

## Une description bilingue des temps verbaux : étude contrastive en corpus

Cristina Grisot & Bruno Cartoni  
Département de Linguistique  
Université de Genève  
{cristina.grisot | bruno.cartoni}@unige.ch

### Résumé

*Cet article présente une analyse contrastive et empirique des temps verbaux du passé et du présent en français et en anglais. Dans cette recherche, nous commençons par passer en revue les descriptions théoriques sur les usages des temps verbaux dans une perspective monolingue dans les langues considérées, puis nous regardons ce que les corpus parallèles révèlent au sujet des divergences d'usage entre l'anglais et le français. Enfin, nous proposons un premier modèle de représentation formelle de ces divergences. L'étude en corpus que nous présentons révèle une disparité entre les descriptions théoriques des usages des temps verbaux et leurs possibilités de traduction et la manière dont les temps verbaux sont réellement utilisés dans divers registres. La formalisation des temps verbaux que nous proposons pour une perspective bilingue allie la Théorie de la Pertinence, les coordonnées de Reichenbach, et inclut des distinctions plus fines entre usage narratif et non narratif, subjectif et non subjectif.*

### 1. Introduction

Les recherches présentées ici s'inscrivent dans le cadre du projet COMTIS<sup>1</sup> qui a pour but l'amélioration de la qualité de la traduction automatique en termes de cohérence par la modélisation des relations interphrases. La cohérence d'un texte traduit est essentiellement influencée par un certain nombre de phénomènes linguistiques, comme les pronoms, les expressions référentielles, les temps verbaux ou les connecteurs. Ces marqueurs sont en lien direct avec les relations du discours et leur mauvaise traduction diminue la cohérence textuelle du texte traduit.

Dans cet article, nous présentons la première étape de notre recherche pour l'amélioration de la traduction des temps verbaux, par l'entremise d'une modélisation de l'équivalence des traductions des

---

<sup>1</sup> Le projet COMTIS « *improving the COherence of Machine Translation output by modeling InterSentential relations* » est une recherche en cours qui a débuté en mars 2010 et est financé par le Fond National Suisse pour la Recherche : Projet Sinergia COMTIS, CRSI22-127510. (<http://www.idiap.ch/project/comtis>)

temps verbaux en français et en anglais. Cette première étape regroupe la construction d'un corpus parallèle qui permet l'identification des divergences de traduction des temps verbaux les plus fréquents, une première analyse de la sémantique et la pragmatique de ces temps et une première proposition des critères de désambiguïsation ainsi que leur modélisation formelle. Dans les étapes suivantes, nous testerons cette modélisation lors d'expériences d'annotation à large échelle de corpus bilingues, pour permettre d'éprouver les critères de désambiguïsation (traits linguistiques) qui permettront ensuite d'améliorer les modèles de langue qui sont à la base des systèmes de traduction automatique statistique.

Un simple regard d'ensemble suffit à se rendre compte que la littérature anglophone et francophone ne décrit pas les temps verbaux de la même manière. Là où les descriptions traditionnelles (les grammaires Riedel et al. 1994, Grevisse & Goose 2008) décrivent les temps verbaux en français selon une valeur de base à laquelle s'ajoutent des usages contextuels, les grammaires anglo-saxonnes (Quirk et al. 1985) décrivent les temps verbaux en anglais en termes d'aspect grammatical exprimé par des états, des événements et des habitudes, avec une forme simple, perfective ou progressive dans le passé, le présent et le futur.

Dans les descriptions monolingues, tous les « emplois » ou « valeurs » sont exemplifiés par des phrases collectées sporadiquement dans la littérature, des journaux ou tout simplement créés par des locuteurs natifs. Une telle approche descriptive donne davantage de place à des emplois très rares et empêche certaines généralisations, plus pertinentes du point de vue de la fréquence d'apparition. Ainsi, dans le cadre de notre projet d'amélioration des systèmes statistiques de traduction automatique, il est nécessaire de disposer d'outils théoriques et méthodologiques qui puissent décrire les systèmes verbaux dans ces deux langues de la même manière et qui soit basés sur une étude en corpus.

Dans ce contexte, nous avons mené une *étude contrastive* sur la base de corpus parallèles bilingues contenant des textes appartenant à divers registres (littéraire, journalistique, législatif et administratif).

Notre recherche porte principalement sur l'étude des temps verbaux qui expriment le passé et le présent en français et en anglais. Dans le cadre de cet article, nous nous limitons à la présentation de notre recherche sur le *simple past* (SP) et sur le *present perfect* (PresPerf) anglais et leur traduction en français. Cependant, dans le but de fournir une image d'ensemble sur le fonctionnement des temps verbaux en anglais comme langue source, nous ferons souvent

référence à d'autres temps qui ne sont pas explicitement présentés dans cet article.

Dans ce qui suit, nous commençons par décrire le cadre méthodologique dans lequel s'inscrit notre étude: la linguistique contrastive.

### 1.1. La linguistique contrastive

La *linguistique contrastive*, connue aussi sous le nom de *l'analyse contrastive*, est définie par Johansson (2003 : 31) comme étant « une comparaison systématique de deux ou plusieurs langues, dans le but de décrire les similarités et les différences » (notre traduction).

Notre étude s'inspire de la méthodologie de la linguistique contrastive (James 1980 et Krzeszowski 1990), qui consiste en une première étape d'observation monolingue, une deuxième étape de juxtaposition et une troisième étape de comparaison, c'est-à-dire d'analyse des données d'après un *tertium comparationis*.

La première étape, dite de *description*, consiste en l'étude des phénomènes concernés d'un point de vue monolingue, en regardant, notamment dans chaque langue, comment ils sont présentés dans les grammaires et les descriptions. La section 2 résume les différentes études monolingues sur les temps verbaux.

La deuxième étape, dite de *juxtaposition*, consiste en la mise en parallèle des données et prépare le terrain pour la troisième étape, qui est l'analyse en elle-même conformément à une grille d'analyse. C'est dans cette étape que nous récoltons, en corpus, les occurrences des phénomènes étudiés. Dans ce projet, nous utilisons des corpus parallèles (c'est-à-dire contenant des textes en langue originale et leur traduction), ce qui permet de repérer précisément les phénomènes dans les deux langues<sup>2</sup>.

Enfin, la troisième étape est la *comparaison* en elle-même. James (1980, 65) souligne le fait que la condition de base de cette étape est d'établir sur quelle base (selon quels critères) la comparaison interlinguistique sera faite. Selon James, cette étape est très pertinente pour l'analyse contrastive, car pour trouver les différences entre deux ou plusieurs éléments, il faut d'abord que ces éléments aient un ou plusieurs points en commun. Cette partie en commun sera considérée comme une grille sur laquelle les différences ressortiront mieux. Cette

---

<sup>2</sup> L'analyse contrastive en corpus s'effectue traditionnellement dans des corpus comparables (c'est-à-dire des textes dans deux langues originales appartenant au même registre). Dans le cas des temps verbaux, l'influence de la traduction nous semble moins importante, et le besoin d'équivalence précise est plus que nécessaire. Ainsi, les corpus parallèles demeurent un terrain propice à ce type d'investigation.

grille avec des informations partagées par les éléments à comparer est souvent désignée par le terme *tertium comparationis*, c'est-à-dire une représentation tierce, abstraite, vis-à-vis de laquelle les deux langues en contraste peuvent être comparées. A ce propos, Lefer (2009) souligne le fait que les éléments à comparer seront comparés au *tertium comparationis*, et non pas les uns aux autres.

Dans la section suivante, nous présentons les descriptions théoriques des usages des temps verbaux en anglais. Puis, nous présentons notre corpus parallèle et l'analyse que nous avons effectuée à partir de ce corpus. Nous terminons par proposer un modèle bilingue de représentation des temps verbaux (notre *tertium comparationis*).

## 2. Les descriptions théoriques sur les usages des temps verbaux

Les descriptions traditionnelles que nous avons consultées (Quirk et al. 1985 pour l'anglais), ainsi que d'autres études sur la linguistique contrastive et la traduction (Guillemin-Flescher, tomes 1-6, 1992-2002) donnent un répertoire très large et détaillé des usages de temps verbaux dans une perspective monolingue et bilingue dans le but de la traduction. Ces études ont comme but une description schématique ou plus détaillée des typologies des usages des temps verbaux pour mettre en évidence les dissymétries les plus remarquables dans l'organisation aspectuo-temporelle de l'anglais et du français, mais elles ne fournissent pas d'informations sur ce qu'il y a de commun entre les deux langues. Ces descriptions ne sont donc pas suffisamment précises pour désambiguïser les usages des temps verbaux dans la perspective d'une implémentation informatique.

Par la suite, nous résumons les descriptions fournies par les grammaires de deux temps de l'anglais (à savoir le *simple past* et le *present perfect*) et, dans la section 3, nous présentons ce que les corpus parallèles révèlent au sujet des usages de ceux-ci, ainsi que leurs divergences de traduction.

### 2.1. *Simple Past*

Le *simple past* anglais (appelé aussi prétérit simple) est un temps qui peut prendre diverses valeurs, suivant le contexte. La valeur de base est en référence avec le passé, quand l'événement ou l'état s'est produit dans le passé et qu'il existe un laps de temps entre le moment de l'événement et le moment de la parole. Selon Quirk et al. (1985, 184), dans la plupart des cas, le locuteur a en tête un moment précis, défini quand l'événement s'est produit (exemple (1)). Cette référence peut être donnée par des adverbes de temps (exemple (2)) ou alors cette information peut être récupérée du contexte immédiat (usage situationnel dans l'exemple (3)), plus large (connaissances générales

dans l'exemple (4)), ce qui a été dit dans la phrase précédente (usage anaphorique dans l'exemple (5)).

- (1) EN : I *stayed* in Africa for several months (I am no longer in Africa).  
FR : Je suis resté en Afrique pendant plusieurs mois.
- (2) EN : Prices slumped last winter / yesterday.  
FR : Les prix ont chuté l'hiver dernier / hier
- (3) EN: *Did* you lock the front door?  
FR: *As-tu fermé* la porte d'entrée ?
- (4) EN : Byron *died* in Greece.  
FR : Byron *mourut / est mort* en Grèce.
- (5) EN : Then we *entered* the city...the square *was* deserted...  
FR : Alors, nous *entrâmes* dans la ville ... la place *était* déserte.

Le *simple past* peut aussi être utilisé en référence aux temps présent et futur. Quirk et al. (1985, 187) notent trois usages : (a) dans le discours indirect libre (*back shifting*) dans l'exemple (6), (b) avec des verbes de état mental (*attitudinal past*) dans l'exemple (7) et (c) dans des situations hypothétiques (*if-clauses*).

- (6) A: Did you say you have/had no money?  
B: Yes, I am completely broke.  
A: Tu disais que tu n'avais/as pas d'argent?  
B : Oui, je suis complètement fauché.
- (7) Do/Did you want to see me?  
Tu veux/voulais me voir?
- (8) If I were you, I wouldn't go there.  
Si j'étais toi, je n'irais pas là-bas.

## 2.2. Present Perfect

Quirk et al. (1985, 192) notent les usages suivants du *present perfect* : un état pertinent dans le moment de la parole (exemples (9) et (10)), un événement dans un passé indéfini (exemple (11)), un événement récurrent dans une période qui peut inclure le moment de la parole (exemple (12)), ou un événement qui s'est produit récemment par rapport au moment de la parole (exemple (13)).

- (9) EN : That house has been empty for ages.  
FR : Cette maison est vide depuis des années.
- (10) EN : Have you known my sister for long?  
FR : Connais-tu ma sœur depuis longtemps ?
- (11) EN : All our children have had measles.  
FR : Tous nos enfants ont eu la rougeole.
- (12) EN : Mr. Terry has sung in this choir ever since he was a boy.  
FR : Mr. Terry chante dans ce chœur depuis qu'il est enfant
- (13) EN : The plane has just landed.  
FR : L'avion vient d'atterrir.

Selon Quirk et al. (1985), l'anglais est ambigu entre le *simple past* et le *present perfect*, quand les deux temps expriment un état ou un événement dans le passé, mais c'est seulement le *present perfect* qui

peut exprimer aussi un état résultant dans le présent. Il est à noter qu'en français, le passé composé regroupe les deux usages du *present perfect* (tiroir verbal d'antériorité et résultatif) ainsi qu'une partie des usages du *simple past* (événement achevé dans le passé).

Cette section montre que les descriptions des temps verbaux dans les grammaires ont un degré de finesse très élevé et que les usages sont construits et jugés comme corrects par des locuteurs natifs. Aucune indication n'est donnée sur la fréquence de l'un ou l'autre des usages. Nous avons donc mené une étude en corpus pour identifier les usages de ces deux temps verbaux, que nous présentons dans la section suivante.

### 3. Le corpus parallèle

Dans cette section, nous présentons le corpus que nous avons construit<sup>3</sup> et sur lequel nous avons basé notre étude contrastive.

#### 3.1. Description de notre corpus

Notre corpus parallèle (de traduction) est constitué de textes appartenant à des registres variés. La partie de ce corpus<sup>4</sup> qui présente l'anglais comme langue source (LS) contient les registres suivants : littéraire (232 verbes analysés), journalistique (226), Euro Parl<sup>5</sup> (399) et législatif (418). Au total, 1'275 verbes au passé et au présent ont été analysés, sur un total de 1'664 de verbes prédicatifs présents dans notre corpus.

Parmi tous les temps verbaux, nous nous sommes concentrés sur les temps de l'indicatif exprimant le passé et le présent (à savoir *Simple Past* (SP), *Past Continuous*, *Past Perfect* (PP), *Present Perfect*, *Present Continuous*, *Simple Present*) en laissant de côté les auxiliaires modaux, le futur de l'indicatif, le subjonctif, le conditionnel et les temps verbaux non-prédicatifs.

Nous avons analysé notre corpus en deux étapes : (a) une première étape monolingue pour estimer la fréquence des temps verbaux visés dans la LS et (b) une deuxième étape, bilingue, pour analyser les possibilités de traduction en français des temps de l'anglais considéré. Calculer la fréquence des temps en LS est important pour vérifier si

<sup>3</sup> Les chiffres présentés dans cet article concernent l'état du corpus en automne 2011.

<sup>4</sup> Les corpus que nous avons utilisés sont en partie construits par d'autres chercheurs et mis à disposition pour consultation et utilisation. Une autre partie est recueillie et construite (alignée) par nous-mêmes. Dans la bibliographie, nous donnons une description exacte de chaque corpus, l'auteur et la source.

<sup>5</sup> La langue utilisée dans le corpus EuroParl est une langue de débat parlementaire, oral mais quand même retranscrite, assez particulière. Nous avons donc décidé de la considérer comme un registre en soi, dans un sous-corpus portant son nom.

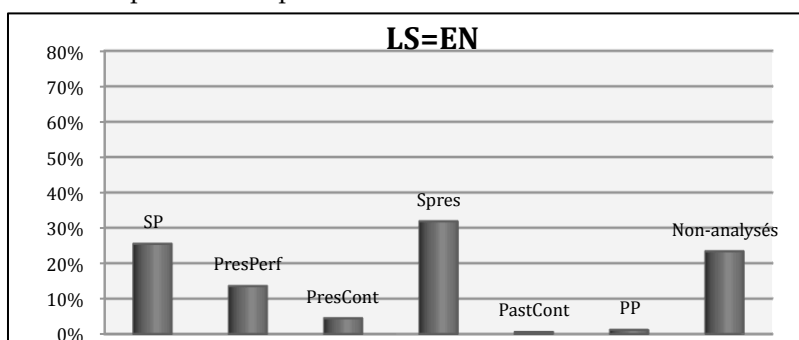
les temps qui posent une divergence de traduction du point de vue de descriptions théoriques sont fréquents dans le corpus. Si tel est le cas, alors le corpus nous fournit assez de contexte pour ces temps afin de les analyser et tirer des généralités pertinentes du point de vue statistique.

### 3.2. Résultats

Dans cette section, nous présentons les résultats de notre analyse, d'abord dans une perspective monolingue (graphiques 1 et 2) et ensuite, dans une perspective bilingue (graphiques 3 et 4), en nous concentrant plus particulièrement sur le *simple past* et le *present perfect*.

#### 3.2.1. La fréquence des temps verbaux dans la LS

Le graphique 1 montre que les temps les plus représentés dans notre corpus sont le *simple present* (31,8%), le *simple past* (25,4%) et le *present perfect* (13,5%). Ce graphique montre également que les valeurs du *past perfect* (PP) et le *past continuous* (PastCont) sont en dessous de 1,15% de la totalité du corpus. Le *present continuous* (PresCont) est aussi assez peu représenté (4,45%). Ces trois temps posent une divergence du point de vue théorique mais à cause de leur faible fréquence dans notre corpus, nous avons décidé de ne pas les analyser dans cette première étape de notre recherche.

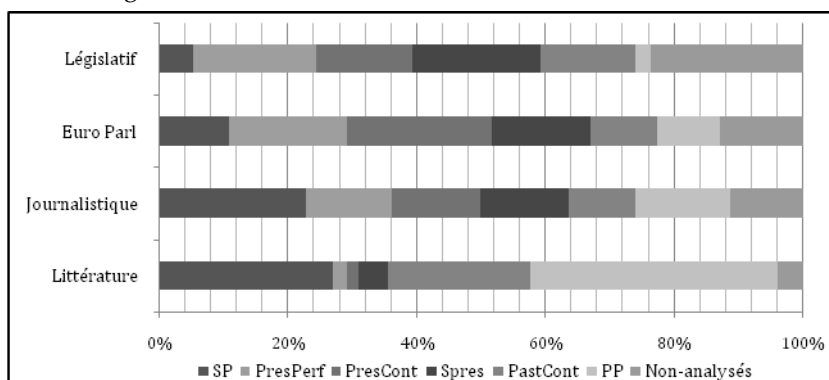


Graphique 1 : La fréquence des temps verbaux dans la partie anglais du corpus

#### 3.2.2. La fréquence des temps verbaux par type de corpus

Le graphique 2 ci-dessous présente la fréquence des temps verbaux en anglais en fonction du type de corpus : littérature, journalistique, EuroParl et législatif. Le *simple past* est bien représenté dans tous les types de corpus (40% dans le littéraire, 24% dans le journalistique, 23% dans EuroParl et, un peu moins, 11% dans le législatif), alors que le *present perfect* est utilisé surtout dans les domaines législatif (43%),

administratif (39%) et journalistique (14%). Le graphique montre aussi la distribution des temps moins utilisés parmi les corpus, comme le *past perfect*, qui se retrouve dans 33% des cas dans le corpus littéraire et législatif. Parmi les occurrences de *present continuous*, 48% sont dans EuroParl et 33% dans le registre législatif. Pour le cas du *past continuous*, il est utilisé le plus souvent dans les registres littéraires (33%) et législatifs (33%).



Graphique 2 : La fréquence des temps verbaux du passé et du présent dans la partie anglaise des corpus<sup>6</sup>

### 3.2.3. Les possibilités de traduction selon les corpus

Notre corpus a montré plusieurs possibilités de traduction pour chaque temps verbal analysé, une partie avec des pourcentages plus importants, qui se détachent largement du reste. Nous avons donc pris en considération comme possibilités de traduction uniquement les temps les plus présents.

Les graphiques 3 et 4 présentent les possibilités de traduction du *simple past* et du *present perfect* en français. Parmi les possibilités de traduction du *simple past* en français, c'est l'imparfait (IMP) et le passé composé (PC) qui se partagent en grande partie la traduction (dans EuroParl, le registre législatif et journalistique), tandis que le passé simple (PS) reste limité au registre littéraire. Ces possibilités de traduction sont mentionnées aussi par l'étude contrastive faite par Wyld (2002) mais qui sont considérées comme des variations de l'*aoriste*<sup>7</sup> au passé. Cette étude décrit une partie du domaine

<sup>6</sup> Pour plus de lisibilité, la légende reprend l'ordre d'apparition des temps par tranche dans les corpus.

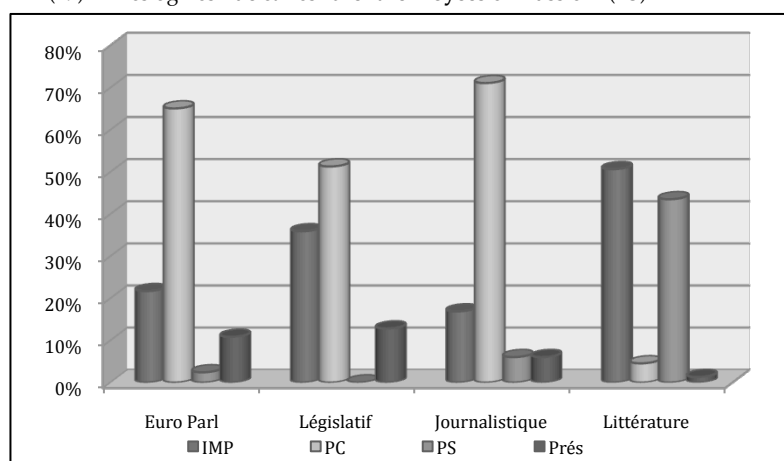
<sup>7</sup> Wyld (2002, 10) décrit l'*aoriste* comme renvoyant aux deux propriétés suivantes : une rupture entre le moment du procès et de l'énonciation, et le procès est vu comme un tout et insécable.



sémantique du *past simple* et les temps verbaux correspondant en français, sans proposer pour autant, des traits qui pourraient être utilisés pour la désambiguïsation des usages du *simple past* dans la perspective d'une implémentation informatique.

Les exemples suivants illustrent la divergence de traduction du *past simple* (un exemple est donné pour chaque possibilité de traduction).

- (14) He **seemed** about seventeen years of age, and was of quite extraordinary personal beauty, though somewhat effeminate. (The portrait of Mr. W.H.)
- (15) « Il **paraissait** avoir dix-sept ans, et il était d'une beauté absolument extraordinaire, quoique manifestement un peu efféminée. » (IMP)
- (16) Madam President, particularly out of respect for you, for myself and for this House, I **did not want** to explicitly refer to the issue in question. (Euro Parl)
- (17) « Madame la Présidente, par respect pour vous surtout, respect que je vous dois et par respect pour ce Parlement, je n'ai pas **voulu** me référer explicitement au sujet dont je parlais. » (PC)
- (18) Nuclear warheads **were returned** to Russia. (The New York Times)
- (19) « Les ogives nucléaires **furent renvoyées** en Russie. » (PS)



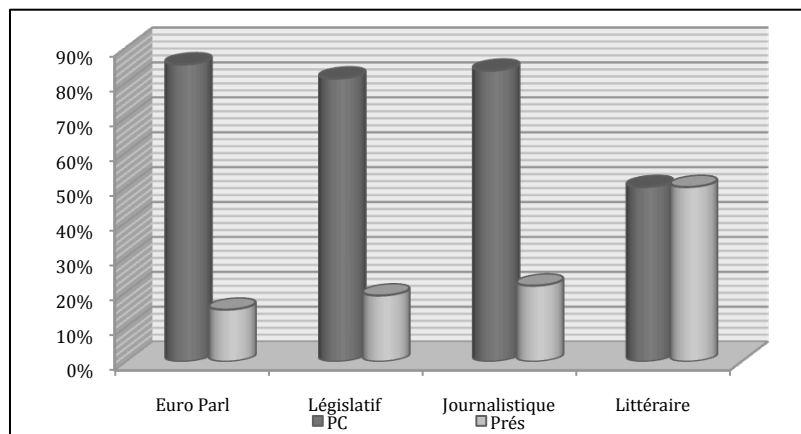
Graphique 3 : Les possibilités de traduction pour le *simple past* anglais parmi les types de registres envisagés

La traduction du *present perfect* (graphique 4) est en grande partie (plus de 80%) couverte par le passé composé dans EuroParl, les registres législatif et journalistique. Dans le registre littéraire, la situation change, car le présent a une proportion égale (50%) avec celle du passé composé. Caudal (2002) note lui aussi les possibilités de traduction en français pour le *present perfect* comme étant le passé composé et le présent pour rendre compte des différences aspectuelles. L'étude contrastive de Caudal est une description des

possibilités de traduction du *present perfect* en français mais qui n'identifie pas de critères de désambiguïsation pour les usages de ce temps verbal. Notre étude en corpus décrit et analyse les contextes d'usage du *present perfect* dans le but d'identifier des indices qui influencent le choix de traduction dans la langue cible.

Cette distribution souligne le fait que le *present perfect* peut renvoyer le plus souvent à un événement antérieur qu'à un état résultant au moment de l'énonciation. Les exemples suivants illustrent la divergence de traduction du *present perfect* (un exemple est donné pour chaque possibilité de traduction).

- (20) Furthermore, the Portuguese Presidency **has proved** that it does not want to see a shift towards an intergovernmental approach. (EuroParl)
- (21) « La présidence portugaise **a montré** qu'elle ne voulait pas d'une dérive intergouvernementale. »
- (22) After all, the victory of the secular Awami National Party (ANP) in Pakistan's North-West Frontier Province is a strong indication that people there have **rejected** religious political forces and violence. (News Commentaries)
- (23) « La victoire de l'Awami National Party (ANP), un parti laïc, dans la province frontière du Nord-Est est après tout un signe clair que ses habitants **rejetent** les forces politiques religieuses et la violence. »



Graphique 4 : Les possibilités de traduction pour le *present perfect* anglais parmi les types de registres envisagés

Dans cette section, nous avons montré qu'il existe une disparité entre ce qui est projeté comme pertinent du point de vue des descriptions théoriques et ce que le corpus montre. Nous avons également montré le peu d'occurrences trouvées pour le *past perfect*, *past continuous*, et le *present continuous*. La disparité entre la description théorique et le corpus se manifeste également par le fait que certains usages décrits dans les études théoriques n'apparaissent

simplement pas dans le corpus. Nous avons montré grâce à l'exemple de ces deux temps verbaux qu'il existe un réel problème de traduction à cause de l'ambiguïté de ces temps. Pour mieux représenter ces divergences, il faut pouvoir disposer d'un cadre théorique et d'un modèle de représentation complet.

#### **4. Vers une représentation multilingue du système verbal**

Notre représentation des temps verbaux s'inscrit dans le cadre général de la théorie de la pertinence (Sperber & Wilson 1986). Selon ce cadre, les temps verbaux sont considérés comme des catégories référentielles, qui ont un rôle très important dans le processus de reconnaissance de l'intention communicative de l'interlocuteur. Le destinataire doit inférer le moment de référence des événements présentés, l'ordre temporel des événements, s'il en existe un, et une éventuelle relation causale ou d'autre nature.

Nous présentons dans cette section notre modèle de représentation des divergences de traduction, ainsi que les critères de désambiguïsation. C'est un modèle bilingue pour le moment, mais qui pourrait, si besoin, être adapté à plusieurs langues. Ce modèle constitue le *tertium comparationis* que nous avons construit et utilisé pour analyser en parallèle les données extraites de nos corpus.

##### **4.1. Cadre théorique : les coordonnées de Reichenbach**

Reichenbach (1947) est le premier à proposer une description formelle des temps verbaux, description qui a été largement acceptée et adoptée par la communauté linguistique. Selon lui, chaque temps peut être décrit en termes d'une combinaison entre le moment de la parole S (*speech point*), le moment de l'événement E (*event point*) et le point de référence R (*reference point*). Les temps du passé ont une combinaison du type E précédent à S avec R simultané à S pour le passé composé français et le *present perfect* anglais et R simultané à E pour le passé simple français et le *simple past* anglais.

En ce qui concerne le temps du présent, les trois coordonnées sont présentées comme étant simultanées. Un des problèmes du modèle de Reichenbach est que les temps progressifs ou imperfectifs, comme l'imparfait français, le *present* et *past continuous* anglais sont assez difficilement représentés par ces coordonnées. Reichenbach propose le terme d'*étendu* pour expliquer la progressivité, ce qui reste relativement difficile à représenter.

Dans notre modèle, nous adoptons les descriptions de temps du passé et du présent avec les combinaisons des coordonnées E, R et S pour les temps perfectifs. En ce qui concerne les temps progressifs, décrits par Reichenbach comme étant des *temps étendus*, le corpus montre un usage moins fréquent que les autres temps. Nous ne les

avons donc pas encore étudiés de très près. Nous allons, par la suite, chercher la meilleure manière de les décrire et de les modéliser.

#### 4.2. Notre proposition de représentation

Suite à de longues observations du paysage linguistique des temps verbaux en français et en anglais, nous sommes arrivés à la conclusion que les systèmes verbaux dans ces deux langues ne diffèrent pas seulement au niveau aspectuel, mais également au niveau de leur description. Pour le français, on parle souvent de *tiroir verbal de l'accompli* (Caudal 2002), *focalisation*, *point de vue* et *usage interprétatif* pour le passé simple (Tahara 1998), *effet de peintre* pour l'imparfait (Riegel 1994), etc. En ce qui concerne l'anglais, on parle des *usages du simple present*, par exemple, *en lien avec le présent, le passé, le futur*, des « *habitudes* » dans le présent, le passé, des *verbes qui acceptent ou pas le progressif* (Quirk et al. 1985) etc.

Nous nous sommes rendu compte que pour désambiguïser les temps verbaux afin d'améliorer la traduction automatique, nous avons besoin de décrire les systèmes verbaux dans ces deux langues selon le même cadre, c'est-à-dire de créer un *tertium comparationis*, présenté plus loin dans la figure 1.

Notre proposition de représentation des temps verbaux est développée en collaboration avec Jacques Moeschler et elle constitue une nouvelle perspective sur les temps verbaux. Dans un autre projet de recherche avec Anne Reboul (L2C2, USC Lyon) et Izumi Tahara (Meiji Gakuin University, Tokyo) sur le *discours indirect libre* et le *présent historique* dans différentes langues, Moeschler propose un squelette de base pour représenter les temps verbaux. Dans une étude en cours avec Moeschler, nous avons adopté et adapté ce squelette pour représenter les temps verbaux dans le but de représenter les divergences de traduction.

Nous considérons que les temps verbaux ont des usages par défaut (une valeur de base) et des usages contextuels, qui peuvent apparaître seulement sous certaines conditions. Nous proposons comme critères de désambiguïstation, la distinction entre usage *narratif* (événements avec ordre temporel) et *non narratif* (sans ordre temporel) (Reboul et Tahara 2010, Moeschler, Grisot et Cartoni 2011). Ce n'est pas la distinction en elle-même qui est innovatrice mais l'usage de cette distinction pour la désambiguïstation du *simple past* et sa traduction en français. Il en va de même pour la deuxième distinction que nous utilisons, c'est-à-dire celle entre usage *subjectif* (avec un point de vue, qui peut être celui du personnage, narrateur ou de l'auteur lui-même) et *non subjectif* (Reboul et Tahara 2010, Moeschler, Grisot et Cartoni, 2011). Cette distinction est basée sur l'idée de *perspective* et d'*effet de*

*subjectivisation* proposée par Fleischmann (1991), Sthioul (1998) et Tahara (1998) pour les temps verbaux français.

Dans notre modèle, nous allions donc les coordonnées de Reichenbach (contenu sémantique) aux informations pragmatiques décrites ci-dessus. Par exemple, l'imparfait français reçoit les mêmes coordonnées que le passé simple (E=R<S) comme le décrit Reichenbach, sauf qu'il est par défaut "non narratif subjectif", tandis que le passé simple est par défaut "non narratif non subjectif".

L'aspect grammatical (perfectif vs. imperfectif) est un trait qui a beaucoup intéressé les linguistes qui ont comparé les systèmes verbaux français et anglais. Suite à l'analyse de notre corpus, qui a montré un usage très limité du *past* et *present continuous* anglais, nous avons décidé de ne pas inclure, pour le moment, cette distinction dans la base de notre modèle de représentation.

Notre description sémantique et pragmatique des temps verbaux est représentée dans la figure 1, qui contient les critères de désambiguïsation, à savoir les coordonnées de Reichenbach qui représentent la base de notre modèle (lecture horizontale), sur laquelle nous ajoutons des critères pragmatiques, notamment les usages narratifs, non narratifs, subjectifs et non subjectifs (lecture verticale). Parmi les différentes combinaisons des coordonnées de Reichenbach, les relations R<S<E (conditionnel) et S<R=E (futur) ne sont pas analysées dans cet article.

E< R	R<S	Non-narratif		Narratif	
		Subjectif	Non-subjectif	Subjectif	Non-subjectif
		IMP- <b>SP</b>	IMP- <b>SP</b>	PS- <b>SP</b>	PS - <b>SP</b>
		PC - <b>SP</b>			
	R=S	PC - PresPerf			
E=R=S		Prés - <b>PresPerf</b>			

Figure 1 *Tertium comparationis* pour les temps verbaux (usages des temps verbaux en anglais comme SL et leurs possibilités de traduction en français ; Grisot, 2011)

La figure 1 est une carte d'équivalence de traduction pour le *simple past* et le *present perfect*, qui montre les différentes traductions suivant le type de contexte. Dans la langue source, il faut identifier dans le contexte certaines conditions (par exemple, des marqueurs de subjectivité ou l'existence d'un ordre temporel entre les événements)

pour choisir la meilleure possibilité de traduction. Ces distinctions aident, par exemple, à la désambiguïsation du *simple past* en anglais quand il est traduit en français, avec le choix entre l'imparfait (en contextes non narratifs subjectifs) et le passé simple (en contextes narratifs non subjectifs).

Les coordonnées de Reichenbach permettent de distinguer le passé composé des deux autres temps du passé en français : le passé simple et l'imparfait. Le passé simple et l'imparfait ont une combinaison de type E<R<S tandis que pour le passé composé, le point de référence R est simultanément au moment de la parole S.

En ce qui concerne la traduction du *present perfect*, les coordonnées de Reichenbach ont un rôle primordial pour sa désambiguïsation. Le *present perfect* est traduit par le passé composé dans des contextes où E précède R et par le présent dans des contextes où E est simultanément à R.

Le modèle de critères de désambiguïsation reste encore à développer pour les autres temps verbaux, ainsi que l'établissement des tests pour identifier les deux types d'usages. Le modèle que nous proposons n'est pas simplement descriptif mais aussi prédictif : ces critères de désambiguïsation seront transformés en étiquettes dans une prochaine étape d'annotation. Un temps verbal donné avec une étiquette donnée conduit forcément dans la langue cible à la bonne possibilité de traduction. Par exemple, un *simple past* avec étiquette « usage subjectif » conduit à l'utilisation de l'imparfait en français. Une prochaine étape de notre recherche est donc l'organisation des expériences d'annotation avec ces traits pragmatiques en vue de leur validation.

## 5. Conclusion

Nous avons montré dans cet article la nécessité d'une analyse contrastive en corpus pour pouvoir proposer un modèle de description sémantique et pragmatique multilingue des temps verbaux. Nous avons aussi souligné la nécessité d'une perspective pragmatique dans la compréhension et la traduction des temps verbaux.

Notre analyse contrastive en corpus nous a permis d'identifier les divergences de traduction les plus importantes basées sur la fréquence des temps verbaux dans le corpus et les pourcentages de temps utilisés comme possibilités de traduction dans la langue cible.

Dans le but de désambiguïser les usages de ces deux temps, nous avons proposé un modèle de représentation dans une perspective bilingue, en nous appuyant sur les coordonnées de Reichenbach, et en incluant des distinctions plus fines entre usage narratif et non narratif, subjectif et non subjectif. Nous avons souligné le fait que le degré de

finesse de notre analyse des usages des temps verbaux tant dans la langue source que dans la langue cible, n'est pas trop élevé pour nous permettre de travailler avec les usages les plus fréquents. En outre, des usages, comme l'imparfait historique, qui aurait un usage narratif subjectif, ne sont pas pris en compte, pour le moment, dans cette représentation.

Par la suite, ces critères de désambiguïsation seront utilisés pour annoter des corpus bilingues, permettant d'implémenter des modèles de langue utiles aux systèmes de traduction automatique statistique. Pour ce faire, tous les usages et les possibilités de traduction de tous les temps verbaux du passé et du présent en français et en anglais seront investigués selon le modèle présenté dans cet article.

### Remerciements

Nous remercions particulièrement le professeur Jacques Moeschler pour les discussions enrichissantes et les conseils tout au long de ce travail. Cette « recherche est financée par le Fond National Suisse pour la Recherche scientifique, par l'entremise du projet Sinergia COMTIS CRSI22-127510 (<http://www.idiap.ch/project/comtis>).

### Bibliographie

- CARSTON R. (1998), "The semantics/pragmatics distinction: a view from the relevance theory". Working papers in linguistics 10, University College London.
- CAUDAL P. (2002), « "Perfect", résultativité et contraste français/anglais. » In Guillemin-Flescher (éd.), *Linguistique contrastive et traduction*, Tome 6, Gap: Ophrys.
- CHESTERMAN A. (1998), *Contrastive Functional Analysis*. Amsterdam: Benjamins.
- FLEISCHMAN S. (1991), « Verb tense and point of view in narrative » In *Discourse-pragmatics and the verb: the evidence from Romance*, 26-54.
- GUILLEMIN-FLESCHER J. (1992-2002), *Linguistique contrastive et traduction*, Tomes 1-6, Gap: Ophrys.
- GRISOT, C. (2011) - *Disambiguation of Verb Tenses for Machine Translation: Semantic and Pragmatic Description and Annotation*. Présentation de poster lors de l'école doctorale CUSO à Champéry, Suisse.
- GREVISSE, M. (2008). *Le bon usage: grammaire française* (14e éd.). Bruxelles [etc.]: De Boeck.
- JAMES C. (1980), *Contrastive analysis*. Londres: Longman.
- JOHANSSON, S. (2003), "Contrastive linguistics and corpora", In S. Granger, J. Lerot et S. Petch-Tyson (Eds). *Corpus-based Approaches to Contrastive Linguistics and translation Studies* Amsterdam and New-York: Rodopi, 31-45
- KRZESZOWSKI T.P. (1990), *Contrasting Languages. The Scope of Contrastive Linguistics*. Berlin & New York: Mouton de Gruyter.

- LEFER, M. A. (2009), *Exploring lexical morphology across languages: a corpus-based study of prefixation in English and French writing*. Thèse de doctorat, Université catholique de Louvain, Belgique.
- MOESCHLER J., JAYEZ J., KOZLOWSKA M., LUSCHER J-M., SAUSSURE L. de, STHIOUL B. (1998), *Le Temps Des Événements: Pragmatique De La Référence Temporelle*. Paris: Kimé.
- QUIRK, R., GREENBAUM, S., LEECH, G., & SVARTVIK, J. (1985). *A comprehensive grammar of the English language*. Londres: Longman.
- REBOUL A., TAHARA I. (2010), "The pragmatics of represented speech and thought: a comparative approach to French/English/Japanese". *13<sup>th</sup> Annual Meeting of the Pragmatics Society of Japan*, Kansai University, 4-5 December 2010.
- REICHENBACH H. (1947), "The tenses of Verbs". Section 51 of *Elements of Symbolic Logic*, New York, The McMillan Company, 287-298.
- RIEGEL M., PELLAT J.-C., RIOUL R. (1994), *Grammaire méthodique du français*. Quadrige Manuels/PUF.
- SPERBER D., WILSON D. (1986/1995), *Relevance Communication and Cognition*. Oxford: Basil Blackwell Ltd.
- STHIOUL B. (1998), « Temps verbaux et point de vue ». In Moeschler J. Jayez J., Kozłowska M., Luscher J-M., Saussure L. de, Sthioul B. *Le Temps Des Événements: Pragmatique De La Référence Temporelle*. Paris: Kimé, 197-220.
- TAHARA I. (2000), « Le passé simple et la subjectivité ». *Cahiers de Linguistique Française* 22, 189-218.
- WILSON D., SPERBER D. (1998), "Pragmatics and Time". In R. Carston & S. Uchida (eds) *Relevance Theory: Applications and implications*, Amsterdam and New-York : John Benjamins, 1-22.
- WYLD H. (2002), « Passé simple, passé composé à valeur dite temporelle, prétérit simple : Variations linguistiques sur l'aoristique au passé ». In Guillemin-Flescher, *Linguistique contrastive et traduction*, Tome 6, Gap: Ophrys, 5-76.

### Référence des corpus

#### Le sous corpus littéraire

*The portrait of Dorian Gray*, O. Wilde, traduction en français par Jean Gattégno, Editions Gallimard, 2000. Formatage électronique et alignement par C. Grisot, 2011.

*Sense and sensibility*, J. Austin, traduction française disponible sur [http://www.gutenberg.org/wiki/Main\\_Page](http://www.gutenberg.org/wiki/Main_Page). Alignement par C. Grisot, 2011.

#### Le sous corpus journalistique

*News Commentaries*. Corpus journalistique En-Fr, utilisé dans des défis de traduction automatique « International Workshop on Spoken Language Translation ». Disponible sur <http://iwslt2010.fbk.eu/node/34> date de consultation : 30-04-2011.



*The New York Times*. Corpus construit par le Laboratoire FORELL en collaboration avec Philippe Rivière. Disponible sur [www.chabal.rezo.net](http://www.chabal.rezo.net), date de consultation : 30-04-2011.

*Presseurop*, articles recueillis (<http://www.presseurop.eu/fr>, date de consultation : 30-04-2011) et alignés par C. Grisot, 2011.

#### **Le sous corpus législatif**

*The JRC-Acquis Multilingual Parallel Corpus*. Disponible sur <http://langtech.jrc.it/JRC-Acquis.html>, date de consultation : 30-04-2011.

*EuroConst*, constitué par Joerg Tiedemann. Disponible sur [www.opus.lingfil.uu.se](http://www.opus.lingfil.uu.se) date de consultation : 30-04-2011.

#### **Le sous corpus Euro Parl**

*EuroParl*, constitué par Joerg Tiedemann. Disponible sur [www.opus.lingfil.uu.se](http://www.opus.lingfil.uu.se), date de consultation : 30-04-2011.