

Time in Translation : une recherche quantitative sur la sémantique du parfait anglais et du parfait français

Mandy Woelk

3746763

M.Woelk@uu.nl

Langue et culture françaises

19-06-2017

Dr. Bert Le Bruyn

Prof. dr. Henriëtte de Swart

Résumé

Les études de corpus pour découvrir comment une langue est utilisée sont de base qualitative, mais cette méthode ne peut pas prouver statistiquement les différences entre langues. Cette recherche a testé une méthode quantitative nouvelle en linguistique afin d'étudier la sémantique du parfait anglais et du parfait français. Des études précédentes ont suggéré que le parfait français représente le present perfect et le simple past en anglais. Dans cet article, nous avons montré que cela n'est pas le cas, puisque le présent joue également un rôle important dans la traduction du parfait. Ces nouvelles notions peuvent aider les enseignants et les traducteurs à améliorer et faciliter leur travail. Futures études devraient tester cette nouvelle méthode avec des données supplémentaires sur le parfait ainsi que sur d'autres temps.



Contenu

Remerciements	3
Introduction	4
Méthode.....	6
Création des données	6
Sélectionner les contextes	6
Visualisation	7
Analyse	8
Préparation des données.....	8
R.....	9
Résultats	14
Statistiques descriptives.....	14
Analysis of similarities	14
Discussion	14
Références	18

Remerciements

Premièrement, j'aimerais remercier Bert le Bruyn de m'avoir donné la possibilité d'intégrer mes connaissances sur la statistique et mes connaissances sur la langue et la linguistique française dans un nouveau projet très intéressant.

Deuxièmement, j'aimerais remercier Kirsten Schutter de rechercher des informations sur l'analyse et d'investir du temps à découvrir et expliquer comment cette analyse pourrait être appliquée sur nos données.

Time in Translation : une recherche quantitative sur la sémantique du parfait anglais et du parfait français

Introduction

La recherche dans la linguistique peut être divisée en deux domaines principales ; la recherche théorique, qui étudie comment on peut utiliser une langue, et la recherche pratique, qui se focalise surtout sur l'utilisation des langues. En linguistique, traditionnellement, la plupart des études sont donc de base théorique (Biber, Conrad, & Reppen, 1998). Il est important d'étudier comment on peut utiliser une langue selon les théories, par exemple pour mieux comprendre les mécanismes derrière l'utilisation. Pourtant, l'aspect pratique mérite également d'être étudié. Une façon d'étudier l'usage d'une langue est avec des études de corpus, qui se sont développées de plus en plus depuis les années '60 (Aijmer & Altenberg, 2014). Le développement dans les deux domaines de recherche linguistique est important puisque si on sait comment une langue peut être utilisée théoriquement et comment une langue est utilisée en réalité, il est possible de faire des théories intégrées. C'est-à-dire, des théories qui comprennent les connaissances théoriques ainsi que les connaissances pratiques.

A l'aide des corpus, on peut non seulement étudier l'utilisation d'une langue, mais aussi découvrir les disparités et les similarités entre les langues. Ces découvertes peuvent également aider à mieux comprendre les différences de sens entre des mots qui sont utilisés de la même façon dans certaines langues, mais pas dans d'autres. Par exemple, Simon-Vandenberg (2013) a étudié les différences concernant les adverbes de base entre l'anglais, le français et le néerlandais. En analysant les traductions en français et en néerlandais, cette étude a clarifié les différences entre les adverbes '*basically*', '*essentially*' et '*fundamentally*'. Ces mots ont à peu près le même sens en anglais, mais sont utilisés différemment en français et en néerlandais. Egan (2013) a analysé le concept de '*betweenness*' dans les traductions de plusieurs livres du norvégien en anglais et en français. Le mot '*mellom*' en norvégien se traduit le plus souvent comme '*between*' et '*entre*' selon cette étude, sauf dans l'utilisation spatiale. Cela peut indiquer une différence de sens dans cette catégorie de '*betweenness*' et une ressemblance entre les autres catégories. Lavid, Arús, & Moratón (2013) ont étudié l'utilisation de l'anglais et de l'espagnol dans des journaux et ont conclu que cette utilisation (par exemple le choix des mots) dépend du genre de l'article, et non nécessairement de la langue même.

Malgré ce développement et la croissance des études de corpus faites pour découvrir les disparités et ressemblances entre langues, ce domaine nécessite encore de l'extension.

C'est-à-dire, on peut montrer que les langues diffèrent l'une de l'autre en comparant des phrases et leur utilisation de façon qualitative, mais cela ne suffit pas. Les études précédentes ont déjà utilisé des méthodes quantitatives en montrant leurs résultats dans des graphiques, mais cela est trop limité. Le projet *Time in Translation : the semantics of the PERFECT* vise à étendre les connaissances des langues en connectant la recherche qualitative et quantitative. Le projet se focalise sur le parfait avec auxiliaire (appelé parfait dans le reste de cet article), qui équivaut dans sa forme un passé composé (PC) en français et un present perfect (PP) en anglais. C'est-à-dire, il consiste en un auxiliaire (être ou avoir) et un participe passé.

Il existe toujours un manque de clarté en ce qui concerne ce temps et sa sémantique, partiellement à cause de la concurrence avec d'autres formes du passé, comme l'imparfait (Schaden, 2009). Par exemple, en français on peut dire « Napoléon a transformé Paris », alors qu'en anglais on ne peut pas utiliser le parfait ici, puisque Napoléon n'est plus vivant (Ritz, 2012). Yoon (2012) montre que l'anglais fait une différence entre l'utilisation du PP et du simple past (SP), vu que ce dernier temps peut être utilisé avec un adverbe défini du passé (comme *yesterday*), mais le PP n'accepte pas de mots dans cette catégorie. Le français, en revanche, permet d'utiliser un adverbe défini du passé dans une phrase écrite au PC. Par exemple, on peut dire « Hier, maman est morte », alors que l'on ne peut pas dire « Mom has died yesterday ». Dans ce cas, il faut le SP : « Mom died yesterday ». Vet (1992 ; 1999 ; 2001 ; cité dans De Swart, 2007) émet l'hypothèse que le PC est un temps polysémique qui est utilisé de la même façon que le PP et le SP. Cela implique que le PC est l'équivalent du PP et du SP ensemble. Cela voudrait alors dire que tous les PPs et les SPs se traduisent par un PC. Pourtant, Grisot et Moeschler (2014) ont trouvé que le PP en anglais se traduit non seulement par le PC en français, mais aussi par le présent. Cela paraît correct, vu que le PP est utilisé afin d'exprimer un événement qui est toujours d'actualité, contrairement au SP (De Swart, 2007). En effet, Loáiciga, Meyer et Popescu-Belis (2014) ont trouvé que 25% des traductions du SP en français contiennent un présent.

En analysant l'utilisation du parfait en différentes langues, on peut combler peu à peu le vide concernant ce temps et sa sémantique. Dahl (2014) a déjà fait une étude sur le parfait et sa sémantique entre langues, mais sa méthode est restreinte à cause de l'approche inductive. C'est-à-dire, il se focalise sur les passés composés prototypiques et par conséquent, il n'arrive pas à découvrir la sémantique entière de ce temps. Le projet *Time in Translation* essaye de résoudre ce problème et de trouver davantage de pièces du puzzle sémantique en utilisant une nouvelle approche. Au lieu des données qualitatives, qui sont utilisées dans la plupart des études précédentes, nous avons créé des données quantitatives basées sur des phrases tirées

d'un corpus. Nos données portent sur le parfait de cinq langues ; l'anglais, le néerlandais, le français, l'espagnol et l'allemand. Dans cet article, nous visons à donner une interprétation aux données par rapport à la sémantique du parfait en ce qui concerne la langue française et anglaise.

Cette recherche a pour but de vérifier si le PC équivaut en effet le PP et le SP, comme suggéré par Vet (1992 ; 1999 ; 2001 ; voir De Swart, 2007) Cela résulte en la question de recherche suivante : « Le PC en français est-il l'équivalent du PP et du SP en anglais ? » Nous nous attendons à ce que la réponse soit non et que le PP ainsi que le SP se traduisent non seulement par le PC, mais aussi par le présent. L'hypothèse nulle correspondant à cette attente est que le PP et le SP se traduisent toujours par un PC. Aussi, nous nous attendons à ce que le PC se traduise non seulement par un PP ou un SP, mais aussi par un simple présent. L'hypothèse nulle de cette attente est que le PC se traduit toujours par un PP ou un SP.

Méthode

Création des données

Afin de créer les données pour la recherche, nous avons repris la méthode de Wälchli et Cysouw (2012). Dans leur article, ils présentent une carte sémantique comparant plusieurs langues. C'est-à-dire, ils comparent la signification des mots entre langues et étudient leurs traductions. Avec ces informations, ils créent une distance, exprimée en valeur de divergence. Par exemple, le mot *come* en anglais peut être traduit de différentes façons en français ; s'approcher, venir, arriver, se diriger ou encore entrer. Si dans deux contextes le mot *come* est traduit différemment, on peut dire que la différence est de 1/2 (le mot anglais reste le même, mais le mot français change). La valeur de divergence sera alors 0.5 (Wälchli & Cysouw, 2012). De cette manière, on peut comparer un nombre de langues illimité dans différents contextes afin de collectionner de plus en plus de pièces du puzzle. Cela crée toute une matrice comprenant les distances entre contextes. Ces distances peuvent à leur tour être visualisées dans un graphique que l'on appelle une carte sémantique. Les données utilisées dans ce mémoire sont créées sur la base de cette méthode que l'on appelle Exploration de données de traduction. Contrairement à Wälchli et Cysouw (2012), nous n'utilisons pas des mots pour comparer, mais des phrases contenant un parfait dans au moins une des cinq langues comprises dans ce projet.

Sélectionner les contextes

Pour trouver les contextes, nous avons utilisé différents corpus parallèles multilingues. Dans l'analyse décrite dans cet article, le corpus Europarl (Koehn, 2005) est utilisé, contenant des

contextes formels. Nous avons sélectionné des contextes comprenant la combinaison auxiliaire (être ou avoir) + participe passé en tenant compte de plusieurs difficultés (Van der Klis, Le Bruyn, & De Swart, 2017) ; 1. des mots entre l’auxiliaire et le participe passé, 2. des restrictions lexicales pour ‘être’ en français, en néerlandais et en allemand et 3. un ordre inversé dans les subordonnées en allemand et en néerlandais (Van der Klis, Le Bruyn & De Swart, 2015, voir Van der Klis et al., 2017). Une personne maîtrisant toutes les cinq langues a vérifié les données en utilisant l’application Alignement des verbes. De cette manière, les non-parfaits et les mauvaises traductions ont été filtrés de sorte que seulement les phrases correctes restent dans l’ensemble des données. Ces phrases ont été marquées avec un des temps présentés dans le Tableau 1 (Van der Klis et al., 2017). Cela résulte en une matrice où chaque ligne présente les temps utilisés dans un certain contexte (dans chacune des cinq langues), où chaque contexte contient au moins un parfait (voir le Tableau 2 ; Van der Klis et al., 2017). Cette matrice est à son tour transformé en une matrice de distance comme décrit ci-dessus. Une distance de 0 veut dire que toutes les cinq langues utilisent le même temps dans les deux contextes comparés, une distance de 1 veut dire que toutes les langues utilisent un temps différent. Dans le cas où, par exemple, quatre temps sur cinq sont pareils, la distance est 1/5.

Tableau 1 : *Les différents temps accordés aux phrases.*

Generic tense	DE	EN	ES	FR	NL
PERFECT	Perfekt	present perfect present perfect continuous	pretérito perfecto compuesto	passé composé	vtt
PRESENT	Präsens	present	presente	présent	ott
PAST	Präterium	simple past	pretérito imperfecto pasado receinte pretérito perfecto simple	imparfait passé récent	ovt
PAST PERFECT	Plusquamperfekt	past perfect	pretérito pluscuamperfecto	plus-que-parfait	vvt
OTHER	Futur I/II	-	participio	futur antérieur	-

Tableau 2 : *Exemple des temps utilisés en chaque langue, présenté par contexte*

#	DE	EN	ES	FR	NL
1	Perfekt	present perf.	passé comp.	pretérito perf. comp.	vtt
2	Präterium	simple past	passé comp.	pretérito perf. comp.	vtt
3	Perfekt	present perf.	passé récent	pasado receinte	vtt

Visualisation

Afin de visualiser les données et créer la carte sémantique, nous avons utilisé la matrice de distance et appliqué une méthode que l’on appelle *multidimensional scaling* (Van der Klis et al., 2017). Cette carte sémantique montre les différents temps et leur place parmi les autres

temps (voir l'Image 1) et permet de comparer les temps dans une langue ainsi que les temps entre les langues, puisque les couleurs correspondent aux mêmes temps¹ dans toutes les langues. Par exemple, le PC est représenté en bleu, comme le PP en anglais. En cliquant sur un point sur la carte, on peut voir les données derrière. C'est-à-dire, on peut voir les phrases sur lesquelles le point est basé.

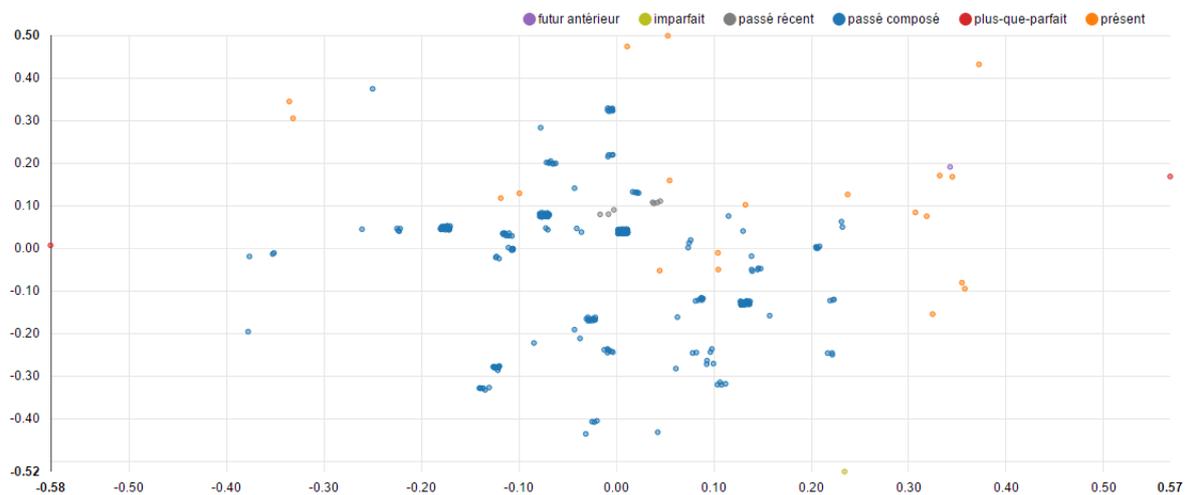


Image 1 : Exemple d'une carte sémantique

Analyse

Nous avons sélectionné 512 contextes d'Europarl dans lequel au moins une des langues utilise un parfait, quelle que ce soit la langue source (c'est-à-dire, quelle que soit la langue dans laquelle nous avons trouvé un parfait). Comme cet article ne traite que les différences entre l'anglais et le français, le nombre de contextes est réduit pour l'analyse. Pour les traductions du PP et du SP, 490 contextes sont inclus et pour les traductions du PC il y en a 416.

Préparation des données

Afin de répondre à la question de recherche, deux analyses seront faites. Premièrement, dans les 512 contextes, toutes les phrases ayant soit un PP soit un SP en anglais seront retirées et alignées avec leur traduction française. Comme l'hypothèse porte sur les deux temps ensemble (PP et SP) et n'indique pas d'effets individuels, nous ne ferons pas de distinction entre ces deux temps en anglais. Les traductions en français seront encodées. Toutes les traductions au PC auront le code '0', toutes les traductions au présent auront un '1' (voir le Tableau 3). De cette manière, les données comprennent deux groupes à comparer. Les

¹ « Les mêmes temps » dans cette phrase concerne seulement la forme du temps et non la sémantique.

traductions en d'autres temps que le PC ou le présent seront enlevées, vu que ces temps ne font pas partie de l'hypothèse. Cela résulte en 480 contextes pour l'analyse finale.

Tableau 3 : *Exemple de codage*

Id	Langue	Temps	Langue	Temps	Codes
11	En	PP	Fr	PC	0
1658	En	PP	Fr	Présent	1
92	En	SP	Fr	PC	0

Deuxièmement, tous les passés composés en français seront sélectionnés des 512 contextes et alignés avec leur traduction anglaise. Comme pour la première analyse, le PP et le SP seront mis dans un seul groupe, en accord avec l'hypothèse. Ce groupe sera comparé avec les traductions du PC au simple present afin de découvrir si le présent joue également un rôle dans les traductions du PC en anglais. Les autres traductions du PC en anglais dans notre ensemble de données, comprenant seulement deux cas du present perfect continuous, ne seront pas incluses dans l'analyse, vu que l'hypothèse ne porte que sur le PP et le SP et on ne sait pas encore quel rôle d'autres temps pourraient jouer dans la traduction du parfait en n'importe quelle langue. Cela résulte en 414 contextes, donnés le code '0' pour les traductions du PC en PP et en SP, et code '1' pour le simple present (de la même manière que pour la traduction de l'anglais au français montrée dans le Tableau 3).

Les matrices de codage seront alignées avec leur matrice de distance, qui sera adaptée selon les contextes sélectionnés pour chaque analyse. C'est-à-dire, les contextes non-utilisés seront enlevés de la matrice de distance selon leur identifiant (id). Afin d'analyser les données, nous ferons une Analyse de ressemblances ('Analysis of Similarities', ANOSIM ; Clarke, 1993) à l'aide de R (Chambers, 2008). Le procédé sera expliqué dans la section suivante.

R

R est une langue que l'on peut utiliser pour analyser des données. Avant de pouvoir commencer à travailler avec R, il faut installer deux choses. Premièrement, on a besoin de la langue même, à trouver sur <https://www.r-project.org/>. Après, il est nécessaire d'installer RStudio (RStudio Team, 2016; à trouver sur <https://www.rstudio.com/>), le logiciel qui sait lire

R. Après avoir installé ce qu'il faut, on peut commencer à écrire le code nécessaire pour faire l'analyse dans le cadre indiqué en bleu sur l'Image 2.

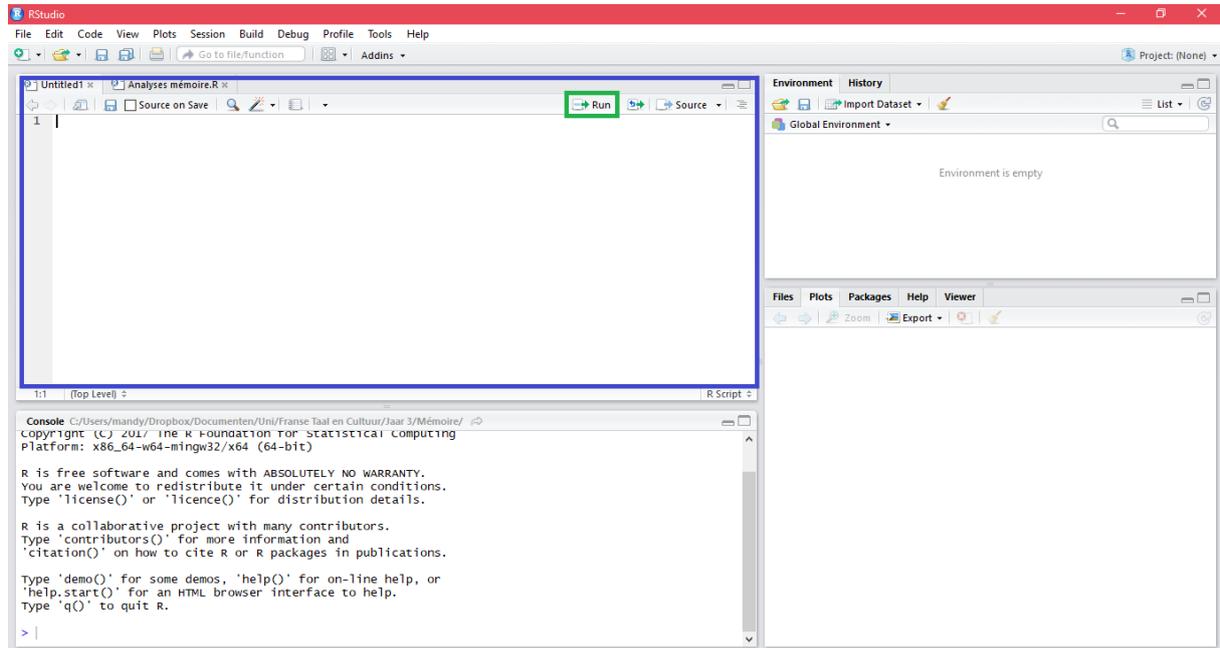


Image 2 : Ecran de R pour écrire le code.

Afin que R marche, il faut des *packages* qui contiennent différentes fonctions. Pour l'ANOSIM, on a besoin du *package* 'vegan' (Oksanen et al., 2017). Si l'on n'a jamais utilisé un *package* encore, il est nécessaire de l'installer. Pour faire cela, on écrit « `install.packages("vegan")` » dans le cadre en bleu (Image 2) et on appuie sur 'run' (dans le cadre vert sur l'Image 2). Après que le *package* est installé, il faut le charger dans le logiciel, ce que l'on fait en écrivant « `library(vegan)` » et en appuyant sur 'run'. Maintenant on est prêt à utiliser ANOSIM, mais il faut également un *package* pour charger les données dans RStudio. Comme les données de cette étude sont dans un fichier Excel, il nous faut le *package* 'readxl', que l'on peut installer et charger de la même manière que 'vegan'.

Pour charger les données, on écrit « `*nom que l'on veut donner aux données dans R* <- read_excel("*lieu où se trouve le fichier avec les données*", col_names = TRUE)`. Par exemple, comme on le voit dans l'Image 3, la matrice de distance dans R sera nommée 'matrix3' puisqu'on a écrit ce nom dans la première partie de la ligne 41, dans le cadre jaune, avant le « `<-` ». Cette 'flèche' a toujours comme résultat que l'on donne ce nom à un élément (par exemple un ensemble de données). Après que l'on appuie sur 'run', on voit 'matrix3' apparaître à droite, dans le cadre rouge sur l'Image 3. En ce qui concerne la partie « `col_names = TRUE` », cela est seulement nécessaire si on a une ligne dans les données

originales qui indique le nom des variables. Si la première ligne contient déjà les données nécessaires pour l'analyse, on n'a pas besoin de mettre cette partie.

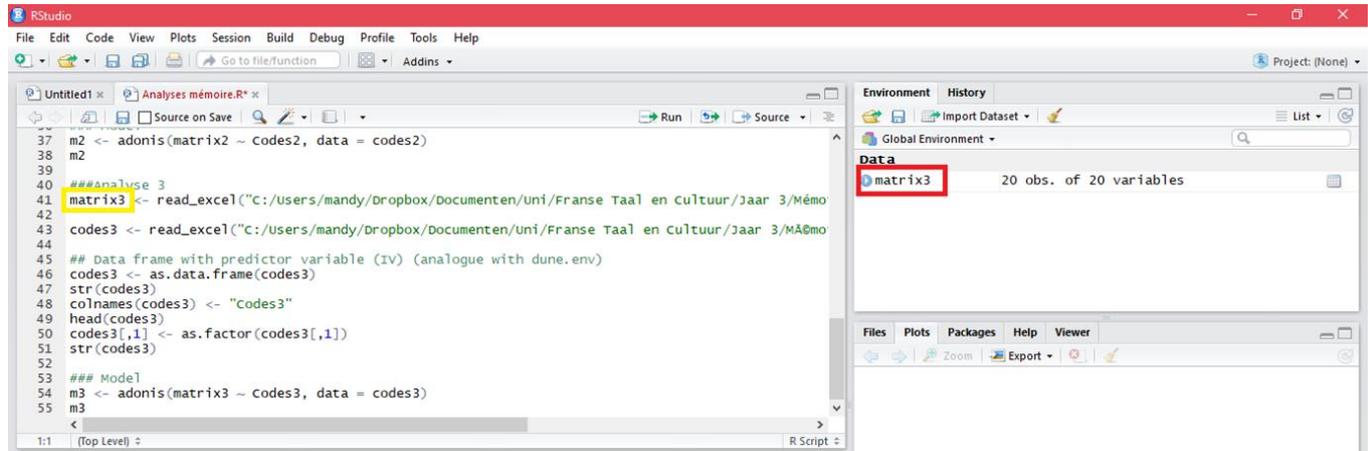


Image 3 : Ecran de R qui montre la matrice créée.

Maintenant que l'on a chargé la matrice de distance dans R, il nous faut la matrice de codage afin de les aligner pour l'analyse. Cela se fait de la même manière que pour la matrice de distance. On a besoin d'un fichier Excel avec les codes que l'on donne un nom dans R (nous l'avons appelé 'codes3'), on écrit où se trouve ce fichier et on indique éventuellement si la première ligne contient les noms des variables. Lorsque les codes sont chargés, il faut transformer la forme du fichier, puisqu'ANOSIM demande un certain type de fichier que l'on appelle un facteur (*factor* dans le code). Pour devenir un facteur, on doit d'abord en faire une trame de données (*data frame* dans le code). Le code que l'on doit utiliser est écrit ci-dessous.

```

codes3 <- as.data.frame(codes3)
str(codes3)
colnames(codes3) <- "Codes3"
head(codes3)
codes3[,1] <- as.factor(codes3[,1])
str(codes3)

```

Ici, on appelle la trame de données 'codes3', comme on l'a appelé en le chargeant dans R d'Excel. De cette manière, on remplace le fichier original et on ne garde que ce dont on a besoin pour l'analyse. Les 'str(codes)' y sont seulement pour résumer les données et pour vérifier si le remplacement a réussi. Dans la troisième ligne, on donne un nom à la colonne. Ceci est nécessaire pour faire référence aux données dans l'analyse (ce qui sera expliqué plus

tard). La ligne d'après montre les 6 premières valeurs de l'ensemble de données 'codes3', encore pour vérifier si on a réussi à donner un nom à la colonne. Après, on remplace la trame de données par un facteur que l'on donne le même nom, encore pour ne pas avoir trop de données dans le logiciel. Maintenant on a tout pour faire l'analyse.

L'analyse même n'exige qu'une ligne de code : « `m3 <- adonis(matrix3 ~ Codes3, data = codes3)` » On utilise la commande 'adonis', qui fait l'analyse sur 'matrix 3' avec les données sous 'Codes3', la colonne à laquelle on a donné ce nom dans l'étape précédente. On indique où R doit trouver les données sous la colonne 'Codes3', ce qui est dans 'codes3' (il faut faire attention aux majuscules et minuscules). On appuie sur 'run' et R crée une liste avec les résultats. On n'a qu'à les appeler en écrivant 'm3' et en cliquant sur 'run', et R donne les valeurs correspondant aux résultats. Cela forme le code entier suivant, qui peut être adapté selon les noms des données. Par exemple, si on a appelé notre matrice de distance « `matrice.xlsx` », on met ce nom au lieu de « `Matrix analyse 3.xlsx` ».

```
library(vegan)
library(readxl)

###Analyse 3

## Indiquer où se trouvent les données

matrix3 <- read_excel("Matrix analyse 3.xlsx", col_names = TRUE)
codes3 <- read_excel("Codes3.xlsx", col_names = TRUE)

## Faire un 'data frame' des données afin que R puisse les lire

codes3 <- as.data.frame(codes3)
str(codes3)
colnames(codes3) <- "Codes3"
head(codes3)
codes3[,1] <- as.factor(codes3[,1])
str(codes3)

### Analyse

m3 <- adonis(matrix3 ~ Codes3, data = codes3)
m3
```

En traduisant cette analyse en langue française, adonis teste si la distance entre les traductions est suffisante pour conclure que, dans la première analyse, le PP et le SP ne sont en effet pas toujours traduits par un PC, mais aussi régulièrement par un présent. C'est-à-dire, les traductions au présent forment leur propre groupe et sont différentes de celles au PC *et* assez

nombreuses pour pouvoir détecter cette différence. L'Image 4 montre cela dans un graphique fait avec des données hypothétiques. Ici, on voit clairement la différence entre les deux groupes de traductions.

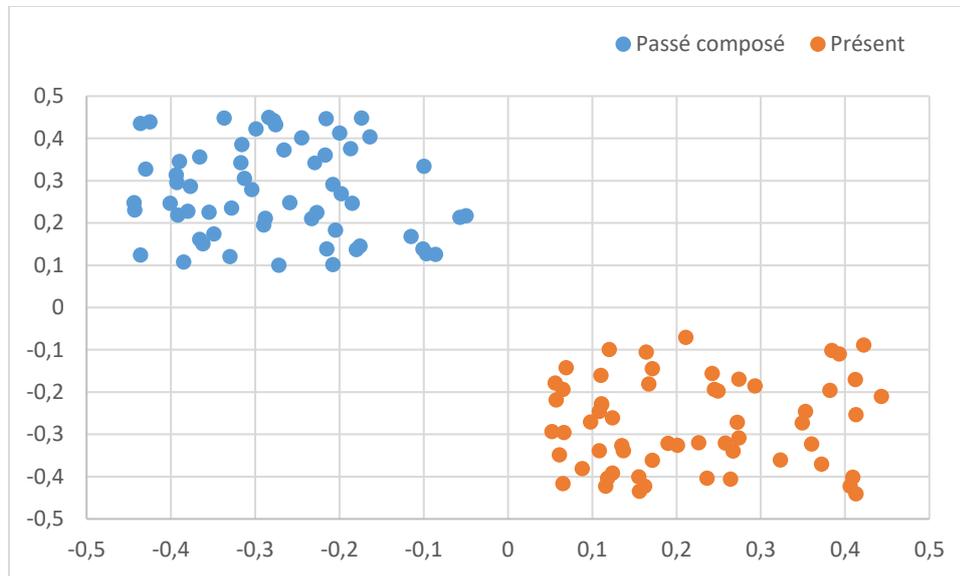


Image 4 : Exemple de points représentant les traductions du PP et SP au PC et au présent.

L'analyse de ressemblances qui nous dit si ce type de groupement est également présent dans les vraies données résultera en plusieurs valeurs, dont les valeurs les plus importantes sont la valeur F et la valeur p. La valeur F est basée sur la variance des groupes qui sont testés. La variance est une mesure qui montre la dispersion de valeurs dans un certain groupe. C'est-à-dire, à quel point les valeurs sont différentes les unes des autres dans un groupe. La valeur F, à son tour, montre la différence de la variance *entre* les différents groupes. Si la valeur F est grande, cela veut dire que cette différence entre variances est grande et que les groupes diffèrent l'un de l'autre. La valeur p indique la probabilité que l'on trouve une certaine valeur F si l'hypothèse nulle est vraie. C'est-à-dire, s'il n'existe pas de différence entre groupes. Donc, plus la valeur F est grande, plus la valeur p est petite, puisque la probabilité de trouver une valeur F si grande alors qu'il n'existe pas de différence est très petite. En général, on rejette l'hypothèse nulle (on suppose qu'il existe une différence) si la valeur p équivaut ou est plus petite que 5%. Pour notre recherche, cela veut dire que si le type de groupement comme dans l'exemple dans l'Image 4 est présent dans les vraies données, R nous donnera une valeur $p \leq .05$. Si cette différence n'est pas présente, R nous donnera une valeur $p > .05$.

Résultats

Statistiques descriptives

Dans l'ensemble des données concernant les traductions du PP et du SP, on voit que 465 sur 480 contextes sont traduits par un PC en français, ce qui représente 96.9% des cas. Quinze contextes (3.1%) sont traduits par un présent. En ce qui concerne l'ensemble des données avec les passés composés en français et leur traduction, 403 sur 414 contextes (97.3%) sont traduits par soit un PP, soit un SP et 11 contextes (2.7%) sont traduits par un simple present.

Analyse de ressemblances

La première analyse indique un effet significatif de la traduction de l'anglais au français, $F(1,478) = 19.327, p < .001$. Cela veut dire que la distance entre les groupes de traduction (PC versus présent) est suffisamment grande pour conclure que le présent joue également un rôle dans la traduction du PP et du SP. La première hypothèse est alors confirmée, ce qui veut dire que la première hypothèse nulle est rejetée.

La deuxième analyse concernant la traduction du PC en anglais indique également un effet significatif, $F(1,412) = 23.265, p < .001$. Cela implique que la distance entre le PP/SP et le simple present est tellement grande que l'on peut conclure qu'en anglais, le simple present joue aussi un rôle dans la traduction du PC en anglais. Ainsi, la deuxième hypothèse est également confirmée et la deuxième hypothèse nulle est rejetée.

Discussion

Cette étude a appliqué une nouvelle méthode pour analyser la sémantique du parfait de deux langues ; le français et l'anglais. Tandis que des études précédentes ont utilisé des données et des approches qualitatives, nous avons créé des données quantitatives qui ont été analysées avec une méthode nouvelle en linguistique : une analyse de ressemblances. En utilisant des contextes au PC/PP, leurs traductions et les distances entre les contextes, nous avons pu analyser la sémantique du parfait français par rapport à celle du parfait anglais. Les analyses ont donné une réponse à la question de savoir si le PC est l'équivalent du PP et du SP. Afin de répondre à cette question, nous avons fait deux analyses ; une qui prenait comme langue de base l'anglais, et une qui avait le français comme langue de base.

Premièrement, nous nous attendions à ce que le PP et le SP ne soient pas toujours traduits par un PC, mais que le présent joue également un rôle important dans cette traduction. Cette hypothèse a été confirmée. La distance entre le PC et le présent dans les traductions du PP et du SP est suffisamment grande pour conclure que le présent prend une place importante

dans la traduction du PP et du SP en français. Cela veut donc aussi dire que ces deux temps en anglais ne sont pas toujours traduits par un PC, comme prévu.

Deuxièmement, nous nous attendions à ce que le PC soit traduit non seulement par un PP ou un SP, mais aussi par un simple present. Cette hypothèse a également été confirmée. La distance entre les contextes est assez grande pour conclure que le PC se traduit régulièrement par ce temps et non seulement par un PP ou un SP.

Les résultats peuvent paraître incorrects si on considère le nombre limité de traductions en présent (première hypothèse) et en simple present (deuxième hypothèse). Pourtant, si la taille d'effet, représenté par la distance dans cette recherche, est très grande, il est possible de trouver des résultats pareils, même si le nombre d'observations est limité. En effet, si un effet est très grand, il est plus facile de le détecter. L'analyse même ne donne pas de tailles d'effets clairs, mais nous avons calculé la taille de la valeur d de Cohen (1992) à l'aide de la valeur F et le nombre d'observations dans les deux groupes de traduction. Nous avons trouvé une valeur $d = 1.156$ pour la première analyse et une valeur $d = 1.477$ pour la deuxième analyse. Les effets avec une valeur d plus large que 0.8 sont considéré comme grands. Cela implique que les distances testées dans cette recherche sont tellement grandes que l'on peut les détecter même si un des groupes de traductions ne contient pas beaucoup de contextes. Pourtant, il faut interpréter ces tailles d'effets avec prudence. La valeur d de Cohen peut être utilisé avec une analyse de la variance (ANOVA). Ce type d'analyse appartient à la même famille que l'analyse de ressemblances comme nous l'avons utilisée dans notre analyse, mais il n'est pas encore clair si on peut utiliser la valeur d de Cohen avec cette analyse. Néanmoins, cela donne une indication de la taille d'effet dans notre étude.

Vu que cette analyse est nouvelle dans le domaine de la linguistique, on peut se demander si elle convient pour notre genre de données. Il se peut, par exemple, que cette analyse donne des résultats significatifs pour chaque ensemble de données. Afin de vérifier si cela est le cas, nous avons fait une analyse supplémentaire. Nous avons créé des données selon la même méthode décrite avant. C'est-à-dire, nous avons retiré tous les contextes contenant un présent en français, ce qui a résulté en 19 contextes au total. Quinze de ces contextes (78.9%) sont traduits par un PP et 4 contextes (21.1%) par un simple present. L'hypothèse testée dans cette troisième analyse est que tous les présents sont traduits par un PP. Comme le présent représente, évidemment, aussi un temps qui parle des choses qui se passent en ce moment, nous nous attendons à ce que cette hypothèse soit rejetée et donnera une valeur p plus large que .05. Cela est en effet le cas, $F(1,17) = 2.289$, $p = .13$, $d = 0.9$. Cela

veut dire qu'il n'y a pas de distance significative entre les traductions du présent. C'est-à-dire, les traductions au simple present ne forment pas leur propre groupe et ne sont pas différentes de celles au PP. Les résultats indiquent que cette variabilité en traductions est dû au hasard. Pourtant, la taille d'effet semble être grande. Cela peut être dû au fait que le nombre de contextes contenant un présent en français est limité, à savoir 19 au total. Si on compare ce nombre aux nombres de contextes disponibles pour les deux analyses concernant le parfait dans cet article, à savoir 480 et 414, on peut comprendre que les résultats de cette analyse supplémentaire ne peuvent pas tout à fait être comparés aux résultats obtenus pour les deux autres analyses. De plus, nos données sont basées seulement sur le parfait, ce qui peut résulter en une déformation des résultats. Cette analyse devra alors être refaite quand on aura plus de données qui ne sont pas basées seulement sur un temps, mais qui comprend par exemple le présent et sa forme correspondante en d'autres langues afin d'obtenir des résultats plus fiables. Ce que prouve cette analyse surtout est qu'elle ne donne pas toujours un résultat significatif, des informations valables à savoir quand on essaye une nouvelle analyse.

Malgré la base des données limitée au parfait, cette étude a contribué à la compréhension de la sémantique de ce temps d'une façon quantitative. De plus, cette recherche a testé une méthode jamais utilisée encore en linguistique qui semble être adéquate pour les données de notre genre. Quand on aura plus de données, l'analyse doit être refaite pour vérifier si l'on trouve toujours le même effet. De plus, cette méthode peut être utilisée pour tester beaucoup d'autres hypothèses concernant la sémantique non seulement du parfait, mais aussi d'autres temps. Cela n'est pas restreint aux langues qui sont incluses dans ce projet jusque maintenant. Dans l'avenir, on peut ajouter des langues pour analyser les différences *et* les similarités. De cette manière, on peut trouver de plus en plus de pièces du puzzle sémantique et compléter notre connaissance sur l'utilisation des temps en différentes langues.

Ces nouvelles connaissances peuvent contribuer à l'amélioration des méthodes d'enseignement puisque plus on sait sur les différences et les similarités entre langues, plus facile ce sera d'enseigner cela aux autres. Aujourd'hui, la plupart des méthodes d'enseignement à l'école utilisent de l'exposition aux traductions plutôt littérales et des exemples prototypiques (Malt, Jobe, Li, Pavlenko, & Ameel, 2015). Cela aide à étendre le vocabulaire de la langue étrangère, mais ne résultera pas en un apprentissage complet. En intégrant les nouvelles notions obtenues de cette recherche et des recherches futures similaires, on peut résoudre ce problème et faciliter l'apprentissage d'une langue étrangère. Malt et collègues (2015) suggèrent que la lecture peut aider à mieux apprendre une langue, vu

que l'on est exposé non seulement aux mots au niveau de la traduction, mais aussi à la grammaire. Pourtant, il est important qu'un livre soit bien traduit afin d'en profiter le plus. Les nouvelles connaissances du projet *Time in Translation* peuvent également aider les traducteurs à améliorer la qualité et réduire la difficulté de leur travail.

Pour conclure, cette recherche nous a fourni des nouvelles connaissances sur la sémantique du parfait en utilisant des données et une méthode d'analyse quantitatives. Elle a montré que le présent joue également un rôle important dans la traduction du parfait, en français ainsi qu'en anglais. Ces notions peuvent être appliquées dans le domaine de l'enseignement et de la traduction, afin de faciliter l'apprentissage d'une langue étrangère.

Références

- Aijmer, K., & Altenberg, B. (2014). *English Corpus Linguistics*. New York: Routledge.
- Biber, D., Conrad, S., & Reppen, R. (1998). Introduction: Goals and methods of the corpus-based approach. In *Corpus Linguistics: Investigating Language Structure and Use* (pp. 1–18). Cambridge: Cambridge University Press.
- Chambers, J. M. (2008). *Software for Data Analysis Programming with R*. New York: Springer. <https://doi.org/10.1007/978-0-387-75936-4>
- Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology*, *18*, 117–143. <https://doi.org/10.1111/j.1442-9993.1993.tb00438.x>
- Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, *112*, 155–159.
- Dahl, Ö. (2014). The perfect map: Investigating the cross-linguistic distribution of TAME categories in a parallel corpus. In B. Szmrecsanyi & B. Wälchli (Eds.), *Aggregating Dialectology, Typology, and Register Analysis: Linguistic Variation in Text and Speech* (pp. 268–289). Berlin: De Gruyter.
- De Swart, H. (2007). A cross-linguistic discourse analysis of the Perfect. *Journal of Pragmatics*, *39*, 2273–2307. <https://doi.org/10.1016/j.pragma.2006.11.006>
- Egan, T. (2013). Tertia comparationis in multilingual corpora. In K. Aijmer & B. Altenberg (Eds.), *Advances in Corpus-Based Contrastive Linguistics: Studies in Honour of Stig Johansson* (pp. 8–24).
- Grisot, C., & Moeschler, J. (2014). How Do Empirical Methods Interact with Theoretical Pragmatics? The Conceptual and Procedural Contents of the English Simple Past and Its Translation into French. *Yearbook of Corpus Linguistics and Pragmatics 2014*, *2*, 7–33. https://doi.org/10.1007/978-3-319-06007-1_2
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. *Machine Translation Summit X*, 79–86. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.459.5497&rep=rep1&type=pdf>
- Lavid, J., Arús, J., & Moratón, L. (2013). Thematic variation in English and Spanish newspaper genres. In K. Aijmer & B. Altenberg (Eds.), *Advances in Corpus-based*

Contrastive Linguistics: Studies in honour of Stig Johansson (pp. 261–286).

- Loáiciga, S., Meyer, T., & Popescu-Belis, A. (2014). English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling. In *The Ninth Language Resources and Evaluation Conference* (pp. 1–8). Reykjavik.
- Malt, B. C., Jobe, R. L., Li, P., Pavlenko, A., & Ameel, E. (2015). What constrains simultaneous mastery of first and second language word use? *International Journal of Bilingualism*, 20, 684–699. <https://doi.org/10.1177/1367006915583565>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2017). *vegan: Community Ecology Package*. R package version 2.4-2. www.R-Project.org. Retrieved from <https://cran.r-project.org/package=vegan>
- Ritz, M.-E. (2012). Perfect Tense and Aspect. In R. I. Binnick (Ed.), *The Oxford Handbook of Tense and Aspect* (pp. 881–907). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780195381979.013.0031>
- RStudio Team, -. (2016). RStudio: Integrated Development for R. [Online] RStudio, Inc., Boston, MA URL [Http://www.Rstudio.com](http://www.Rstudio.com), RStudio, Inc., Boston, MA. <https://doi.org/10.1007/978-81-322-2340-5>
- Schaden, G. (2009). Present perfects compete. *Linguistics and Philosophy*, 32, 115–141. <https://doi.org/10.1007/s10988-009-9056-3>
- Simon-Vandenberghe, A.-M. (2013). English adverbs of essence and their equivalents in Dutch and French. In K. Aijmer & B. Altenberg (Eds.), *Advances in Corpus-Based Contrastive Linguistics: Studies in Honour of Stig Johansson* (pp. 83–102). Retrieved from [https://books.google.nl/books?hl=nl&lr=&id=6LzUeK8RquQC&oi=fnd&pg=PA83&dq=English+adverbs+of+essence+and+their+equivalents+in++Dutch+and+French&ots=IHVTvvCLo3&sig=qJ35IPCP9SE5dEymYEFdV1QGKaU#v=onepage&q=English adverbs of essence and their equ](https://books.google.nl/books?hl=nl&lr=&id=6LzUeK8RquQC&oi=fnd&pg=PA83&dq=English+adverbs+of+essence+and+their+equivalents+in++Dutch+and+French&ots=IHVTvvCLo3&sig=qJ35IPCP9SE5dEymYEFdV1QGKaU#v=onepage&q=English%20adverbs%20of%20essence%20and%20their%20equ)
- Van der Klis, M., Le Bruyn, B., & De Swart, H. (2017). Mapping the PERFECT via Translation Mining. *EACL 2017*, 497–502.
- Wälchli, B., & Cysouw, M. (2012). Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics*, 50, 671–710. <https://doi.org/10.1515/ling->

2012-0021

Yoon, Y. (2012). The English present perfect and simple past tense. *Linguistic Research*, 29, 485–513.