible, notamment dans les cas attribuables à des variantes de la forme superficielle d'un terme. La tâche à accomplir, ici, est très claire : les conditions exigeant une identité formelle complète dans la définition de cohérence terminologique dans TransCheck doivent être assouplies, afin que le système puisse reconnaître au moins certaines de ces variantes formelles comme des représentations valides du terme cible complet. Dans quelle mesure nous parviendrons à le faire sans que nous échappe aucune des incohérences terminologiques réelles présentement détectées par le système, cela reste à voir. Mais il serait beaucoup plus difficile de concevoir et de rendre opérationnelles des stratégies permettant au système de distinguer entre les cas acceptables et inacceptables de termes cibles qui ont été omis, ou encore remplacés par un autre terme ou une périphrase. En fait, on pourrait se demander s'il n'est pas un peu abusif, dans ce cas, de décrire les cas acceptables comme des cas de cohérence *terminologique*. Quoi qu'il en soit, nous croyons qu'il est préférable de laisser ces cas au jugement du réviseur, afin qu'il puisse les examiner tour à tour et prendre les décisions pertinentes.

L'objectif visé par cette étude était de déterminer si oui ou non il est possible de vérifier la cohérence terminologique d'ébauches de traduction au moyen d'un outil comme TransCheck, en

12. Notre impression est que les utilisateurs tolèrent un niveau de bruit qui est souvent aussi élevé dans le cas des vérificateurs d'orthographe monolingues courants.

utilisant essentiellement la même approche que celle employée par le système pour détecter les problèmes d'interférence de la langue source. Nous étions conscients, bien sûr, qu'en appliquant notre définition naïve d'incohérence terminologique à des textes réels, nous relèverions à coup sûr des cas où la forme du terme cible, sans correspondre exactement à celle donnée dans le glossaire des équivalences, serait néanmoins acceptable. Certains travaux récents sur l'extraction automatique de termes décrivent de façon détaillée les diverses variantes d'un terme susceptibles d'être observées; voir en particulier (Daille 1994). Le but n'était pas de vérifier s'il existe de telles variantes, mais plutôt quelles formes elles prennent dans de véritables traductions, et avec quelle fréquence. Cette étude, nous l'espérons, a fourni un début de réponse à cette question, et nous permet de repérer un sous-ensemble de problèmes se prêtant à des solutions à court terme.

Notre façon d'aborder la vérification de la cohérence terminologique comporte (au moins) deux faiblesses qui n'ont pas encore été mentionnées. La première a trait à la directionnalité implicite de l'algorithme, c'est-à-dire le fait que TransCheck effectue sa recherche d'une paire de termes donnés en repérant d'abord une occurrence du terme source, puis en vérifiant le segment cible aligné pour y détecter la présence du terme cible. Mais, bien sûr, les termes sources peuvent aussi subir les mêmes variations de forme que les termes cibles : eux aussi sont l'objet d'ellipses, d'omissions, de pronominalisations, d'erreurs d'orthographe, etc. Dans chacun de ces cas, étant donné la stratégie de recherche d'équivalences exactes de TransCheck, le système passera outre au segment aligné sans s'occuper de vérifier le terme cible, réduisant ainsi le champ de couverture. Il y a une solution à laquelle on pense tout de suite pour résoudre ce problème : il s'agirait de convertir les entrées du glossaire des équivalences en requêtes bidirectionnelles, c'est-à-dire que pour chaque énoncé «e-terme1 = f-terme1», le système scruterait le bi-texte pour y trouver toute paire alignée dans laquelle le «f-terme» apparaît sans le «e-terme» correspondant. Cette stratégie, toutefois, présente un certain nombre de difficultés. L'une d'elle tient au fait que les traductions sont souvent plus explicites que les textes sources. Nous avons mentionné à la note 8, par exemple, que le terme «tireur d'élite» apparaît plus souvent dans le texte cible que «sniper» dans le texte source. On peut penser, par conséquent, que les requêtes inverses produiraient un niveau de bruit encore plus élevé que les requêtes initiales. Autre difficulté, plus fondamentale celle-là : il se peut que ces requêtes inverses ne correspondent pas aux équivalences terminologiques que le réviseur souhaitait vérifier. En stipulant qu'un «e-terme1» donné doit être traduit par «f-terme1», le réviseur ne voulait peut-être pas exclure la possibilité que d'autres termes sources aient aussi f-terme1 comme équivalent en langue cible. Or, la requête «f-terme1 *MAIS NON* e-terme1» entraînerait le marquage non désiré de telles équivalences.

Un autre problème général que pose notre approche de vérification terminologique, c'est qu'il oblige le réviseur à fournir au système un énoncé explicite de toutes les équivalences terminologiques à vérifier. Sans aucun doute, il surviendra des cas où le réviseur n'aura pas le temps de le faire, mais souhaitera quand même faire valider la cohérence terminologique de la traduction avant de l'envoyer au client. Ce genre de vérification est possible pour les erreurs d'interférence de la langue source, car TransCheck incorpore une base préexistante d'équivalents interdits connus. On pourrait peut-être procéder de la même façon pour la vérification de la cohérence

terminologique, quand le texte à valider appartient manifestement à un domaine précis pour lequel un glossaire terminologique fiable existe déjà.

Une autre possibilité intéressante a récemment été soulevée dans (Dagan et Church 1994). Leur système *Termight* extrait des termes candidats d'un texte source (comme notre programme F-TERM), puis repère automatiquement un équivalent probable dans la traduction, grâce à un programme d'alignement de mots qu'ils ont mis au point. *Termight* pourrait être adapté à la vérification de la cohérence terminologique, suggèrent les auteurs, si on lui confiait la tâche d'attirer l'attention de l'utilisateur sur les paires de termes correspondant aux cas où il existe plus d'un équivalent pour un même terme source. Notons que ces non-correspondances incluraient toutes les variantes formelles qui amènent actuellement TransCheck à signaler de faux positifs. En outre, l'utilisateur ne voudrait pas que le système omette de signaler les cas où toutes les occurrences d'un terme source ont été systématiquement traduites par un équivalent identique, mais **inexact**; nous avons, en fait, observé plusieurs cas de ce genre dans le texte PPCI. Bien que l'intervention du réviseur pour établir le glossaire d'équivalences terminologiques comporte des inconvénients, elle assure non seulement la cohérence de la terminologie vérifiée, mais aussi son exactitude.

REMERCIEMENTS

Plusieurs collègues du groupe de traduction assistée par ordinateur du CITI m'ont apporté leur soutien, aussi bien technique que linguistique, tout au long de la réalisation de cette étude : Pierre Isabelle, directeur du groupe, Michel Simard, Marie-Louise Hannan, François Perrault et Jean-Marc Jutras. J'ai beaucoup apprécié leur aide, et aucun d'eux, évidemment, n'est responsable de toute erreur de fait ou de jugement que pourrait contenir ce rapport. Je tiens aussi à remercier spécialement le réviseur anonyme qui m'a aidé à obtenir les traductions analysées et qui a pris le temps, avec beaucoup de patience, de répondre à mes nombreuses questions.

RÉFÉRENCES

- AHMAD, Khurshid et al. (1994), «What is a term? The semi-automatic extraction of terms from text», M. Snell-Horby et F. Pochhacker (dir.), *Translation Studies: An Interdiscipline*, Selected Proceedings of the Vienna Conference, pp.9-12.
- BÉDARD, Claude (1986), *La traduction technique : principes et pratique*, Montréal, Linguatex.
- DAILLE, Béatrice (1994), *Approche Mixte pour l'extraction de terminologie : statistique lexicale et filtres linguistique*, Thèse de doctorat en informatique fondamentale, Université Paris VII.
- DAGAN, Ido et Ken CHURCH (1994), «*Termight: Identifying and Translating Technical Terminology*», in Proceedings of EACL.
- HANNAN, Marie-Louise (1995), *The Use of Semantic Classes for Automatic Terminology Extraction*, Mémoire de maîtrise en linguistique, Université de Montréal.
- HARRIS, Brian (1988), «Bi-text: A New Concept in Translation Theory», dans *Language Monthly*, no 54.
- ISABELLE, Pierre et al.(1993), «Translation Analysis and Translation Automation», dans les Actes de la *Fifth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-93)*, Kyoto, pp.201-217.
- JUSTESON, John et Slava KATZ (1993), «Technical Terminology: some linguistic properties and an algorithm for identification in text», *Technical Report #RC 18906 (82591)*, IBM T.J. Watson Research Center, Yorktown Heights, New York, 13 p.
- MACKLOVITCH, Elliott (1994), «Using Bi-textual Alignment for Translation Validation: the TransCheck System», dans les Actes de la *First Conference of the Association for Machine Translation in the Americas (AMTA-94)*, Columbia, Maryland, pp. 157-168.
- SIMARD, Michel, G. FOSTER et P. ISABELLE (1992), «Using Cognates to Align Sentences in Parallel Corpora», dans les Actes de la *Fourth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-92)*, Montréal, pp.67-81.

Ventilation des non-correspondances : le texte «Tireurs d'élite»

Tableau 1:

Terme Source (TS)	Nbre	Terme Cible (TC)	TS =/= TC	Tête seul.	Pron.	Autre TC	TS non traduit	Divers
sniper	1277	tireur d'élite	105	32	35		38	
observation post OP	52	poste d'observation PO	1					1
magazine	45	chargeur	0					
trigger	40	détente	4			1	1	
point of aim	33	point de visée	1					1
telescopic sight	32	lunette de visée de tir	11	9		2		
spotting telescope scope	30	lunette d'observation	2	2				
windage	25	dérive	2			1	1	
chamber	25	chambre	2				2	
ghillie suit	22	tenue caméléon	2			1		2
crosshair	22	fil croisé	0					
indirect fire	18	tir indirect	2					2
night vision device	17	dispositif appareil de vision nocturne	2					2
sight picture	16	image de visée	1				1	
zeroing	16	zérotagage	1				1	
eye relief	15	dégagement oculaire de l'oeil	4	1			2	1
cocking piece	15	talon de percuteur	0					
stripping	14	démontage	3				3	
bolt sleeve	13	manchon de culasse	4			3	1	
range card	12	croquis de repérage	0					
mean point of impact MPI	11	point moyen d'impact PMI	0					
bore	10	âme	4			1	3	
fire support	7	feux d'appui (?)	1			1		
main body	6	gros des troupes	0					1
handstop hand stop	6	cale-main	1				1	
Total (excluding "sniper")	502		48	12	0	10	16	10

Ventilation des non-correspondances : le texte «Lait»

Tableau 2:

Terme Source (TS)	Nbre	Terme Cible (TC)	TS =/= TC	Tête seul.	Pron.	Autre TC	TS non traduit	Divers
dairy product	133	produit laitier	47	10	2		34	1
processing	71	transformation transformer	20			7	13	
raw milk	58	lait cru	12	3			9	
competitive position	30	position concurrentielle	10			4	6	
dairy farm	25	exploitation laitière x 5	16	5		5	6	
		exploitation agricole x 4						
supply management	23	gestion de l'offre	6	2	1	1	1	1
industrial milk	22	lait de transformation	14	1		6	3	4 conj
agri-food	18	agroalimentaire	0					
fluid milk	17	lait de consommation	3	1			1	1 conj
cost competitiveness	15	compétitivité des/... coûts	7	4		2	1	
market share	14	part du marché	1	1				
milk powder	13	poudre de lait	2			1	1	
economy of scale	13	économie d'échelle	2				2	
factor condition	11	facteur relatif à la production	4			1	3	
milk fat butterfat	11	matière grasse	2				2	
premium	10	haut de gamme	4			2	2	
demand condition	9	évolution de la demande	3	2		1		
supporting industry	9	secteur auxiliaire	1			1		
chance events	8	impondérables	1					1
linkages	8	ramifications	1			1		
Industry Competitiveness Group	8	Groupe de la compétitivité du secteur	3	2	1			
pharmaceuticals	6	médicaments	2			1	1	
(bovine) semen	6	sperme	1			1		
exchange rate	5	taux de change	2			2		
regional focus	4	caractère régional	1				1	
Total	547		165	31	4	36	86	8

Ventilation des non-corrépondances : le texte «PPCI»

Tableau 3:

Terme Source (TS)	Nbre	Terme Cible (TC)	TS ==/= TC	Tête seul.	Pron.	Autre. TC	TC non traduit	Divers
post(s)	47	mission diplomatique	18	14	1	1		2
DFAIT	47	MAECI	1				1	
recruitment / recruit	26	recrutement / recruter	13			10	3	
PIBD	26	PPCI	1					1
trade fair	24	foire commerciale	1			1		
direct cost	18	frais direct	5			3	2	
cost sharing	16	partage des frais	2		1	1		
cost sharing fee	5	frais de participation?	2			2		
briefing session / meeting	13	séance / réunion d'information	2	1		1?		
reporting officer	12	agent chargé de rédiger les rapports	12			12		
ITBP	12	PPCE	1					1
OGD(s)	12	autre(s) ministère(s)	0					
incoming (trade) mission	12	mission (commerciale) étrangère au Canada	5	1			2	2 conj
outgoing (trade) mission	9	mission (commerciale) canadienne à l'étranger	5			1		4 conj
provincial government	11	gouvernement provincial	5			5		
market information	11	information commerciale	1				1	
geographic bureau(x)	11	direction générale géographique	1			1		
geographic division	9	direction géographique	5			5		
TAMS	9	TAMS x 4						
		SGAC x 5						
fact sheet	8	fiche d'information	1			1		
WIN	8	WIN	0					
project manager	7	gestionnaire de projet	1			1		
project officer	4	agent de projet	1			1		
TOS	4	TOS	2		1		1	
cost recovery	3	recouvrement des coûts	1		1			
TOC	3	TOC	1		1			
Total	367		87	16	5	46	10	10

Ventilation des non-correspondances : le texte «Grappes»

Tableau 4:

Terme Source (TS)	Nbre	Terme Cible (TC)	TS ==/= TC	Tête seul.	Pron.	Autre TC	TS non traduit	Divers
cluster	43	grappe	0					
western Canada	15	Ouest canadien	4			4		
western Canadian	10	Ouest canadien	2			2		
linkages	12	liens	0					
economic development	11	développement économique	0					
sponsor	10	parrain	1			1		
industrial cluster	9	grappe industrielle	0					
enabling technology	8	technologie clé	1					1
system of innovation	8	système ... d'innovation	0					
proponent	8	auteur de/d'une proposition	8			1	7	
jurisdiction	7	entité politique	7			6	1	
comparative advantage	6	avantage comparatif	3			3		
cluster analysis	6	analyse des grappes	6			6		
strategy development	6	élaboration de(s) stratégies	4		1	3		
technology cluster	6(5)	grappe de technologie	5			5		
cluster strategy	5	stratégie de groupement	2				2	
advanced materials - and / - manufacturing	5	matériaux nouveaux / fabrication avancée	4					4
global economy	5	économie mondiale	1			1		
industrial structure	5	structure industrielle	0					
cluster opportunity	5	possibilité de formation de grappes	5				5	
competitive advantage	4	avantage concurrentiel	1					1
technology focus	4	pôle technologique	2				2	
technology infrastructure	4	infrastructure technologique	2					2 conj
stakeholder	3	gens concernés	3			3	1	
project budget	3	budget du projet	2				1	1
new economy	3	nouvelle économie	1		1			
Total	211		64		2	35	17	10



1575, boulevard Chomedey, Laval (Québec), H7V 2X2 Téléphone: (514) 973-5700 Télécopieur: (514) 973-5757

1575, Chomedey Boulevard, Laval, Quebec, H7V 2X2 Telephone: (514) 973-5700 Facsimile: (514) 973-5757