



Master

2018

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Intégration de la TA au domaine du jeu-vidéo : Spécialisation de Lucy LT
pour World of Warcraft

Fasel, Adélie

How to cite

FASEL, Adélie. Intégration de la TA au domaine du jeu-vidéo : Spécialisation de Lucy LT pour World of Warcraft. Master, 2018.

This publication URL: <https://archive-ouverte.unige.ch/unige:113752>

Adélie Fasel

*Intégration de la TA au
domaine du jeu vidéo :*

*Spécialisation de Lucy LT pour
World of Warcraft*

Directrice : Prof. Pierrette Bouillon

Jurée : Johanna Gerlach

*Mémoire présenté à la Faculté de Traduction et d'Interprétation
(Département Technologies de la traduction) pour l'obtention du
Master en traduction, mention technologies de la traduction*

Université de Genève

Juin 2018



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DE TRADUCTION
ET D'INTERPRÉTATION

Déclaration attestant le caractère original du travail effectué

J'affirme avoir pris connaissance des documents d'information et de prévention du plagiat émis par l'Université de Genève et la Faculté de traduction et d'interprétation (notamment la Directive en matière de plagiat des étudiant-e-s, le Règlement d'études de la Faculté de traduction et d'interprétation ainsi que l'Aide-mémoire à l'intention des étudiants préparant un mémoire de Ma en traduction).

J'atteste que ce travail est le fruit d'un travail personnel et a été rédigé de manière autonome.

Je déclare que toutes les sources d'information utilisées sont citées de manière complète et précise, y compris les sources sur Internet.

Je suis consciente que le fait de ne pas citer une source ou de ne pas la citer correctement est constitutif de plagiat et que le plagiat est considéré comme une faute grave au sein de l'Université, passible de sanctions.

Au vu de ce qui précède, je déclare sur l'honneur que le présent travail est original.

Nom et prénom :

Fasel Adélie

Lieu / date / signature :

Genève, mai 2018 *Atarel*

Remerciements

Je souhaiterais avant tout remercier ma directrice de mémoire, Mme Pierrette Bouillon, pour son soutien, ses précieux conseils, et sa motivation sans faille dès l'ébauche de ce projet. Je remercie également Mme Johanna Gerlach d'avoir accepté d'être ma jurée et de suivre ce projet.

Merci à mes parents Simon et Anne, ma sœur Clarie et mon frère Etienne pour tout le soutien et la patience dont ils ont su faire preuve, ainsi que l'intérêt qu'ils ont porté à ce projet passionnant, et tout particulièrement à mon père, qui m'a ouvert les portes de cet univers fantastique qu'est le MMORPG. Merci également à ma grand-mère Christiane de m'avoir tenue à bout de bras durant ces longs mois et d'être si souvent venue illuminer mes journées d'étude.

Je remercie également mes amies et collègues Marie et Tatiana pour le réconfort qu'elles ont su m'apporter dans les moments de doute, ainsi que l'équipe de l'École Mosaic qui m'a permis de prendre toutes les dispositions nécessaires pour combiner mon travail et mes études durant de nombreuses années.

Merci également à mes amis et joueurs francophones de *World of Warcraft* qui ont pris sur leur temps libre afin de m'apporter leurs conseils et suggestions quant aux traductions les plus courantes de certains mots, puis ont participé à l'évaluation du fruit de ce travail : Alhyia, Anianta, Codéine, Crapolion, Crunshy, Dunette, Eckoh, Eiko, Emmaskull, Eolia, Hdeusø, Jonas, Kah, Kiwi, Lionel, Louis, Michgasiglia, Mistake, Poldari, Rêvelune, Shindra, Stigmard, Stylard, Tismael, Vincent, Yacinth, et de nombreux anonymes.

Enfin, je souhaiterais témoigner ma reconnaissance à l'ensemble des joueurs de la guilde Rébellion (serveur privé « Nostalgeek ») pour leur chaleureux accueil de ces derniers mois, leur patience à toute épreuve et l'immersion qu'ils m'ont permis de réaliser au sein du jeu *World of Warcraft*. Merci tout particulièrement à Atchiko, Baøc, Barbehache, Izyon, Kamalphy, Magedubled, Schnau et Soradaï pour leur soutien et l'intérêt qu'ils ont porté à ce travail.

À Émile. Que j'aurais aimé que tu puisses suivre ce projet à mes côtés.

Table des matières

I. Introduction.....	6
I.1 Objectifs	6
I.2 Question de recherche et méthodologie	6
I.3 Les jeux-vidéo	7
I.4 Plan du mémoire.....	10
II. La traduction automatique (TA)	11
II.1 Présentation de la TA.....	11
II.2 L’histoire de la TA.....	12
II.3 Les différents systèmes de TA	14
II.3.1 Les systèmes linguistiques.....	15
II.3.2 Les systèmes statistiques traditionnels (par segments).....	20
II.3.3 Les systèmes neuronaux	21
II.4 Les problèmes rencontrés en TA	22
II.5 Présentation de Lucy LT	24
II.5.1 Lucy LT : les règles lexicales	27
II.5.2 Lucy LT : les règles semi-structurales.....	29
II.6 Présentation de DeepL	31
II.7 Conclusion	32
III. Développement du corpus et spécialisation de Lucy LT	33
III.1 Introduction	33
III.2 Développement du corpus de texte.....	33
III.3 Méthode	38
III.4 Spécialisation de Lucy LT	43
III.4.1 Mots simples	44
III.4.2 Mots complexes.....	48
III.4.3 Choix des traductions	51

III.4.4 Structures complexes.....	55
III.4.5 Synthèse de la spécialisation	58
III.5 Conclusion	60
IV. Évaluation	61
IV.1 Introduction	61
IV.2 Phrases-tests	61
IV.3 Évaluation humaine	62
IV.3.1 Méthodologie	62
IV.3.2 Évaluation comparative (qualité)	65
IV.3.3 Évaluation comparative (fluidité)	69
IV.4 Évaluation machine	72
IV.5 Qualité de la source et prédiction	75
IV.6 Conclusion.....	78
V. Réflexion générale.....	79
Bibliographie.....	83

I. Introduction

I.1 Objectifs

Si le développement des programmes de traduction automatique (communément appelée en français « TA ») est aujourd'hui globalement répandu, il y aura encore de nombreuses années avant que le processus puisse produire une traduction parfaite. Afin de nous pencher sur le sujet, nous avons choisi de spécialiser un programme de TA indirect, Lucy LT, avec le vocabulaire propre au jeu vidéo *World of Warcraft*, jeu en ligne multijoueur sorti en 2004 et toujours en activité. Il nous paraît en effet que la possibilité d'y intégrer un programme de TA créerait un engouement supplémentaire auprès des joueurs monolingues.

L'une des grandes nouveautés avec l'arrivée des jeux vidéo MMORPG (*Massively Multiplayer Online Role-Playing Game*) est qu'il est désormais important que les joueurs interagissent : il ne s'agit plus de jouer « en solo », la communication et l'aspect social jouent un rôle crucial. De fait, un lien social se crée entre les joueurs, qui peuvent s'ajouter en « ami » (fonction de messages privés au sein du jeu, sous forme de discussion instantanée) ou créer des clans appelés « guildes ». Nous sommes donc persuadée que l'intégration d'un programme de TA au sein de *World of Warcraft* permettrait de donner aux joueurs une opportunité supplémentaire de découvrir d'autres cultures, de nouer des amitiés au-delà des frontières physiques et linguistiques.

De plus, à la suite à la mise en activité il y a quelques mois d'un nouveau type de TA (système neuronal), nous souhaitons vérifier si la spécialisation d'un programme de TA linguistique est encore à même de battre un système neuronal (ou généraliste). Cette question nous paraît d'autant plus intéressante que le langage du jeu vidéo *World of Warcraft*, s'il ne peut être totalement considéré comme un *sublangage*, diffère de la langue générale et n'est pas toujours intelligible pour les « non-joueurs ». En outre, il ne semble pas exister à ce jour de corpus parallèle pour le jeu en question, ce qui laisse imaginer qu'un système neuronal ou statistique ne serait peut-être pas des plus efficaces.

I.2 Question de recherche et méthodologie

Afin de déterminer si un programme de TA linguistique est capable d'offrir une meilleure traduction qu'un programme de TA neuronal, il nous sera nécessaire de créer notre propre corpus monolingue de termes propres à *World of Warcraft*, puis de spécialiser Lucy LT afin d'intégrer ces derniers aux dictionnaires du système. Il nous faudra ensuite vérifier quel type

de programme, Lucy LT ou DeepL, mène à la meilleure traduction. Pour cela, nous aurons recours à une évaluation automatique (score BLEU) ainsi qu'une évaluation humaine de la part de joueurs francophones.

I.3 Les jeux-vidéo

Les jeux-vidéo ont longtemps été considérés comme un divertissement réservé au jeune public. Pourtant, Tomblaine rapporte qu'en 2013 en France, « plus de 83% des 28,6 millions de joueurs actifs ont plus de 18 ans » (Tomblaine, 2015 : 7), et la moyenne d'âge avoisine même plutôt les 35 ans. Le marché du jeu vidéo s'adresse donc principalement à des adultes qui, généralement indépendants financièrement, pourront déboursier une certaine somme pour l'achat de leurs jeux favoris.

De ce fait, l'industrie du jeu vidéo, toujours en plein essor, génère en France un chiffre d'affaire de 2,7 milliards d'euros, bien plus élevé que ceux de l'industrie cinématographique (1,3 milliards d'euros), vidéo (1,2 milliards d'euros) ou musicale (617 millions d'euros). Ces chiffres sont certes faibles comparés à d'autres secteurs de l'industrie française, notamment les transports (51 milliards d'euros pour PSA Automobiles), les télécommunications (23 milliards d'euros pour Orange) ou encore l'alimentaire (15 milliards d'euros pour Carrefour)¹, mais ils démontrent cependant qu'en matière de divertissement, l'industrie du jeu vidéo est un marché florissant et de plus en plus intéressant d'un point de vue financier.

De plus, selon Tomblaine, l'expérience *online* se partage désormais « amplement » (Tomblaine, 2015 : 8) : on en discute de vive-voix, à travers le jeu ou via les forums dédiés, on lit la presse spécialisée et les événements mondiaux axés sur les jeux vidéo sont plus nombreux chaque année. Ainsi, les univers présentés dans les jeux vidéo, leurs communautés et vocabulaires spécifiques sont un champ d'étude à la fois actuel et passionnant pour le traducteur, qui doit faire face aux néologismes, aux anglicismes, et à l'atmosphère même des jeux vidéo, qui ne lui est pas forcément familière.

Nous avons choisi, dans cette étude, de nous concentrer sur l'un des jeux vidéo qui a le plus fait parler de lui, *World of Warcraft*.

World of Warcraft est l'un des premiers MMORPG, commercialisé en novembre 2004 en Amérique du Nord, en Australie et en Nouvelle-Zélande. Il donne suite aux jeux de stratégie

¹ *Verif.com*, Data Altares & BFM TV (2016).

créés par Blizzard, *Warcraft : Orcs and humans*, sorti en 1994 et vendu à plus de 300'000 exemplaires en une année (Tomblaine 2015), et *Warcraft II* qui, grâce à l'avancée du Web en 1995, permet aux joueurs de créer des parties en réseau. L'engouement est tel qu'en 1996, selon Cario, les ventes de *Warcraft II* ont dépassé le million (Cario 2011).

Selon les concepteurs de Blizzard, comme le relate Ichbiah, le but de *World of Warcraft* était de réunir des joueurs qui avaient été habitués au MMORPG par le biais d'autres jeux, tels que *Half Life* et *Everquest*, mais aussi d'attirer de nouveaux joueurs potentiels, encore occasionnels (Ichbiah 2004). Le développement de *World of Warcraft* a nécessité un budget de 50 millions de dollars, pour un jeu dont la durée de vie était alors estimée à 4-5 ans, et qui devait engendrer un million d'abonnés (ce qui était, à cette époque, le record pour un jeu en ligne). Le 24 novembre 2004 au soir, date de la sortie du jeu en Amérique du Nord, en Nouvelle-Zélande et en Australie, 250'000 exemplaires sont déjà écoulés. Lorsque Blizzard lance *World of Warcraft* en France, en Italie, en Espagne et en Allemagne, le 11 février 2005, 350'000 exemplaires sont achetés dans les premières vingt-quatre heures. Le jeu connaît un succès tel, précise Ichbiah, qu'en janvier 2007 sort une première extension du jeu, *The Burning Crusade*, suivie dès novembre 2008 d'une deuxième : *Wrath of the Lich King* (Ichbiah 2004).

Selon Cario, si le joueur peut « s'émerveiller » (Cario 2011 : 171), c'est notamment en raison du style graphique proche du *cartoon* et de l'univers *fantasy* du jeu, qu'il estime « très soignés » (Cario 2011 : 171). Les chiffres présentés par Tomblaine le confirment : en 2008, plus de 10 millions de joueurs sont enregistrés avec un abonnement actif, et 70% des joueurs de MMORPG jouent à *World of Warcraft*. En octobre 2010, le jeu atteint un pic, avec 12 millions de joueurs (Tomblaine 2015). Selon Cario, ce chiffre se maintient jusqu'en 2011, soit 7 ans après la sortie du jeu et après une troisième extension en décembre 2010, *Cataclysm* (Cario 2011).

Depuis, *World of Warcraft* a connu trois extensions supplémentaires : *Mist of Pandaria* en septembre 2012, *Warlords of Draenor* en novembre 2014, alors que 7,6 millions de joueurs étaient encore actifs en 2013, et *Legion* en août 2016². Grâce à celles-ci, le nombre d'abonnés actifs a repassé la barre des 10 millions. Enfin, lors de la Blizzcon (rassemblement annuel organisé par la société Blizzard Entertainment Inc.) de novembre 2017, une septième extension,

² *Timeline (World of Warcraft)*, Wikia (collectif) (2017).

Battle for Azeroth, a été annoncée pour septembre 2018³, preuve que le jeu a encore plusieurs années devant lui.

Concernant l'univers de *World of Warcraft*, le continent principal, Azeroth, est peuplé de races qui s'affrontent en deux clans : la Horde (taurens, morts-vivants, trolls et orcs) contre l'Alliance (humains, elfes de la nuit, nains et gnomes). À travers le jeu, et au fil des extensions, les joueurs rencontrent de nouvelles races (jouables ou non), de nouveaux paysages et continents.

Metzen déclare, dans la préface du recueil de nouvelles « *World of Warcraft : Chronicles of War* », que l'intérêt de ces histoires est de donner aux joueurs et lecteurs un aperçu supplémentaire de la culture et de l'univers des personnages de *World of Warcraft* : l'accent est mis sur l'immersion potentielle dans un univers pourtant fictif, ce qui démontre chez les concepteurs une certaine volonté de rendre ledit univers complet, « réel » : « going back and adding characters enriched the universe, making it realer and more cohesive » (Metzen, *in Golden, Grubb & Rosenberg*, 2010 : VIII). Metzen évoque d'ailleurs une volonté de planter un décor dans lequel le jeu prend place (« to show [...] the “behind the scenes” events of the First War », Metzen, *in Golden, Grubb & Rosenberg*, 2010 : VII), et de proposer l'histoire de certains personnages (non-joueurs) de façon plus poussée (« a more intimate look at [some characters] », Metzen, *in Golden, Grubb & Rosenberg*, 2010 : VII).

D'un point de vue plus technique, il faut savoir qu'en Europe, *World of Warcraft* est divisé en plusieurs serveurs selon les langues suivantes : allemand, anglais, espagnol, français, italien, portugais et russe. Si ces serveurs ont longtemps été totalement séparés les uns des autres, il est, depuis le 30 novembre 2011⁴ (durant l'extension *Cataclysm*), possible de former un groupe à travers les différents serveurs. Il n'est donc pas rare de se retrouver en jeu face à un groupe de personnes dont on ne parle pas la langue. Si, généralement, les joueurs communiquent alors en anglais, il leur serait certainement plus aisé de pouvoir converser dans leur propre langue, ou de pouvoir bénéficier d'un système de traduction automatique directement implémenté dans le jeu. Les échanges humains seraient ainsi améliorés, et la dynamique de jeu multijoueur renforcée tout en faisant tomber la barrière de la langue.

³ *World of Warcraft, Battle for Azeroth*, Blizzard Entertainment Inc. (2018).

⁴ *World of Warcraft.com, Nethaera* (pour Blizzard Entertainment Inc.) (2011).

I.4 Plan du mémoire

Nous présenterons d’abord, dans le chapitre II, le concept de la traduction automatique (Partie II.1) et donnerons un aperçu historique du développement de celle-ci (Point II.2). Nous nous pencherons ensuite sur les différents systèmes existants (Point II.3) et les problèmes auxquels ils se retrouvent confrontés (Point II.4). Enfin, nous présenterons le programme de TA que nous avons retenu dans le cadre de notre étude, Lucy LT (Point II.5) et le programme DeepL, qui nous permettra de comparer les traductions obtenues (Point II.6).

Dans le chapitre III, nous exposerons un compte-rendu de notre travail pratique, à savoir le développement d’un corpus de texte en vue d’une traduction (Point III.2), notre méthode de recherche pour le choix des traductions (Point III.3), puis la spécialisation des *patches* de Lucy LT afin d’entraîner le système de TA pour le vocabulaire propre à *World of Warcraft* (Point III.4).

Nous évaluerons ensuite, dans le chapitre IV, les résultats de notre spécialisation, en la comparant avec les traductions proposées par le programme de TA neuronal DeepL. Après avoir présenté notre choix de phrases-tests (Point IV.2), nous présenterons les résultats obtenus lors de l’évaluation humaine d’une part (Point IV.3), puis lors de l’évaluation automatique d’autre part (Point IV.4). Enfin, nous discuterons de la qualité du texte source et de l’utilité de la prédiction (Point IV.5).

Enfin, nous répondrons dans le chapitre V à notre question de recherche, avant de soumettre une réflexion plus générale sur la TA et les problématiques auxquelles elle se retrouve encore aujourd’hui confrontée.

II. La traduction automatique (TA)

Dans ce travail, nous allons déterminer si le programme de TA choisi, Lucy LT, est à même de traduire de façon optimale des phrases et textes issues du jeu-vidéo *World of Warcraft*. Nous présenterons donc dans ce chapitre le concept de traduction automatique (Point II.1) et donnerons un bref aperçu historique du procédé (Point II.2). Nous listerons ensuite les différents systèmes de TA (Point II.3), puis les principaux problèmes auxquels se confrontent les programmes de TA (Point II.4). Enfin, nous présenterons le programme de TA retenu pour notre étude, Lucy LT (Point II.5), et le programme neuronal DeepL, qui nous permettra de comparer les traductions obtenues (Point II.6).

II.1 Présentation de la TA

La traduction automatique (TA) consiste à appliquer l'informatique au processus de traduction, dans le but d'obtenir d'un texte en langue source sa traduction en langue cible. Pour Arnold, un système de TA doit au moins être capable de produire une traduction correcte (« reasonable draft translations », Arnold & al., 1994 : I). Cependant, Hutchins et Somers estiment que, dans le cas où le texte produit par la TA ne peut être acceptable en l'état, il peut tout de même soutenir le traducteur en lui servant de premier jet (« pre-translation », Hutchins & Somers, 1992 : 3). Jurafsky et Martin partageaient ce point de vue en 1999, soulignant que la TA se concentre alors sur des tâches où une traduction brute est suffisante, ou lorsqu'elle est ensuite améliorée par un post-éditeur (« a rough translation is adequate, [...] where a human post-editor can be used to improve MT output », Jurafsky & Martin, 1999 : 800). Arnold précisait en revanche en 1994 que, dans le but d'obtenir une traduction automatique de qualité, en vue de la dissimilation du texte cible, le texte source doit se plier à quelques restrictions quant au sujet, au style et au vocabulaire employés (« provided the input observes certain restrictions on subject matter, style, and vocabulary », Arnold & al., 1994 : I). Il confirmait ainsi le point de vue de Nirenburg, qui soutenait qu'il n'existait pas encore de système de TA qui puisse traduire des textes de domaines variés (« reasonably broad domain of discourses », Nirenburg, 1987 : XIII). Ceci semble encore vrai aujourd'hui puisque, comme le souligne Aikawa, la mise en place d'un langage contrôlé est nécessaire à une traduction de meilleure qualité et à l'augmentation de la productivité du programme de TA employé (Aikawa & al. 2007). Or, le vocabulaire que nous emploierons au sein de notre étude, s'il n'est pas contrôlé, restera malgré tout focalisé sur un domaine précis, celui du jeu-vidéo *World of Warcraft*, ce qui laisse donc présager qu'une traduction automatique correcte est envisageable.

Selon Bowker en 2002, la TA a grandement progressé ces cinquante dernières années (Bowker 2002), mais elle n'est pas encore entièrement automatique, puisqu'elle requiert une action humaine, et le terme « traduction automatique » sous-entend parfois la « traduction automatique assistée par l'homme » (TAAH). Malgré les avancées technologiques de ces dernières années, ceci est encore vrai aujourd'hui. En effet, d'après L'Homme, la TA « nécessite invariablement des corrections [et l'] enjeu consiste à mettre au point des mécanismes pour en réduire le nombre » (L'Homme, 2008 : 15). Les mécanismes de cette assistance humaine peuvent s'effectuer sur trois niveaux : en préédition (travail en amont de la TA, sur le texte en langue source), en post-édition (travail en aval, sur le texte en langue cible produit par la machine), ou encore de façon interactive (l'édition a lieu en court de traduction) (Fuchs 1993).

Comme le soulignait O'Brien en 2002 déjà, puisque la demande en traduction ne cesse de croître, la TA et la post-édition deviennent nécessaires pour répondre à ce besoin (« MT and post-editing are viable solutions to meet the growing translation demand », O'Brien 2002 : 99). Ce point de vue est confirmé en 2012 par Aikawa, qui considère que coupler la TA et la post-édition est désormais une pratique commune pour limiter le coût des traductions tout en conservant sa qualité (« the use of machine translation (MT) together with post-editing (PE) has become one of the common practices to achieve cost-effective and high quality translation », Aikawa & al. 2012 : 2).

II.2 L'histoire de la TA

Si l'idée d'une machine traduisant à la place de l'homme peut donner à certains l'impression d'entrer dans un univers de science-fiction futuriste, il faut pourtant savoir qu'Arnold situe les prémices de la TA en 1947, dans une correspondance entre Booth et Weaver, membre de la Fondation Rockefeller (Arnold & al. 1994). Selon Hutchins et Somers, l'idée de créer des dictionnaires basés sur un code universel (« creation of dictionaries based on universal numerical codes », Hutchins & Somers, 1992 : 5) remonterait même au XVII^{ème} siècle, et aurait été abordée par Descartes et Leibniz, notamment dans le but de créer un « langage "mathématisé" pour interpréter les diverses langues » (Clas, *in* Bouillon & Clas, 1993 : 11). En 1933, deux brevets relatifs à la traduction automatique voyaient le jour : d'une part, l'inventeur arménien devenu français, Artsrouni, proposait une machine qui pouvait donner l'équivalent d'un mot dans une autre langue, et d'autre part le Russe Smirnov-Troyanskii élaborait un projet qui consistait à extraire l'analyse logique d'un texte en langue source, transformer celle-ci en

une représentation équivalente pour la langue cible, puis qui convertissait le résultat de la machine en un texte de langue (Hutchins & Somers 1992). Ces diverses propositions créèrent un engouement soudain dans un univers déjà encouragé par le décryptage informatique mis en place durant la Seconde Guerre mondiale, et de nombreux groupes de recherche virent alors le jour en Europe et aux États-Unis d'Amérique. Comme l'écrit Nirenburg, les chercheurs étaient partis du principe que si les hommes étaient capables d'effectuer des traductions (« since translation is such a common everyday task [...] with relative ease by humans », Nirenburg, 1987 : 1), il était forcément possible d'automatiser le procédé.

Cependant, comme le nuance Arnold, dès 1959, le philosophe Bar-Hillel démontra qu'il était impossible de créer un système de TA complet, puisqu'il était invraisemblable qu'une machine puisse comprendre les ambiguïtés sémantiques et lexicales de la même façon qu'un homme et, de fait, les traduire convenablement (Arnold & al. 1994). Hutchins et Somers soulignent en revanche que les ambitions de la TA devraient être revues à la baisse (« should adopt less ambitious goals », Hutchins & Somers, 1992 : 7), afin que les coûts liés à sa création soient raisonnables, et qu'une interaction entre l'homme et la machine puisse naître. Malheureusement, cette perte d'intérêt pour la TA devient plus importante encore en 1966, lorsque le groupe ALPAC (*Automatic Language Processing Advisory Committee*) présenta un rapport qui démontrait qu'un système de TA serait non seulement trop coûteux à développer, mais encore superflu face aux besoins en traduction de l'époque.

Actuellement, comme l'explique Ladmiral, les détracteurs de la TA sont encore légion en raison de leur « angoisse devant la menace d'une concurrence [...] informatique » (Ladmiral, 1994 : 6), mais la traduction assistée par ordinateur (TAO) connaît un succès de plus en plus répandu, et les programmes de TA fleurissent depuis leur « renaissance » (Arnold & al., 1994 : 14) à la fin des années 1970, après une période de « stagnation » d'environ dix ans (Fuchs 1993 : 195). En effet, pour Hutchins et Somers, même si les États-Unis d'Amérique se sont désintéressés de la TA suite à l'évaluation du groupe ALPAC, la TA est restée un enjeu important pour les communautés plurilingues, telles que le Canada et la Communauté Économique Européenne : « [...] was demanding translations of scientific, technical, administrative and legal documentation from and into all the Community languages » (Hutchins & Somers, 1992 : 7).

Et de fait, il existe aujourd'hui de nombreux programmes de TA, bien des systèmes entièrement automatiques ne soient pour l'instant possibles que pour des sujets spécialisés, comme par exemple le système TAUM-METEO, créé et utilisé par le groupe TAUM de 1977 à 2001 pour

traduire les bulletins météorologiques en anglais et en français au Canada. Il s'agit d'un programme fonctionnant uniquement par transfert lexical et structural, comme le rappelle Tucker (Tucker, *in* Nirenburg, 1987), dont le succès de traduction sans post-édition est de 80%. Cette moyenne est due au champ d'étude restreint du système, comme le souligne Nirenburg : le vocabulaire du programme compte environ 1500 entrées, dont la moitié correspond à des noms de lieux (Nirenburg 1987). L'Homme estimait d'ailleurs en 2008 que ce système était, dans son domaine, un des rares cas de TEAHQ (Traduction Entièrement Automatique et de Haute Qualité) (L'Homme 2008).

Dans un monde où les échanges entre différents pays et différentes langues sont de plus en plus courants, il devient évident, malgré les critiques, que la TA est et sera un outil d'un poids exceptionnel. Pour reprendre les exemples d'Arnold, la TA offre en effet un intérêt à la fois commercial (productions de manuels dans la langue des utilisateurs, réduction des retards de traduction, et donc des coûts inhérents), scientifique (défi technologique toujours d'actualité, puisqu'il n'existe à l'heure actuelle aucun système de TA « parfait »⁵), et philosophique (création d'une machine capable d'utiliser toutes les connaissances humaines nécessaires) (Arnold & al. 1994). Il reste pourtant clair que l'enjeu commercial reste un atout principal pour la défense de la TA puisque le besoin en traduction dépasse largement les capacités humaines, comme le souligne Kay : « to produce more translations, to produce them faster [...] at lower cost » (Kay, *in* Hutchins & Somers 1992 : XII).

II.3 Les différents systèmes de TA

Il existe deux types de programmes de TA : les systèmes linguistiques d'une part, qui cherchent à lever les ambiguïtés lexicales et syntaxiques grâce à leurs connaissances linguistiques (par le biais de dictionnaires et de grammaires) et d'autre part les systèmes statistiques, dont les connaissances sont extraites de corpus de textes. Ces derniers ne possèdent donc aucun niveau de compréhension du texte, source ou cible. Arnold classe ces mêmes systèmes sous les termes « rule-based systems » et « knowledge-based systems » pour les programmes linguistiques et « example-based systems » et « statistical systems » pour les programmes statistiques (Arnold & al., 1994 : 174 et 188).

⁵ Nous entendons ici par « parfait » un système de TEAHQ qui ne serait pas limité à un domaine de spécialité précis.

Bouillon précise en outre que les programmes de TA peuvent être conçus pour une paire de langue (systèmes bilingues) ou plus (systèmes multilingues), et fonctionner dans une direction seulement (systèmes unidirectionnels) ou dans les deux sens (systèmes bidirectionnels) (Bouillon, *in* Bouillon & Clas 1993).

Enfin, sont apparus dernièrement les systèmes dits « neuronaux », qui fonctionnent également par statistiques et permettent de représenter le sens des mots (*word embedding*). Selon Tilde, ces programmes fonctionnent à l'aide d'une intelligence artificielle, qui examine la phrase dans son ensemble afin de produire une traduction. Celle-ci est alors généralement plus fluide que celle proposée par les systèmes statistiques purs (ou « traditionnels ») basés sur les segments ou par les systèmes linguistiques. En revanche, la traduction produite se trouve parfois bien plus éloignée du texte source⁶.

II.3.1 Les systèmes linguistiques

Les programmes de TA linguistiques, dont fait partie Lucy LT, le programme que nous utiliserons principalement dans cette étude, sont habituellement répartis en deux sous-catégories : d'un côté, les systèmes « directs », aussi appelés « minimalistes », et de l'autre les systèmes « indirects », ou « maximalistes ». Arnold présente une troisième appellation pour chacun de ces deux sous-ensembles : « transformer architecture » pour les systèmes directs, et « linguistic knowledge architecture » pour les systèmes indirects (Arnold & al., 1994 : 2). Ces derniers termes, quoiqu'anglais, illustrent peut-être davantage notre propos, puisqu'ils explicitent la relation de compréhension que la machine possède avec le texte.

Les systèmes minimalistes furent les premiers à être développés, ce qui fait d'eux les programmes de TA les moins performants. En effet, ils se limitent à une analyse lexicale, puisqu'il s'agit du processus le plus facile à implémenter dans un système, et sont « fondés sur des listes d'équivalence terme à terme » (Bouillon, *in* Bouillon & Clas, 1993 : 16). Ils ne sont donc capables de résoudre que des ambiguïtés catégorielles (aussi appelées « homographes catégoriels », L'Homme, 2008 : 36), puisqu'ils divisent simplement la phrase en mots, puis tentent d'attribuer à chaque mot sa catégorie grammaticale la plus probable en utilisant des statistiques (fréquence) ou des règles (choix par défaut). Nous pouvons, pour ce type de systèmes minimalistes, citer les programmes Reverso, que nous utiliserons pour éclairer nos phrases d'exemple, ou encore Systran, créé en 1976 par la Communauté Européenne (Bouillon

⁶ *Neural Machine Translation*, Tilde (2017).

1998). Ce dernier, comme l'explique Fuchs, fonctionne uniquement à l'aide de plusieurs dictionnaires, tels « mots de base » ou « expressions idiomatiques » (Fuchs, 1993 : 203).

Avec la phrase (1) ci-dessous, nous souhaitons montrer la façon dont « réfléchira » un système de TA minimaliste pour déterminer la catégorie du mot *cry*. La question que se posera le programme pour lever l'ambiguïté catégorielle sera la suivante : est-il plus fréquent qu'une phrase comporte le segment [Pron.+N_{sg}] ou [Pron.+V_{conj}] ?

(1) I cry when I am sad.

Statistiquement, un pronom est bien sûr plus souvent suivi par un verbe conjugué que par un substantif, raison pour laquelle le système traduira le segment *I cry* par « Je pleure » plutôt que par *« Je cri »⁷. Pourtant, nous pouvons voir ci-dessous dans la Figure (1) que la traduction de *cry* [N_{sg}] par « cri » [N_{sg}] est disponible dans le dictionnaire de Reverso, signe qu'il ne s'agit pas d'une traduction « par défaut », mais bien du résultat de la désambiguïstation.

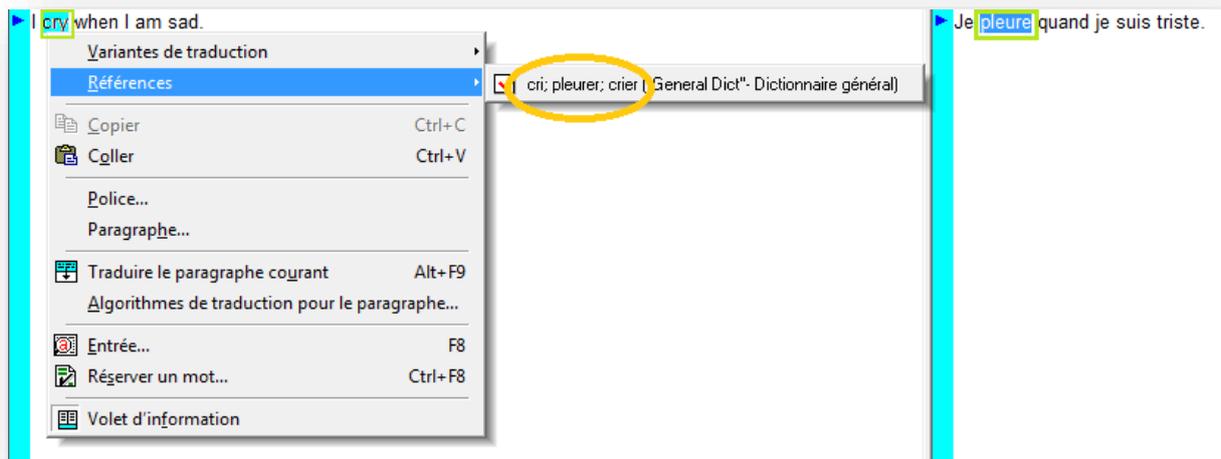


Figure (1) : Traduction de la phrase (1) par Reverso

Cependant, puisque ce type de système ne dispose que de dictionnaires, et non de grammaires, les fautes de désambiguïstation sont nombreuses, et aucune interprétation de la structure grammaticale de la phrase ne se fait. La traduction se fait donc plus ou moins mot-à-mot, et ne lève pas les ambiguïtés de structures.

Les systèmes maximalistes, en revanche, possèdent une compréhension plus poussée, puisque leurs grammaires (généralement une pour la langue source et une pour la langue cible) leur

⁷ Nous utiliserons l'astérisque (*) pour indiquer une phrase agrammaticale, et donc inacceptable.

permettent d'analyser la phrase au niveau syntaxique et sémantique, puis de la générer. La désambiguïsation est donc bien plus précise, car le programme va effectuer une « analyse du texte source » et en donner « une représentation intermédiaire » (Bouillon, *in* Bouillon & Clas, 1993 : 16). Cette représentation est généralement représentée par un arbre syntaxique de la phrase en question, qui permettra d'attribuer à chaque mot (et à chaque groupe de mots) sa catégorie. Ainsi, le système détecte l'ambiguïté de structure et peut sélectionner la bonne traduction. De plus, il est généralement possible de spécialiser le système, par exemple en fonction du domaine du texte, et d'attribuer de nouvelles préférences de traduction.

Selon Arnold, le processus de TA d'un système maximaliste se répartit en deux étapes : l'analyse (ou « parsing », Arnold & al., 1994 : 50), puis la synthèse (ou « generation », Arnold & al., 1994 : 50). Dans un premier temps, le système de TA découpe la phrase (*parsing*) du texte source et y applique les règles de grammaire qu'il connaît, afin de vérifier la façon dont les mots se combinent entre eux pour créer la phrase, et de s'assurer que celle-ci soit grammaticalement correcte (*analyzing*). Lors de cette phase, le programme crée un premier arbre syntagmatique pour exprimer les relations entre les mots et groupes de la phrase. Dans un second temps, le programme se focalise sur la représentation syntagmatique créée, puis, par un effet de miroir, génère (*generating*) l'arbre syntagmatique correspondant en langue cible, duquel il peut ensuite synthétiser (*synthesizing*) la phrase traduite. Hutchins et Somers précisent que l'analyse comme la génération peuvent s'effectuer sur trois niveaux distincts : d'abord au niveau morphologique, puis syntaxique, et enfin sémantique (Hutchins & Somers 1992).

Les systèmes indirects peuvent être divisés en deux sous-catégories : les systèmes d'approche par transfert (« transfer approach », Arnold & al. 1994 : 2, ou « transfer model », Jurafsky & Martin, 1999 : 807) et les systèmes d'approche par interlangue (« interlingual approach », Arnold & al. 1994 : 2, ou « interlingua », Jurafsky & Martin, 1999 : 812).

Les systèmes d'approche par transfert établissent « des correspondances [...] entre les mots et les structures syntaxiques » (L'Homme, 2008 : 257) à travers une représentation abstraite de la phrase. Cette représentation est ensuite transférée selon les structures syntaxiques et lexicales typiques de la langue cible (Tucker, *in* Nirenburg 1987). Ainsi, les systèmes d'approche par transfert fonctionnent, comme l'expliquent Jurafsky et Martin, par leurs connaissances de la différence relative à deux langues (« knowledge about the differences between the two languages », Jurafsky & Martin, 1999 : 807).

Les systèmes d’approche par interlangue, quant à eux, s’éloigne davantage des langues. En effet, comme le souligne Tucker, l’interlangue est un concept universel et indépendant de la langue, qui est (supposément) directement utilisable pour la phase de génération (Tucker, *in* Nirenburg 1987). Il s’agit donc d’une représentation au niveau conceptuel, qui nécessite une compréhension plus en profondeur du texte, ce que Jurafsky et Martin appellent une représentation du sens : « meaning representation, [...] in a language-independent canonical form » (Jurafsky & Martin, 1999 : 812).

Si nous reprenons la phrase (1), nous pouvons donc voir qu’un programme de TA maximaliste tel Lucy LT traduira lui aussi *I cry* par « Je pleure », mais cette fois-ci car le système va relier le pronom *I* et le verbe *cry* dès la création de l’arbre syntagmatique (en langue source) en les plaçant ensemble devant la phrase adverbiale (*when I am sad*). Les deux mots seront donc traduits selon leur équivalent français grammaticalement correct, et placés au bon endroit dans la phrase grâce à l’arbre de génération, comme ci-dessous dans la Figure (2).

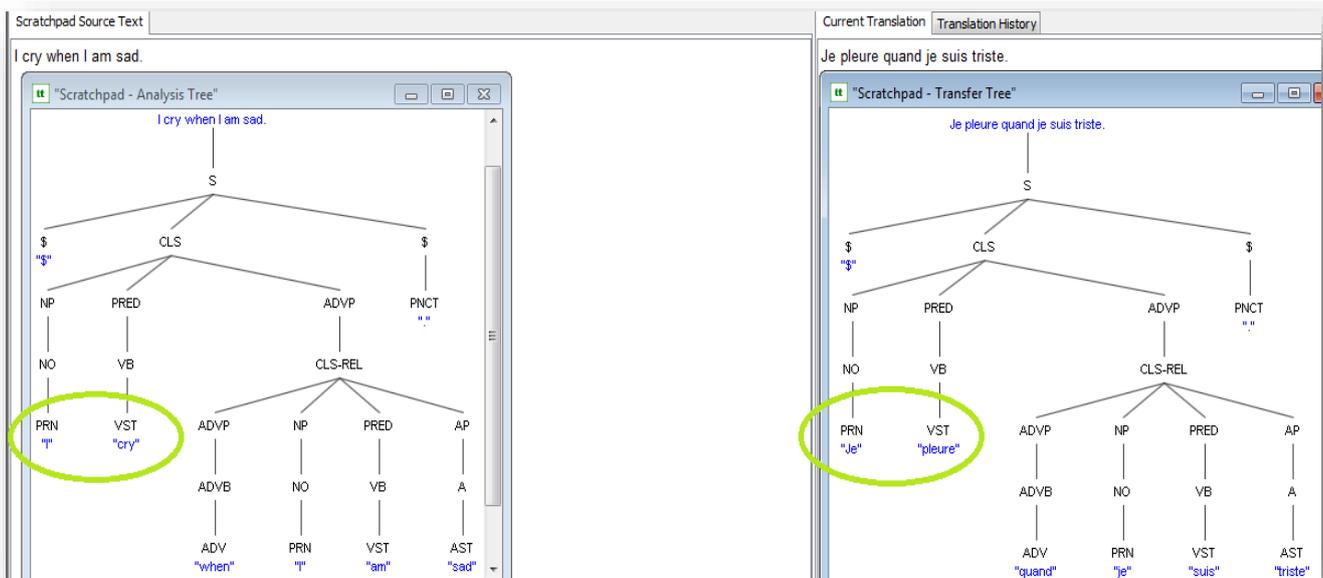


Figure (2) : Traduction de la phrase (1) par Lucy LT

Ainsi, les programmes de TA maximalistes, dont Lucy LT fait partie, sont de meilleure qualité que les minimalistes. Le fait de passer d’une traduction de mots à une traduction de la représentation de ceux-ci permet une traduction de meilleure qualité. En effet, si nous prenons une phrase qui contient une ambiguïté catégorielle plus complexe, comme dans la phrase (2), un système maximaliste tel que Lucy LT sera capable de reproduire un arbre syntagmatique

pour attribuer à chaque mot sa catégorie grammaticale. Au contraire, Reverso, programme minimaliste, ne parviendra pas à déterminer la catégorie des mots.

(2) Fumer nuit à la santé.

Lucy LT attribue correctement au mot « nuit » la catégorie de verbe (« nuire », présent de l'indicatif, 3^{ème} personne du singulier), comme nous pouvons l'observer dans la Figure (3) ci-dessous. En revanche, Reverso en fait un nom (« nuit », substantif féminin au singulier). Il est cependant intéressant de souligner que le verbe « nuire » est présent dans le dictionnaire de Reverso, puisqu'il existe la forme verbale *harm* en tant que référence, comme nous pouvons le voir ci-dessous dans la Figure (4). Il est donc probable que le système ait également été induit en erreur par le verbe infinitif employé comme sujet de phrase (« Fumer »).

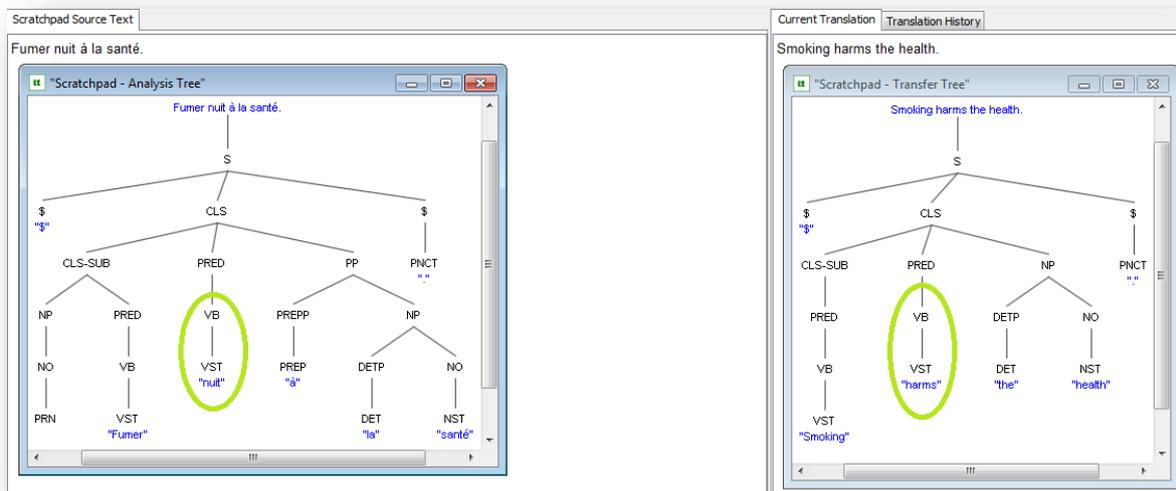


Figure (3) : Traduction de la phrase (2) par Lucy LT

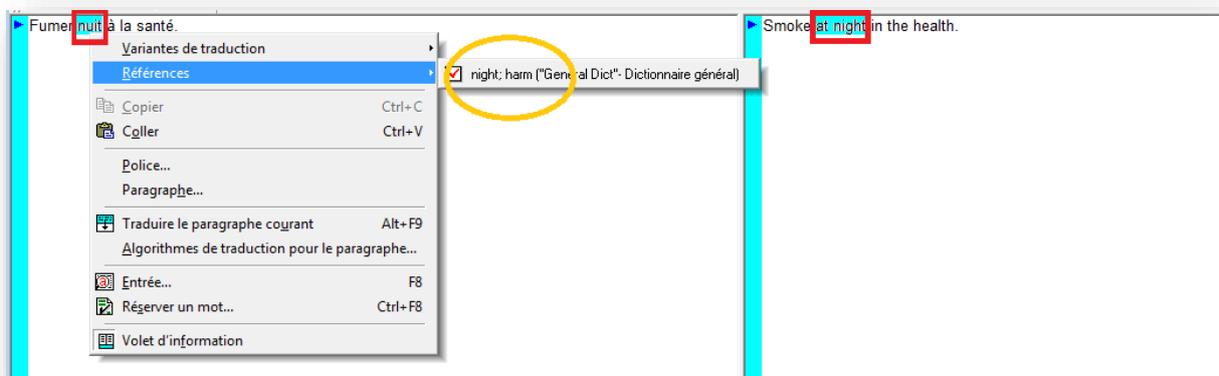


Figure (4) : Traduction de la phrase (2) par Reverso

II.3.2 Les systèmes statistiques traditionnels (par segments)

Comme nous l'avons évoqué précédemment, les systèmes linguistiques ne sont pas la seule alternative de traduction automatique. Selon Jurafsky et Martin, les systèmes statistiques traditionnels se basent moins sur le processus de traduction que sur son résultat (« to focus on the result, not the process », Jurafsky & Martin, 1999 : 819). L'idée principale de ces systèmes est donc avant tout d'offrir une traduction compréhensible dans la langue et la culture cible, même si pour cela elle doit s'éloigner de la phrase produite en langue source.

La principale différence avec les systèmes linguistiques est que le processus de traduction ne passe plus par des arbres syntagmatiques : comme le soulignent Hutchins et Somers, les systèmes statistiques ne passent par aucune analyse sémantique ou syntaxique de la phrase (« neither is now restricted to rule-based analyses », Hutchins & Somers, 1992 : 319). Ils sont cependant basés sur des corpus monolingues et bilingues, et fonctionnent selon un algorithme reflétant à la fois la fidélité au texte source et la fluidité du texte cible (« faithfulness and fluency », Jurafsky & Martin, 1999 : 819).

Leur but est donc, comme l'explique Somers, de « trouver pour une phrase [source] donnée la phrase [cible] dont la probabilité $Pr(S|C)$ est maximale » (Somers, *in* Bouillon & Clas, 1993 : 157), où « Pr » signifie « probabilité », « S » la source et « C » la cible. Il souligne cependant que cette méthode ne prend en compte qu'une correspondance exacte, et qu'un énoncé qui ne correspondrait pas séquentiellement ne serait pas pris en compte, même s'il est identique sémantiquement parlant. Hearne et Way précisent que les systèmes statistiques fonctionnent en deux étapes principales : tout d'abord l'entraînement (*training*), qui extrait un modèle de traduction d'un corpus parallèle, puis une phrase de décodage (*decoding*), qui détermine quelle traduction est la plus probable pour une séquence donnée (Hearne & Way 2011).

Cependant, il reste donc clair que les systèmes statistiques traditionnels, en raison de leur absence d'analyse grammaticale de la phrase, n'offrent pas toujours la traduction la plus adéquate à une phrase source. Ainsi, la fluidité apparente d'une traduction effectuée à l'aide d'un système de TA statistique peut camoufler une ou plusieurs erreurs de traduction qu'un post-éditeur ne remarquera pas forcément. En effet, comme le précisent Hearne et Way, la fluidité et la justesse (*adequacy*) des phrases traduites sont générées séparément (« adequacy and fluency [...] are modelled completely separately », Hearne & Way 2011 : 207).

II.3.3 Les systèmes neuronaux

Les systèmes neuronaux (NMT, de *Neural Machine Translation*), et plus récemment les systèmes neuronaux profonds (Deep NMT) sont apparus sur le marché dans le courant des dernières années. Le vif succès qu'ils connaissent depuis est principalement dû à Google, qui en vantait les mérites et estimait que son niveau de traduction égalait le niveau humain⁸. Burchardt confirme l'engouement commercial impressionnant (« striking improvement of the commercial online system », Burchardt & al., 2017 : 169) de l'évolution neuronale des systèmes statistiques, à laquelle a largement contribué l'intelligence artificielle.

Les systèmes neuronaux profonds, tout comme les systèmes neuronaux « simples », fonctionnent selon un système de « couches » pour leur réseau neuronal. Le terme « profond » signifie que le système possède plusieurs couches, rendant son réseau beaucoup plus large et efficace. En effet, plus un système possède de couches, plus son processus comprend les mots, et plus il est capable d'analyser une interaction entre mots éloignés au sein de la phrase. Cependant, Omniscien Technologies estime que ces systèmes dominent pour le moment uniquement sur le marché des traductions de domaines fixes, et sont beaucoup moins efficace en langue générale (« out-of-domain content »)⁹. Ce point de vue nous paraît en revanche discutable, étant donné que, dans le cadre de notre étude, le système neuronal DeepL s'est révélé très efficace sur le vocabulaire de langue générale (LG) et beaucoup moins apte à traduire un vocabulaire spécifique comme celui de *World of Warcraft*.

L'un des principaux avantages évoqués par l'*Association for Computational Linguistics* est que les systèmes neuronaux nécessitent dans un premier temps des données uniquement monolingues en quantité, quoiqu'ils doivent ensuite être entraînés à l'aide de corpus parallèles. De ce fait, ils sont plus faciles à mettre en place, car les corpus monolingues sont abondants dans la plupart des langues (« only require (sufficient) monolingual data to train models, which is available in abundance for many languages », Association for Computational Linguistics, 2016 : 162). Microsoft ajoute que le problème des systèmes statistiques était son absence d'analyse grammaticale, et son contexte limité à quelques mots avant et après le mot ciblé (N-grammes, « N » représentant le nombre de mots pris en compte lors de l'analyse d'un mot). Or, dans le cas du *DeepLearning* (et du même fait, de la traduction neuronale profonde), il s'agit

⁸ *Riding the Machine Translation Hype Cycle – From SMT to NMT to Deep NMT by Dion Wiggins*, Omniscien Technologies (2017).

⁹ *Deep Neural Machine Translation*, Omniscien Technologies (2018).

d'une réflexion comme celle du cerveau humain : le système passe par une étape de représentation « mentale » du sens complet de la phrase avant d'en entamer la traduction.

D'un point de vue technique, les systèmes neuronaux fonctionnent en deux parties : d'abord, une analyse (*encoding*) de la phrase source, qui est ensuite transformée en vecteurs, puis une phase de transfert (*decoding*) qui présente une phrase en langue cible, sans phase de génération préalable. La représentation du sens des mots, transformée en vecteurs, permet aux programmes de TA neuronaux de choisir une traduction adéquate selon le contexte des mots en langue source comme en langue cible (Blog Systran 2016).

II.4 Les problèmes rencontrés en TA

La TA, bien que très fonctionnelle d'un point de vue théorique, rencontre encore plusieurs problèmes. Comme le souligne Bourquin, « vouloir automatiser, c'est [...] s'astreindre à tout expliciter » (Bourquin, *in* Bouillon & Clas, 1993 : 27) au sein des systèmes linguistiques, ce qui se révèle un défi non seulement chronophage, mais encore complexe.

Ensuite, les systèmes basés sur corpus, notamment les systèmes statistiques, font encore débat puisque, comme le soulignent Jurafsky et Martin, un score élevé de fréquence ne signifie pas que l'équivalent traduit sera forcément le bon, mais seulement le plus fréquent (Jurafsky & Martin 1999). En effet, les systèmes statistiques, à l'instar des systèmes linguistiques minimalistes, ne possèdent pas de compréhension de la phrase à traduire : comment peuvent-ils alors décider quel équivalent utiliser pour un terme en langue source qui contient plusieurs équivalents en langue cible ? Cette problématique ne se pose pas pour les systèmes neuronaux qui, de ce point de vue, présentent alors un réel avantage, aux côtés des systèmes linguistiques maximalistes.

En revanche, sur le plan linguistique, Arnold démontre que la TA pose encore problème sur trois niveaux de connaissance principaux¹⁰ : la connaissance sémantique (*semantic knowledge*), qui permet de déterminer ce qu'un mot ou une expression signifie, quel que soit son contexte d'apparition ; la connaissance pragmatique (*pragmatic knowledge*), qui permet de comprendre ce qu'une expression signifie dans un contexte en particulier, et enfin le sens commun (*real world knowledge* ou *common sense knowledge*) (Arnold & al. 1994). Ce dernier niveau correspond aux connaissances extralinguistiques, que L'Homme appelle « connaissances du

¹⁰ Les connaissances lexicales et syntaxiques ont été écartées de cette liste, puisque la plupart des systèmes de TA linguistiques sont actuellement à même de résoudre, au moins partiellement, ce type d'ambiguïtés.

monde » (L'Homme, 2008 : 13). Or, c'est justement ces deux dernières catégories qu'il est difficile d'implémenter dans un programme de TA : si ces connaissances se développent rapidement chez l'homme au cours de sa vie, elles requièrent un niveau de compréhension que l'on peine encore à intégrer à l'informatique.

Parmi les problèmes de traduction les plus courants au sein des systèmes de TA, Vandooren cite plusieurs cas de divergences, ainsi que ce qu'elle nomme les « décalages »¹¹ (Vandooren, in Bouillon & Clas, 1994 : 77). Ces derniers ne peuvent être résolus en traduction car ils expriment en langue source une réalité inexistante en langue cible. En revanche, les divergences sont le résultat d'un « contenu sémantique [...] exprimé par des formes morphologiques ou syntaxiques différentes en langue cible » (Vandooren, in Bouillon & Clas, 1994 : 77). Présentons comme exemples les cas de la divergence syntagmatique et de la divergence de densité lexicale.

La divergence syntagmatique a lieu lorsqu'une recatégorisation des syntagmes est nécessaire lors de la traduction, en raison des collocations imposées aux différents syntagmes. Par exemple, l'expression anglaise *I'm hoping for the weather to improve* se traduira en français par « J'espère que le temps s'améliorera ». Le groupe prépositionnel *for the weather to improve* est transformé en français par une proposition subordonnée conjonctive complétive (« que le temps s'améliorera »). Lorsqu'un terme isolé nécessite un changement de catégorie morpho-syntaxique, on parle de divergence catégorielle.

La divergence de densité lexicale signifie qu'un terme en langue source doit se traduire par plusieurs mots en langue cible, ou inversement, plusieurs mots en langue source sont traduits par un seul mot en langue cible. Dans le premier cas, on parle d'expansion (une unité lexicale se traduit par plus d'un syntagme). Par exemple, le mot allemand *Schadenfreude*, qui signifie « la joie que l'on éprouve face au malheur d'autrui », pourrait certes se traduire en français par un syntagme simple tel « cruauté » ou « jubilation », mais sans réussir à en rendre le sens complet. Au contraire, lorsqu'on réduit plusieurs syntagmes à une seule unité lexicale en langue cible, on parle de réduction. C'est le cas de l'anglais *wooden floor*, qui se traduira en français par un syntagme unique : « plancher ».

¹¹ Le décalage correspond à ce que Jurafsky et Martin nomment le « *lexical gap* », soit lorsqu'il existe en langue cible un vide lexical pour exprimer le concept de la langue source (Jurafsky & Martin, 1999 : 806).

Jurafsky et Martin ajoutent que l'ordre des mots varie en fonction du langage, et que l'ordre [Sujet-Verbe-Objet] utilisé en français ou en anglais appelle généralement des prépositions, tandis que l'ordre [Sujet-Objet-Verbe] utilisé notamment en japonais appelle des postpositions. Ils précisent en outre que certaines langues, telles l'arabe classique ou l'hébreu biblique utilisent l'ordre [Verbe-Sujet-Objet] (Jurafsky & Martin 1999 : 801). Cette modification de l'ordre des mots lors de la traduction peut être mis en lien avec la divergence thématique présentée par Vandooren, soit lorsque le thème et le rhème de la phrase ne sont pas les mêmes en langue cible. C'est le cas de l'expression anglaise *I miss you*, qui devient « Tu me manques » en français (nous assistons donc à une inversion du thème et du rhème lors de la traduction).

En plus de ces nombreux problèmes de divergences et de décalages, Hutchins et Somers évoquent d'autres problèmes récurrents dans les systèmes de TA : les ambiguïtés lexicales (catégorielles ou sémantiques), les ambiguïtés structurelles, les constructions agrammaticales ou elliptiques (ou « incompleteness », Sinclair, 2004 : 272) ou encore les reprises anaphoriques (Hutchins & Somers, 1992). Ils concluent que la problématique reste donc d'extraire le sens (*meaning*) de la phrase et de le restituer à travers l'écriture en langue cible (« in another set of linguistic symbols with an equivalent meaning », Hutchins et Somers, 1992 : 3), même si les codes langagiers diffèrent d'une langue à l'autre.

Ce sont donc ces différents problèmes auxquels nous devons rester attentive tout au long de notre étude, et nous verrons comment ils sont traités dans les systèmes de TA linguistiques et neuronaux.

II.5 Présentation de Lucy LT

Comme nous l'avons précédemment expliqué, Lucy LT fait partie des systèmes de traduction maximalistes et est basé sur une architecture de connaissance linguistique (*linguistic knowledge architecture*). Ce programme de TA peut donc aller jusqu'à une compréhension syntaxique de la phrase, en traduisant non pas des mots mais une représentation de la phrase, sous forme d'arbres syntagmatiques : l'un pour la phase de transfert, en langue source, et l'autre pour la phase de génération, en langue cible. Si Lucy LT est considéré comme le meilleur programme de l'évaluation dirigée par Burchardt, atteignant un taux de réussite de 83% sur un corpus de phrases-tests (« the RBMT system performs best of all systems on average, reaching 83% », Burchardt & al., 2017 : 167), c'est notamment en raison de la complexité des phrases qu'il est capable de traduire à l'aide des arbres syntagmatiques, ainsi que grâce aux différents dictionnaires qu'il emploie.

En effet, Lucy LT se base sur deux dictionnaires monolingues (langue source et langue cible), ainsi qu'un dictionnaire de transfert (langue source – cible). Ce dernier lui permet de ne traduire plus directement les mots dans leurs équivalents respectifs en langue cible, mais de traduire la représentation du mot, du concept ou de la phrase elle-même. Ainsi, nous approchons davantage du but évoqué ci-dessus par Hutchins et Somers : un système de TA qui devrait, dans l'idéal, extraire le sens de la phrase en langue source pour générer un sens équivalent en langue cible, au-delà des simples mots (Hutchins & Somers 1992). Comme le présentent les Figures (5) et (6) ci-dessous, les entrées monolingues (Figure (5)) et bilingues (Figure (6)) de Lucy LT sont classées dans des dictionnaires différents. Chaque terme est pourvu de différents attributs selon sa catégorie lexicale.

CAN="elf"		Feature Name	Value	CAN="elfe"		Feature Name	Value
l-src	elf (NST)	✓ CAN	"elf"	l-src	elfe (NST)	✓ CAN	"elfe"
l-tgt	elf	CAT	NST	l-tgt	elfe	CAT	NST
l-xfr	elves	✓ ALO	"elf"	l-xfr		✓ ALO	"elfe"
edi-src		ABB		edi-src		ABB	
edi-tgt		ABB-DOT		edi-tgt		ABB-DOT	
def-src		ANA		def-src		ANA	
def-tgt		ARGS		def-tgt		ARGS	
im-src		CA		im-src		✓ CL	(PF-S PM-S SF-0 SM-0)
im-tgt		✓ CL	(S-01)	im-tgt		✓ CMT	"Race WoW (PJ)"
im-xfr		✓ CMT	"Race WoW (PJ)"	im-xfr		DR	
i-cf-src		DR		i-cf-src		FW	
i-cf-tgt		FW		i-cf-tgt		✓ GD	(F M)
		GEN				GEN	
		✓ KN	CNT			✓ KN	CNT
		✓ NU	(SG)			NU	
		✓ ON	VO			✓ ON	VO
		✓ PLC	(WF)			ON...	
		✓ SX	(F M)			PLC	
		✓ TYN	(HUM)			TAL	
		!AUTHOR	"faselad0"			TOWN	
		!DATE	Tue, Feb 27, 2018			✓ TYN	(HUM)
		!OWNER	"FTI-Uniqe"			ZONE	
						!AUTHOR	"faselad0"
						!DATE	Tue, Feb 27, 2018
						!OWNER	"FTI-Uniqe"

Figure (5) : Entrées monolingues du terme « elf – elfe », langue source (anglais) à gauche et langue cible (français) à droite

Active Entries		Feature Name	Value
l-src	elf (NST)	✓ SLCAN	"elf"
l-tgt	★ elfe (NST)	SLCAT	NST
l-xfr		✓ TLCAN	"elfe"
edi-src		TLCAT	NST
edi-tgt		✓ CMT	"Race WoW (PJ)"
def-src		CUST	
def-tgt		P-S	
		PRF	10000
		✓ TAG	(CV)
		TESTS	
		✓ XFMS	((XFR-NST-MAP :SL-CAT NP :TL-PREPCA
		!AUTHOR	"faselad0"
		!DATE	Thu, Feb 22, 2018
		!OWNER	"FTI-Uniqe"

Figure (6) : Entrée bilingue du terme « elf – elfe » (dictionnaire de transfert)

Les attributs des mots disponibles sur Lucy LT portent sur les caractéristiques des termes entrés dans les dictionnaires, tels le nombre et le genre des substantifs, et sont représentés dans les Figures (5) et (6) ci-dessus dans les colonnes **Feature Name** et **Value**. Le Tableau (1) présenté ci-dessous résume les attributs principaux et leur signification. Notons que ce tableau reprend les attributs des Figures (5) et (6), qui présentent des substantifs (NST). Il existe des attributs propres aux autres catégories grammaticales, notamment les verbes (VST) et les adjectifs (AST).

Nom de l'attribut	Exemples de valeur	Signification
CAN	"elf", "elfe"	Forme canonique
CAT	AST, NST, VST	Catégorie grammaticale
ALO	"elf", "elv"	Allomorphes avec des racines différentes
ABB / ABB-DOT	T / OBL, POS, IMPOS	Abréviation / Avec ou sans point
ANA	T	Allomorphe uniquement utilisé pour la phase d'analyse
ARGS	(N1 (PREP "de"))	Arguments (compléments)
CA (anglais)	O, S, P	Cas (rôle)
CL	PM-S, SM-0	Classes morphologiques (relatives au nombre)
CMT	Race WoW	Commentaire (auteur)
DR	T	Présence de déterminant
FW	T	<i>Function word</i> (déterminants, prépositions, etc.)
GEN	T	Allomorphe utilisé uniquement lors de la génération
KN / NU	CNT, MS / SG, PL	(In)comptable / Nombre
ON / ON-SPEC-CO (français)	CO, VO / T	Son voyelle ou consonne / « H » aspiré
SX (anglais) / GD (français)	M, F, N	Genre
PLC	NF, NI, WF, WI	Place de l'allomorphe dans le mot (préfixe et suffixe)
TAL	X-XX, C-QU	Modifications morphologiques (relatives au genre)
TOWN / ZONE	T	Informations sur le type de lieu
TYN	ABS, LNG, NAT, SOC	Type sémantique

Tableau (1) : Présentation des attributs de Lucy LT et de leur signification

Le système de traduction de Lucy LT nous permet également de créer des règles de transfert (*patches*) qui se situent sur trois niveaux : des règles lexicales (afin de donner une traduction par défaut des mots), des règles semi-structurales (qui changent la structure de la phrase en présence de certains mots), et enfin des règles structurales (qui traduisent les éléments de structure, tel le rôle grammatical). Cependant, cette dernière catégorie est fournie par défaut, raison pour laquelle nous ne la détaillerons pas ci-dessous. Nous allons en revanche brièvement démontrer comment fonctionnent les règles lexicales et semi-structurales, que nous appliquerons ensuite lorsque nous spécialiserons Lucy LT (Point III.4, « Spécialisation de Lucy LT »). Soulignons que, bien que l'objet de notre étude soit uniquement la spécialisation du programme Lucy LT, nous avons choisi d'illustrer chacun de nos exemples par une comparaison avec le système minimaliste Reverso, que nous avons présenté plus tôt.

II.5.1 Lucy LT : les règles lexicales

Les règles lexicales sont les règles qui visent à traduire un terme par défaut, c'est-à-dire à attribuer une équivalence source-cible par défaut (sans lien avec le contexte d'énonciation). Il s'agit donc de règles focalisées sur les informations morphosyntaxiques des entrées lexicales (Huë, *in* Bouillon & Clas, 1993), qu'il s'agisse d'un mot, ou d'une expression qui devra se traduire « telle quelle » et non mot à mot.

Si nous prenons par exemple le substantif anglais *dragon*, son équivalent français sera « dragon » par défaut : aucun test ou aucune action n'a été ajouté au couple source-cible, comme présenté dans la Figure (7) ci-dessous.

Feature Name	Value
✓ SLCAN	"dragon"
SLCAT	NST
✓ TLCAN	"dragon"
TLCAT	NST
CMT	
CUST	
P-S	
PRF	10000
✓ TAG	(GV)
TESTS	
XFMS	
IAUTHOR	"sys_LC7.17"
IDATE	Fri, Mar 18, 2016
IOWNER	"sys"

Figure (7) : règle lexicale du terme « dragon – dragon » dans Lucy LT (dictionnaire de transfert)

En revanche, nombre d'entrées ne peuvent se contenter d'une règle lexicale par défaut. Prenons par exemple le verbe anglais *take off*, illustré par les phrases (3) et (4) ci-dessous :

(3) *The plane takes off.*

(4) *I take off the cup.*

Il est clair que le verbe *take off* ne peut se traduire par « décoller » dans les deux cas. En effet, dans la phrase (4), *take off* devrait être traduit par « enlever » ou « retirer ». Comme le démontrent ci-dessous les Figures (8) et (9), Reverso (Figure (8)) et Lucy LT (Figure (9)) ont traduit correctement les deux phrases. Précisons cependant que Reverso laisse une option pour que le traducteur décide si le verbe de la phrase (3) doit se traduire par « décoller » ou « partir », bien que les deux cas soient grammaticalement corrects.

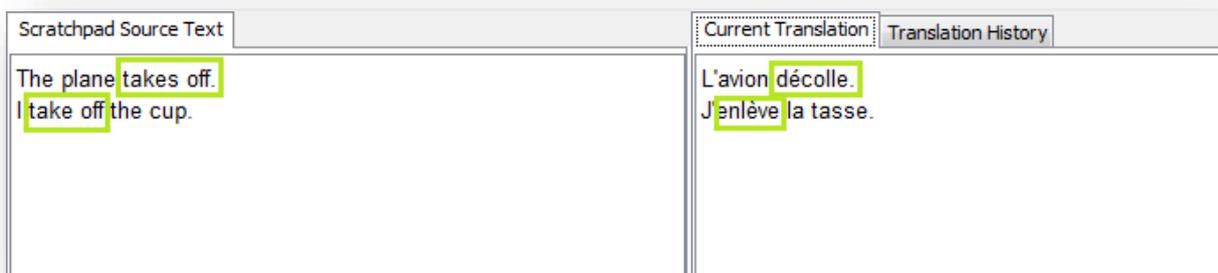


Figure (9) : Traduction des phrases (3) et (4) par Lucy LT

Comme nous pouvons l'observer, la compréhension de Lucy LT sur ces deux phrases est plus poussée que celle de Reverso, puisque le programme a été codé par des tests en langue source sur les différents sens du mot, et va donc au-delà de la règle lexicale (un mot source pour un mot cible).

Pour l'entrée de transfert de « *take off*-décoller », les tests de Lucy LT sont les suivants : (XFR-VST-SL-ARGS :INDEX '(\$SUBJ N1 (TYN HUM SOC CNC ABS)) OPT (\$DOBJ N1 (TYN CNC TMP HUM SOC ABS)))) (XFR-VST-TEST :TEST '(\$SUBJ N1 (TYN CNC ABS))). La première parenthèse signifie que Lucy LT effectue une première analyse du verbe, afin de déterminer si le sujet

(\$SUBJ) de la phrase est de type humain (HUM)¹², social (SOC)¹³, concret (CNC) ou abstrait (ABS), et s'il existe un objet direct (\$DOBJ) optionnel (OPT) de type concret, temporel (TMP)¹⁴, humain, social ou abstrait. La seconde partie montre que, dans le schéma (*frame*) préanalysé, si le sujet est de type concret ou abstrait, la traduction du verbe *take off* par « décoller » sera choisie.

Pour la traduction en français par « enlever », le premier test qu'effectue Lucy LT est le même que pour « décoller » : (XFR-VST-SL-ARGS :DEFAULT T :INDEX '\$(\$SUBJ N1 (TYN HUM SOC CNC ABS)) OPT (\$DOBJ N1 (TYN CNC TMP HUM SOC ABS))'), mais il possède également l'option :DEFAULT T, ce qui signifie qu'en cas d'échec du test, c'est cette traduction qui sera préférée. Les tests suivants pour la traduction par « enlever » sont (XFR-VST-TEST :TEST '\$(\$DOBJ N1 (TYN CNC))'), qui présente que l'objet direct doit être de type concret, et (XFR-VST-TEST :TEST '\$(\$SUBJ N1 (TYN HUM SOC))'), pour que le sujet soit de type humain ou social.

II.5.2 Lucy LT : les règles semi-structurales

Comme nous l'évoquions plus haut, les règles semi-structurales permettent de modifier partiellement la structure de la phrase en cas de divergence. Par exemple, d'une langue à l'autre, il est courant que le thème de la phrase (et, de fait, son rhème) soient modifiés, ellipsés, voire intervertis : il s'agit d'une divergence thématique (cf. Point II.4, « Les problèmes rencontrés en TA »). C'est ce dernier cas que nous pouvons par exemple illustrer avec le verbe anglais *miss*, qui se traduit souvent en français par « manquer ». Or, ce verbe se comporte différemment (et devra donc être traduit différemment) selon son complément. Observons les phrases (5) et (6) suivantes :

(5) *I missed my train this morning.*

(6) *I missed you so much!*

La phrase (5) se traduira par « J'ai raté/manqué mon train », tandis que la phrase (6) se traduira par « Tu m'as tant/tellement manqué ! ». Si ces phrases peuvent certes mener à d'autres traductions, il est clair que, pour la phrase (6), *« J'ai tant manqué toi/de toi » n'est pas une proposition acceptable. En effet le verbe *miss*, selon qu'il soit suivi par un complément animé

¹² Selon l'onglet d'aide du LT LexShop, le type humain comprend les êtres humains ou les groupes d'êtres humains.

¹³ Selon l'onglet d'aide du LT LexShop, le type social comprend les institutions et les organisations.

¹⁴ Selon l'onglet d'aide du LT LexShop, le type temporel comprend les spécifications temporelles, telles *Saturday*, *fortnight*, ou encore « épiscopal » en français.

ou inanimé, ne se comporte pas de la même façon : dans la phrase (6), il est nécessaire que le sujet et le complément (objet direct) soit inversés.

Comme le montre la Figure (10) ci-dessous, Reverso traduit correctement la première phrase, mais n'a pas été capable d'effectuer l'inversion sujet/complément de la phrase (6). Nous pouvons également noter une erreur (minime, certes) au niveau de la ponctuation : en français, il faudrait placer un espace entre la fin du mot et le point d'exclamation, ce que Reverso a omis de faire.



Figure (10) : Traduction des phrases (5) et (6) par Reverso

Lucy LT, en revanche, a été capable de traduire les deux phrases correctement (notons au passage que la ponctuation est correcte cette fois-ci). Comme indiqué dans la Figure (11) ci-dessous, nous avons choisi de modifier les options par défaut de la paire de langue « anglais/français » : nous avons préféré le passé composé pour la traduction du *past tense* et la forme du singulier (« tu ») pour *you*. Avant ces modifications, la phrase (6) était correcte grammaticalement parlant, mais la formulation était un peu lourde : ?« Vous me manquiez tellement ! »¹⁵.

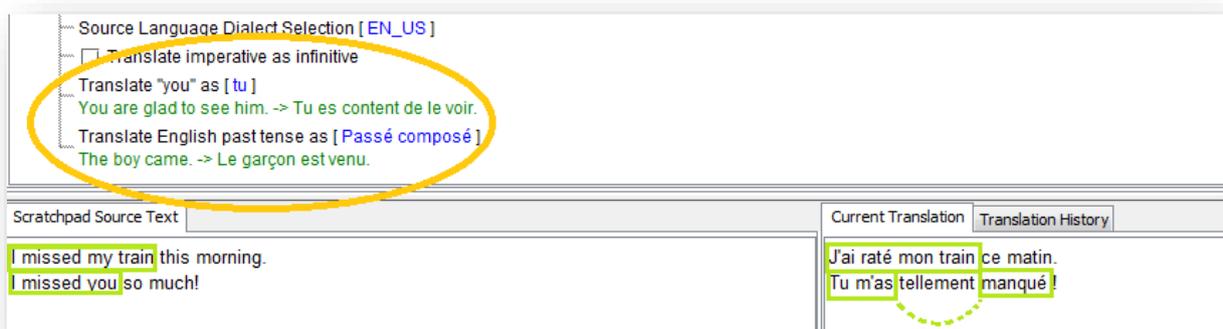


Figure (11) : Traduction des phrases (5) et (6) par Lucy LT

¹⁵ Le point d'interrogation (?) signifie qu'une phrase est grammaticalement correcte, mais un peu étrange pour un locuteur natif.

Si Lucy LT a été capable de traduire les phrases de façon correcte, ce n'est pas dû au hasard : le programme a été codé afin d'effectuer l'inversion entre le sujet et l'objet lors de la phase de génération. Dans la partie dictionnaire de Lucy LT (LT LexShop), l'entrée « *miss* – manquer » du dictionnaire de transfert comprend le test suivant : (XFR-VST-TEST :TEST '(\$DOBJ N1 (TYN HUM SOC))) et les actions correspondantes : (XFR-VST-MAP :SL-ROL \$DOBJ :TL-ROL \$SUBJ) (XFR-VST-MAP :SL-ROL \$SUBJ :TL-ROL \$DOBJ). Le test porte sur la langue source et signifie que lorsque le verbe *miss* est suivi d'un objet direct (\$DOBJ) de type (TYN) humain ou social, les actions qui y sont liées doivent être effectuées. Les actions, elles, portent sur la traduction et signifient que le syntagme qui avait en langue source le rôle d'objet (:SL-ROL \$DOBJ) devra être transformé en sujet pour la langue cible (:TL-ROL \$SUBJ) en langue cible, et inversement (:SL-ROL \$SUBJ :TL-ROL \$DOBJ).

II.6 Présentation de DeepL

En août 2017, l'entreprise allemande de *deep learning* DeepL, anciennement Linguee, a lancé sur le marché un nouveau système de TA neuronal : DeepL, de son nom complet DeepL Traducteur. Le programme, toujours gratuit, est disponible à l'adresse www.deepl.com/translator.

Son succès a été tel que, dès mars 2018, l'entreprise a créé un service de TA élargi, DeepL Pro, qui peut être intégré aux programmes de la TAO, notamment SDL Trados Studio (version 2017). Le tarif de la version pro est de 20€ par mois pour un million de caractère, puis les caractères supplémentaires sont facturés en sus.

Actuellement, le programme prend en charge 7 langues (français, anglais, allemand, espagnol, italien, néerlandais et polonais), qui peuvent être indifféremment choisies comme langues source ou cible. Selon l'entreprise DeepL, leur programme de TA posséderait une « puissance de calcul » capable de traduire « un million de mots en moins d'une seconde »¹⁶.

Si l'aspect technique du programme DeepL Traducteur n'a pas été présenté par ses créateurs, nous savons cependant qu'il s'agit d'un puissant système de TA neuronal (cf. Point II.3.3, « Les systèmes neuronaux »), et la qualité des traductions fournies sera donc certainement très proche d'une traduction humaine. Cependant, l'entreprise DeepL n'ayant aucun lien avec Blizzard Entertainment Inc. et *World of Warcraft*, il sera intéressant de tester leur programme

¹⁶ *Communiqué de presse*, DeepL (2017).

de TA afin de voir s'il est capable de rivaliser avec Lucy LT, que nous allons spécialiser dans le domaine du jeu-vidéo, et puis spécifiquement de *World of Warcraft*.

II.7 Conclusion

Ce chapitre nous a permis de présenter brièvement le concept de traduction automatique, de ses origines à aujourd'hui. Nous avons ensuite dressé une liste des différents systèmes de TA disponibles sur le marché, avant d'expliquer les problèmes auxquels ils se confrontent actuellement, notamment les systèmes linguistiques. Nous avons finalement présenté Lucy LT, programme de TA linguistique, et démontré de quelle façon il utilise les capacités de ses trois dictionnaires afin de donner une traduction optimale. Enfin, nous avons brièvement présenté le programme de TA neuronal DeepL, que nous utiliserons pour comparer les traductions proposées par Lucy LT. Ainsi, nous serons à même de déterminer si, sur un domaine aussi précis que l'univers de *World of Warcraft*, un programme de TA linguistique spécialisé est encore capable d'offrir une meilleure traduction qu'un système neuronal.

III. Développement du corpus et spécialisation de Lucy LT

III.1 Introduction

Afin de déterminer si Lucy LT, le programme de TA linguistique que nous avons choisi et présenté dans le chapitre précédent, est capable d'offrir une traduction de meilleure qualité que le système de TA neuronal DeepL, il nous faut avant tout le spécialiser. Pour cela, nous avons choisi de développer un corpus de texte propre au vocabulaire de *World of Warcraft* (Point III.2). Dans un second temps, nous présenterons notre méthode pour le choix des différentes traductions nécessaires (Point III.3), qui nous a permis de spécialiser de façon optimale Lucy LT pour notre étude (Point III.4). Enfin, nous ferons une brève synthèse de notre spécialisation (Point III.5).

III.2 Développement du corpus de texte

Afin de spécialiser notre programme de TA linguistique, Lucy LT, nous cherchons dans un premier temps à trouver les mots les plus fréquemment utilisés au sein du jeu *World of Warcraft*. Ainsi, notre spécialisation s'effectuera non pas par catégories de mots (« Races et classes », « Monnaies », etc.) mais par liste de priorité. En effet, nous souhaitons que notre spécialisation de Lucy LT soit utile aux joueurs francophones, raison pour laquelle nous devons d'abord déterminer quels sont les mots qu'ils utilisent et lisent le plus, soit les mots les plus récurrents au sein du jeu (conversations et contenu à proprement parler). Pour mener à bien cette phase de spécialisation, nous devons donc d'abord développer un corpus de texte propre au vocabulaire de *World of Warcraft*.

Un corpus, selon Sinclair, est une collection de fragments d'une langue, sélectionnée selon des critères explicites pour servir d'échantillon de la langue en question (Sinclair 1996). Bowker et Pearson précisent en outre que le corpus peut également servir d'échantillon pour un sous-ensemble de la langue étudiée (« subset of that language », Bowker & Pearson, 2002 : 10), et que l'auteur du texte devrait être un expert du domaine (« acknowledged subject-field expert », Bowker & Pearson, 2002 : 51). La question est donc de savoir si l'on peut considérer le vocabulaire propre à *World of Warcraft* comme un domaine en soi, et qui peut être considéré comme un expert du domaine en question.

D'après Bowker et Pearson, un expert est une personne qui a été formée ou travaille dans le domaine d'expertise en question, et qui, de fait, est familière avec la langue de spécialité employée dans ce domaine (« have received training in and/or work in a specialized subject

field, [...] familiar with the LSP used to communicate in that field », Bowker & Pearson, 2002 : 17). De fait, les employés de Blizzard Entertainment Inc. peuvent être considérés comme des experts du domaine : leurs messages sur les forums, leur communication en jeu et le contenu qu'ils développent correspond donc à une langue de spécialité. Cependant, une langue de spécialité peut également inclure un loisir, tant qu'elle traite d'un sujet restreint (« can include [...] hobbies, as long as they treat a restricted subject », Bowker & Pearson, 2002 : 39). Nous sommes donc en droit d'estimer qu'un joueur est également en quelque sorte « expert » du domaine, d'autant que Kittredge estime qu'un *sublanguage* est employé par une communauté de locuteurs (« "community" of speakers », Kittredge, in Nirenburg, 1987 : 59) sans qu'il ne précise s'il s'agit d'experts ou non.

Enfin, Bowker et Pearson soulignent l'intérêt de prendre en compte dans un corpus des textes écrits par des experts à l'intention d'autres experts, et de ne pas se limiter à l'idiosyncrasie d'un seul auteur (Bowker & Pearson 2002). En soi, puisqu'il s'agit du vocabulaire proposé par une seule entreprise, il serait possible de s'y limiter : nous pourrions considérer Blizzard Entertainment Inc. comme des experts ayant le monopole du domaine en question. Cependant, nous avons choisi de nous rapprocher davantage du vocabulaire employé par les joueurs, que nous estimons également spécialistes.

Notre recherche doit donc nous permettre d'obtenir un corpus spécialisé pour le domaine précis de *World of Warcraft*. Selon la typologie proposée par Bower et Pearson, notre corpus se doit donc d'être créé *ad hoc* pour les besoins de notre étude (*DIY corpora*¹⁷), et synchronique : en effet, même si le jeu a évolué depuis 2004, son vocabulaire est resté très similaire. De plus, il s'agira d'un corpus monolingue, duquel nous extrairons le vocabulaire source uniquement, puisque, comme nous le détaillerons plus loin, aucun corpus parallèle n'est à ce jour disponible ou suffisamment complet.

Puisque les textes généralement produits par les joueurs sont échangés par messagerie instantanée, et sont donc inaccessibles en-dehors des sessions de jeu, nous avons choisi d'utiliser pour notre corpus le site WowHead¹⁸ (version originale : anglaise). S'il ne s'agit pas du forum officiel, WowHead a la particularité de regrouper d'une part la plus grande base de données du jeu, des articles régulièrement mis à jour, et d'autre part les messages et

¹⁷ *Do It Yourself corpora*.

¹⁸ WowHead (anglophone) www.wowhead.com.

commentaires des joueurs (forums et commentaires des diverses données de la base), ce qui le rend extrêmement complet et efficace. L'impact des joueurs est en outre plus important sur ce site, raison pour laquelle nous l'avons préféré aux forums officiels, qui offrent une moins grande marge d'expression. Enfin, nous estimons que la cohabitation de textes longuement révisés par Blizzard Entertainment Inc. et des messages bruts (dont l'orthographe et la syntaxe laissent parfois à désirer) des joueurs offre une fourchette plus large de textes, et soulèvera également la question de la nécessité d'une pré- ou post-édition.

Notons que le site WowHead existe aussi en version francophone, mais qu'il ne nous a pas semblé recommandé de l'utiliser pour créer un corpus parallèle. En effet, d'une part les commentaires anglophones sont accessibles depuis la version francophone mais ne sont pas traduits, hormis par quelques joueurs souvent amateurs, comme le montre la Figure (12) ci-dessous.

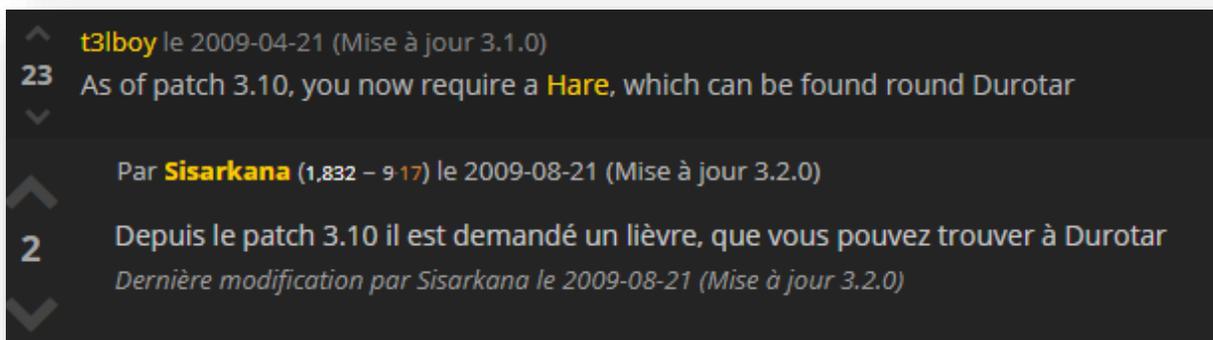


Figure (12) : Traduction « calque » de la version anglaise, à laquelle le renvoi (lien jaune) n'a pas été ajouté

D'autre part, comme dans les Figures (13) et (14) présentées ci-dessous, certains liens sont problématiques lors de la traduction en française : retour à la carte « parent » (Figure (13)), page inexistante (Figure (14)), ou deux traductions différentes pour un même objet¹⁹. Soulignons cependant que certaines erreurs de traduction concernent le jeu *World of Warcraft* lui-même, et ne sont pas seulement une erreur au niveau de la base de données de WowHead, comme nous pouvons l'observer dans la Figure (15) ci-dessous.

¹⁹ Pendant longtemps, un haut-fait (objectif supplémentaire à atteindre) demandait au joueur de créer l'objet [Pommeau ciselé], alors que l'objet créé par le joueur était nommé [Poignée ciselée] (*Engraved Scimitar Hilt*). Notons qu'en raison de notre propre contribution au site WowHead, l'erreur a depuis été corrigée.

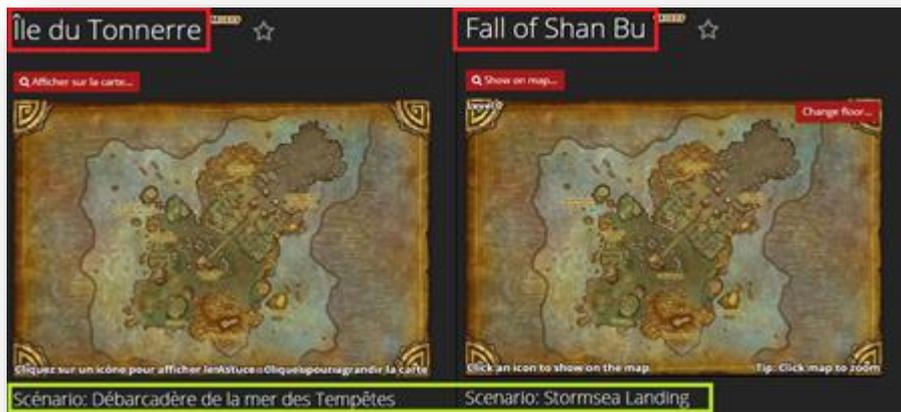


Figure (13) : Bien qu'il s'agisse du même scénario, le titre de la carte en français est trop vague : le scénario se déroule dans un point plus précis de l'île, à la « Chute de Shan Bu » (comme le présente la version anglaise)

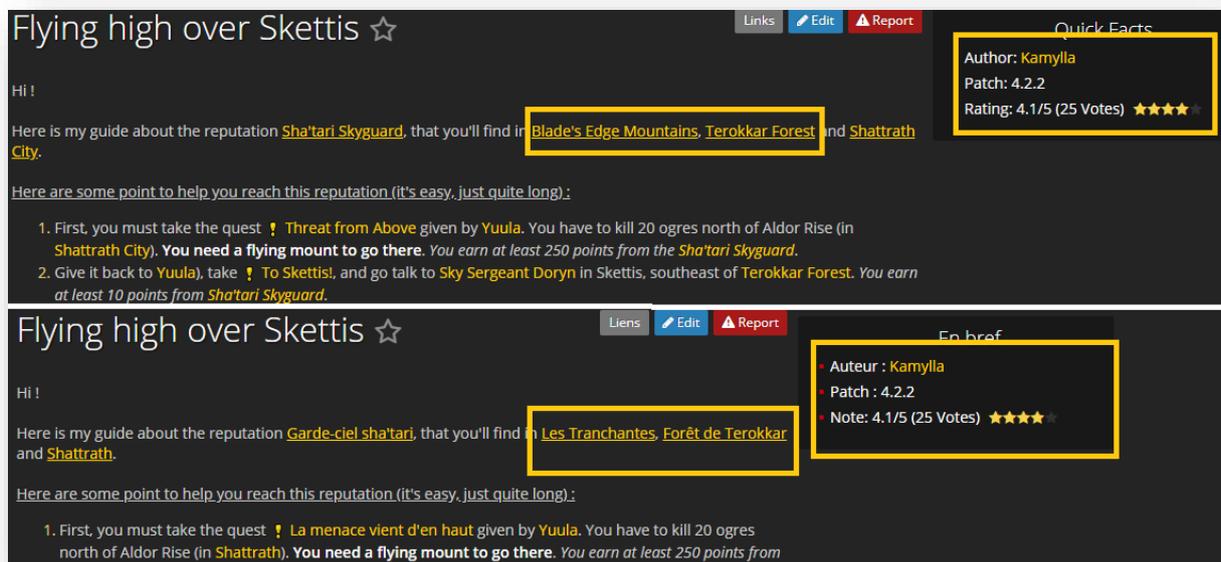


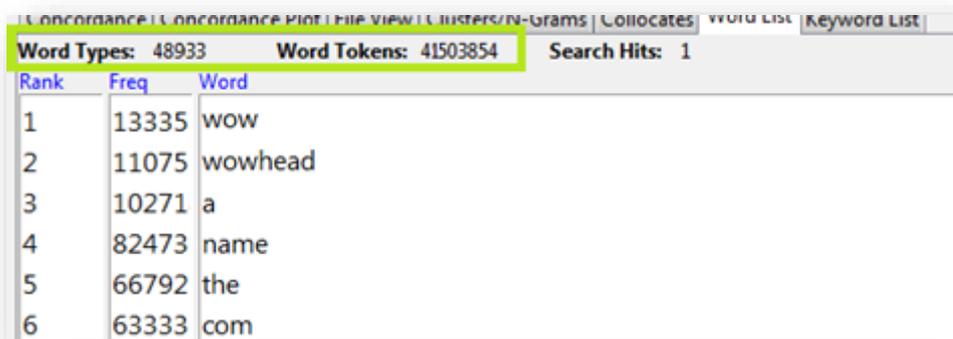
Figure (14) : Le guide n'est pas disponible en version française, pourtant, nous voyons bien que les champs « Auteur » et « Note » sont traduits, de même que le nom des renvois



Figure (15) : Le terme anglais seal est censé se traduire en français par « sceau », et non par « seau »

Afin de créer notre corpus, nous avons choisi d'utiliser l'outil HTTrack pour compiler nos données issues de WowHead au format HTML. Notre compilation comprend toute l'arborescence descendante de la base de données (www.wowhead.com/database), les commentaires et le contenu relatifs, les forums (www.wowhead.com/forum), ainsi que les articles (*news*) du site. Ce choix nous permettra, comme nous l'avons évoqué, d'avoir un corpus représentatif à la fois des textes de Blizzard Entertainment Inc. et des joueurs.

Une fois la compilation de notre corpus complétée (626'347 phrases pour 41'503'854 mots), nous avons navigué dans celui-ci à l'aide du programme AntConc, pour en exploiter les données, comme présenté dans la Figure (16) ci-dessous.



Rank	Freq	Word
1	13335	wow
2	11075	wowhead
3	10271	a
4	82473	name
5	66792	the
6	63333	com

Figure (16) : Données obtenues suite à la compilation de notre corpus, effectuée avec AntConc

Cette compilation nous a permis, après suppression du bruit (termes tels « url », « id », « com », etc.) de trouver les mots les plus utilisés par les joueurs selon leur rang (*rank*) dans la *Word List*, ainsi que de lister les noms officiels des zones, sorts, personnages, objets et quêtes de *World of Warcraft* à l'aide de la *Keyword List* et de l'index de la base de données du site WowHead. Les mots les plus fréquents sont présentés dans le Tableau (2) ci-dessous : après suppression du bruit à gauche, et après un tri pour ne conserver que le vocabulaire de *World of Warcraft* à droite.

Mots les plus fréquents (corpus)			Mots les plus fréquents (<i>World of Warcraft</i>)		
Rang	Fréquence	Mot	Rang	Fréquence	Mot
1	133'351	<i>wow</i>	1	4'445	<i>target</i>
2	110'754	<i>wowhead</i>	2	4'357	<i>item</i>
3	95'344	<i>was</i>	3	3'805	<i>screenshot</i>
4	82'473	<i>and</i>	4	3'621	<i>addon</i>
5	68'478	<i>name</i>	5	3'428	<i>wipe</i>
6	66'792	<i>the</i>	6	3'340	<i>class</i>
7	62'148	<i>is</i>	7	3'223	<i>spell</i>
8	8'299	<i>this</i>	8	2'687	<i>buff</i>
9	4'966	<i>if</i>	9	2'007	<i>quest</i>
10	4'798	<i>for</i>	10	1'564	<i>side</i>

Tableau (2) : Les 10 mots les plus fréquents de notre corpus de texte, après suppression du bruit (à gauche) ;
Les 10 mots les plus fréquents de notre corpus de texte, après tri pour ne garder que le vocabulaire spécifique à *World of Warcraft* (à droite)

III.3 Méthode pour les choix traductifs

Le vocabulaire spécifique à *World of Warcraft* a été créé par la société Blizzard Entertainment Inc., et ce sera donc ce vocabulaire que nous devons traduire avant tout, notamment le nom de tous les objets, lieux, sortilèges et personnages de cet univers très riche. Cependant, nous remarquons rapidement que certains mots sont très peu utilisés par les joueurs, qui leur préfèrent leurs propres expressions, qu'il s'agisse d'anglicismes ou non, voire même d'abréviations. Or, le but de notre recherche sera effectivement de trouver un juste milieu entre le discours théorique (narration), vocabulaire proposé par Blizzard Entertainment Inc. (qu'il s'agisse du texte en jeu ou de celui de leurs produits dérivés), et le discours en situation, discours de *chat* ou de commentaire de forums, employé par les joueurs et parfois fort éloigné de ce que proposent les développeurs.

Bien que nous soyons familière avec le jeu et son vocabulaire, nous souhaitons rester au plus près du langage actuellement employé par les joueurs. Nous avons donc choisi de confirmer nos intuitions de traduction par le biais du site WowHead (version française)²⁰, directement en jeu sur *World of Warcraft*, ou encore à l'aide d'un questionnaire Google Forms distribué auprès d'un panel de 53 joueurs francophones, recrutés dans notre entourage et celui de nos

²⁰ Wowhead (francophone) fr.wowhead.com.

connaissances. Cette méthode de *crowdsourcing* nous a semblé être une approche intéressante puisque, comme le rappelle Aikawa, si les personnes interrogées ne sont pas expertes en linguistique, elles possèdent malgré tout des connaissances solides dans des domaines précis (« highly knowledgeable in specific domains », Aikawa & al. 2012 : 2). Or, dans le cas d'une spécialisation du vocabulaire de *World of Warcraft*, les joueurs sont généralement les plus au courant des termes utilisés au sein de leur communauté.

D'un point de vue statistique, précisons que la moyenne d'âge des joueurs interrogés se situe aux alentours de 30 ans, et que seulement 10% d'entre eux sont des femmes. Si tous sont de langue maternelle française, 65% des joueurs interrogés sont français, et 35% sont suisses. Afin de s'assurer une traduction optimale et représentative de l'ensemble des joueurs francophones, nous avons également posé aux personnes interrogées les questions suivantes :

- Quel type de joueur estimez-vous être ?
- Depuis quand jouez-vous à *World of Warcraft* ?
- Vous retrouvez-vous parfois dans des groupes anglophones (ou, à défaut, des groupes où l'on vous parle en anglais) ?
 - Si oui, à quel niveau se situe votre difficulté de compréhension lors de ces échanges ?

Pour la question « Quel type de joueur estimez-vous être ? », dont les résultats sont présentés dans la Figure (17) ci-dessous, les joueurs interrogés pouvaient choisir entre trois options : « Ancien joueur », « Raideur », ou « *Casual* ».

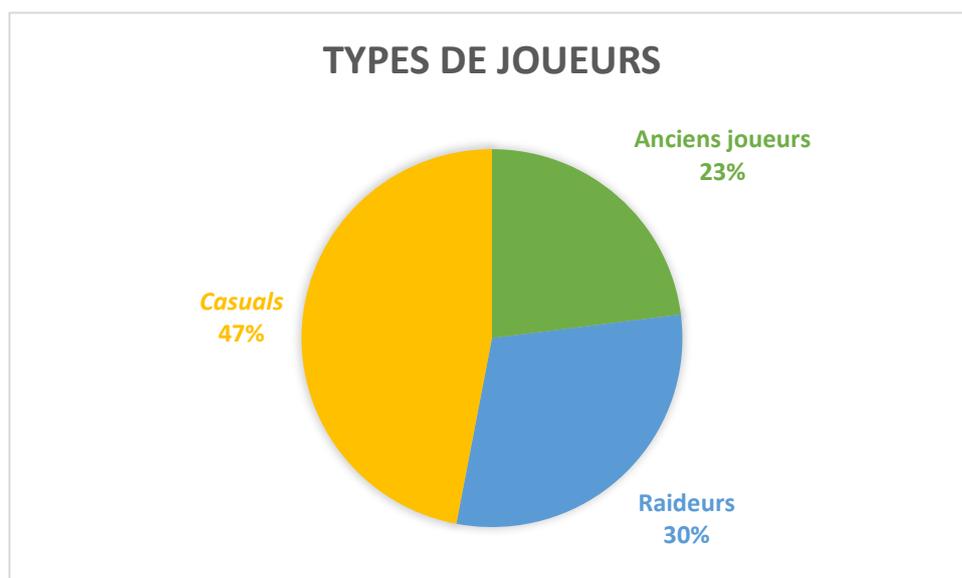


Figure (17) : Réponses à la question « Quel type de joueur estimez-vous être ? »

La catégorie « Ancien joueur » signifie que le joueur a cessé de jouer (interruption de son abonnement), de façon temporaire ou définitive. Puisque les personnes situées dans cette catégorie ne connaissent pas le contenu du jeu le plus récent (soit les termes très récents, issus généralement de la dernière mise à jour), leurs réponses sur le contenu en question n'ont pas été prises en compte.

La catégorie « Raideur » regroupe les joueurs qui cherchent à accomplir un objectif de type PVE (*Player versus Environment*) ou PVP (*Player versus Player*) important. Ces personnes tendent à se connecter très régulièrement, principalement en soirée, et accomplissent des combats plus difficiles qui nécessitent un groupe de joueurs et une cohésion plus importante. Leurs réponses concernant des termes propres aux raids (notamment l'interjection « Héro ! » et les abréviations « LFR » et « LFG », que nous présenterons plus loin) ont été pondérées à double, compte tenu qu'ils en sont les principaux utilisateurs.

Enfin, la catégorie « *Casual* » prend en compte les joueurs plus occasionnels, qui se connectent généralement pour jouer seuls ou en petit groupe. S'il leur arrive de faire des donjons ou des raids, c'est bien plus rare : ils font généralement leurs quêtes, montent leurs niveaux de profession, ou accomplissent des objectifs secondaires du jeu (combats de mascottes, haut-faits d'anciens raids, etc.). Leurs réponses n'ont bénéficié d'aucune modification de score.

Pour la question « Depuis quand jouez-vous à *World of Warcraft* ? », dont les réponses sont présentées dans la Figure (18) ci-dessous, les joueurs interrogés ont dû donner le nom de l'extension lors de laquelle ils ont commencé à jouer à *World of Warcraft*. Nous avons demandé aux joueurs de prendre en compte leur première arrivée sur le jeu. Les personnes ayant arrêté de jouer pendant plusieurs extensions ont donc donné la première extension à laquelle ils aient joué, bien que dans un certain cas le joueur se soit interrompu pendant près de dix ans avant de revenir.

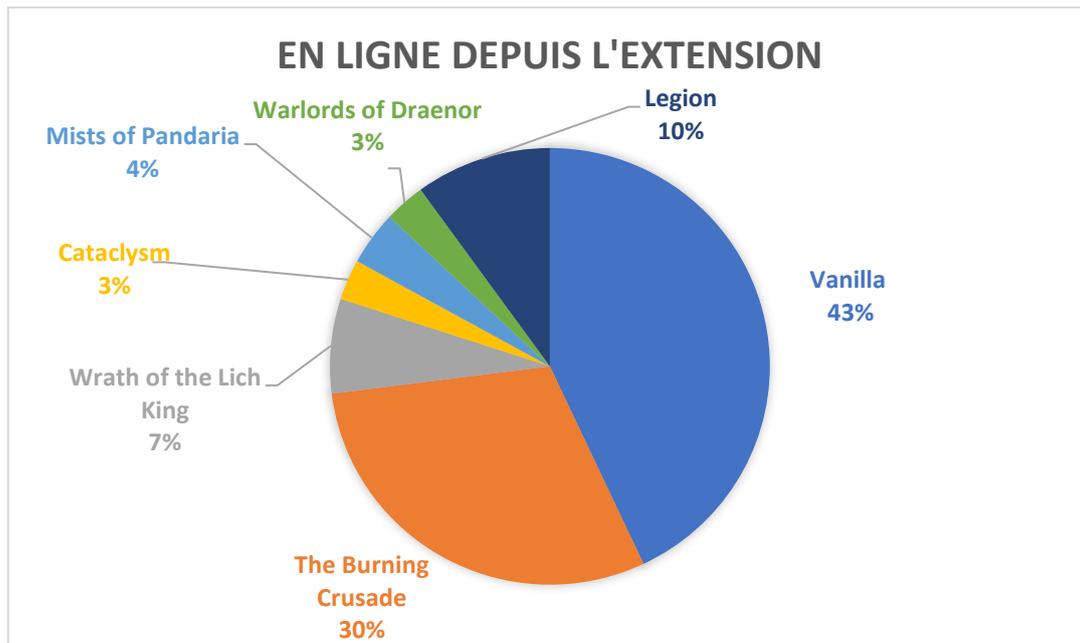


Figure (18) : Réponses à la question « Depuis quand jouez-vous à World of Warcraft ? »

Cette donnée nous a permis d'estimer l'habitude de jeu des personnes interrogées, et nous pouvons voir que près de 75% des joueurs jouent depuis (ou ont joué durant) les premières années de *World of Warcraft*. Ainsi, nos traductions auront pour objectif de couvrir les termes régulièrement utilisés par les joueurs depuis maintenant près de quatorze ans.

Enfin, notre question « Vous retrouvez-vous parfois dans des groupes anglophones (ou, à défaut, des groupes où l'on vous parle en anglais) ? » nous a permis de relever que 70% des joueurs interrogés se retrouvent dans des groupes anglophones ou, à défaut, des raids en langue étrangère où l'on communique en anglais. C'est donc à ces joueurs-ci que nous avons posé une dernière question les concernant, à savoir « [...] à quel niveau se situe votre difficulté de compréhension lors de ces échanges ? », dont les réponses sont présentées dans la Figure (19) ci-dessous. Les joueurs interrogés pouvaient choisir entre les options : « Difficulté modérée », « Difficulté importante », et « Incompréhension (quasi) totale ».

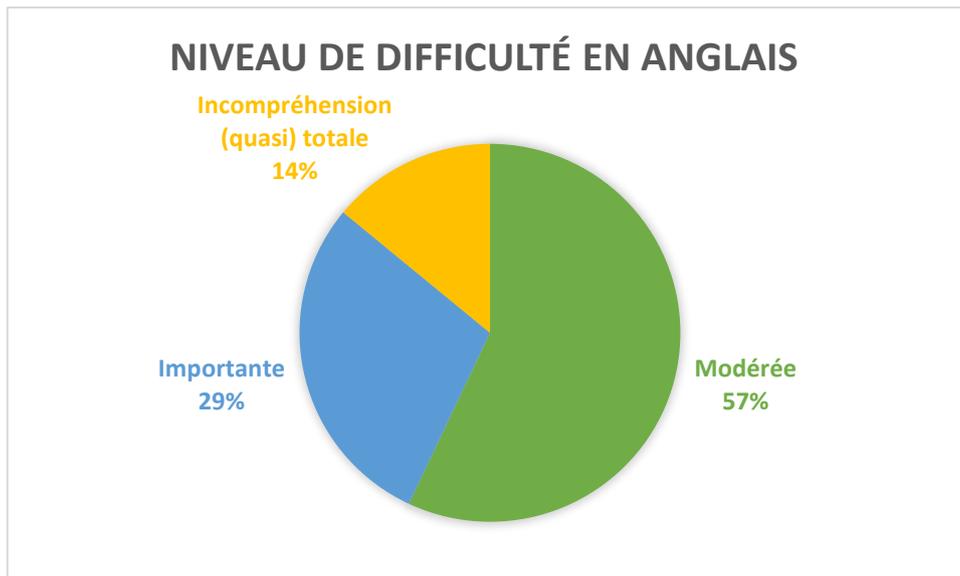


Figure (19) : Réponses à la question « [...] à quel niveau se situe votre difficulté de compréhension lors de ces échanges ? »

Le niveau de difficulté est bien sûr très subjectif, raison pour laquelle nos plages de réponses étaient plutôt vagues. Précisons que nous n’avons aucune personne bilingue anglaise dans notre étude, raison pour laquelle le niveau le plus faible de difficulté de communication en anglais reste « modérée », et non pas « aucune ». En revanche, ces réponses nous permettent d’estimer qu’un outil de TA intégré au jeu *World of Warcraft* pourrait bien se révéler utile, puisque 43% des joueurs interrogés sur cette question (soit près d’un tiers du total des joueurs interrogés) éprouvent de grandes difficultés à communiquer en anglais.

Après ces quelques précisions sur les joueurs interrogés, nous leur avons soumis une liste de questions à l’aide d’un questionnaire Google Forms concernant les termes « problématiques », à savoir ceux qui possèdent en français plusieurs équivalents (généralement l’un issu de l’anglais et l’autre du français). Afin de s’assurer une totale objectivité de la part des joueurs interrogés, nous ne leur avons pas fourni le terme anglais, mais leur avons demandé de choisir le terme français qu’ils utiliseraient pour les descriptions que nous leur fournissions. Les joueurs ont ainsi donné leur avis sur les 67 termes parmi les plus récurrents de notre corpus (Point III.2, « Développement du corpus de texte »), pour lesquels il n’existait pas d’entrées sur le site de WowHead : il s’agissait donc des termes propres au vocabulaire des joueurs, et non du contenu du jeu proposé par Blizzard Entertainment Inc. Le Tableau (3) ci-dessous présente quelques exemples de questions posées et les statistiques de réponses.

Termes anglais évalués	Questions	Réponses proposées (pourcentage de réponses en faveur du terme)				
<i>heal</i> (substantif)	Comment appelez-vous la personne qui vous soigne ?	heal (46,7%)	healer (53,3%)	healeur (0%)	soigneur (0%)	autre (0%)
<i>gear</i> (substantif)	Comment appelez-vous l'armure que vous portez ?	équipement (6,7%)	stuff (90%)	gear (0%)	autre (3,3%)	-
<i>drop</i> (verbe)	Si un monstre vous laisse un objet à sa mort, c'est qu'il le...	droppe (53,3%)	loot (36,7%)	lâche (10%)	fait tomber (0%)	autre (0%)

Tableau (3) : Exemples de questions concernant les termes problématiques

III.4 Spécialisation de Lucy LT

Lors de notre spécialisation de Lucy LT, nous avons classé le vocabulaire en plusieurs catégories : « Classes et races », « Extensions », « Factions et personnages », « Hauts-faits et titres », « Monnaies et ressources », « Objets », « Professions et spécialisations », « Quêtes », « Vocabulaire des joueurs », « Zones (scénarios, donjons et raids inclus) ». Au cours de ce chapitre, nous présenterons plusieurs tableaux des termes que nous avons intégrés aux dictionnaires de Lucy LT. Dans ceux-ci, les termes ont été répartis sur différentes colonnes pour des raisons de présentation et de lisibilité, en fonction de leur statut : « Termes identiques », « Termes traduits par défaut », et « Termes à implanter / modifier ».

La catégorie « Termes identiques » présente les termes qui n'ont pas nécessité de traduction car ils sont identiques en anglais et en français. Cela implique qu'ils n'ont pas eu à être traduits par Lucy LT pour fonctionner. Notre choix de ne pas traduire les termes de cette catégorie s'explique par le fait que nos dictionnaires étaient déjà passablement volumineux, et nous souhaitions éviter une surcharge de termes traduits lorsque cela ne s'avérait pas nécessaire.

Cette catégorie concerne uniquement les zones, et reste minime : 49 contre 338 entrées (modifiées et créées), soit environ 12% du total. En effet, la plupart des noms de zones jeu nécessitent en français un déterminant que la version anglaise ne possède pas, comme le montrent la phrase (7) et sa traduction de référence (7') ci-dessous.

(7) *Mulgore is a wonderful place.*

(7') *Le Mulgore est un merveilleux endroit.*

Si nous n'entrons pas le terme « *Mulgore* – Mulgore » dans les dictionnaires de Lucy LT afin de préciser que le terme prend un déterminant en français, le programme traduira l'unité, alors inconnue, comme un simple nom propre sans y ajouter un déterminant, puisqu'il se base sur la version anglaise.

La catégorie « Termes traduits par défaut » présente les unités que Lucy LT traduit correctement par défaut, sans qu'une modification de la traduction ne soit nécessaire. En revanche, dans quelques cas particuliers, nous avons dû modifier les entrées monolingues, par exemple pour le féminin de « moine » (*monk*), qui devient dans la version française de *World of Warcraft* « moniale », mais n'était pas implanté dans le programme de TA. Cette catégorie représente cependant seulement 13% du total de nos entrées.

Enfin, la catégorie « Termes à planter / modifier » regroupe les entrées lexicales que nous avons dû créer car inexistantes, dans un ou plusieurs dictionnaire(s). C'est notamment le cas du terme « *warlock* – démoniste » : si l'entrée *warlock* existe en anglais, elle était traduite par « sorcier », et le terme français « démoniste » n'existait pas et a donc dû être créé. Cette catégorie représente près de 87% du total de nos entrées et reste donc la plus conséquente.

III.4.1 Mots simples

World of Warcraft est un jeu dont l'univers est extrêmement développé, et il possède donc un vocabulaire très complet. Notre première étape consistera donc à ajouter aux dictionnaires de Lucy LT (monolingues et transfert) les entrées lexicales qui correspondent aux noms du jeu : zones (continents, villes, donjons et raids), classes, races (jouables ou non), objets, quêtes, sorts personnages et factions.

La plupart des termes que nous avons relevés dans notre corpus ont dû être entrés manuellement dans les dictionnaires de Lucy LT. Cependant, certains lieux ou noms n'ont pas nécessité de traduction, puisque le terme est identique en langue source et cible. C'est le cas, entre autres, des villes de Gnomeregan, Shattrath et Darnassus, mais aussi des continents d'Azeroth et de Draenor, des donjons tels Zul'Gurub, Drak'Tharon et Ahn'Qiraj, ou encore de personnages comme Chromie, Nomi et il en serait de même, bien sûr, pour le nom des joueurs eux-mêmes. Notons cependant que le nom de certaines villes, dont la traduction est pourtant identique, est réutilisé pour le nom d'autres zones du jeu, généralement pour préciser les différents endroits

d'une carte (« Égoûts de Dalaran », « Arène d'Orgrimmar », etc.). Nous avons donc dû malgré tout créer ces entrées dans les dictionnaires de Lucy LT, généralement afin que la préposition « de » relie correctement les deux noms.

Nous avons ensuite entré les unités lexicales simples qui devaient être traduites car leur équivalent francophone ne correspond pas à l'anglais. C'est par exemple le cas des donjons *Arcway* (« Arcavia ») ou de *Mechanar* (« Méchanar »), des zones de *Felwood* (« Gangrebois ») ou d'*Ashenvale* (« Orneval »), de villes telles *Ironforge* (« Forgefer »), *Stormwind* (« Hurlevent ») ou *Undercity* (« Fossoyeuse »), ainsi que des races comme les *nightborne* (« sacrenuit ») ou les *tuskarr* (« rohart »).

De nombreuses entrées que nous avons effectuées portaient sur une traduction qui joue sur le sens porté par le terme anglais (par exemple, *Aeda Brightdawn* devient « Aeda Brillelaube »). En revanche, dans certains cas, nous avons dû créer une entrée dans les lexiques de Lucy LT uniquement pour ajouter un accent : dans la version francophone du jeu, *Desolace* devient « Désolace » et *Azjol-Nerub* devient « Azjol-Nérub ».

Notons que, principalement dans le cas des villes, nous avons coché l'option *Town* dans le lexique de Lucy LT, ce qui signifie que ces noms ne peuvent pas prendre de déterminant, alors que l'option *Zone* le permet. Prenons deux exemples, illustrés dans la Figure (20) ci-dessous. *Azjol-Nérub*, ancienne cité des Nérubiens et actuellement donjon de l'extension *Wrath of the Lich King*, doit être prise comme une ville (**TOWN**) dans les options, afin que Lucy LT n'y ajoute pas de déterminant et utilise la bonne préposition (généralement « à » plutôt que « en » ou « dans »). En revanche, le Bassin de Scholazar, jungle primitive du Norfendre, est considéré comme une zone (**ZONE**) : il prend un déterminant et nécessite la préposition « au » ou « dans ».

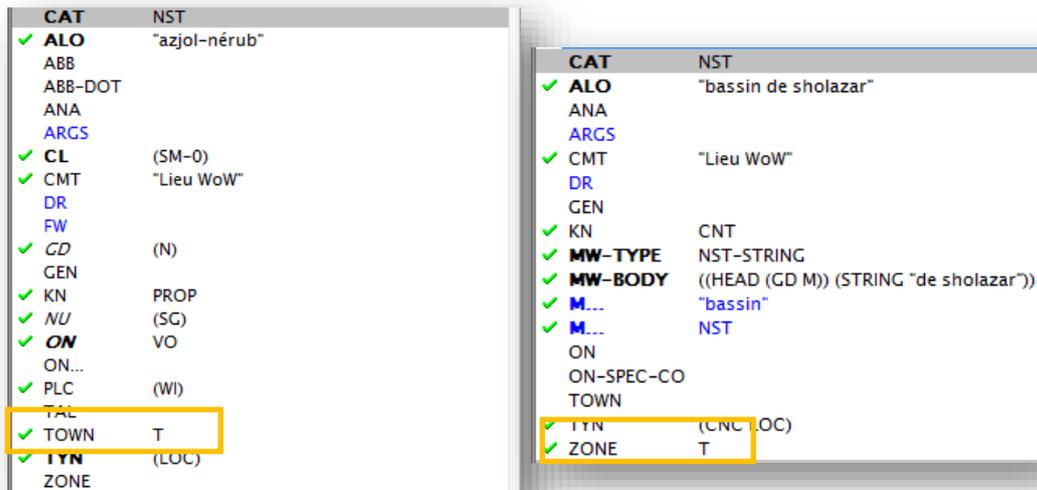


Figure (20) : Les options (**TOWN**) et (**ZONES**), disponibles pour les entrées du lexique français de Lucy LT

De plus, pour la plupart des traductions de mots commençant par la lettre « H », Blizzard Entertainment Inc. a choisi de faire des « H » aspirés. Cette option peut également être modifiée dans le lexique français de Lucy LT, afin que les termes tels « Hurlevent » et « Haut-Roc » soient considérés comme débutant par des sons consonnes (« de Hurlevent », *« d’Haut-Roc »).

Comme le montre la Figure (21) ci-dessous, la ville de Hurlevent, capitale des humains située au nord de la Forêt d’Elwynn, se prononce avec un « H » aspiré. Sur Lucy LT, cette option est rendue possible en appliquant à l’entrée lexicale un son initial « consonne » (**ON (CO)**) (le programme traduira alors par exemple *Stormwind City* par « Cité de Hurlevent » et non *« Cité d’Hurlevent »).

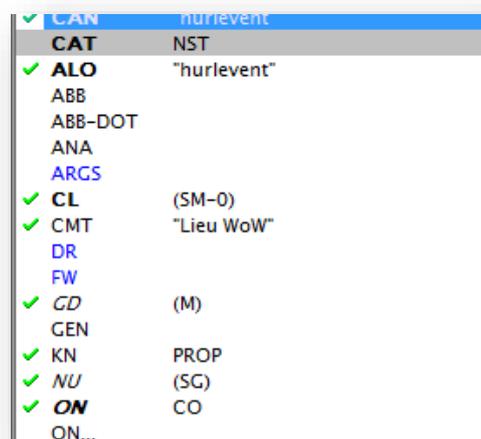


Figure (21) : L’options (**ON**), disponible pour les entrées du lexique français de Lucy LT

Nous avons également choisi d'entrer autant de données que possible au singulier (forme canonique), afin que Lucy LT puisse reconnaître toutes les flexions du mot, même si le terme n'existe pas au singulier dans le jeu. En effet, si certains noms ne sont utilisés qu'au pluriel dans *World of Warcraft*, entrer un terme au singulier dans le lexique permet à Lucy LT de traduire le terme en question, même s'il correspond pas au contexte du jeu, comme l'illustrent dans les phrases (8) et (8') ci-dessous.

(8) *I'm tired of questing in the Wetlands.*

(8') *?I'm tired of questing in the Wetland.*

Dans le contexte du jeu, la seconde phrase n'est pas correcte, puisque la zone n'existe qu'en tant que nom pluriel (« *Wetlands* – Paluns »). La traduction de référence serait donc « Je suis fatigué de quêter dans les Paluns ». Cependant, cela permet au moins au système de donner une traduction qu'un joueur francophone sera capable de comprendre (« Je suis fatigué de quêter dans le Palun »), tandis que si Lucy LT avait été incapable de donner une traduction, la phrase proposée aurait été *« Je suis fatigué de quêter au Wetland », ce qui est bien plus difficile à inférer pour un joueur francophone. Dans un second temps, nous avons ajouté une simple action dans le dictionnaire de transfert (XFR-NST-PUT :TL-NU PL), qui nous a permis de demander au programme de toujours traduire le terme *Wetland(s)* par le pluriel en français (« Paluns »). Ainsi, même si le joueur anglophone écrit *Wetland* au lieu de son pluriel, le joueur francophone recevra la traduction avec le terme « Paluns » au pluriel, comme indiqué sur sa carte ou dans ses quêtes, améliorant considérablement sa compréhension du jeu, surtout dans le cas d'un nouveau joueur, généralement moins familiarisé avec le nom des zones.

Dans de nombreux cas, notre travail a été facilité par l'ajout du nom d'une zone, qui a suffi à la traduction de plusieurs groupes. En créant l'entrée « *Alterac* – Alterac », Lucy LT a été capable de traduire sans modifications supplémentaires les termes *Alterac Mountains* et *Alterac Valley* (respectivement « Vallée d'Alterac » et « Montagnes d'Alterac »). De même, certains termes de noms composés étaient déjà correctement traduits par défaut par Lucy LT, et l'ajout d'un seul mot a permis une traduction complète du terme. Nous avons utilisé cette opportunité autant que possible afin d'éviter de surcharger nos dictionnaires d'entrées composées lorsque celles-ci ne s'avéraient pas nécessaires. Ainsi, l'ajout de l'adjectif « *borean* – boréen » a permis la traduction de la zone *Borean Tundra* par Toundra Boréenne, puisque l'entrée « *tundra* – toundra » existait déjà au sein du lexique de Lucy LT.

Lorsque nous avons implanté les noms des races dans le lexique de Lucy LT, la question du féminin des termes s'est posée. En effet, le féminin n'est pas visible en anglais de façon graphique, et ne se comprend que selon le contexte. Nos termes seront donc au féminin lorsque le pronom *she* est utilisé (*She's a pretty tauren* – « Elle est une jolie taurène »), mais le programme ne sera pas capable de traduire au féminin uniquement en fonction du contexte (*I play a pretty draenei* – « Je joue un joli draeneï », même s'il s'agit d'un personnage féminin). En revanche, nous pouvons imaginer que les développeurs de *World of Warcraft* pourraient insérer dans le code du jeu une ligne de commande qui spécifierait le sexe du personnage joué et, de fait, permettre à notre programme d'inférer si le personnage en question doit parler au féminin ou non.

De plus, si certaines races sont invariables en français comme en anglais (« *draenei* – draeneï »), d'autres sont invariables uniquement en anglais (« *vrykul* – vrykuls », « *naga* – nagas », « *taunka* – taunkas »). Nous avons modifié les options en français afin que le pluriel puisse apparaître, cependant Lucy LT ne traduira par le nom au pluriel que lorsque le verbe est inféré correctement par le programme, soit uniquement au présent de l'infinitif, aux temps composés, ou encore le verbe *be*, comme l'illustrent les phrases (9), (9') et (9'') ci-dessous.

(9) *The taunka dies too quickly.* – « Le taunka meurt trop rapidement. »

(9') *The taunka die too quickly.* – « Les taunkas meurent trop rapidement. »

(9'') *The taunka died too fast.* – ?« Le taunka est mort trop rapidement. »

Les deux premières phrases ci-dessus sont traduites correctement car Lucy LT reconnaît la différence du verbe *die* conjugué au présent de l'indicatif selon qu'il s'agisse de la troisième personne du singulier (ajout du -s final) ou non. En revanche, au *past simple*, rien ne permet au programme d'inférer s'il s'agit du singulier ou du pluriel, aussi le syntagme *taunka* sera traduit par défaut au singulier.

III.4.2 Mots complexes

Afin de lever les ambiguïtés lexicales, nous avons suivi une proposition de Slodzian : entrer les unités lexicales complexes comme des mots composés ou expressions figées (Slodzian, *in Segond 2002*). C'est notamment la norme que nous avons mise en place dans des cas comme celui de la race (non-jouable) des *fel orcs* (« gangr'orcs »). Nous avons en effet décidé d'entrer ce terme comme une entrée *multiword* (ci-dessous « MW ») en anglais, mais simple en français (*single word*, ci-dessous « SW ») en raison de son apostrophe (dans le cas d'une entrée MW

avec apostrophe, un espace est ajouté à l'intérieur du mot, ce que nous voulions éviter autant que possible). Ainsi, la phrase **(10)** ci-dessous

(10) Fel orcs and taunka cannot be played: they are only NPC.

sera traduite par : « Les gangr'orcs et les taunkas ne peuvent être joués : ils sont seulement PNJ » et non par *« Les orcs et les taunkas gangrenés [...] » : le terme *fel* est uniquement attaché aux orcs.

La plupart de nos entrées SW correspondent à des noms propres composés, comme par exemple le donjon de *Grim Batol* (également « Grim Batol » en français). En créant une entrée à part dans le lexique, nous avons empêché que la traduction de *grim* (« sinistre ») soit utilisée par Lucy LT, qui analyse désormais le terme comme un ensemble. En revanche, nous n'en avons pas fait une entrée MW puisqu'elle ne possède pas de « tête » (**HEAD**) à proprement parler en français, et qu'il nous aurait alors fallu ajouter une entrée simple (*grim* ou *batol*) dans les dictionnaires monolingues.

Concernant la traduction des noms des extensions de *World of Warcraft*, il nous a fallu laisser les termes proposés en anglais, puisqu'ils n'ont pas été traduits dans la version française²¹. Tous les termes ont donc été entrés comme des mots simples (SW) et non des *multiwords* étant donné qu'ils ne s'accordent ni en nombre ni en genre. De plus, notre volonté était de ne pas surcharger nos dictionnaires déjà conséquents et de n'y ajouter que des mots qui font sens. Or, les termes sont uniquement en anglais même lors de la traduction et il aurait dès lors fallu créer une entrée simple séparée pour le mot employé comme « tête » du *multiword* en français. Par exemple, dans le nom d'extensions *Wrath of the Lich King*, : le terme *wrath* peut servir de tête pour un *multiword* anglais (et *of the Lich King* serait le corps (**STRING**) du mot composé), mais en français ce terme n'existe pas et aurait donc dû être ajouté.

Nous avons également parfois dû remplacer une traduction existante par celle qui est utilisée en jeu. Bien sûr, nous pouvons également ajouter des tests afin qu'un mot anglais puisse se traduire par plus d'un équivalent en français, mais après vérification, certains termes proposés par Lucy LT ne sont absolument pas utilisés au sein du jeu. C'est par exemple le cas du terme *Badlands*, traduit par défaut en français par « bad-lands ». Or, dans *World of Warcraft*, les *Badlands* sont une zone intitulée en français « Terres ingrates ». Puisque le terme francophone

²¹ Les différents mots employés dans le nom des extensions ont souvent été traduits en jeu (le nouveau continent de « *Pandaria – Pandarie* », par exemple), mais les noms des extensions elles-mêmes n'ont pas été traduits.

« bad-lands » n'est jamais utilisé dans l'univers du jeu, nous avons préféré supprimer la traduction proposée et imposer la nôtre.

De même, nous avons parfois dû empêcher la traduction pour conserver un mot identique à la version anglophone de *World of Warcraft*, bien qu'il s'agisse parfois presque d'une traduction « calque ». C'est le cas de l'*Arena of Annihilation*, pour laquelle Lucy LT proposait la traduction « Arène de l'Anéantissement ». Si le sens de cette traduction est correct, la zone s'appelle pourtant « Arène de l'Annihilation » en français. Nous avons donc ajouté une entrée « *annihilation* – annihilation » et y avons appliqué un test qui en fasse la traduction préférée (XFR-NST-TEST :PREF-ALT T), que nous présenterons plus en détail au point III.4.4, « Structures complexes ».

Précisons aussi que nous avons rencontré certains cas de traduction « biaisée ». En effet, la zone *Arathi Blizzard* est traduite en français par « Bassin Arathi » par Blizzard Entertainment Inc., tout comme *Arathi Basin*. Afin de ne pas interloquer les joueurs en recréant une traduction plus exacte que celle généralement utilisée, nous avons bloqué la traduction en créant une entrée MW « *Arathi Blizzard* – Bassin Arathi », même si cette traduction est clairement fautive et souligne que certains points du jeu n'ont pas encore été complètement pris en charge dans la version française.

Enfin, notons que, malheureusement, la TA n'est actuellement pas encore de taille à traduire les « jeux de mots » ou, plus généralement, les noms proposés par *World of Warcraft* qui contiennent un sens au-delà du simple nom propre. Par exemple, même si Lucy LT connaît et sait traduire les termes *wild* et *hammer* ou encore *frost* et *wolf*, le programme n'est pas capable de comprendre que le clan *Wildhammer* se traduira par « Marteau-Hardi », et le peuple *Frostwolf* par « Loup-de-Givre ». Pourtant, les mots existent séparément (en unités simples) dans les lexiques de Lucy LT et pourraient donc être « compris » et traduits. Si nous pouvons arguer qu'un traducteur humain ne traduirait pas forcément les mots de la même façon que la section francophone de Blizzard Entertainment Inc. a choisi de le faire, il lui serait cependant aisé d'en comprendre le sens et de le conserver. De ce point de vue, le programme de TA reste bel et bien une machine, incapable d'inférer ce qu'un humain pourrait comprendre et considérer comme une sorte de « néologisme » propre au jeu *World of Warcraft*.

III.4.3 Choix des traductions

Dans le vocabulaire des joueurs, nous pouvons remarquer de nombreux cas où la traduction des termes anglais reste mitigée pour les joueurs francophones, et il n'est pas rare que ceux-ci utilisent alors plusieurs équivalents en français pour un seul mot anglais.

À la suite du sondage effectué auprès de notre panel de 53 joueurs francophones, nous avons choisi de ne conserver que la traduction préférée, soit celle qui avait obtenu le pourcentage de réponses le plus élevé. Cependant, les scores sont parfois serrés, et ce pour deux raisons. La première, c'est qu'il s'agit de synonymes « purs », puisqu'il n'existe aucune nuance de sens, même minime, entre les différents mots employés : de fait, nous n'aurons aucune perte sémantique à déplorer en ne conservant que l'un des termes. La seconde, c'est qu'il est impossible pour le système de déterminer le « mot préféré » du joueur qui utiliserait le programme.

Rappelons que le langage contrôlé rend d'autant plus efficace les systèmes de TA qu'ils contiennent un nombre de mots restreints et, s'il nous est impossible d'imposer un tel langage aux joueurs, nous pouvons néanmoins en contrôler les traductions. Cependant, si nous devons inverser les langues source et cible, il faudrait alors pointer plusieurs mots français vers une seule traduction anglaise, puisque c'est généralement en français que sont nés de (trop ?) nombreux synonymes.

En raison du nombre très important de termes propres aux joueurs francophones que nous avons dû traiter, nous n'en présenterons ci-dessous qu'un petit échantillonnage, afin d'éclairer les différentes questions que nous avons pu nous poser durant la réalisation de notre travail. Ces exemples nous serviront à illustrer les divers obstacles rencontrés lors de notre spécialisation de Lucy LT, ainsi qu'à poser certaines problématiques concernant les traductions choisies par les joueurs.

Heal (verbe et nom) : si la forme verbale *heal* est largement conservée en français (plus de 90%)²², le substantif est davantage sujet à controverse. Il existe en effet deux formes en français pour le terme *heal* lorsqu'il désigne une personne qui soigne : « healer » et « heal ». Dans le premier cas, préféré à 53%, le mot a été grandement « francisé » car il existe au féminin

²² Nous entendons par là que le verbe *heal* est largement préféré au verbe français déjà existant « soigner ». Cependant, il est régi par les règles habituelles des verbes français du premier groupe et s'utilise le plus souvent sous sa forme « francisée » : « healer » (je heale, tu heales, etc.).

(« healeuse »), ainsi qu'au pluriel (« healers », « healeuses »). En revanche, le second cas, utilisé par 47% des joueurs interrogés, est beaucoup plus proche de sa forme anglaise puisqu'il ne s'accorde qu'en nombre (« heal », « heals »). Cependant, ce cas induit ce que Fuchs appelle une « polycatégorie » (Fuchs, 1993 : 91), c'est-à-dire une dérivation impropre du concept : le terme « heal » peut également désigner le sort de soin (« Tu me lances un heal, stp ? »).

Drop (verbe et nom) : Le « drop » signifie l'objet qu'un monstre donne à sa mort (et qui se ramasse généralement sur son cadavre). Le verbe qui en est issu (« dropper ») est utilisé par 53% des joueurs interrogés, et seulement 10% utilisent le verbe « lâcher » (« Le boss a enfin lâché l'épée dont j'ai besoin ! »). En revanche, 37% des joueurs utilisent plutôt le verbe « looter », ce qui pose un problème, puisque ce verbe est également utilisé pour exprimer l'action de récupérer un objet sur un cadavre (action du joueur) ; cela crée donc une polysémie pour le verbe « looter ».

Loot (verbe et nom) : Le « loot » est l'objet qu'un joueur ramasse (généralement sur un cadavre), et le verbe « looter », qui en est issu, signifie donc « ramasser un objet ». Ces mots sont préférés par 93% de nos joueurs, contre 7% pour le terme « ramasser ». Cependant, comme nous le précisons ci-dessus, le verbe est également utilisé pour décrire l'action d'un ennemi qui, en mourant, laisse un objet pour les joueurs. Cette polysémie est moins problématique dans une traduction de l'anglais au français, mais si nous devons inverser les deux langues de travail, il faudrait trouver une action permettant au programme de savoir si, par « looter », le joueur veut dire *to drop* (« Ce mob n'a à nouveau rien looté ! ») ou *to loot* (« J'ai enfin looté un légendaire. »).

Buff (verbe et nom) : Le substantif « buff » et son verbe « buffer » signifient respectivement « amélioration » et « améliorer », sous-entendu « par le biais d'un sort bénéfique appliqué au joueur, qui augmentera l'une ou plusieurs de ses caractéristiques pour un temps donné ». Les buffs sont notamment appliqués par des sorts, des élixirs, des runes, de la nourriture, ou encore des événements spéciaux. Il s'agit du seul mot de notre liste pour lequel les joueurs tombent entièrement d'accord : ce terme est utilisé par 100% de nos joueurs interrogés, alors que l'équipe francophone de traduction et de localisation Blizzard Entertainment Inc. utilise encore et toujours le terme « amélioration ».

Kick (verbe et nom) : Le terme « kick » vient du sort anglais du même nom qui permet à un voleur d'interrompre le sort que prépare un adversaire. Cependant, de nombreuses autres classes possèdent leur propre version de ce sort : [Cisaille des Vents] (*Wind Shear*) pour les chamans,

[Gel de l'Esprit] (*Mind Freeze*) pour les chevaliers de la mort, [Réprimandes] (*Rebuke*) pour les paladins, [Volée de Coups] (*Pummel*) pour les guerriers, [Coup de Crâne] (*Skull Bash*) pour les druides, [Contresort] (*Counterspell*) pour les mages, [Tir de Riposte] (*Counter Shot*) pour les chasseurs, etc.

Il s'agit du terme de notre liste sur lequel les joueurs s'accordent le moins concernant sa traduction. 44% des joueurs interrogés utilisent « kick » et le verbe « kicker », alors que 43% préfèrent le terme « interruption », et que 13% utilisent le nom du sort (dont 10% utilisent le nom du sort en anglais, indépendamment du fait que leur jeu soit installé en français ou en anglais). En revanche, les joueurs qui utilisent le terme « interruption » utilisent comme verbe une forme figée, « interrupt », et non le verbe « interrompre » : « J'interrupt le boss au prochain coup. », « Pourquoi t'as pas interrupt ? ».

Respawn – « Repop » : Le terme anglais *respawn*, qui signifie le retour à la vie d'un adversaire, personnage joueur ou non, est utilisé tel quel en français seulement par 25% des joueurs interrogés. 60% des joueurs utilisent plutôt le verbe figé « repop » (« Fais gaffe, le monstre va repop à côté de toi ! », « Il repop tous les combien de temps ? »), mais 14% de ceux-ci précisent que, s'il s'agit d'un joueur, ils utilisent le verbe « ressusciter » ou l'abréviation « rez » (« repop » concerne donc les personnages non-joueurs). Enfin, les derniers 15% utilisent une expression diverses autres expressions (« revenir à la vie », notamment).

Gear – « Stuff » : étonnamment, le terme anglais *gear* (l'équipement du personnage, comprenant son armure, ses accessoires et ses armes) a été traduit par les joueurs francophones par un autre terme anglais : « stuff ». Ce terme est très largement répandu, puisqu'il est utilisé par plus de 93% des joueurs, le reste d'entre eux préférant l'équivalent français « équipement ». Le verbe réfléchi « se stuffer » est également utilisé la plupart du temps.

Gank – « Camper » : le verbe anglais *gank* désigne l'action d'un joueur ennemi qui reste près du cadavre du joueur adverse pour le tuer à nouveau dès son retour à la vie. 7% de nos joueurs l'utilisent sous sa forme « francisée » (« ganker »), et 3% utilisent plutôt le terme « harceler » qui, s'il n'est pas un équivalent exact de *gank*, exprime l'action d'un point de vue psychologique plutôt que physique²³. Enfin, le terme largement préféré des joueurs (90%) reste le verbe « camper ». Cette traduction très imagée peut se comprendre comme le fait que le joueur

²³ Bien entendu, le terme « physique » est ici à comprendre comme physique du point de vue du personnage, pourtant virtuel, par opposition au point de vue « psychologique », qui atteint, lui, le joueur.

ennemi « s'installe », « place son campement » devant le joueur qu'il vient de tuer, afin d'être prêt à le tuer à nouveau dès son retour.

Les abréviations (PVP, LFR et LFG) : de façon générale, les joueurs semblent préférer les abréviations aux termes complets. Ainsi, les termes *LFR* et *LFG* (respectivement, *looking for raid* et *looking for group*) sont préférés à 67% aux termes « outil raid » et « outil donjon ». Notons que pour ces deux termes, aucune abréviation française n'a été proposée par la communauté. En revanche, le terme *PVP* (*Player versus Player*) peut se traduire par « JCJ » en français (« Joueur contre Joueur ») et est d'ailleurs utilisé par Blizzard Entertainment Inc. pour la version francophone du jeu. Cependant, pour 87% des joueurs interrogés, « PVP » reste l'option la plus fréquemment employée.

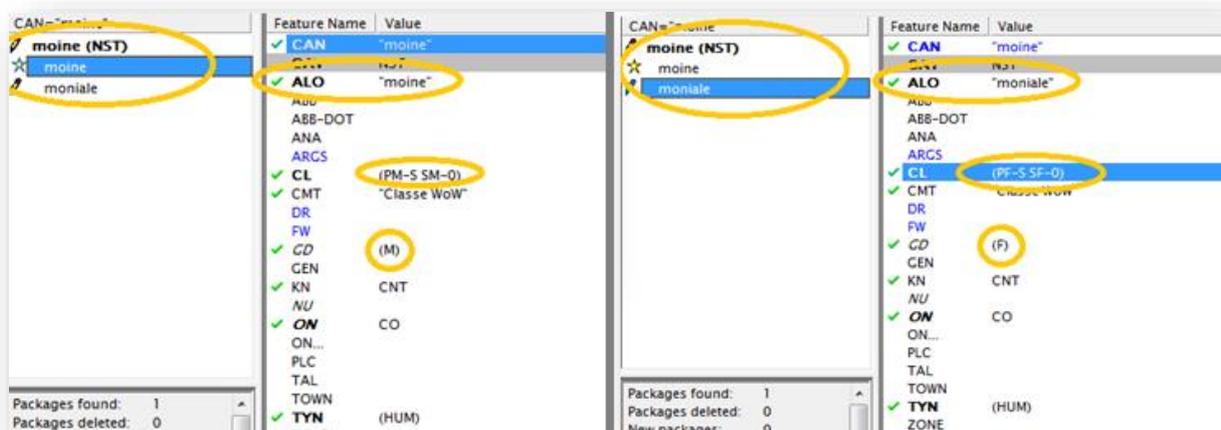
L'héroïsme et ses variantes : Le terme « héro » est sujet à de nombreuses controverses, et ce depuis le début du jeu. En effet, il s'agit d'un sort qui « augmente de 30% la hâte des membres du groupe ou du raid pendant 40 secondes »²⁴ et qui, à l'époque de la première version de *World of Warcraft* (*Classic*, aussi appelé « Vanilla »), n'était disponible que dans la horde puisque seuls les chamans (classe alors réservée aux orcs de la Horde) pouvaient le lancer. Le sort se nommait alors [Furie Sanguinaire] (*Bloodlust*). Cependant, dès la première extension (*The Burning Crusade*), les chamans arrivèrent également dans l'Alliance avec la création de la race des draeneï, et leur sort s'appelait alors [Héroïsme] (*Heroism*). S'agissant d'un sort très puissant et très prisé en raid, il a été décidé durant l'extension *Cataclysm* d'accorder un sort similaire, nommé [Distorsion Temporelle] (*Time Warp*) aux mages, également présents dans les deux factions. Enfin, les chasseurs peuvent désormais dompter des Chiens du Magma (*Core Hound*) et des Raies du Néant (*Nether Ray*), pour combattre à leurs côtés. Ces animaux ont la particularité de pouvoir lancer un sort similaire, respectivement [Hystérie Ancienne] (*Ancient Hysteria*) et [Vents du Néant] (*Netherwinds*).

Ce sort a donc possédé divers noms au court des différentes extensions du jeu, selon la faction et la classe du personnage, et bien évidemment selon la langue employée par les joueurs. À la suite de notre sondage, nous avons pu déterminer que trois termes reviennent le plus fréquemment chez les joueurs : 54% utilisent le terme « héro », soit « héroïsme » sous sa forme raccourcie, 36% utilisent l'abréviation « BL », issue de *Bloodlust*, 7% utilisent « TW », qui vient du terme *Time Warp*, et 3% utilisent divers autres termes.

²⁴ *Héroïsme*, WowHead (version française) (2018).

III.4.4 Structures complexes

Dans les lexiques monolingues français, nous avons parfois dû faire des entrées plus complexes en raison d'une divergence morphologique entre le nom masculin et féminin, aussi appelés des morphèmes flexionnels (L'Homme 2008). Dans le cas de la race *troll*, le féminin français privilégié par Blizzard Entertainment Inc. est « trollesse », et la classe *monk* (« moine ») devient au féminin « moniale ». Dans le lexique anglais, nous avons donc laissé les entrées déjà existantes, en vérifiant simplement que les deux genres soient disponibles pour ces termes (**SX (F M)**). En revanche, pour le français nous avons modifié les entrées « troll » et « moine » afin de vérifier qu'elles ne contenaient plus que le masculin (**GD (M)**). Nous avons ensuite ajouté un deuxième allomorphe pour chaque entrée lexicale, respectivement « trollesse » et « moniale », auquel nous avons ajouté uniquement le genre féminin (**GD (F)**), comme le montre la Figure



(22) ci-dessous.

Figure (22) : La même entrée (« moine (NST) ») possède les deux allomorphes (« moine » et « moniale »), avec leur genre respectif.

Durant notre travail, nous avons également eu recours à des modifications plus complexes, en raison de la structure d'un terme, ou de son contexte d'énonciation. Ainsi, nous avons dû utiliser plusieurs tests et actions afin de proposer plusieurs traductions pour un même mot. En effet, il est rare qu'un terme anglais ne possède qu'un équivalent en français.

Dans le cas d'une traduction déjà existante dans le lexique de transfert de Lucy LT, mais peu ou pas utile dans le contexte de *World of Warcraft*, nous avons utilisé la ligne de test (:**PREF-ALT T**). Celle-ci permet de ne pas altérer la traduction préalablement utilisée par notre programme, afin de ne pas perdre de données déjà écrites. Prenons comme exemple dans le cas de « *lair* – tanière », paire de traduction proposée par Lucy LT. Dans le jeu *World of Warcraft*,

le terme *lair* se traduit par le plus souvent par « repaire », comme le présente le Tableau (4) ci-dessous :

Catégorie	Occurrences « Repaire »	Autres occurrences
Zones	8	1x tanière
Quêtes et missions	21	0
Hauts-faits	24	0
Objets	7	1x antre, 1x tanière
Personnages	2	0

Tableau (4) : Occurrences du terme anglais *lair* et ses traductions respectives

En définitive, la traduction « *lair* – repaire » doit donc être implantée dans Lucy LT avec la ligne de test (:**PREF-ALT T**) afin de devenir l’option par défaut, puisqu’elle est valable dans 95% des cas. Pour les cas restants, comme la zone *Ursoc’s Lair* (« Tanière d’Ursoc »), nous avons créé une entrée lexicale MW séparée.

De la même façon, nous avons dû ajouter une ligne de test pour certains adjectifs, lorsque nous avons besoin que sa traduction change en fonction du nom auquel il se rapportait. La ligne de test (**XFR-AST-TEST :MOD-CAN "XXX"**) permet de changer l’adjectif lorsqu’il modifie (**MOD-CAN**) le syntagme nominal représenté par (**XXX**).

Ainsi, même si la traduction par défaut de Lucy LT est « *old* – vieux », pour l’instance d’*Ahn’Kahet, the Old Kingdom*, la traduction ajoutée est « *old* – ancien » (« Ahn’Kahet, l’Ancien Royaume »), avec la ligne de test (**XFR-AST-TEST :MOD-CAN "kingdom"**). Dans le cas du donjon *Black Morass*, le « Noir Marécage », nous avons non seulement dû entrer cette même ligne de test pour que l’adjectif *black* soit bien traduit par « noir », mais il nous fallait également préciser au programme que, dans ce cas particulier, l’adjectif en français devait se placer devant le nom qu’il modifie. De ce fait, nous avons ajouté une ligne dans les actions du dictionnaire de transfert : (**XFR-AST-PUT :TL-PO BEFMOD**). Grâce à cette commande, la position de l’adjectif en langue française (:**TL-PO**) se trouve bien avant le nom auquel il était lié (**BEFMOD**).

Enfin, dans de nombreux cas, nous avons dû ajouter des prépositions aux syntagmes nominaux que nous utilisons. Si, en anglais, les noms sont souvent juxtaposés (*Blood Furnace, Bloodmyst Isle, Sunwell Plateau, Icecrown Citadel*, etc.), ils doivent généralement en français être reliés par une ou plusieurs prépositions (« Fournaise de Sang », « Île de Brume-Azur », « Plateau du Puits de Soleil », « Citadelle de la Couronne de Glace »). Cette distinction entre les deux

langues est bien intégrée dans Lucy LT, qui ajoute le plus souvent la préposition « de », à condition d'arriver à déterminer si les noms sont bien censés être liés. Précisons que, lorsque les noms font partie du lexique de Lucy LT, la traduction et l'ajout de la préposition est bien plus efficace.

Cependant, il existe dans *World of Warcraft* de nombreux noms, notamment de zones (dont principalement des scénarios), où les prépositions sont également présentes en anglais : *Assault on Zan'Veess*, *Battle for Gilneas*, *Battle of Thunder Pass*, *Battle on High Seas*. Pour ces cas, nous avons dû créer (ou compléter) un mappage qui permet à Lucy LT de traduire les différentes traductions en situation. En effet, il ne serait pas correct d'estimer qu'un simple transfert « *for* – pour » ou « *on* – sur » fonctionnerait dans la multitude de combinaisons possibles avec ces deux prépositions. Ainsi, nous avons entré plusieurs lignes d'actions dans le dictionnaire de transfert de notre programme, afin de s'assurer que, selon le syntagme nominal auquel elle se rattache, la préposition sera correctement traduite :

Assault – « **Assaut** » : (XFR-NST-MAP :SL-PREPCAN '("to") :TL-PREPCAN "à") (XFR-NST-MAP :SL-PREPCAN '("on") :TL-PREPCAN "de"). La fonction (MAP) permet au mappage de se faire, et de lier une préposition anglaise (:SL-PREPCAN) à une préposition française (:TL-PREPCAN). Si les prépositions en anglais sont précédées d'une apostrophe et encadrées de parenthèses ('("to")), c'est parce que plusieurs prépositions peuvent mener vers une même traduction. En revanche, l'inverse n'est pas possible, raison pour laquelle les prépositions en français sont simplement signalées entre guillemets ("à"). Comme nous l'évoquions plus haut, si généralement *on* se traduit par « sur », il est ici nécessaire de le faire traduire à Lucy LT par « de » afin qu'*Assault on Zan'Veess* soit correctement traduit par « Assaut de Zan'Veess ».

Précisons que, tout comme pour nos traductions « préférées » (:PREF-ALT T), nous avons calculé les probabilités afin que nos traductions d'un mot soient toujours celles qui sont le plus représentées sur *World of Warcraft*. Si, dans le cas d'un autre assaut, la préposition *to* ou *on* doit être traduite différemment en français, nous choisirons de créer une entrée lexicale à part, afin de conserver par défaut les traductions les plus fréquentes d'un mot. Par exemple, le syntagme *assault on* se traduit sept fois par « assaut de », et seulement deux fois par « assaut sur » (« sur la Cime de Socrethar », « sur la Fosse »).

Battle – « **Bataille** » : (XFR-NST-MAP :SL-PREPCAN '("of""for") :TL-PREPCAN "de") (XFR-NST-MAP :SL-PREPCAN '("against") :TL-PREPCAN "contre") (XFR-NST-MAP :SL-PREPCAN

'("between") :TL-PREPCAN "entre") (XFR-NST-MAP :SL-PREPCAN ("on") :TL-PREPCAN "en"). Comme nous l'expliquions dans le point ci-dessus, deux prépositions anglaises ("of""for") pointent ici vers la même traduction ("de"). Ici, le mappage est bien plus conséquent car une grande quantité de zones (scénarios), de quêtes, voire même d'objets traitent du combat : le mot battle revient donc très fréquemment, et avec plus de prépositions possibles.

Bien que le mappage « *between* – entre » ne soit pas utilisé dans le contexte de *World of Warcraft*, nous l'avons laissé afin de montrer que Lucy LT est déjà un programme avec une base de traduction importante. Le couple de transfert « *against* – contre » ne correspond à aucun nom de zone ou de quête non plus, mais il est utilisé à plusieurs reprises dans la description de quêtes ou de hauts-faits.

Le couple de transfert « *of* – de » est le plus utilisé, et ne contient aucune exception de traduction (« *Battle of Snowblind Mesa* - Bataille de la Mesa du Voile Blanc »). La traduction « *for* – de » est un peu plus controversée au sein du jeu, mais elle représente quand même environ 70% des traductions, et reste donc notre traduction par défaut (« *Battle for Shattrath* – Bataille de Shattrath », mais « *Battle for Andorhal* – Bataille pour Andorhal »).

Enfin, le syntagme *battle on* n'est utilisé que trois fois au sein du jeu, et dans l'un des trois cas, *on* se traduit par « sur ». Cependant cela ne pose pas de problème dans notre contexte « battle – bataille », car dans ce cas précis *Battle on the Broken Isle* a été traduit par « Combat sur les Îles Brisées » et n'est donc pas compris dans ces lignes d'actions qui sont propres à la traduction « *battle* – bataille ».

III.4.5 Synthèse de la spécialisation

Comme nous l'avons pressenti, la plus grande part de notre travail a été d'entrer des syntagmes nominaux. Certes, nous avons aussi présenté ci-dessus un certain nombre de verbes et d'adjectifs, qu'il nous a fallu entrer dans les lexiques de Lucy LT, mais les noms restent la part principale de notre spécialisation du programme. Le nombre d'entrées effectuées, que nous n'avons pu présenter entièrement dans nos annexes en raison de l'importance du vocabulaire spécifique à *World of Warcraft*, peut se découper comme dans le Tableau (5) présenté ci-dessous.

Catégorie	Nombre d'entrées ²⁵
Classes et races	23 entrées
Extensions	7 entrées
Factions et personnages	350 entrées
Hauts-faits et titres	329 entrées
Monnaies et ressources	68 entrées
Objets	350 entrées
Professions et spécialisations	37 entrées
Quêtes	180 entrées
Vocabulaire des joueurs	67 entrées
Zones (scénarios, donjons et raids inclus)	284 entrées
TOTAL	2'843 entrées

Tableau (5) : Présentation du nombre d'entrées réalisées dans les dictionnaires de Lucy LT dans le cadre de notre étude

Si nous avons souhaité être aussi exhaustive que possible dans ce travail, nous devons cependant admettre que de nombreux termes n'ont pu être traduits. C'est notamment le cas des noms de certaines quêtes et d'objets, ainsi que de nombreux personnages non-joueurs : ces trois catégories additionnées représentaient à elles seules un total de 228'790 entrées. Nous avons donc choisi d'en laisser une grande partie de côté, afin de nous concentrer, dans le cadre de cette étude, sur le vocabulaire le plus utilisé et surtout sur les termes employés par les joueurs, puisque ces derniers ne se trouvent généralement dans aucune base de données, et de fait dans aucun corpus parallèle. Notons également que le vocabulaire des joueurs de *World of Warcraft*, même s'il présente des similarités avec celui d'autres jeux-vidéo, reste un vocabulaire spécifique à part entière.

Nous estimons cependant que ce travail, malgré sa non-exhaustivité, reste digne d'intérêt et que les *patches* créés sur Lucy LT sont en grande partie opérationnels. Il est évident que la section francophone de traduction et de localisation de Blizzard Entertainment Inc. n'aurait probablement aucun besoin de se servir de notre base de données de traduction, mais celle-ci pourrait tout à fait être mise à disposition des joueurs qui utilisent et contribuent au site

²⁵ Par « entrée » nous entendons les termes que nous avons dû entrer au moins dans l'un des lexiques de Lucy LT, qu'il s'agisse du lexique anglais, français, ou de transfert. Les termes qui ont pu être traduits par défaut, ou qui étaient identiques dans les deux langues n'ont pas été comptabilisés dans ce tableau.

WowHead, dont les informations restent principalement disponibles en anglais (notamment au niveau des commentaires, des messages du forum, des *news* et des guides).

III.5 Conclusion

Ce chapitre nous a permis dans un premier temps de déterminer que le langage propre au jeu-vidéo *World of Warcraft* pouvait bel et bien être considéré comme une langue de spécialité, et d'en constituer un corpus de texte qui allait ensuite nous permettre de spécialiser Lucy LT. Nous avons également pu présenter notre groupe de joueurs interrogés et exposer notre méthode de recherche pour les choix de traduction les plus problématiques.

Lors de la spécialisation de Lucy LT, nous avons présenté notre démarche, des entrées les plus simples aux plus complexes, puis nous avons résumé notre réflexion sur les divers choix traductifs que nous avons dû effectuer selon les résultats proposés par notre groupe de joueurs. Enfin, nous avons démontré les possibilités offertes par les nombreux tests et actions proposés par le programme Lucy LT.

Pour clore ce chapitre, nous avons brièvement synthétisé le travail effectué. Celui-ci ne demande qu'à être modifié, amélioré et poursuivi, puisque l'univers de *World of Warcraft* offre un vocabulaire extrêmement large, que les joueurs ont encore développé à leur tour. Cependant, nous avons pu estimer que notre spécialisation de Lucy LT offre une base suffisamment solide pour être soumise à une évaluation de son potentiel.

IV. Évaluation

IV.1 Introduction

Le programme de TA choisi, Lucy LT, étant désormais spécialisé par nos soins au vocabulaire spécifique à *World of Warcraft*, nous allons maintenant évaluer la qualité et la fluidité de la spécialisation effectuée.

Au cours de ce chapitre, nous présenterons d'abord comment ont été choisies nos phrases-tests pour l'évaluation (Point IV.2), puis nous effectuerons une évaluation humaine (Point IV.3), qui permettra de comparer dans un premier temps la qualité des traductions proposées par Lucy LT et DeepL, puis dans un second temps qui en évaluera la fluidité. Ensuite, nous réaliserons une évaluation automatique (Point IV.4), afin de pouvoir comparer les résultats obtenus lors des deux méthodes d'évaluation. Finalement, nous discuterons de l'importance de la qualité de la source et de l'utilité d'une prédiction (Point IV.5).

IV.2 Phrases-tests

Afin de tester la spécialisation effectuée sur le programme de TA linguistique Lucy LT, nous avons récolté 213 phrases en anglais à propos du jeu *World of Warcraft* : 33 phrases nous ont été proposées par 4 joueurs anglophones de notre connaissance (aucun d'entre eux ne maîtrise le français), et les 180 phrases restantes ont été prises sur le site WowHead. Ces dernières ont été choisies aléatoirement à l'aide du générateur de pages aléatoires proposé par le site (www.wowhead.com/random). Afin d'assurer une totale impartialité dans le choix des phrases, nous avons demandé à deux personnes de notre entourage de copier une à deux phrases sur les pages générées. Nous ne leur avons donné aucune autre consigne ou explication, pour que ce choix soit le plus neutre possible, et qu'il ne soit pas basé sur les entrées lexicales que nous avons spécialisées. Cependant, puisque ces 180 phrases ont été prises sur le site WowHead, elles se trouvaient inmanquablement dans le corpus que nous avons compilé (cf. Point III.2, « Développement du corpus de texte »). En revanche, nous avons uniquement effectué une spécialisation du vocabulaire par mots et n'avons pas traité de phrases entières, raison pour laquelle nous estimons que le choix de phrases-test sur le site de WowHead n'influencera en rien notre évaluation.

Les 213 phrases n'ont pas été retouchées, afin d'être au plus près de la réalité rencontrée par les joueurs, et de déterminer si une pré- ou post-édition s'avèreraient nécessaires pour le bon fonctionnement de notre programme.

Une fois ces phrases compliées, nous les avons traduites à l'aide de Lucy LT, dont le vocabulaire est désormais spécialisé pour le jeu-vidéo *World of Warcraft*. Nous les avons également traduites avec le programme de TA neuronal DeepL²⁶, afin de pouvoir comparer les deux traductions, l'une effectuée par un programme spécialisé dans le domaine de *World of Warcraft* (Lucy LT) et une seconde sans spécialisation au domaine (DeepL). Ainsi, nous serons à même d'évaluer si les traductions de notre système spécialisé sont dignes d'intérêt, ou si aujourd'hui un système de TA neuronal est forcément plus performant qu'un programme linguistique.

Comme nous le précisons, les phrases-tests n'ont pas été retouchées, et leurs auteurs étaient donc libres quant à l'orthographe, la grammaire et les opinions qu'ils ont exprimées. Cependant, Arnold présente les règles suivantes pour qu'un programme de TA fonctionne de façon optimale (Arnold & a., 1994 : 26) :

- a) Faire des phrases courtes.
- b) S'assurer une grammaire correcte.
- c) Éviter les constructions grammaticales complexes.
- d) Éviter au possible les mots qui ont plusieurs équivalents en langue cible.

Il est donc évident que plus les phrases recueillies s'éloigneront de ces lignes de conduites, plus la traduction se révélera incorrecte : « the less a text conforms to the rules [...], the worse the raw translation output is likely to be » (Arnold & al., 1994 : 28).

IV.3 Évaluation humaine

IV.3.1 Méthodologie

Pour cette évaluation, nous avons repris le panel de 53 joueurs présentés au point III.3 « Méthode », et qui nous avaient aidé à déterminer les traductions préférées des joueurs de *World of Warcraft*. Nous avons choisi d'effectuer deux types d'évaluations humaines, l'une portant sur la qualité et la seconde sur la fluidité, afin de nous assurer suffisamment de données authentiques, précises et analysables. Nous avons donc transmis à tous les joueurs interrogés deux documents, créés à l'aide de Google Forms, afin de récolter leurs évaluations. Nous avons donc généralement obtenu 53 évaluations par phrase, bien que certains joueurs

²⁶ <https://www.deepl.com/translator> [utilisé du 5 au 8 mars 2018]

aient choisi de ne pas évaluer la totalité des phrases présentées. Ainsi, le nombre d'évaluation par phrase varie entre 47 et 53.

Dans le premier document, pour l'évaluation de la qualité des traductions, nous avons présenté ensemble les deux versions traduites d'une même phrase anglaise (l'une par Lucy LT, l'autre par DeepL), ainsi qu'une option « Les deux sont équivalentes ». Les évaluateurs devaient donc déterminer s'ils préféraient la phrase traduite par Lucy LT, par DeepL, ou si les deux leur semblaient équivalentes (que ce soit en positif ou en négatif). Précisions qu'afin de ne pas biaiser les résultats, les trois choix des réponses étaient aléatoirement mélangés par le système de Google Forms. Ainsi, nos évaluateurs ne savaient pas quelle phrase était issue de quel système de TA.

Dans le second document, pour l'évaluation de la fluidité des traductions, nous avons présenté séparément les deux versions traduites d'une même phrase anglaise (l'une par Lucy LT, l'autre par DeepL), accompagnées que la phrase en langue source. Nous avons alors demandé à nos évaluateurs de noter la fluidité des traductions de Lucy LT et DeepL à l'aide d'un barème imposé, présenté dans le Tableau (6) ci-dessous.

Score de 0	La traduction est incompréhensible ou induit le joueur francophone en erreur.
Score de 1	La traduction est peu lisible, mais le joueur francophone peut deviner le sens global de la phrase.
Score de 2	La phrase « sonne mal » aux oreilles d'un locuteur francophone, mais elle est bien compréhensible.
Score de 3	Le locuteur n'est pas capable de déterminer s'il s'agit d'une traduction automatique ou humaine.

Tableau (6) : Barème imposé aux évaluateurs pour l'évaluation humaine (comparaison de la fluidité)

Si les phrases en anglais ont été laissées à disposition des évaluateurs, rappelons-le, aucun d'eux n'est bilingue. Cependant, nous tenions à ce que les joueurs puissent déterminer la phrase la plus fluide tout en ayant à l'esprit le sens original de la phrase (dans les limites de leur compréhension), notamment pour qu'ils puissent donner le score de 0 en connaissance de cause. Par exemple, dans la phrase (11) et sa traduction de référence (11') ci-dessous,

(11) Male draenei laugh cracks me up.

(11') Le rire des draeneï mâles me fait rire.

la traduction proposée par Lucy LT est « Le rire de draeneï mâle fait me rire ». Cette traduction est certes syntaxiquement incorrecte (inversion verbe/pronom et oubli du pluriel à « mâle », le terme « draeneï » étant invariable), mais le sens global de la phrase est plus proche de l'original que la version proposée par DeepL, pourtant plus lisible : « Les draenei mâles me font rire ».

Dans le Tableau (7) ci-dessous, nous présentons quelques exemples des phrases que les joueurs interrogés ont évaluées. Notons que la première colonne n'était disponible que dans la seconde évaluation comparative (fluidité), et que l'évaluation comparative de qualité était accompagnée d'une option « Les deux sont équivalentes » pour les phrases jugées semblables par les évaluateurs.

Phrases sources <i>(Evaluation de fluidité uniquement)</i>	Phrases cibles (Lucy LT) <i>(Evaluation de qualité et fluidité)</i>	Phrases cibles (DeepL) <i>(Evaluation de qualité et fluidité)</i>
<i>Can I change race of my human to Lightforged Draenei?</i>	Est-ce que je peux changer de la race de mon humain en Draeneï Sancteforge ?	Puis-je changer la race de mon être humain en Draenei Forgé de Lumière?
<i>Goblins are better than Worgen in my opinion.</i>	Les gobelins sont meilleurs que le Worgen dans mon avis.	Les gobelins sont meilleurs que Worgen à mon avis.
<i>As a Mage at level 100 I just nuked him before any spores appeared and got the achievement.</i>	Comme Mage au niveau 100 je viens de le bombarder avant que toute spore soit apparue et ait eu le haut-fait.	En tant que Mage au niveau 100, je l'ai juste atomisé avant l'apparition des spores et j'ai obtenu le résultat.
<i>All Blood Elves are played by kids and they are all noobs.</i>	Tous les Elfes de Sang sont joués par des enfants et ils sont tous les noobs.	Tous les Elfes du Sang sont joués par des enfants et ce sont tous des noobs.
<i>I'm fed up of farming the same thing each week.</i>	Je suis fatigué de farmer la même chose chaque semaine.	J'en ai marre de cultiver la même chose chaque semaine.
<i>The only instance in which fel orcs were active on Azeroth was during the Third War.</i>	Le seul exemple dans lequel les gangr'orcs ont été actifs sur Azeroth a été pendant la Troisième Guerre.	Le seul cas dans lequel les orcs de fellation ont été actifs sur Azeroth a été pendant la Troisième Guerre.
<i>Great, we get to pwn Deathwing's wife...</i>	Grand, nous avons pour l'épouse de pwn Aile-De-Mort...	Super, on a la femme d'Aile de mort....

Tableau (7) : Exemples de phrases évaluées par les joueurs interrogés

Dans la suite du chapitre, nous présenterons et analyserons les résultats obtenus suite à cette évaluation.

IV.3.2 Évaluation comparative (qualité)

Rappelons que dans cette première phase, nos 53 évaluateurs ont pu choisir entre trois options : voter pour la phrase traduite par Lucy LT, voter pour celle traduite par DeepL, ou juger que les deux phrases sont équivalentes (que ce soit en positif ou négatif). La Figure (23) ci-dessous résume les résultats obtenus (moyenne des votes sur les 213 phrases).

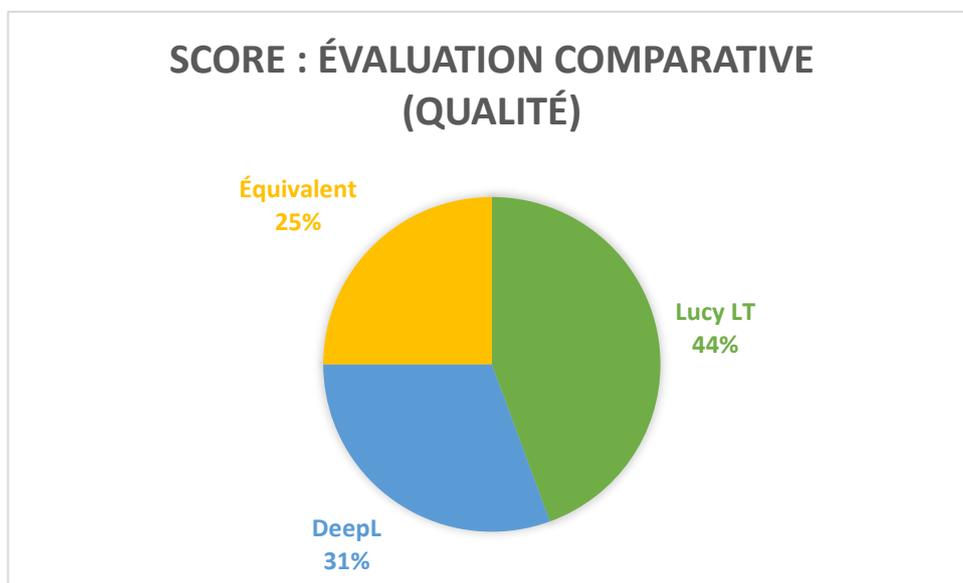


Figure (23) : Résultats moyens de l'évaluation comparative de la qualité des traductions de 213 phrases (Lucy LT et DeepL) ; le nombre de jugements obtenus varie entre 47 et 53 selon les phrases

Les phrases traduites par Lucy LT sont donc préférées à 44,4% par les joueurs interrogés, tandis que les traductions produites par DeepL ne sont préférées que dans 30,6% des cas, et 25% des phrases sont jugées par nos évaluateurs comme équivalentes. L'accord interjuges (score Kappa de Fleiss²⁷) de cette évaluation s'élève à 0,74 (la valeur maximale est 1). Selon Landis et Koch (1977), nous pouvons donc estimer que ces données sont relativement fiables et que les évaluations rendues par les joueurs interrogés concordent. Ces statistiques montrent en outre une prédominance de Lucy LT, notre programme spécialisé, face à au programme neuronal DeepL, comme dans l'échantillon présenté dans le Tableau (8) ci-dessous.

²⁷ <http://www.ccitonline.org/jking/homepage/kappa1.xls> [utilisé en avril 2018]

N°	Phrases cibles (Lucy LT)	Phrases cibles (DeepL)	Score des votes (Lucy LT)	Score des votes (DeepL)	Score des votes (équivalentes)
12	Les taurens commencent à quêter dans Mulgore.	Tauren commence à chercher dans Mulgore.	100%	0%	0%
13	Est-ce que nous pouvons échanger sur trois stacks la fois prochaine ?	Peut-on échanger trois piles la prochaine fois ?	100%	0%	0%
14	Les draeneï peuvent être des Chevaliers de la Mort, des Chasseurs, des Mages, des Paladins, des Prêtres, des Chamans, des Guerriers, ou des Moines.	Draeneï peut être chevaliers de la mort, chasseurs, mages, paladins, prêtres, chamans, guerriers ou moines.	100%	0%	0%

Tableau (8) : Exemple de traductions évaluées en faveur de Lucy LT, avec le pourcentage de votes des différentes options (phrase produite par Lucy LT, phrase produite par DeepL, phrases jugées équivalentes)

Si les phrases de Lucy LT ne sont pas forcément fluides, elles ont au moins l'avantage de contenir un vocabulaire spécialisé pour *World of Warcraft*, et sont donc généralement plus compréhensibles pour nos évaluateurs. Notre seconde évaluation humaine nous permettra, elle, de nous pencher davantage sur la fluidité des phrases traduites. D'un point de vue qualitatif, les fautes de grammaire sont pour le plupart moins dérangeantes que la mauvaise traduction de certains termes. C'est notamment le cas du verbe *questing* (phrase (12)), traduit par « quêter » par Lucy LT grâce à notre spécialisation du vocabulaire des joueurs (contre « chercher » pour DeepL), et du substantif *stack* (phrase (13)), traduit par « pile » par DeepL. Or, la traduction proposée par DeepL pour cette phrase est pourtant plus fluide et proche d'une traduction humaine, puisque l'adjectif « prochaine » a été mis avant le nom « fois », alors que Lucy LT les a inversés (?« fois prochaine »). Notons également que le manque de spécialisation de DeepL empêche le programme de reconnaître certains mots, notamment le substantif *draenei* (phrase (14)), qu'il considère alors comme un nom propre et non comme un nom commun, ce que Lucy LT traite correctement.

En revanche, comme certains termes anglophones sont très répandus parmi les joueurs francophones, les termes non-traduits n'ont parfois pas posé de problèmes à nos évaluateurs, qui ont alors préféré la fluidité offerte par DeepL, comme l'illustre le Tableau (9) ci-dessous.

N°	Phrases cibles (Lucy LT)	Phrases cibles (DeepL)	Score des votes (Lucy LT)	Score des votes (DeepL)	Score des votes (équivalentes)
15	Les taurens de Haut-Roc honorent les esprits de terre, rivière, et ciel.	Les taurens de Highmountain honorent les esprits de la terre, du fleuve et du ciel.	33%	50%	17%
16	Aussi s'il se trouve que vous voyiez Brann Barbe-de-Bronze...	Et si vous voyez Brann Bronzebeard...	0%	83%	17%
17	Beaucoup de personnes disent que les nains femelles n'existent pas.	Beaucoup de gens disent que les naines n'existent pas.	8%	57%	35%
18	La caverne du sanctuaire du serpent est complètement duoable et partiellement soloable.	Serpentshrine Cavern est entièrement duoable et partiellement solitaire.	79%	0%	21%

Tableau (9) : Exemple de traductions évaluées en faveur de DeepL ou Lucy LT, avec le pourcentage de votes des différentes options (phrase produite par Lucy LT, phrase produite par DeepL, phrases jugées équivalentes)

Comme nous pouvons le remarquer, les scores de ces phrases-ci sont beaucoup plus serrés. En effet, les phrases traduites par DeepL sont généralement plus fluides et proposent une traduction qui paraît parfois presque humaine. Dans la phrase (15) notamment, l'ajout par DeepL des pronoms réfléchis et de la répétition de la préposition « du » rend le texte beaucoup plus fluide. De plus, comme le nom de la zone, *Highmountain* (« Haut-Roc ») est bien connu des joueurs francophones, la moitié d'entre eux ont préféré cette proposition et, même si 33% de nos évaluateurs ont voté pour la phrase produite par Lucy LT (très probablement les joueurs qui ne connaissaient pas le terme anglais), les 17% restants ont jugé les deux phrases équivalentes, estimant que le manque de fluidité de Lucy LT était compensé la non-traduction du terme *Highmountain* par DeepL. La phrase (16), répétée ci-dessous, a été évaluée de la même façon par nos joueurs francophones :

(16) *Also if you happen to see Brann Bronzebeard...*

En effet, l'expression *Also if you happen* a été traduite de façon très laborieuse par Lucy LT (*« Aussi s'il se trouve que »), alors que DeepL la modifie, sans en changer pour autant le sens, afin de la rendre bien plus fluide (« Et si »). Puisque *Brann Bronzebeard*, nain archéologue emblématique de l'univers de *World of Warcraft*, est connu tant sous son nom anglais que

français (« Barbe-de-Bronze »), il n'est pas surprenant que les joueurs francophones interrogés aient préféré à 83% la traduction proposée par DeepL, plus naturelle, malgré un terme non-traduit. Notons que pour cette phrase également, 17% des évaluateurs ont estimé que la traduction peu naturelle de Lucy LT compensait la non-traduction de *Bronzebeard* par DeepL.

La phrase (17), répétée ci-dessous, nous permet d'observer une nouvelle fois une traduction « calque » de la part de Lucy LT, qui a parfois dérangé nos évaluateurs.

(17) *Many people say female dwarves don't exist.*

En anglais, la phrase nécessitait la précision *female* puisque l'anglais ne spécifie pas le genre directement sur le mot concerné (hormis quelques exceptions telles *men* et *women*). En revanche, la précision « femelle » n'est pas utile en français, puisque le terme « naine » signifie bien que le nain est de sexe féminin. Si Lucy LT n'a pas les ressources pour traiter cette information et propose une traduction très calquée sur l'anglais (?« les nains femelles »), DeepL parvient à reconnaître et traiter cette information (« les naines »). Notons que si 57% des évaluateurs ont préféré cette solution, 35% d'entre eux ont malgré tout estimé que ces phrases étaient équivalentes quant au sens qu'elles transmettent : il ne s'agissait pas en effet de tester lors de cette évaluation la fluidité des traductions, mais de les comparer afin de chercher si le sens apporté était le même ou non.

Enfin, la phrase (18) de notre tableau nous permet de nous pencher sur un point qu'il nous paraissait particulièrement important d'aborder quant aux traductions proposées par DeepL. Puisque ce système de TA cherche ses informations dans une gigantesque base de données, le programme va parfois traduire « par défaut » certains termes en estimant qu'il s'agit de la séquence la plus probable, là où Lucy LT laisse généralement les mots inconnus en anglais dans la phrase cible. Si la façon de traduire de DeepL permet généralement d'offrir une traduction au moins partiellement correcte de la phrase, elle crée parfois des incompréhensions plus importantes que ne l'aurait été la phrase que partiellement traduite. En effet, les termes *duoable* et *soloable*, qui signifient qu'une quête, un donjon ou un raid peuvent être réalisés respectivement à deux ou seul, s'écrivent en français « duotable » et « solotable ». Là où la sur-traduction de DeepL transforme le terme *soloable* en « solitaire », la phrase cible perd davantage de son sens que la phrase proposée par Lucy LT, qui conserve les deux termes tels quels. Cette perte de sens due à la sur-traduction proposée par DeepL, couplée avec la non-traduction du terme *Serpentshrine Cavern* a encouragé les évaluateurs à voter pour la

phrase proposée par Lucy LT, qui est bien plus lisible du point de vue du vocabulaire, puisqu'il ne manque qu'une lettre à chacun des deux termes, ce qui pourrait tout à fait arriver lorsqu'un joueur fait une faute de frappe ou n'orthographe pas correctement un mot.

En conclusion, cette première évaluation souligne l'importance de spécialiser le programme de Lucy LT, faute de quoi les traductions proposées seraient certainement moins réussies que celles proposées par DeepL. Cependant, grâce à notre spécialisation du vocabulaire spécifique à l'univers de *World of Warcraft* et ses joueurs, Lucy LT se révèle plus efficace et produit des phrases généralement plus compréhensible pour notre panel d'évaluateurs.

IV.3.3 Évaluation comparative (fluidité)

Lors de cette seconde étape, les phrases de Lucy LT et DeepL ont été évaluées séparément par nos joueurs francophones. Le but ici n'était en effet pas de comparer les deux phrases, mais de les lire, détachées l'une de l'autre, afin de se focaliser uniquement sur la fluidité de la traduction. Le jugement de nos évaluateurs reste bien évidemment subjectif, mais nous leur avons demandé de se concentrer avant tout sur la syntaxe, l'orthographe, et la spécificité du vocabulaire employé. Afin de les soutenir dans cette tâche, nous leur avons demandé d'imaginer que les phrases qu'ils lisaient leur était fournies dans le texte de l'une de leurs quêtes, ou qu'il s'agissait d'un message envoyé par un autre joueur. Nous leur avons ensuite demandé de donner aux phrases évaluées un score de 0 à 3 (barème présenté en détail au point IV.3.1, « Méthodologie »).

Le score, présenté dans le Tableau (10) ci-dessous, est bien plus serré pour cette évaluation, ce qui nous prouve bien que la fluidité de DeepL est extrêmement importante et appréciée par nos évaluateurs. L'accord interjuges (score Kappa de Fleiss²⁸) de cette évaluation s'élève à 0,59 (la valeur maximale est 1). Ce score plus bas s'explique facilement par le choix plus large de réponses qui était offert à nos évaluateurs. Cependant, selon Landis et Koch (1977), ce score correspond à un accord interjuges « moyen » : nous estimons donc que les données sont analysables en l'état.

²⁸ <http://www.citonline.org/jking/homepage/kappa1.xls> [utilisé en avril 2018]

Score moyen (Lucy LT)	Score moyen (DeepL)
1,7925	1,7835

Tableau (10) : Score moyen (maximum = 3) de l'évaluation humaine (fluidité) de Lucy LT et DeepL, divisé par le nombre de jugements obtenus (entre 47 et 53 selon les phrases)

Comme nous l'évoquions précédemment, certains termes anglais qui n'ont pu être traduits par DeepL sont bien connus des joueurs francophones, tels la ville d'*Ironforge* (« Forgefer »), même si d'autres termes ont été traduits de façon si surprenante qu'ils ont laissé nos évaluateurs perplexes, comme certaines des exemples du Tableau (11) ci-dessous.

N°	Phrases sources	Phrases cibles (Lucy LT)	Phrases cibles (DeepL)	Score moyen (Lucy LT)	Score moyen (DeepL)
19	The dwarves make their home in their ancestral seat of Ironforge.	Les nains font leur maison dans leur siège ancestral de Forgefer.	Les nains s'installent dans leur siège ancestral d'Ironforge.	1,5	2
20	They took the name "Nightborne" or "Shal'dorei".	Ils ont pris le nom "Sacrenuit" ou "Shal'dorei".	Ils ont pris le nom de "Nightborne" ou "Shal'dorei".	3	2
21	When you turn your worgen to human form...	Quand tu transformes ton worgen en forme humaine...	Quand vous transformez votre ver en forme humaine...	3	0
22	I'm fed up of farming the same thing each week.	Je suis fatigué de farmer la même chose chaque semaine.	J'en ai marre de cultiver la même chose chaque semaine.	3	0
23	Do you know how to go to Ulduar?	Est-ce que vous savez comment aller à Ulduar ?	Savez-vous comment aller à Ulduar ?	3	3
24	Great, we get to pwn Deathwing's wife...	Grand, nous avons pour l'épouse de pwn Aile-De-Mort...	Super, on a la femme d'Aile de mort....	0	0

Tableau (11) : Exemple de traductions évaluées, avec la moyenne du score de fluidité obtenu

La phrase (19) démontre bien que DeepL séduit par sa fluidité, malgré des termes non-traduits, alors que Lucy LT obtient un score moindre avec une traduction plus aboutie, mais bien plus calquée sur la phrase source (*make their home* – *« font leur maison »). En revanche, la phrase (20) nous confirme que pour des mots anglais moins connus des joueurs francophones, la

traduction de Lucy LT est préférée à celle de DeepL : les joueurs francophones ne connaissent pas forcément le terme *nightborne*, d'autant que la traduction française proposée par Blizzard Entertainment Inc. est « sacrenuit » : il n'est pas possible à nos évaluateurs de le traduire « mot à mot » afin d'en deviner l'équivalent français.

Cependant, la spécialisation de Lucy LT a parfois été largement préférée par les joueurs que face aux phrases produites par DeepL, comme par exemple pour les phrases (21) et (22). En effet, la phrase (21) n'a pas été comprise par DeepL, qui a traduit le terme *worgen* par « ver », la rendant totalement incompréhensible pour nos évaluateurs. Rappelons à cette occasion que les traductions produites par Lucy LT et DeepL ont toujours été données séparément à nos évaluateurs, afin qu'ils ne puissent comprendre l'une des deux phrases « à l'aide de l'autre », ce qui aurait biaisé notre évaluation. La phrase (22), quant à elle, a souvent semblé totalement correcte aux joueurs, qui ont hésité à lui donner un score de 3. Cette confusion vient du fait qu'il existe une zone au sein de *World of Warcraft*, implantée durant l'extension *Mists of Pandaria*, où il était possible pour les joueurs de cultiver la terre et d'y planter diverses graines qui, chaque jour, fournissaient des composants nécessaires à la progression de leurs métiers. Après avoir vérifié à l'aide de la phrase source, les évaluateurs ont choisi d'attribuer à la traduction de DeepL un score de 0, puisque, selon notre barème, « la traduction [...] induit le joueur francophone en erreur ».

Enfin, les deux traductions ont parfois été évaluées de façon équivalentes, comme dans le cas des phrases (23) et (24). La phrase (23) a en effet été jugée comme excellente dans ses deux traductions, puisqu'hormis le changement de tournure de phrase interrogative (« Est-ce que » pour Lucy LT, et une inversion sujet-verbe pour DeepL), nos évaluateurs ne savaient dire s'il s'agissait d'une traduction humaine ou automatique. En revanche, dans le cas de la phrase (24), les deux propositions ont été jugées tout aussi mauvaises l'une que l'autre. L'expression *pwn*, orthographe auparavant fautive du terme *own* (« posséder », soit quand un joueur gagne ses combats, et par extension quand il combat et « massacre » aisément son adversaire), est désormais adoptée telle quelle dans l'univers du jeu-vidéo. Cependant, aucun des deux programmes de TA n'a compris ce dont il s'agissait, et leur technique de résolution du problème est alors fort différente : si DeepL a choisi d'effacer ce mot de sa traduction pour donner une phrase malgré tout fluide (que certains évaluateurs souhaitaient noter avec un score de 3 jusqu'à ce qu'ils prennent connaissance de la phrase source), Lucy LT a tenté de le traduire tel quel comme un substantif.

En conclusion, le score beaucoup plus serré de cette seconde évaluation montre bien que seule la spécialisation au vocabulaire de *World of Warcraft* de Lucy LT lui permet de proposer des phrases plus fluides qu'un système de TA neuronal comme DeepL. En effet, les phrases traduites par Lucy LT retransmettent généralement mieux le sens porté par la phrase source, mais sa syntaxe trop souvent fautive ne séduit pas les joueurs francophones, qui donnent souvent un score aussi bon (voire meilleur) aux traductions proposées par DeepL. Si Lucy LT garde une légère avance, c'est donc que sa spécialisation lui permet de mieux analyser les phrases en langue source, tandis que DeepL fait face à une incompréhension totale pour la plupart des termes, pour lesquels il n'a pas été entraîné. En fin de compte, la spécialisation du vocabulaire de Lucy LT lui permet de conserver une avance sur le système de TA neuronal DeepL, à la fois au niveau de la qualité (*adequacy*) de la traduction, et au niveau de la fluidité, quoique le score soit beaucoup plus serré sur ce second point.

IV.4 Évaluation automatique

Pour réaliser une évaluation automatique des capacités de Lucy LT une fois spécialisé, nous avons réutilisé les 213 phrases en anglais que nous avons récoltées auprès de joueurs anglophones et sur le site WowHead, ainsi que leurs traductions par Lucy LT et DeepL. Tout comme pour l'évaluation humaine, ces phrases n'ont fait l'objet d'aucune retouche afin que le résultat de notre étude soit le plus authentique possible. Les traductions de référence ont été d'abord cherchées sur la version francophone de WowHead, lorsqu'il s'agissait de textes proposés par Blizzard Entertainment Inc. (texte de quêtes, info-bulles et description des objets, etc.). Pour les 178 phrases en langue source qui ne possédaient pas de traduction (phrases proposées par les joueurs anglophones, commentaires et messages sur les forums), les traductions de référence ont été effectuées par un joueur bilingue de notre connaissance, afin d'éviter que nous ne proposons notre propre traduction, ce qui aurait pu biaiser le résultat de notre évaluation.

Nous avons choisi d'évaluer les traductions de Lucy LT et de DeepL de façon automatique, à l'aide du calculateur de score BLEU (*Bilingual Evaluation Understudy*, Papineri 2001) en ligne proposé par Tilde²⁹. Comme l'expliquent Hearne et Way, le score BLEU permet d'évaluer la qualité d'un système de TA en comparant une traduction machine à sa traduction humaine de

²⁹ <http://www.letsmt.eu/Bleu.aspx> [utilisé en avril 2018]

référence en analysant le nombre de chaînes de mots, les N-grammes, qui se révèlent identiques (« comparing output translations to their reference translations in terms of the numbers co-occurring n-grams », Hearne & Way, 2011 : 219). Puisque l'évaluation est un score de précision, la phrase issue de la TA est également pénalisée lorsqu'elle est plus longue que cette dernière par un BP, *brevity penalty*, pour éviter qu'une phrase qui est un sous-ensemble de la phrase de référence obtienne un score de 100%.

À l'issue de cette évaluation, il apparaît que le score BLEU obtenu par Lucy LT est supérieur à celui de DeepL, comme le présente le Tableau (12) ci-dessous. Contrairement à DeepL, le score de Lucy LT dépasse même le seuil moyen de 50 (le score BLEU donnant un résultat entre 0 et 100). De plus, 68% des phrases proposées par Lucy LT ont obtenu un meilleur score BLEU que DeepL, qui ne comptabilise que 32% avec un score BLEU plus élevé.

Score BLEU (Lucy LT)	Score BLEU (DeepL)
56,75562863	39,90963761

Tableau (12) : Score BLEU moyen de l'évaluation automatique de Lucy LT et DeepL

Cette « victoire » assez imposante de Lucy LT face à DeepL s'explique avant tout par notre travail de spécialisation des syntagmes nominaux utilisés dans *World of Warcraft*. En effet, si DeepL traduit de façon plus fluide, il ne traduit pas (ou mal) les termes employés dans *World of Warcraft* : le raid *Serpentshrine Cavern* (« Caverne du Sanctuaire du Serpent »), la faction *Timbermaw* (« Grumegueules »), ou encore le sort *Summon Lightforged Warframe* (« Invocation d'Exossature Sancteforge ») n'ont pas été traduits, tandis que la monture *Voidtalon of the Dark Star* (« Serre du Vide de l'Étoile Noire ») ou la race des *fel orcs* (« gangr'orcs ») ont été traduites partiellement (*« Voidtalon de [...] l'Etoile Noire ») ou de façon complètement incorrecte (*« orcs de fellation »).

Or, si dans nos traductions de référence nous avons toujours utilisé les noms francophones des différents termes du jeu, un joueur humain connaît souvent plusieurs noms également en anglais (notamment les capitales, telles *Undercity*, *Ironforge*, *Thunder Bluff*, etc.). C'est principalement pour cela que l'évaluation humaine diffère de façon conséquente de l'évaluation « machine ». Si nous analysons par exemple les phrases présentées dans le Tableau (13) ci-dessous, le score BLEU est en faveur des traductions de Lucy LT. Pourtant, les seules différences notables sont

l'absence de traduction de certains termes, ou leur traduction incorrecte, probablement parce que DeepL ne les a pas encore rencontrés et ne sait donc comment les traduire.

N°	Phrase source	Phrase Cible (Lucy LT)	Phrase Cible (DeepL)	BLEU Score (Lucy LT)	BLEU Score (DeepL)
25	<i>Void Elves have two forms, and can shift into their void form in combat.</i>	Les elfes du vide ont deux formes, et peuvent se déplacer dans leur forme de vide dans le combat.	Les Elfes nuls ont deux formes, et peuvent se déplacer dans leur forme vide au combat.	72.42448	39.829976
26	<i>I need to go to the Black Temple as I don't have all the achievements yet.</i>	Je dois aller au Temple Noir comme je n'ai pas tous les haut-faits encore.	J'ai besoin d'aller au Temple Noir car je n'ai pas encore toutes les réalisations.	51.294971	50.317476
27	<i>I spent 2 hours in Siege of Orgrimmar yesterday, what a waste of my time.</i>	J'ai dépensé 2 heures dans le Siège d'Orgrimmar hier, quelle perte de mon temps.	J'ai passé 2 heures à assiéger Orgrimmar hier, quelle perte de temps.	45.788314	61.626071

Tableau (13) : Exemple de traductions évaluées, avec le score BLEU obtenu

Dans la phrase (25), le syntagme nominal *void elves* (« elfes du vide ») a été traduit par *« elfes nuls » par DeepL, qui ne pouvait reconnaître une race de personnages pour laquelle son vocabulaire n'avait pas été spécialisé. En revanche, l'expression *in combat* a été mieux traduite par DeepL (« au combat ») que par Lucy (?« dans le combat »), ce qui nous prouve une fois encore que DeepL est plus à même de faire des phrases fluides. Enfin, précisons que le syntagme verbal *shift into* a été traduit par « se déplacer dans » par les deux programmes de TA, alors que notre traduction de référence lui préfère le terme « se métamorphoser dans ».

Le score BLEU de la phrase (26) est beaucoup plus serré entre les deux programmes. En effet, si Lucy LT traduit correctement les syntagmes nominaux *Black Temple* (« Temple Noir ») et *achievements* (« haut-faits »), la syntaxe de la phrase laisse à désirer : *as* a été traduit par « comme » et *yet* (« encore ») a été placé en fin de proposition. Au contraire, DeepL n'a pas pu traduire correctement *achievements* (*« réalisations »), mais la phrase est plus fluide et ne pose aucun problème d'un point de vue syntaxique.

Cependant, DeepL est également capable d'obtenir un score BLEU plus élevé que Lucy LT, comme pour la phrase (27). En effet, même si le programme n'a pas été capable de traduire

correctement le nom du raid (« *Siege of Orgrimmar* – Siège d’Orgrimmar »), le verbe *spent* a été mieux traduit (« passé » au lieu de *« dépenser »), de même que l’expression *waste of my time*, qui ne prend pas de pronom possessif en français (« perte de temps »). Ainsi, même la spécialisation de Lucy LT ne permet pas de résoudre tous les problèmes que le programme peut rencontrer, principalement d’un point de vue syntaxique. Évidemment, il s’agit de petites erreurs syntaxiques qu’il est possible de corriger au sein du programme, mais nous ne souhaitons pas les corriger durant l’évaluation, afin de ne pas en biaiser le résultat.

IV.5 Qualité de la source et prédiction

Comme le rappelle King, pour « déterminer si un système peut répondre à [nos] besoins », il faut juger « l’exactitude » des traductions obtenues (King, *in* Bouillon & Clas, 1993 : 265). Mais, dans un cas où l’orthographe et la syntaxe en langue source laissent à désirer, peut-on estimer que la traduction est inexacte ? Un joueur francophone doit parfois faire face à une phrase écrite par un autre joueur francophone dont l’orthographe et/ou la syntaxe sont déplorable : les erreurs ne sont pas uniquement l’apanage des systèmes de TA. Or, dans ce cas-là, il n’existe aucun programme qui rectifiera les erreurs, et le joueur doit parfois « déchiffrer » le message envoyé par son congénère.

D’ailleurs, comme le souligne Arnold, il est nécessaire que le texte soumis à la TA soit soigné, afin que la traduction proposée par le programme soit la plus lisible possible (« tries to ensure that text which is submitted to an MT system is written in a way which helps to achieve the best possible raw MT output », Arnold & al., 1994 : 25). Si nous avons eu l’occasion de discuter avec de nombreux joueurs adultes et dont l’orthographe n’était nullement déplorable, il ne faut pas oublier que de plus en plus de joueurs tendent à écrire en mode « SMS » et sans respecter la grammaire et l’orthographe. Il est donc évident que notre programme aura bien plus de peine à traduire une phrase dont la présentation est moins soignée, comme dans les phrases présentées dans le Tableau (14) ci-dessous.

N°	Phrase source	Phrase Cible (Lucy LT)	Phrase Cible (DeepL)
28	<i>Oddly enough, when created, Nightborne representant to Sivermoon is neutral.</i>	Singulièrement, quand eu créé, le représentant de Sacrenuit à Sivermoon est neutre.	Curieusement, une fois créé, le représentant de Nightborne dans Sivermoon est neutre.
29	<i>But i still dont like the knowing that worgens dont have tail</i>	Mais je n'aime pas encore le savant que worgens n'ont pas de queue	Mais je n'aime toujours pas savoir que les worgens n'ont pas de queue
30	<i>Come Cataclysm, you can finally be a Human Hunter.</i>	Viens Cataclysm, tu peux finalement être un Chasseur Humain.	Viens Cataclysm, tu peux enfin être un chasseur humain.

Tableau (14) : Exemple de traductions dont la phrase source a induit le système de TA en erreur

La phrase (28) contient une faute de frappe pour la ville *Silvermoon* (oubli du « L »), raison pour laquelle Lucy LT est incapable de traduire correctement ce terme par « Lune d'Argent ». En revanche, comme nous le remarquons précédemment, il est peu probable que DeepL, non spécialisé au vocabulaire de *World of Warcraft*, ait été capable de le traduire si le terme avait été correctement écrit. Le reste de la phrase source est syntaxiquement correcte, mais elle n'a pas été traduite de façon optimale par Lucy LT. La traduction de DeepL est donc bien plus fluide, même si elle ne traduit pas non plus le terme *Nightborne* que Lucy LT traduit correctement grâce à la spécialisation effectuée.

Dans la phrase (29), le pronom défini *the* a été omis (bien qu'il soit moins nécessaire en anglais qu'en français), raison pour laquelle Lucy LT n'a pas été capable de traduire le syntagme nominal de façon plus propre (ajout du pronom défini « les »). La traduction fautive de *the knowing* en « le savant » ne peut en revanche être imputée à l'auteur, puisqu'en anglais cette tournure de phrase est correcte, quoique peu commune. D'ailleurs, DeepL traduit celle-ci correctement. Notons que l'oubli de la ponctuation finale a été répercuté sur les traductions et, si nous n'avions pas séparé nos phrases-tests une par une, il est fort probable que l'adjonction de la phrase suivante aurait mené à une traduction bien plus erronée de la part des deux programmes.

Enfin, la phrase (30) ne contient pas de faute grammaticale à proprement parler, mais la tournure familière de la phrase a également induit nos deux systèmes de TA en erreur. En effet, l'expression anglaise *Come Cataclysm* signifie « Avec l'arrivée de Cataclysm » de façon peu littéraire, ce qui n'a pas été perçu par nos programmes, qui ont donc traduit cette expression de façon très littérale. Signalons que, même si la phrase de DeepL a été jugée comme équivalente

à celle de Lucy LT par plus de 90% de nos évaluateurs francophones, le terme *Cataclysm* ne prend pas de « E » final en français, puisqu'il s'agit du nom de la troisième extension de *World of Warcraft* et non de la catastrophe naturelle.

La question de la préédition se pose donc naturellement, car de légères corrections sur ces phrases sources auraient probablement mené à de meilleures traductions, bien que dans le cas de Lucy LT la création de certaines expressions, telles *oddly enough*, puissent s'avérer nécessaires. Nous tenons également à préciser qu'en jeu, compte tenu de la simultanéité d'un grand nombre d'événements, il n'est pas rare que les joueurs soient « dépassés » (notamment lors de combats de raids), raison pour laquelle le langage « SMS » est encore souvent privilégié, quoique de plus en plus de joueurs utilisent désormais des canaux audios pour communiquer. En revanche, sur les forums ou les sites contributifs, dont WowHead.com est l'exemple le plus connu, les joueurs utilisent généralement une langue plus soignée, dans le but d'être compris même par des joueurs débutants (qui sont d'ailleurs souvent ceux qui cherchent des informations par le biais de ces sites).

Il est bien sûr évident que, dans ces cas de messageries instantanées lors des sessions de jeu, aucune post-édition n'est possible, et la seule préédition imaginable est celle du joueur lui-même, qui prendrait le temps de relire son message afin de s'assurer qu'il soit le plus irréprochable possible. Cependant, dans le cadre du jeu, il est impossible de demander aux joueurs une telle implication : la rapidité des actions (combats, événements chronométrés, etc.) force parfois les joueurs à taper rapidement. Pour ces raisons, nous estimons que notre spécialisation de Lucy serait bien plus fonctionnelle dans le cadre d'une traduction de messages sur des forums ou de bases de données que dans le cadre d'un *chat*. Cependant, de nombreux joueurs estiment que puisqu'il s'agit d'un jeu, le soin de leur orthographe n'est pas une nécessité (« on n'est pas à l'école ! »), y compris sur les forums. Pour ce dernier cas de figure, nous estimons qu'une post-édition est davantage envisageable, qu'il s'agisse d'implanter un programme automatique de correction d'orthographe, ou de faire appel à une équipe de modérateurs.

Pourtant, comme l'évoquait Bowker, la technologie se fait de plus en plus présente dans nos vies et les traducteurs sont de plus en plus prisés pour les traductions de médias, y compris dans le cadre des messageries instantanées : « As the number of electronic products [...] increases, there is a growing demand for translators who are able to translate media such as [...] even online chat sessions » (Bowker, 2002 : 131). Il paraît donc évident qu'il s'agit là d'un sujet

d'actualité, et qui ne se fera que plus présent au cours des années à venir, qu'il s'agisse de la discussion instantanée des jeux-vidéo, des *chats* tels WhatsApp, HangOut ou Messenger, ou encore de discussion en ligne (forums, messagerie instantanée sur les sites de ventes, etc.).

IV.6 Conclusion

Au cours de ce chapitre, nous avons mis en place une méthodologie d'évaluation pour tester les traductions produites par Lucy LT une fois spécialisé. Celle-ci nous a permis d'évaluer d'une part la qualité des traductions, puis dans un second temps leur fluidité. Puis, nous avons réalisé une évaluation automatique afin de faire également face à une évaluation très objective, qui permet de créer un contraste avec la subjectivité de nos évaluateurs. Enfin, nous avons pu souligner que la qualité du texte source reste primordial pour le bon fonctionnement d'un système de TA, et qu'une certaine forme de prédiction reste envisageable dans le cadre de *World of Warcraft*, par le biais du soin que les joueurs peuvent apporter à leurs messages.

En conclusion, dans le cadre de notre étude, nos deux évaluations humaines (qualité et fluidité) ainsi que notre évaluation automatique (score BLEU) penchent plutôt en faveur de Lucy LT. Cependant, le nombre d'entrées effectuées lors de la spécialisation de Lucy LT et de celui des phrases-tests traduites reste très faible au vu de la masse de vocabulaire que représente l'univers de *World of Warcraft*. Il est donc évident que notre travail mériterait d'être poursuivi, amélioré et corrigé afin d'être plus représentatif de l'ensemble des discussions possibles au sein du jeu. Les corrections que nous pourrions faire suite à la traduction de nos phrases-tests ne seraient donc qu'un début, tant l'univers créé par Blizzard Entertainment Inc. est riche et varié, et sans cesse réactualisé par les néologismes des joueurs.

V. Réflexion générale

Au cours de ce travail, nous avons spécialisé un programme de TA linguistique, Lucy LT, afin de vérifier si, une fois la spécialisation effectuée, il pouvait se révéler plus efficace qu'un programme de TA neuronal, ceux-ci étant pourtant désormais considérés comme les systèmes de TA les plus efficaces. Le langage du jeu-vidéo, et de *World of Warcraft* plus particulièrement, reste une langue de spécialité en soi, puisqu'il est difficilement intelligible pour un néophyte. De fait, il était intéressant de baser notre étude sur un vocabulaire souvent très éloigné de la langue générale, puisqu'un système neuronal non-entraîné n'est pas forcément à même de produire une traduction optimale sur ce genre de phrases.

Rappelons que, suite à notre spécialisation de Lucy LT, nous avons obtenu 2'843 entrées lexicales, toutes catégories confondues. Chacune de ces entrées a été implantée (ou, le cas échéant, modifiée) dans au moins l'un des trois dictionnaires de Lucy LT : monolingues source et cible, et transfert. Si ce nombre d'entrées est déjà conséquent dans le cadre de notre étude, nous estimons qu'il ne représente qu'un infime pourcentage de la masse de vocabulaire créé par Blizzard Entertainment Inc. et utilisé par les joueurs de *World of Warcraft*. Cependant, puisque les entrées que nous avons développées sont présentées en langue source et cible au sein du dictionnaire de transfert de Lucy LT, elles pourraient également servir de base de données terminologiques. Ce travail nécessiterait certes d'être repris afin de se conformer aux standards des fiches terminologiques (catégories, format, etc.), mais il pourrait servir de base à la création de celles-ci.

Lors de l'évaluation de nos phrases-tests, nous avons pu voir que le système neuronal DeepL manque en effet de spécialisation pour ce vocabulaire si spécifique, et qu'il produit souvent des phrases moins compréhensibles pour les joueurs, raison pour laquelle la première évaluation humaine fait apparaître un score en faveur de Lucy LT (44% des votes, contre 31% seulement pour DeepL et 25% d'équivalence). En revanche, les phrases proposées par DeepL sont de manière générale plus fluides et proches d'une traduction humaine, ce qui séduit parfois davantage les utilisateurs, habitués à se repérer dans le jeu avec de nombreux termes anglais. C'est pourquoi le score de la seconde évaluation humaine est bien plus serré que le précédent (score de fluidité de 1,79 pour Lucy LT contre 1,78 pour DeepL). Enfin, l'évaluation automatique démontre également que la spécialisation de Lucy LT est suffisamment efficace pour produire une bonne traduction, avec un score de 56,75, contre 39,90 pour DeepL (sur un maximum de 100).

Si, lors de nos évaluations, Lucy LT est toujours resté en première place, les scores sont pourtant serrés (notamment lors de l'évaluation humaine portant sur la fluidité), et il ne serait pas étonnant qu'avec une spécialisation de DeepL à l'aide d'un corpus parallèle, celui-ci offre alors des traductions optimales, y compris pour un langage aussi spécifique que celui de *World of Warcraft*.

Dans le futur, il serait donc intéressant d'utiliser Lucy LT pour générer un corpus bilingue spécifique au vocabulaire de *World of Warcraft*, afin de comparer les efforts nécessaires à une spécialisation de programme de TA linguistique et neuronal, ce qui permettrait de déterminer quelle est l'option la plus rentable pour mener à bien le projet ébauché au cours de cette étude.

Or, une question reste, à la suite de notre analyse, prééminente : notre programme serait-il toujours viable pour les autres langues de *World of Warcraft* ? En effet, non seulement les serveurs européens utilisent aussi l'allemand, l'espagnol, l'italien, le portugais et le russe, mais les serveurs basés en Amérique couvrent également l'anglais (US) et l'espagnol (Mexique), mais encore le portugais (Brésil). Enfin, les serveurs asiatiques proposent des serveurs coréens, taiwanais et chinois. Ainsi, il nous faudrait étendre notre spécialisation à d'autres paires de langues, toutes soutenues par Lucy LT, mais ce travail se révélerait fort fastidieux, puisqu'il s'agit bien de paires de langues et non d'une interlingua (cf. Point II.3.1, « Les systèmes linguistiques ») : ne serait-ce que pour traduire de l'anglais vers les autres langues, il faudrait spécialiser le vocabulaire de *World of Warcraft* dans Lucy LT pour 6 paires différentes, en incluant le français³⁰.

De plus, comme le souligne Arnold, un système de TA ne pourra jamais être sûr d'une traduction, puisqu'il lui manque une perspective humaine : « Lacking a human perspective, the system is always uncertain about the correct way to analyse a sentence. » (Arnold & al., 1994 : 26). Malgré les progrès fantastiques de l'intelligence artificielle ces dernières années, la traduction nécessite encore et toujours une intervention humaine, et le but n'est aujourd'hui plus de « choisir entre le traducteur ou l'ordinateur mais de faire cohabiter l'ordinateur et le traducteur », comme le proposait Ladmiraal en 1994 déjà (Ladmiraal, 1994 : 10). Ceci semble encore vrai aujourd'hui, puisque de plus en plus d'outils de TAO (Traduction Assistée par Ordinateur) sont développés et commercialisés.

³⁰ Puisque les royaumes implantés sur les serveurs des différents continents ne peuvent entrer en contact, nous n'avons pas inclus les serveurs asiatiques dans ce chiffre.

Le travail de pré- ou post-édition s'impose alors au traducteur, bien que, comme nous l'évoquions plus haut, cette tâche s'avère impossible dans le domaine de la messagerie instantanée de *World of Warcraft*. Pourtant, selon Carbonell et Tomita, la post-édition serait un gain de temps majeur, puisque le post-éditeur ne doit plus que corriger les fautes subsistantes, et non traduire entièrement le texte source (« a potential for major gains in human efficiency », Carbonell & Tomita, in Nirenburg, 1987 : 72). De fait, même si la messagerie instantanée de *World of Warcraft* ne permet pas un travail de pré- ou post-édition à proprement parler, il est fort probable qu'une traduction, même partiellement correcte, pourra soulager les méninges du joueur francophone.

De plus, l'intégration de la TA aux forums (officiels ou non) de *World of Warcraft* se présente comme une action envisageable. Celle-ci permettrait notamment aux commentaires anglais d'être traduits vers le français (ce qui, rappelons-le, n'est actuellement pas le cas sur WowHead), voire même d'être ensuite étendue aux autres langues employées. Pour évaluer l'utilité d'une API de traduction pour les forums, nous pourrions dans le futur mener une étude comparative entre une post-édition automatique sur l'un des forums de WowHead, et une post-édition humaine après une traduction automatique à l'aide d'un programme de TA spécialisé (Lucy LT et/ou DeepL).

Pour reprendre l'idée de Johnson et Whitelock, la TA peut être utilisée comme un outil pour augmenter la productivité de l'expert : « used as a tool for the expert to increase productivity » (Johnson & Whitelock, in Nirenburg, 1987 : 140). Or, comme nous l'avons précédemment déterminé, nos joueurs sont en quelques sortes experts de domaine de *World of Warcraft* puisqu'ils forment une communauté qui partage un vocabulaire spécifique et difficile d'accès aux non-initiés. Ainsi, notre spécialisation de Lucy LT pourrait, sinon être irréprochable, du moins améliorer la compréhension entre les joueurs de différentes langues et, de fait, permettre la création de groupes plus unis vers le même objectif. De même, comme le constatait Kay, la TA peut toujours être appliquée afin d'obtenir une idée globale du texte d'origine : « seeking only a rough idea of what is being said, so that an extremely rough translation will be adequate » (Kay, in Hutchins & Somers 1992 : XII).

Enfin, il serait nécessaire d'entrer en contact avec Blizzard Entertainment Inc. pour évaluer l'utilité de notre spécialisation de Lucy LT en situation réelle, c'est-à-dire d'implanter un module de TA au sein du jeu *World of Warcraft* afin d'estimer si cette situation est viable et utile aux joueurs. En effet, il est fort probable que seule une utilisation « réelle » du programme

permettrait de fournir une liste d'erreurs à corriger, même s'il nous paraît évident qu'il ne sera sûrement pas possible de rendre Lucy LT totalement indépendant et capable d'offrir des traductions entièrement automatiques de haute qualité (TEAHQ) en raison des limites du système lui-même.

N'oublions pas non plus que de nombreux inconvénients techniques pourraient également réduire à néant un tel projet, d'autant que le jeu est déjà conséquent (place sur le disque dur, graphismes et configuration minimale recommandés, etc.). De plus, un nombre conséquent de joueurs est connecté au même moment sur chacun des serveurs, faisant parfois « crasher » ces derniers : il est fort probable que la mise en place d'un système de TA au sein de jeu monopolise davantage les ressources du serveur, créant des *lags* et des *freezes* supplémentaires, chose très peu appréciée des joueurs.

Finalement, l'analyse que nous avons achevons ici ne fait qu'effleurer la surface de la problématique de la traduction des jeux-vidéo, et ne parcourt que rapidement l'ensemble des possibilités de communication offertes, ne serait-ce qu'au sein de *World of Warcraft*. En effet, chaque jour de nouveaux joueurs arrivent sur le jeu avec leur lot d'expressions, leurs particularités langagières, et il ne nous sera pas possible de spécialiser un programme de TA pouvant couvrir toutes les phrases possibles au sein de leurs discussions. De même, le jeu *World of Warcraft* a encore de nombreuses années devant lui, et la création de nouveaux mondes, objets et personnages nécessitera à nouveau de nombreuses traductions. Nous souhaiterions donc laisser le mot de la fin à Ichbiah, qui souligne que le champ des possibilités de l'univers de *World of Warcraft* est encore bien plus grand que nous ne pouvons le soupçonner à l'heure actuelle : « par définition, un univers virtuel n'est jamais achevé » (Ichbiah, 2004 : 420).

Bibliographie

Monographies

Arnold, Douglas & al. *Machine Translation, An Introductory Guide*. Londres, NCC Blackwell (1994), hors collection, 206 pages.

Bouillon, Pierrette & Clas, André. *La traductique*. Montréal (Québec), Les Presses de l'Université de Montréal (1993), hors collection, 508 pages.

Bouillon, Pierrette. *Traitement automatique des langues naturelles*. Bruxelles, Duculot (1998), Collection Universités Francophones, 245 pages.

Bowker, Lynne & Pearson, Jennifer. *Working with Specialized Language, A practical guide to using corpora*. Londres, Routledge (2002), hors collection, 257 pages.

Bowker, Lynne. *Computer-aided translation technology : A practical introduction*. Ottawa, University of Ottawa Press (2002), Collection Didactics of translation series, 185 pages.

Cario, Erwan. *Start ! La grande histoire des jeux vidéo*. Paris, La Martinière (2011), hors collection, 205 pages.

De Saussure, Ferdinand. *Cours de linguistique générale*. Paris, Payot & Rivages (2005), hors collection, 520 pages.

Fuchs, Catherine. *Linguistique et Traitements Automatiques des Langues*. Paris, Hachette (1993), Collection Hachette Université Linguistique, 303 pages.

Golden, Christie & Grubb, Jeff & Rosenberg, Aaron. *World of Warcraft: Chronicles of War*. New York, Gallery Book (2010), hors collection, 772 pages.

Gray, Charlotte. *A Comparative Evaluation of Localisation Tools : Reverso Localize and SYSTANLinks*. Genève, Université de Genève (2014), Maîtrise. 146 pages.

Hutchins, John W. & Somers, Harold L. *An introduction to machine translation*. Londres, Academic Press (1992), hors collection, 333 pages.

Ichbiah, Daniel. *La saga des Jeux Vidéo*. Paris, Vuibert (2004), hors collection, 520 pages.

Jurafsky, Daniel & Martin, James H. *Speech and Language Processing*. New Jersey, Prentice Hall (1999), Collection Prentice Hall series in artificial intelligence, 934 pages.

L'Homme, Marie-Claude. *Initiation à la traductique*. Montréal, Lingatech (2008), hors collection, 317 pages.

Muller, Evelyne & Schmidt, Barbara. *Traduction anglaise : Maîtriser l'outil Internet*. Paris, Presses Universitaires Françaises (2007), Collection Major, 200 pages.

Nirenburg, Sergei. *Machine translation : Theoretical and methodological issues*. Cambridge, Cambridge University Press (1987), Collection Studies in natural language processing, 350 pages.

Segond, Frédérique. *Multilinguisme et traitement de l'information*. Paris, Hermès (2002), Collection Traité des sciences et techniques de l'information, 253 pages.

Sinclair, John McH. *How to Use Corpora in Language Teaching*. Amsterdam, J. Benjamins (2004), Collection Studies in Corpus Linguistics, 307 pages.

Tomblaine, Philippe. *Jeux vidéo ! Une histoire du 10^e art*. Lyon, Moutons électriques (2015), Collection Bibliothèque des miroirs, 240 pages.

Vreck, Françoise. *Entraînement à la version anglaise*. Paris, Orphys-Ploton (2002), hors collection, 264 pages.

Articles

Aikawa & al. *How Good Is a Crowd Post-Editing? Its potential and Limitations*. In : *AMTA 2012 Workshop on Post-Editing Technology and Practice*. Association for Computational Linguistics. San Diego, AMTA (2012), pp. 69-77.

Aikawa & al. *Impact of controlled language on translation quality and post-editing in a statistical machine translation environment*. In : *Proceedings on the MT Summit XI*. Copenhagen, Bente Maegaard (2007), pp. 1-7.

Burchardt, Alijoscha & al. *A Linguistic Evaluation of Rule-Based, Phrase-Based, and Neural MT Engines*. In : *Prague Bulletin of Mathematical Linguistics*. Univerzita Karlova. Prague, Univerzita Karlova (2017) n°108, pp. 159-170.

Collectif. *Finding of the 2016 Conference on Machine Translation (WMT16)*. In : *Proceedings of the First Conference on Machine Translation*. Association for Computational Language. Berlin, ACL (2016). Volume n°2, pp.131-198.

Habert, Benoît. *Des corpus représentatifs : de quoi, pour quoi, comment ?* In : *Linguistique sur corpus. Études et réflexions*. Bilger, Mireille. Perpignan, Presses Universitaires de Perpignan (2000), pp. 11-58.

Hearne, Mary & Way, Andy. *Statistical Machine Translation: A Guide for Linguists and Translators*. In : *Language and Linguistics Compass*. England, Blackwell Syngery (2011) n°5, pp. 205-226.

Ladmiral, Jean-René. *Le traducteur et l'ordinateur*. In : *Langages*, 28^e année. Ladmiral, Jean-René. Larousse (1994) n°116, pp.5-19.

Landis, Richard J. & Koch, Gary G. *The Measurement of Observer Agreement for Categorical Data*. In : *Biometrics*. The International Biometric Society. Washington DC, The International Biometric Society (1977) n°33, pp.159-174.

O'Brien, Sharon. *Teaching Post-Editing: A Proposal for Course Content*. In : 6th EAMT Workshop Teaching Machine Translation. EAMT. Manchester, EAMT (2002), pp. 99-106.

Papineni, Kishore & al. *Bleu: a Method for Automatic Evaluation of Machine Translation*. In : *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*. ACL. Philadelphia, IBM Research Report (2002), pp. 311-318.

Sites et pages Web

Blizzard Entertainment Inc., *World of Warcraft, Battle for Azeroth*, février 2018, [<https://us.shop.battle.net/en-us/product/world-of-warcraft-battle-for-azeroth>]

Data Altares & BFM TV, *Verif.com*, février 2016, [<https://www.verif.com/Hit-parade/01-CA/00-Pays/0-France>]

Gaelle Bou (Systran Blog), *How does Neural Machine Translation work?*, 17 octobre 2016, [<http://blog.systransoft.com/how-does-neural-machine-translation-work/>]

Microsoft, *Microsoft Translator launching Neural Network based translations for all its speech languages*, 15 novembre 2016, [<https://blogs.msdn.microsoft.com/translation/2016/11/15/microsoft-translator-launching-neural-network-based-translations-for-all-its-speech-languages/>]

Nethaera (pour Blizzard Entertainment Inc.), *World of Warcraft.com*, 30 novembre 2011, [<https://worldofwarcraft.com/fr-fr/news/10040682>]

Omniscien Technologies, *Deep Neural Machine Translation*, 2018, [<https://omniscien.com/deep-neural-machine-translation/>]

Omniscien Technologies, *Riding the Machine Translation Hype Cycle – From SMT to NMT to Deep NMT by Dion Wiggins*, octobre 2017, [<https://omniscien.com/riding-machine-translation-hype-cycle/>]

Omniscien Technologies, *The State of Neural Machine Translation (NMT) by Philipp Koehn*, novembre 2016, [<https://omniscien.com/state-neural-machine-translation-nmt/>]

Tilde, *Neural Machine Translation*, 2017, [<https://tilde.com/products-and-services/machine-translation/features/neural-translation>]

Wikia (collectif), *Timeline (World of Warcraft)*, 18 novembre 2017, [[http://wowwiki.wikia.com/wiki/Timeline_\(World_of_Warcraft\)](http://wowwiki.wikia.com/wiki/Timeline_(World_of_Warcraft))]

WowHead (version française), *Héroïsme*, 2018, [<http://fr.wowhead.com/spell=32182/heroisme>]

Annexe I : Informations aux participants

Intégration de la TA au domaine du jeu-vidéo : **Spécialisation de Lucy LT pour World of Warcraft**

Etude réalisée par Adélie FASEL dans le cadre de l'obtention d'un Master en Traduction, délivré par l'Université de Genève, sous la direction du Prof. Pierrette BOUILLON

Objectifs

Dans le cadre de mon travail de mémoire, je cherche à spécialiser un programme de traduction automatique pour le langage propre à *World of Warcraft* (WoW), puis à évaluer les traductions ainsi obtenues. Je vous demanderai donc quelques précisions sur la façon dont vous utilisez certains termes spécifiques au jeu, puis, dans un second temps, je vous proposerai d'évaluer les traductions obtenues.

Déroulement et durée de la collecte de données

1. Traductions des termes : ce questionnaire vous prendra environ 30 minutes. Ce questionnaire peut être interrompu et envoyé tel quel lorsque vous ne souhaitez plus participer. Les réponses peuvent être rendues jusqu'au 15 février 2018.
2. Évaluation des traductions : ce questionnaire vous prendra environ 2 heures en raison de la quantité d'informations à traiter. Cependant, ce questionnaire peut être complété sur plusieurs sessions, ou interrompu et envoyé tel quel lorsque vous ne souhaitez plus participer. Les réponses peuvent être rendues jusqu'au 1^{er} avril 2018.

Résultats

Le mémoire, dans lequel les résultats seront publiés, devrait être disponible fin juin 2018. Vous pouvez le rechercher directement sur l'archive ouverte de l'Université de Genève (<https://archive-ouverte.unige.ch/structures/view/9>) ou me le demander par courriel : Adelie12@hotmail.com.

Informations supplémentaires

Vous pouvez me contacter par courriel (Adelie12@hotmail.com) ou par téléphone (076.679.16.25) en cas de questions.

Vous pouvez cesser à tout moment de participer à cette étude, sans avoir besoin de vous expliquer.

Votre pseudo en jeu sera inscrit dans la page des remerciements, seulement si vous l'acceptez. Aucune autre information personnelle ne vous sera demandée, en-dehors de votre sexe et nationalité, à des fins statistiques. Ces renseignements sont cependant facultatifs, vous pouvez passer aux questions suivantes sans y répondre.

Annexe II : Formulaire de consentement

Intégration de la TA au domaine du jeu-vidéo : **Spécialisation de Lucy LT pour World of Warcraft**

Etude réalisée par Adélie FASEL dans le cadre de l'obtention d'un Master en Traduction, délivré par l'Université de Genève, sous la direction du Prof. Pierrette BOUILLON

Consentement

Par ma signature, je confirme avoir lu la page « Informations » et compris les conditions de participations. Je sais notamment que je peux cesser à tout moment de participer à cette étude, sans avoir besoin de prévenir Adélie FASEL ou de fournir d'explications.

J'ai en outre compris que les informations que je fournis seront utilisées dans le cadre du mémoire réalisé par Adélie FASEL (cf. intitulé en tête de page), et uniquement dans le cadre de celui-ci.

Prénom/Pseudonyme

.....

Signature

.....

Date

.....

Contact

Adélie FASEL
Adelie12@hotmail.com
076.679.16.25

Annexe III : Instructions des questionnaires

Questionnaire 1 (choix traductifs)

Bonjour, et merci d'avoir accepté de répondre à ce questionnaire !

J'ai besoin de votre aide dans le cadre de mon travail de mémoire : l'intégration de la traduction automatique au domaine du jeu-vidéo, et plus particulièrement de *World of Warcraft (WoW)*. Pour ce faire, j'aimerais avoir votre avis sur certains termes utilisés par les joueurs francophones.

Gardez en tête qu'il n'y a pas de bonnes ou de mauvaises réponses, il s'agit juste de dégager une tendance générale, afin que je puisse traduire certains termes au plus proche de la "réalité" des joueurs.

Si vous n'êtes pas sûrs d'une réponse, vous pouvez passer à la question suivante. Si la réponse que vous souhaitez fournir n'apparaît pas, merci de cliquer sur l'option « Autre » (vous pouvez préciser votre réponse, si vous le souhaitez, dans la section « Commentaires »).

N'hésitez pas à me contacter si vous avez une question.

Merci !

Questionnaire 2 (évaluation humaine : qualité)

Bonjour, et merci d'avoir accepté de répondre à ce second questionnaire !

J'ai encore besoin de votre aide dans le cadre de mon travail de mémoire : l'intégration de la traduction automatique au domaine du jeu-vidéo, et plus particulièrement de *World of Warcraft (WoW)*. Maintenant que la spécialisation du programme est effectuée, je vais vous demander d'évaluer les phrases présentées ci-dessous.

Imaginez qu'un joueur de votre groupe/raid vous parle et vous écrit ces phrases, et choisissez laquelle vous préféreriez lire. Vous pouvez à chaque fois choisir la Phrase 1, la Phrase 2, ou, si vous jugez que les deux phrases sont équivalentes (en bien comme en mal), vous pouvez sélectionner l'option « Les deux sont équivalentes ».

N'hésitez pas à me contacter si vous avez une question.

Merci !

Questionnaire 3 (évaluation humaine : fluidité)

Bonjour, et merci d'avoir accepté de répondre à ce dernier questionnaire !

Dans le cadre de mon travail de mémoire (l'intégration de la traduction automatique au domaine du jeu-vidéo, et plus particulièrement de *World of Warcraft (WoW)*), j'ai à nouveau besoin de votre aide.

Il s'agit maintenant d'effectuer une seconde évaluation, basée cette fois-ci sur la fluidité des phrases. Imaginez qu'un joueur vous parle et vous écrit les phrases présentées ci-dessous, et notez-les selon le barème suivant :

Score de 0 = La traduction est incompréhensible, ou vous induit en erreur.

Score de 1 = La traduction est peu lisible, mais vous en devinez le sens global.

Score de 2 = La traduction « sonne mal » à vos oreilles, mais vous la comprenez.

Score de 3 = Vous ne savez pas s'il s'agit d'une traduction humaine ou automatique (oui, vous vous en doutez vu le sujet du mémoire, mais jouez le jeu !)

Vous avez tous reçu en pièce jointe (avec le lien du questionnaire) un fichier reprenant les phrases en anglais dont sont issues les traductions que vous évaluez. Vous pouvez vous en servir si vous le souhaitez, afin de vérifier le sens de certaines phrases.

N'hésitez pas à me contacter si vous avez une question.

Merci !

Annexe IV : Extrait du corpus de texte développé

Fichier Edition Format Affichage ?

XkcdofDentarg on 2012/02/15 (Patch 4.3.2) @cawfee SAME!!! 1 yoshimario40 on 2014/10/04 (Patch 6.0.1) It's as if he left an open paranthesis. (It drives people insane 16 By Tehslip (621 - 4·18) on 2009/01/06 (Patch 3.0.3) Brann will bug in HoS if you manage to get him killed on the trash before you start the event with him. It requires a soft reset to be able to start the event within HoS. Fix: Don't get him killed. 8 Veratus on 2009/01/27 (Patch 3.0.8) Just clear before you start the escort part of the quest, if you're concerned about that happening. 10 By Cross88 (25,499 - 5·23·229) on 2009/08/15 (Patch 3.1.3)I wonder if he misses Glibb. 5 By Frostalix (751 - 6) on 2009/02/27 (Patch 3.0.9) Brann Bronzebeard will also return for the Ulduar Raid in the upcoming patch, 3.1. As if I couldn't tolerate his voice enough already. 16 Ressenueus on 2009/03/06 (Patch 3.0.9) Must everyone rag on the poor dwarf? I like his voiceover. Maybe because I feel that if he had what everyone calls a "normal dwarf" voice, he would sound JUST LIKE EVERY OTHER DWARF NPC. His voice is somewhat strange, but the way he talks just makes it seem... right. But you know, that's just my opinion. Keep up the good work, Brann! And would it kill ya to say hi to your monkey in Silithus? 4 Ictiv on 2009/04/11 (Patch 3.0.9) Exactly! I think its a good thing that he has a unique voice too. Also: He traveled most of the world, from Khaz Modan, trough Azeroth and into the wildest depths of Kalimdor. Who could keep his "perfect dwarf" accent after meeting so many strange people with so many strange accents?4 By mac605 (828 - 8·10) on 2009/03/14 (Patch 3.0.9) One of his quotes is "The Truth is out there" - a clear X-Files - Fox Mulder reference :)4 By garymce (9,870 - 3·24·84) on 2009/04/06 (Patch 3.0.9) His lines appear to have been re-recorded by a new voice actor in patch 3.1 along with a set of lines for the Ulduar raid wing, he sounds much better in my opinion, his accent is much more consistent with other Ironforge dwarves.Last edited by garymce on 2009/04/06 (Patch 3.0.9) 5 Ludmor on 2009/04/20 (Patch 3.1.0) [quote]his accent is much more consistent with other Ironforge dwarves.[/quote] Therefore he ceased to be unique. :(4 By GameSlave (549 - 5) on 2009/05/16 (Patch 3.1.1) PRISONER? Hah. With It's bindings shattered.. It's influence unchecked.. It's gonna come after us- and we're gonna be the prisoners. Ulduar trailer, patch 3.1, where this little dwarf evades death just barely yet again. 6 Daystar on 2009/05/25 (Patch 3.1.2) For the love of all that's holy, in 3.1 they also FIXED HIS VOICE. He doesn't sound like he's gargling gravel anymore. They even went back and redid his lines in Halls of Stone. 3 By aoeu (13,707 - 1·19·93) on 2010/10/11 (Patch 4.0.1) Looks like Brann will be guiding us through Uldum/Halls of Origination in Cataclysm, too. Looking forward to it. :) Last edited by aoeu on 2010/10/11 (Patch 4.0.1) 2 By semlar (7,071 - 2·14·60) on 2010/07/02 (Patch 3.3.5) Body and Soul will make him run faster. 1 By Supahtoons (15,504 - 4·19·129) on 009/12/13 (Patch 3.2.2) Anyone want to explain how a Dwarf Hunter can Stealth? Last edited by

Annexe V : Extraits des termes entrés dans les dictionnaires de Lucy LT

Note : Les termes en **bleu** ont été entrés en *multiword* (MW), les termes en **vert** ont été entrés en mots simples (*single word*, ou SW). Les noms composés laissés en noir existaient en unités simples dans la base de données et ont simplement été modifiés selon les besoins. Les termes entre parenthèses existaient déjà dans les dictionnaires de Lucy LT, ou ont été entrés séparément. Bien que les entrées soient présentées ici avec des majuscules pour les noms propres, les entrées ont été faites en minuscules dans les dictionnaires de Lucy LT.

Extrait des noms de zones (scénarios, donjons et raids inclus)

Termes identiques	Termes traduits par défaut	Termes à implanter / à modifier
Ahn'Kahet	<i>Abyssal Depths</i>	<i>Abyssal Maw</i> - Gueule des Abysses
Ahn'Qiraj	<i>A Little Patience</i>	<i>Acherus: The Ebon Hold</i> - Achérus : Le Fort d'Ébène
Antorus	<i>Akazamzarak's Hat</i>	<i>Alliance Airship</i> - Vaisseau de l'Alliance
Arcatraz	<i>Ashamane's Fall</i>	<i>Alterac (Mountains + Valley)</i> - (Montagnes + Vallée d')Alterac
Ashran	<i>Auchenai Crypts</i>	<i>Ammen Vale</i> - Val d'Ammen
Auchindoun	<i>Bastion of Twilight</i>	<i>Antonidas Memorial</i> - Monument à Antonidas
Azshara	<i>Black Temple</i>	<i>Antoran Wastes</i> - Étendues Antoréennes
Azsuna	<i>Blood in the Snow</i>	<i>Arathi Basin</i> - Bassin Arathi
Darnassus	<i>Camp Narache</i>	<i>Arathi Highlands</i> - Hautes-Terres arathies
Draenor	<i>Cathedral of Eternal Night</i>	<i>Archmage (Vargoth's Retreat)</i> - (Retraite de l')Archimage (Vargoth)
Durotar	<i>Crypt of the Forgotten Kings</i>	<i>Arcway</i> - Arcavia
Exodar	<i>Deaths of Chromie</i>	<i>(Arena of) Annihilation</i> - (Arène de l')Annihilation
Gnomeregan	<i>Defense of Karabor</i>	<i>Ashenvale</i> - Orneval
Gorggrond	<i>Emerald Nightmare</i>	<i>Assault on (Zan'Vess)</i> - Assaut de (Zan'Vess)
Gundrak	<i>Eye</i>	<i>Azjol-Nerub</i> - Azjol-Nérub
Helheim	<i>Eye of Azshara</i>	<i>Azuremyst (Isle)</i> - (Île de) Brume-Azur
Kalimdor	<i>Eye of Eternity</i>	<i>Badlands</i> - Terres Ingrates
Karazhan	<i>Fall of Shan Bu</i>	<i>Baradin Hold</i> - Bastion de Baradin
Kezan	<i>Great Sea</i>	<i>(Battle of) Thunder Pass</i> - (Bataille du) Passage du Tonnerre
Krokuun	<i>Hour of Twilight</i>	<i>(Battle) on the High Seas</i> - (Bataille) en Haute Mer
Loch Modan	<i>Isle of Quel'Danas</i>	<i>Beyond</i> - Au-delà
Mac'Aree	<i>Isle of Thunder</i>	<i>Bilgewater Harbor</i> - Havre Baille-Fonds
Mardum	<i>Lost Glacier</i>	<i>Black Morass</i> - Noir Marécage
Mulgore	<i>Lost Isles</i>	<i>Black Rook Hold</i> - Bastion du Freux

Extrait des noms de races et classes

Termes traduits par défaut (modifiés ou non)	Termes implantés
<i>blood elf</i> - elfe de sang	<i>broken</i> - roué
<i>druid</i> - druide	<i>death knight</i> - chevalier de la mort
<i>dwarf</i> - nain	<i>demon hunter</i> - chasseur de démons
<i>gnome</i> - gnome	<i>fel orc</i> - grangr'orc
<i>goblin</i> - gobelin	<i>forest troll</i> - troll des forêts
<i>human</i> - humain	<i>gilnean</i> - gilnéen
<i>hunter</i> - chasseur	(<i>Hightmountain</i>) <i>tauren</i> - tauren (de Haut-Roc)
<i>monk</i> - moine	<i>ice troll</i> - troll des glaces
(<i>Northrend</i>) <i>skeleton</i> - squelette (du Norfendre)	<i>lightforged draenei</i> - draenei sancteforge
<i>orc</i> - orc	<i>mage</i> - mage
<i>priest</i> - prêtre	<i>naga</i> - naga
<i>shaman</i> - chaman	<i>night elf</i> - elfe de la nuit
<i>troll</i> - troll	<i>nightborne</i> - sacrenuit
<i>warrior</i> - guerrier	<i>paladin</i> - paladin
	<i>pandaren</i> - pandaren
	<i>rogue</i> - voleur
	<i>taunka</i> - taunka
	<i>tuskarr</i> - rohart
	<i>undead</i> - mort-vivant
	<i>void elf</i> - elfe du vide
	<i>vrykul</i> - vrykul
	<i>warlock</i> - démoniste
	<i>worgen</i> - worgen

Extrait du vocabulaire des joueurs

Termes traduits par défaut	Termes à implanter / modifier
<i>armor</i>	<i>achievement</i> - haut-fait
<i>assassination</i>	<i>affliction</i> - affliction
<i>assault</i>	<i>anglers</i> - hammeçonneurs
<i>balance</i>	<i>arcane</i> - arcane
<i>banquet</i>	<i>arms</i> - armes
<i>battleground</i>	<i>auction house</i> - hôtel des ventes
<i>beast mastery</i>	<i>BG (NST)</i> - BG (NST)
<i>blade</i>	<i>BOE (NST)</i> - LQE (NST)
<i>blood</i>	<i>boss (NST)</i> - boss (NST)
<i>champion</i>	<i>brewmaster</i> - maître brasseur
<i>class</i>	<i>buff (NST)</i> - buff (NST)
<i>copper</i>	<i>buff (VST)</i> - buffer (VST)
<i>damage</i>	<i>cast (NST)</i> - cast (NST)
<i>demonology</i>	<i>cast (VST)</i> - caster (VST)
<i>destruction</i>	<i>char (NST)</i> - perso (NST)
<i>discipline</i>	<i>character</i> - personnage
<i>elemental</i>	<i>cooldown (NST)</i> - CD (NST)
<i>elite</i>	<i>DPS (NST)</i> - DPS (NST)
<i>enhancement</i>	<i>drop (NST)</i> - drop (NST)
<i>faction</i>	<i>drop (VST)</i> - droper (VST)
<i>fire</i>	<i>dungeon</i> - donjon
<i>fish</i>	<i>expansion</i> - extension
<i>food</i>	<i>feral</i> - farouche
<i>fury</i>	<i>frost</i> - givre
<i>gold</i>	<i>gear (NST)</i> - stuff (NST)
<i>guardian</i>	<i>havoc</i> - dévastation
<i>heroic</i>	<i>heal (VST)</i> - healer (VST)
<i>level</i>	<i>healer (NST)</i> - healer (VST)
<i>map</i>	<i>health</i> - vie
<i>mastery</i>	<i>holy</i> - sacré
<i>missions</i>	<i>instance</i> - instance
<i>money</i>	<i>interrupt (NST)</i> - kick (NST)
<i>mythic</i>	<i>interrupt (VST)</i> - kicker (VST)
<i>outlaw</i>	<i>item</i> - objet
<i>pet</i>	<i>legacy</i> - héritage