

Étude analytique de Google Translate des pronoms sujets de 3^e personne du persan vers le français selon la Théorie du Liage de Noam Chomsky (cas d'étude : « une collection d'écrits de Sadegh Hedayat »)

Recherche originale

Roya KHALASI*

Etudiante en Master de Traductologie française, Université Allameh Tabataba'i, Téhéran, Iran.

Marjan FARJAH*

Professeure-assistante, Université Allameh Tabataba'i, Téhéran, Iran.
(Date de réception : 21/10/2021; Date d'approbation : 16/01/2022)

Résumé

Le fait qu'au contraire du persan le français est une langue genrée implique des problèmes pour la traduction automatique des pronoms sujets de 3^e personne depuis le persan vers le français. Dans la présente étude, nous voudrions vérifier l'accord des pronoms sujets de 3^e personne dans le système de traduction en ligne de Google Translate persan-français de 50 extraits tirés d'*une collection d'écrits de Sadegh Hedayat*, choisis et compilés par Mohammad Baharlou (1994). Notre objectif principal est de savoir à quel point la traduction automatique des pronoms sujets de 3^e personne offerte par Google Translate est correcte ou incorrecte, et aussi quelles sont les raisons pour les traductions erronées ; nous supposons que le phénomène de pro-drop et l'ellipse linguistique des pronoms en persan ainsi que le manque de traduction interphrastique aux systèmes de traduction automatique impactent la qualité de traduction des pronoms étudiés. Par conséquent, nous allons mener une recherche basée sur l'analyse statistique-descriptive des données en nous appuyant sur la théorie du Liage chomskienne. D'après les résultats obtenus, Google a

* **E-mail:** khalasi_roya@atu.ac.ir (auteure responsable)

* **E-mail:** farjah@atu.ac.ir

Recherches en langue française, vol 2, n° 4, automne-hiver 2021, pp. 175-205.

traduit ces pronoms avec 44% de précision (y compris des pronoms bien traduits (40%) et certains pronoms non traduits (4%)); en effet, il a offert une traduction correcte dans les cas où l'antécédent était dans la même phrase que le pronom, ou s'il se plaçait dans la phrase précédant celle du pronom, il était traduit par un syntagme nominal masculin.

Mots-clés : traduction automatique, Google Translate®, pronom sujet, théorie du Liage de Noam Chomsky, Sadegh Hedayat.

Introduction

Étant donné que le persan est une langue sans genre ayant son origine dans la branche indo-iranienne de la famille des langues indo-européennes, et que le français est une langue genrée faisant partie des langues romanes de cette même famille, il n'y aura pas lieu de considérer ces deux langues comme « proches ». Pour cette raison, c'est-à-dire l'existence du genre grammatical dans une langue et son absence dans une autre, la traduction entre le persan et le français entraînera certainement des problèmes. À cet égard, la traduction automatique ne compte pas une exception ; or, elle crée des problèmes d'autant plus qu'elle se fait automatiquement. Néanmoins, la proximité des langues genrées en traduction ne garantit pas nécessairement le bon accord des genres dans la langue source et la langue cible : la traduction d'un pronom féminin de la langue source par un pronom féminin de la langue cible en l'absence d'un module capable d'identifier les antécédents des pronoms nous en est témoin (Scherrer et al., 2011 : 6).

Dans la présente étude, nous nous proposons de vérifier l'accord des pronoms sujets de 3^e personne dans Google Translate persan-français d'une collection d'écrits de Sadegh

*Hedayat*¹ (Mohammad Baharlou, 1994) en nous appuyant sur la théorie du Liage chomskyenne. En ce qui concerne la langue persane, elle possède 6 pronoms sujets de 3^e personne (dont 2 personnels et 4 démonstratifs) qui n'ont pas de genre : « او » (personnel, sing.²), « ایشان » (personnel, plur.³), « آن » (démonstratif, sing.), « آنها » (démonstratif, plur.). Au demeurant, le persan est une langue à sujet nul ou **pro-drop**, donc les pronoms sujets sont facultatifs⁴. Nous définissons le phénomène du pro-drop comme l'absence du pronom sujet, qui peut ne pas être exprimé en persan grâce à la richesse de la morphologie verbale de cette langue. En règle générale, les langues à sujet nul expriment l'accord de personne, de nombre et / ou de genre avec le référent sur le verbe, ce qui rend une phrase nominale de sujet redondante⁵. Dans le cadre des principes et paramètres, le sujet nul est contrôlé par le paramètre pro-drop, qui est activé ou désactivé pour une langue particulière⁶. C'est pourquoi les sujets nuls constituent environ une moitié des pronoms sujets de 3^e personne que nous allons analyser depuis nos extraits. Quant à la langue française, elle bénéficie de 20 pronoms sujets de 3^e personne (dont 4 personnels et 16 démonstratifs), ce qui veut dire 14 pronoms plus que le persan – et ceci principalement dû à l'existence du genre grammatical (masculin et féminin) en français. Ces pronoms sont : « il » (personnel, masc.⁷, sing.), « elle » (personnel, fém.⁸, sing.), « ils » (personnel, masc., plur.), « elles » (personnel, fém., plur.), « ce » et « ça » (démonstratif, neutre, sans nombre), « ceci » et

¹ Titre persan : « مجموعه‌ای از آثار صادق هدایت »

² L'abréviation de « singulier » employée par les auteurs.

³ L'abréviation de « pluriel » employée par les auteurs.

⁴ https://fr.qaz.wiki/wiki/Persian_grammar

⁵ https://fr.qaz.wiki/wiki/Null-subject_language

⁶ https://fr.qaz.wiki/wiki/Null-subject_language

⁷ L'abréviation de « masculin » employée par les auteurs.

⁸ L'abréviation de « féminin » employée par les auteurs.

« cela » (démonstratif, neutre, sing.), « celui », « celui-ci » et « celui-là » (démonstratif, masc., sing.), « celle », « celle-ci » et « celle-là » (démonstratif, fém., sing.), « ceux », « ceux-ci » et « ceux-là » (démonstratif, masc., plur.), « celles », « celles-ci » et « celles-là » (démonstratif, fém., plur.). En dépit de la multiplicité des pronoms sujets persans et français, les extraits que nous allons analyser ne contiennent que des pronoms « او آن » et « آن‌ها » dans le corpus persan, et des pronoms « il », « elle », « ils », « elles » et « celui » dans la traduction française.

Le corpus que nous avons choisi en vue d'étude et d'analyse de la traduction automatique des pronoms sujets de 3^e personne est un livre persan intitulé *une collection d'écrits de Sadegh Hedayat*, choisis et compilés par Mohammad Baharlou, qui a paru en 1994 et dont nous avons tiré 50 extraits où surgissent des pronoms sujets de 3^e personne.

Le traducteur automatique que nous avons utilisé en vue de la traduction des extraits du corpus persan vers le français est Google Translate®, qui serait abordé de façon plus élaborée dans la troisième section. Outre ceci, il y a également d'autres traducteurs automatiques en ligne à l'accès libre et gratuit comme Systran, Microsoft Bing, Yandex Traduction, DeepL, et Reverso, dont Systran est le plus ancien, tandis que DeepL est le plus récent. Parmi ces systèmes de traduction, Systran et Yandex, tout comme Google, offrent des traductions depuis ou vers le persan. Pourtant, la raison pour laquelle nous avons préféré Google au Systran et au Yandex est qu'en fonction de la précision et de l'aspect « humain », Google offre de meilleures traductions que ses adversaires, ce que nous avons vérifié au préalable.

La méthode par le biais de laquelle nous allons analyser les pronoms sujets de 3^e personne dans les extraits persans ainsi que dans leur traduction française, c'est la théorie du Liage proposée

par le grand linguiste américain, Noam Chomsky. La théorie du Liage s'actualise à travers la **résolution des anaphores** et la **résolution des pronoms**. La première s'intéresse à trouver l'antécédent des expressions munies d'un contenu référentiel, alors que la deuxième cherche à trouver l'antécédent des pronoms non nécessairement accompagnés d'un contenu référentiel. En ce qui concerne la méthode appliquée à l'analyse des données, nous en parlerons en détail dans une section séparée.

Notre objectif principal dans cette étude est de savoir à quel point la traduction automatique des pronoms sujets de 3^e personne offerte par Google Translate est correcte ou incorrecte, et aussi quelles sont les raisons pour les traductions erronées. Ainsi, nous allons répondre aux questions suivantes au cours de cette étude :

1- Combien la traduction Google persan-français des pronoms sujets de 3^e personne est-elle précise ?

2- Dans quelles conditions Google offre-t-il une traduction correcte de ces pronoms ?

3- Est-ce qu'il existe une différence en termes du nombre de ces pronoms dans le texte-source et le texte-cible ? Pour quelles raisons ?

Voici les réponses hypothétiques que nous pourrions donner à ces questions :

1- Environ 50% des traductions persan-français des pronoms sujets de 3^e personne faites par Google font preuve de précision. En ce qui concerne l'inexactitude de la moitié restante, il faudra l'imputer à l'incapacité des traducteurs automatiques à faire une analyse interphrastique.

2- Google offre une traduction correcte de ces pronoms tant que l'antécédent soit dans la même phrase que le pronom ou,

dans le cas où il se situe dans la phrase précédant celle du pronom, qu'il soit traduit par un SN masculin.

3- Oui, car la structure grammaticale du français exige que l'on traduise vers ce dernier certains de ces pronoms sujets persans omis en vertu de l'ellipse linguistique. En outre, Google a son propre algorithme de traduction qui se veut partiellement responsable de la différence du nombre de ces pronoms dans le texte-source et le texte-cible.

Par la suite, nous avons l'intention de dresser l'état actuel des recherches pertinentes dans le domaine de TA. Puis dans la troisième section, nous allons aborder le corpus et la méthode employés en détail. Enfin, dans la quatrième section, nous cherchons à analyser et à commenter simultanément nos données en nous servant d'exemples sélectionnés et de tableaux.

État de recherche

Weissenbacher & Nazarenko (2007) ont proposé une nouvelle approche pour identifier les pronoms anaphoriques et leurs antécédents, qui repose sur les réseaux bayésiens et qui permet de combiner au sein d'une même représentation ces deux types d'informations hétérogènes et complémentaires. De ce point de vue, ils ont comparé les performances du réseau bayésien à celles des systèmes de l'état de l'art, sur un problème difficile du TAL¹, celui de la résolution anaphorique.

Russo (2010) a abordé la problématique de la traduction automatique des pronoms clitiques en se focalisant sur la traduction de l'italien vers le français et en comparant les résultats obtenus par trois systèmes : Its-2, basé sur un analyseur syntaxique profond ; Babelfish, basé sur des règles

¹ Abréviation pour « Traitement Automatique des Langues ».

linguistiques ; et Google Translate®, caractérisé par une approche statistique.

Dans une recherche menée par Scherrer et al (2011), le système de TA a fait l'objet d'une évolution manuelle de la traduction des pronoms pour cinq paires de langues et sur deux corpus au moyen de la résolution des ambiguïtés d'analyse et de la résolution des anaphores pronominales. Les résultats obtenus montrent que les pourcentages d'erreurs peuvent atteindre 60% selon la paire de langues et le corpus.

Dans une étude inter-langues, Russo et al. (2011) ont décrit la distribution des pronoms selon le style de texte et selon la langue en utilisant l'analyseur syntaxique Fips pour l'étiquetage morpho-syntaxique des pronoms. Effectivement, ils ont évalué une précision moyenne de plus de 95%.

Russo et Wehrli (2011) ont discuté la méthodologie utilisée par Its-2, un traducteur à base de règles, pour la traduction automatique des pronoms clitiques, en se focalisant particulièrement sur les séquences clitiques, pour la traduction automatique entre le français et l'anglais.

Russo (2013), dans le cadre d'une recherche consacrée à l'analyse linguistique comparative italien-français dans le domaine de la traduction automatique, a présenté deux projets de recherche autour d'Its-2 ; le premier est un projet universitaire visant à développer de même qu'à améliorer un traducteur automatique multilingue basé sur des règles appelée Its-2, alors que le deuxième est axé sur la traduction automatique de langues voisines. L'objectif de cette étude a été de comprendre comment améliorer la traduction automatique entre des langues de la même famille, et de souligner l'importance des informations linguistiques afin d'obtenir des traductions grammaticalement correctes.

Loaiciga Sanchez (2013) a mis en œuvre un premier prototype d'une méthode inspirée de la théorie du Liage chomskyenne pour l'interprétation des pronoms dans le but d'améliorer la traduction automatique à base de règles des pronoms personnels entre l'espagnol et le français.

Corpus et méthode

D'abord nous avons choisi *une collection d'écrits de Sadegh Hedayat* (Mohammad Baharlou, 1994) comme corpus en vue d'étude et de recherche parce qu'en premier lieu, ce n'était pas écrit à l'orthographe populaire (en persan : شکسته‌نویسی) et que Google était donc capable d'en analyser la structure morpho-syntaxique. Bien plus, en tant qu'un livre dans le genre narratif appelé « la nouvelle », il profite certainement d'une grande quantité de pronoms sujets, dont ceux de 3^e personne, ce dont nous avions besoin pour mener cette recherche. Enfin, une raison décisive qui nous a poussées à opter absolument pour ces nouvelles est qu'elles sont écrites par Sadegh Hedayat, un grand écrivain iranien connu pour son écriture expressive et explicite et sa connaissance lexicale riche.

D'un nombre total de 30 nouvelles (pp. 35-404), nous avons lu les 20 premières (pp. 35-257) pour en tirer 50 extraits au total. Notre méthode d'échantillonnage a été non-probabiliste, exigeant à choisir des extraits de façon à retenir la meilleure diversité des pronoms sujets de 3^e personne possible, ce qui ne nous a pas épargnées pour autant de la dominance du pronom « او » en fonction de fréquence de l'occurrence. Entre-temps, nous avons évité de sélectionner dans une même nouvelle les extraits manifestant un algorithme de placement analogue pour les pronoms et leurs antécédents, ou encore les extraits où

l’antécédent était un nom propre iranien, traditionnel et autochtone, et concernant ce dernier, puisque naturellement Google Translate n’est pas à même de l’identifier correctement et de le remplacer par le bon pronom sujet. Mais la raison pour laquelle nous nous sommes contentées de 50 extraits, c’est qu’ils ont pu répondre à nos questions, étant donné la présence des paradigmes répétitifs à la fois dans le texte original et dans sa traduction Google, à partir desquels nous avons pu déduire des règles généralisables.

Ensuite, comme il a été indiqué plus haut, nous avons assigné la tâche de traduire nos extraits à Google Translate. Avant tout, il faudrait savoir qu’il existe deux principales approches en TA : approche linguistique (à base de règles syntaxiques) et approche statistique (Russo, 2013 : 10), cette dernière résultant du travail d’IBM® lors du projet Candide, à la fin des années 80 (Koehn, 2009 : 81). Google Translate est système de traduction en ligne à base statistique qui s’appuie sur une approche statistique exploitant des techniques d’apprentissage statistique appliquées au modèle de traduction (Russo, 2013 : 10-11) : des milliards de mots provenant de corpus monolingues ou de corpus alignés – créés à partir de traductions réalisées par des traducteurs professionnels – sont introduits dans le système (Russo, 2010 : 3). Déjà en 1949, Warren Weaver a suggéré que l’exploitation des techniques statistiques pouvait déboucher sur le renforcement des algorithmes de traduction (Ducotterd, 2017 : 22). De surcroît, compte tenu de sa nature statistique, Google Translate se sert d’un algorithme analogique pour trouver parmi les exemples contenus dans son corpus, celui qui a la proximité maximale avec l’entrée en langue source; après avoir trouvé le meilleur exemple de référence, le système substitue, modifie et adapte les séquences différentes afin qu’elles se rapprochent de l’entrée en langue source (Carl, 2003 : 4). Depuis 2016, Google fournit des traductions basées sur l’IA via un système nommé

Google Neural Machine Translation (GNMT) qui utilise un réseau de neurones artificiels pour augmenter la fluidité et la précision de Google Translate. Les modèles de traduction automatique neuronale (NMT) sont formés à l'aide d'exemples de phrases et de documents traduits, qui sont généralement collectés sur le Web public. Par rapport à la traduction automatique basée sur des phrases, la NMT s'est avérée plus sensible à la qualité des données¹. En tant que tel, Google a remplacé l'ancien système de collecte de données par un nouveau mineur de données qui se concentre davantage sur la précision que sur le rappel, ce qui permet la collecte de données de formation de meilleure qualité à partir du Web public.

Enfin, nous avons eu recours à la théorie du Liage chomskyenne pour analyser nos données, que nous allons illustrer à présent. Formulée par Chomsky dans le cadre de la théorie standard étendue (la 2^e version de la théorie de la grammaire générative), la théorie du Liage a pour objectif de décrire et d'analyser la distribution des pronoms et des mots lexicaux dans la phrase (Daas, 2013 : 14). Tout d'abord, il est à noter que les anaphores, les pronoms et les expressions-R – dont nous n'aborderons que les deux premiers – constituent des éléments importants pour la théorie du Liage et sont tous représentés comme un SN dans la phrase. Alors, nous parlons de la dépendance et de l'indépendance des SN (Van Rees, 2007 : 8). En vérité, les **anaphores** sont des mots qui, selon ladite théorie, dépendent d'un autre SN au niveau référentiel qui s'appelle « l'antécédent » (Van Rees, 2007 : 8). En français, l'antécédent est un groupe de mots (SN, pronom personnel ou démonstratif) auquel se rapporte un pronom relatif ou un **pronon**

1 Google AI Blog: Recent Advances in Google Translate (googleblog.com)

anaphorique¹, tandis qu'en persan, l'antécédent ne peut être qu'un SN. Alors, dans les deux langues, l'antécédent est remplacé par le pronom². En guise d'exemple :

Valérie est absente : elle est malade. → *Valérie est l'antécédent du pronom personnel elle*³.

Sur ce point, les études menées par Le Nagard et Koehn (2010) ainsi que par Hardmeier et Federico (2010) au sujet de la résolution des anaphores proposent d'annoter les pronoms dans la langue source avec des informations sur leurs antécédents dans la langue cible (Loaiciga Sanchez, 2013 : 684).

Du reste, l'algorithme des traducteurs automatiques à base de règles pour la résolution des anaphores pour chaque pronom trouvé consiste en 2 étapes suivantes :

1. Vérifier la nature du pronom
2. Retenir l'antécédent (Loaiciga Sanchez, 2013 : 686-687)

« Dans la théorie du Liage, nous observons également les **pronoms**. La différence entre les pronoms et les anaphores est le fait que les pronoms ne sont pas obligatoirement dépendants d'un autre élément dans la phrase. Aussi nous voyons la différence en remplaçant l'anaphore par un pronom (Van Rees, 2007 : 9) » :

- a. Il aimeraït venir demain.
- b. Jacques₁ le₂ voit.

Ces phrases sont grammaticales, donc les pronoms n'ont pas toujours besoin d'un antécédent dans la phrase. Il y a également des phrases dans lesquelles le pronom est lié, c'est-à-dire que le

¹ <https://la-conjugaison.nouvelobs.com/regles/grammaire/l-antecedent-232.php>

² <https://www.jerevise.fr/antecedent-pronom-relatif.html>

³ <https://la-conjugaison.nouvelobs.com/regles/grammaire/l-antecedent-232.php>

pronome a un antécédent dans la phrase, comme dans (c) (Van Rees, 2007 : 9) :

c. Jacques₁ a dit qu'il₁ aurait aimé venir.

Après avoir constaté les anaphores et les pronoms, qui jouent un rôle dans la théorie du Liage, nous souhaiterons évoquer les principes de la théorie du Liage relatifs à ces deux catégories différentes, formulés par Chomsky comme suit :

Principe A : Les anaphores doivent être liées dans leur domaine de liage

Principe B : Les pronoms doivent être libres dans leur domaine de liage (Kerstens, 1994 : 218)

Tout de suite dans la section suivante, nous allons analyser et commenter nos données en nous référant aux deux catégories importantes de la théorie du Liage (pronoms sujets anaphoriques de 3^e personne, pronoms sujets de 3^e personne) tout comme à leurs principes respectifs.

Analyse des données

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرستال جامع علوم انسانی

Avant de nous lancer dans l'analyse d'exemples, il nous semble indispensable d'indiquer les statistiques générales du texte original et de sa traduction (voir tableau 1).

Tableau 1 – Nombre d'« extraits », de « phrases » et de « pronoms sujets de 3^e personne » présents dans le corpus persan et dans sa traduction française effectuée par Google Translate®

Pronoms sujets de 3 ^e personne en corpus persan (y compris des sujets nuls)			Pronoms sujets de 3 ^e personne en traduction française				
او	آن	آن‌ها	Il	Elle	Ils	Elles	Celui
38 (dont 21 sujets nuls)	3 (sujets nuls)	11 (dont 4 sujets nuls)	45	12	10	1 Prenom de 3 ^e pe (y compris sujets nuls en cas de persan)	2
Corpus	Extraits			Phrases			
Persan		50	95			52 (dont 28 sujets nuls)	
Google Translate en français		50	94			70	

En ce qui concerne l'excès des pronoms sujets de 3^e personne en traduction française sur ceux du corpus persan, nous en énumérerons les raisons en détail plus bas.

Dans le tableau suivant, nous avons marqué la distribution numérique de chacun des pronoms sujets de 3^e personne du persan et du français selon la fréquence de leurs occurrences dans les extraits (voir tableau 2).

Tableau 2 – Nombre de chaque pronom sujet de 3^e personne présent dans le corpus persan et dans sa traduction française effectuée par Google Translate®

Comme nous avons illustré plus haut, à la différence du français, le persan se permet d'un phénomène appelé « pro-drop » qui privilégie des sujets nuls. Du moins c'est que ces sujets nuls, malgré leur invisibilité, sont grammaticalement présents dans les phrases et comptent alors parmi l'ensemble des pronoms sujets. Ledit phénomène est tout à fait commun en persan, si bien qu'il s'est produit dans plus de la moitié (54%) des cas d'occurrence des pronoms sujets de 3^e personne dans nos extraits. L'exemple ci-dessous montre un cas d'occurrence d'un sujet nul de 3^e personne singulier dans un extrait persan et sa traduction Google en français. (Dans tous les exemples que nous avons sélectionnés parmi les 50 extraits afin de les étaler ici-bas, les pronoms sujets de 3^e personne sont en **gras**, alors que leurs antécédents sont soulignés ; « Ø » est le symbole du sujet nul ; et des nombres écrits entre parenthèses devant chaque pronom sujet s'accordent en corpus persan et français. De plus, nous n'avons étudié que les pronoms dont l'antécédent était présent dans le même extrait que le pronom, afin que Google Translate ait assez de données pour traduire correctement des pronoms.)

1- خداد مردی شخصت ساله بود. Ø (1) استخوان بندی درشتی داشت.
(داستان شماره 14: لاله، ص 155)

GT¹ : *Khodadad était un homme dans la soixantaine. Il (1) avait un grand squelette.*

1 pronom bien traduit

Pronom persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « مردی شخص ساله » → Tr. fr.² : « un homme dans la soixantaine » (SN, masc., sing.) | Place de l'antécédent : la phrase précédant celle du pronom | Pronom français (1) : « il » (personnel, masc., sing.)

Or, la raison principale pour le surplus des pronoms sujets de 3^e personne en français par rapport au persan reste dans la nécessité de traduire certains pronoms persans omis en vertu de l'**ellipse linguistique** vers le français.

« L'ellipse linguistique peut être divisée, un peu schématiquement, en deux types : elle peut être dite « grammaticale » quand les mots qui manquent peuvent être suppléés grâce à la connaissance des règles syntaxiques de la langue (par ex. *Sommes bien arrivés* > *Nous sommes bien arrivés*) ; elle peut être qualifiée de « situationnelle » quand c'est le co-texte (les énoncés adjacents) ou le contexte (la situation) qui permet de suppléer ce qui manque : *Hier après-midi, j'ai peint* > *Hier après-midi, j'ai peint des tableaux / les murs de mon appartement*, selon qu'il s'agit d'un artiste-peintre ou de quelqu'un qui est en train d'emménager. »³

L'exemple ci-dessous démontre bien un cas de la traduction française d'un pronom démonstratif sujet de 3^e personne singulier omis par le biais de l'ellipse grammaticale en persan.

¹ L'abréviation de « Google Translate » employée par les auteurs.

² L'abréviation de « Traduction française » employée par les auteurs.

³https://www.fabula.org/atelier.php?Ellipse_linguistique_et_ellipse_narratologique

2- ولی روح شخصی هر کسی با تنش می‌میرد، چون محتاج به خوراک است و بعد از تن نمی‌تواند زنده بماند. (داستان شماره 17: گجسته دز، ص 185-186)

GT: *Mais l'âme personnelle de chacun meurt de stress, car elle (1) a besoin de nourriture et ne peut survivre après le corps.*

1 pronom bien traduit

Pronom français (1) : « elle » (personnel, fém., sing.) (Quoique pronom n'ait pas de correspondant en persan, il est la traduction du pronom « آن » omis par le biais de l'ellipse grammaticale en persan, qui se trouve traduit en vertu des règles syntaxiques du français.) | Antécédent : « روح شخصی » → Tr. fr. : « l'âme personnelle » (SN, fém., sing.) | Place de l'antécédent : la même phrase que le pronom

Toutefois, il n'est pas à oublier qu'une partie du surplus des pronoms français par rapport au persan est directement associée aux préférences algorithmiques de Google Translate. Bien sûr, l'absence de pronoms ne se limite pas au persan : des fois, nous avons vu la non traduction de certains pronoms du corpus persan vers le français, en raison des dites préférences algorithmiques. L'exemple (3) contient un pronom en français (pronome français (3)) qui n'a pas de correspondant en extrait persan, tandis que l'exemple (4) comprend un pronom personnel de 3^e personne singulier en persan qui n'a pas été traduit en français.

3- خانم این مرد یک تو به من نگفت... شوهر بیچاره‌ام. Ø (1) ورپرید، او (2) نمرد، او را کشتند. (داستان شماره 8: مردخورها، ص 87)

GT: *Madame, cet homme ne m'a pas dit tu ... mon pauvre mari. Il (1) a sauté, il (2) n'est pas mort, il (3) a été tué.*

3 pronoms bien traduits

Pronom persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « شوهر بیچاره‌ام » → Tr. fr. : « mon pauvre mari » (SN, masc., sing.) | Place de l'antécédent : la phrase précédent celle du pronom | Pronom français (1) : « il » (personnel, masc., sing.)

Pronom persan (2) : « او » (personnel, sing.) | Antécédent : « شوهر بیچاره‌ام » (SN, masc., sing.) | Place de l'antécédent : la phrase précédent celle du pronom | Pronom français (2) : « il » (personnel, masc., sing.)

Pronom français (3) : « il » (personnel, masc., sing.) (En raison du choix algorithmique de Google Translate, ce pronom ne correspond pas à celui du corpus persan, car la dernière proposition de l'extrait persan est active, tandis que celle de la traduction française est passive.) | Antécédent : « mon pauvre mari » (SN, masc., sing.) | Place de l'antécédent : la phrase précédent celle du pronom

4- یکی از برادرهایش که زنده مانده بود او (1) هم لال و احمق بود تا اینکه دو سال پیش مرد. با خوش می‌گفت: «شاید آن‌ها (2) خوشبخت بوده‌اند!» (داستان شماره 6: داود گوژپشت، ص 74)

GT: *Un de ses frères survivants était également stupide et stupide jusqu'à sa mort il y a deux ans. « Peut-être qu'ils (2) ont eu de la chance », se dit-il.*

1 pronom non traduit (pronome persan (1)) | Cause de non-traduction : choix algorithmique de Google Translate ; 1 pronom bien traduit (pronome persan (2))

Pronome persan (1) : « او » (personnel tonique, sing.) | Antécédent : « یکی از برادرهایش که زنده مانده بود » → Tr. fr. :

« Un de ses frères survivants » (SN, masc., sing.) | Place de l'antécédent : la même phrase que le pronom

Pronom persan (2) : « آن‌ها » (démonstratif, plur.) | Antécédent : « برادرهایش که زنده مانده بودند » → Tr. fr. : « ses frères survivants » (SN, masc., plur.) | Place de l'antécédent : la phrase précédant celle du pronom | Pronom français (2) : « ils » (personnel, masc., plur.)

Les raisons énumérées en haut pour le surplus des pronoms sujets de 3^e personne en traduction française par rapport au corpus persan sont étalées dans le tableau ci-dessous, suivies du nombre des pronoms appartenant à chaque catégorie (voir tableau 3).

Raisons pour l'excès des pronoms sujets de 3 ^e personne en traduction française sur ceux du corpus persan		
1. Nécessité de traduire des sujets nuls du persan en français	2. Nécessité de traduire certains pronoms persans omis en vertu de l'ellipse linguistique vers le français	3. Préférences algorithmiques de Google Translate
28	14	4

Total : 46 (surplus des pronoms sujets de 3^e personne en français par rapport au persan) / 98 (tous les pronoms sujets de 3^e personne en français)
≈
47%

Tableau 3 – Nombre de pronoms français traduits en surplus de pronoms persans, distribués selon des raisons différentes

D'après notre analyse statistique et grammaticale, 40% des pronoms sujets de 3^e personne des extraits persans ont été bien traduits dans la langue française. Google Translate a correctement traduit des pronoms dans le cas où l'antécédent se plaçait dans la même phrase que le pronom (ici, il est question de la résolution anaphorique), ou s'il se trouvait dans la phrase précédant celle du pronom (ici, il est question de la résolution pronomiale), son équivalent français choisi par Google Translate était sans doute un SN masculin. Par conséquent, au cas qu'un antécédent féminin se situait dans la phrase précédant celle du pronom, Google n'a pu le remplacer par le bon pronom. Cette incapacité de Google Translate à l'égard de la résolution pronomiale est causée par l'impossibilité de faire une analyse interphrastique dans un système de TAL. Dans les exemples (5) et (6), nous voyons le premier cas d'une traduction correcte des pronoms, où du fait que Google Translate a bien identifié les antécédents (« deux filles » et « une statue délicate et fragile »), le pronom sujet démonstratif pluriel « آن‌ها » a été bien traduit par « elles », ainsi que le sujet nul singulier « او » par « elle ».

5- در این بین دیدم رفیقم باند شد و به دو نفر دختر که به ما نزدیک شدند دست داد و مرا معرفی کرد، آن‌ها (1) هم آمده پهلوی ما روی لبه باند سد نشستند. (داستان شماره 2: مادلن، ص 41)

GT: *En attendant, j'ai vu mon ami se lever et serrer la main de deux filles qui se sont approchées de nous et*

m'ont présenté, et elles (1) sont venues s'asseoir à côté de nous sur le bord haut du barrage.

1 pronom bien traduit

Pronom persan (1) : « آن‌ها » (démonstratif, plur.) | Antécédent : « دو نفر دختر » → Tr. fr. : « deux filles » (SN, fém., plur.) | Place de l'antécédent : la même phrase que le pronom | Pronom français (1) : « elles » (personnel, fém., plur.)

6- از دور که روی تخت دراز کشیده بود مانند مجسمه ظریف و شکننده‌ای به نظر می‌آمد که انسان جرئت نمی‌کرد او را لمس بکند، از ترس این که مبادا Ø (1) کنفت و پیزمرده بشود. (داستان شماره 18: س.گ.ل.ل، ص 199-198)

GT: *Allongée de loin sur le lit, elle ressemblait à une statue délicate et fragile que l'homme n'osait pas toucher, de peur qu'elle (1) ne se fane et se fane.*

1 pronom bien traduit

Pronom persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « مجسمه ظریف و شکننده‌ای » → Tr. fr. : « une statue délicate et fragile » (SN, fém., sing.) | Place de l'antécédent : la même phrase que le pronom | Pronom français (1) : « elle » (personnel, fém., sing.)

Le deuxième cas d'une traduction correcte des pronoms se reflète dans l'exemple suivant où, comme l'antécédent français était masculin, Google Translate a correctement traduit les pronoms par défaut. En fait, il semble que peu importe le genre grammatical de l'antécédent, s'il se place dans la phrase précédant celle du pronom, Google Translate va traduire le pronom persan par un pronom masculin en français. Ce choix

apparemment systémique se justifie en vertu de l'incapacité des traducteurs automatiques à effectuer une analyse interphrastique.

7- روی پرده مردی ظاهر شد که پشت میز بزرگی نشسته بود، جلو او چند لوله شیشه و دواهای مختلف بود. اول مثل این بود که Ø (1) کاغذی را نگاه می‌کند، بعد سرش را بلند کرد و با لحن طبیعی و چهره تودار گفت: (...). (داستان شماره 18: س.گ.ل.ل، ص 207)

GT: *Un homme est apparu sur l'écran, assis à une grande table, avec plusieurs tubes de verre et divers médicaments devant lui. Au début, c'était comme s'il (1) regardait un morceau de papier, puis il (2) leva la tête et dit d'un ton naturel et d'un visage plein: (...).*

2 pronoms bien traduits

Pronom persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « مردی » → Tr. fr. : « un homme » (SN, masc., sing.) | Place de l'antécédent : la phrase précédant celle du pronom | Pronom français (1) : « il » (personnel, masc., sing.)

Pronom français (2) : « il » (personnel, masc., sing.) (Ce pronom est considéré comme la traduction obligée du pronom « او » dans l'extrait persan omis dû à l'ellipse grammaticale.) | Antécédent : « un homme » (SN, masc., sing.) | Place de l'antécédent : la phrase précédant celle du pronom

L'exemple ci-dessous indique le cas d'une mauvaise traduction du pronom où un antécédent féminin se trouve dans la phrase précédant celle du pronom. Nous remarquons que du fait de l'impossibilité de l'analyse interphrastique en TAL, le système a systématiquement remplacé l'antécédent féminin par un pronom sujet masculin.

8- ولی نمی‌دانم چرا مرگ ناز کرد؟ Ø (1) چرا نیامد؟ چرا نتوانستم بروم پی کارم آسوده بشوم؟ (داستان شماره 3: زنده به گور، ص 56)

GT: *Mais je ne sais pas pourquoi la mort était mignonne? Pourquoi n'est-il (1) pas venu? Pourquoi ne pourrais-je pas aller travailler confortablement?*

1 pronom mal traduit | Cause : erreur d'identification de l'antécédent

Pronom persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « مرگ » → Tr. fr. : « la mort » (SN, fém., sing.) | Place de l'antécédent : la phrase précédant celle du pronom | Pronom français (1) : « il » (personnel, masc., sing.) | Remplacement proposé : « elle » (personnel, fém., sing.)

Des vérifications répétées nous ont prouvé que 54% des pronoms sujets de 3^e personne des extraits persans ont été mal traduits dans la langue française. La raison pour la mauvaise traduction des pronoms dans quasiment tous les cas est que Google Translate n'a pu identifier leur antécédent. Les exemples suivants présentent des cas de mauvaise traduction des pronoms sujets de 3^e personne.

9- جلو چشم گوی های سرخ و بنفش چرخ میزد، می رقصید. **(1)**
 یک لحظه محو می شد، دوباره پدیدار می گردید و انعکاس آن به طرز درنگی روی عصب چشم نقش می بست. (داستان شماره 9: سایه مغول، ص 108-109)

GT : *Des sphères rouges et violettes tournent devant ses yeux, il (2) danse. Il (1) s'est évanoui pendant un moment, a réapparu, et son reflet a été douloureusement gravé sur le nerf de son œil.*

2 pronoms mal traduits | Cause : erreur d'identification de l'antécédent

Pronom persan (1) : sujet nul « آن‌ها » (démonstratif, plur.) | Antécédent : « گویی‌های سرخ و بنفش » → Tr. fr. : « des sphères rouges et violettes » (SN, fém., plur.) | Place de l'antécédent : la phrase précédant celle du pronom | Pronom français (1) : « il » (personnel, masc., sing.) | Remplacement proposé : « elles » (personnel, fém., plur.)

Pronom français (2) : « il » (personnel, masc., sing.) (Google Translate n'a pu identifier le sujet du verbe « danser » correctement et a donc lié ce dernier à un sujet hors de cet extrait. Toutefois, étant donné que le SN « des sphères rouges et violettes » est le sujet de « danser », il n'exige aucun pronom ; effectivement, le pronom doit être omis par le biais de l'ellipse grammaticale.)

10- بالاخره بعد از اصرار و پاشاری مادر مهربانو، مادرش راضی شد و پس از کشمکش‌های زیاد یک دست لباس سرخ برای او گرفت. ولی **Ø (1)** هر تکه آن را که می‌برید نفرین و ناله می‌کرد و می‌گفت: (...). (داستان شماره 19: زنی که مردش را گم کرد، ص 231)

GT: *Finalement, après l'insistance de la gentille mère, sa mère était satisfaite et après de nombreuses luttes, **elle (2)** a obtenu une robe rouge pour elle. Mais **il (1)** maudissait et gémissait à chaque morceau que vous coupez, en disant: (...).*

1 pronom mal traduit (pronome persan (1)) | Cause : erreur d'identification de l'antécédent ; 1 pronom bien traduit (pronome français (2))

Pronome persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « مادرش » → Tr. fr. : « sa mère » (SN, fém., sing.) | Place de l'antécédent : la phrase précédant celle du pronom | Pronome français (1) : « il » (personnel,

masc., sing.) | Remplacement proposé : « elle »
(personnel, fém., sing.)

Pronom français (2) : « elle » (personnel, fém., sing.)
(En dépit de l'absence de la correspondance formelle aux pronoms dans le corpus persan, ce pronom français se voit comme la traduction obligée du pronom « او » omis de l'extrait persan à travers l'ellipse grammaticale.)
| Antécédent : « sa mère » (SN, fém., sing.) | Place de l'antécédent : la même phrase que le pronom

Ce qui reste est 6% des cas où les pronoms sujets de 3^e personne des extraits persans n'ont pas été traduits en français. Toutefois, dans 4% de ces cas, la traduction de l'ensemble des phrases a été faite correctement. Ce dernier exemple présente un cas de non-traduction, de même qu'un cas où Google Translate a bien identifié l'antécédent mais à cause du cas précédent, n'a pas placé le pronom français dans la bonne position.

11- در باز شد و دختر رنگپریده‌ای هراسان بیرون آمد. (...)
(1) Ø
دست برادرش را گرفت و Ø (2) در اتاق تاریک کوچک که تا کمرکش
دیوار نم کشیده بود داخل شدند. (داستان شماره 16: چنگال، ص 173)

GT: *La porte s'ouvrit et une fille pâle effrayée en sortit.*
(...) *Ils (2) ont pris la main de son frère et sont entrés dans la petite pièce sombre qui avait un mur humide jusqu'à la taille.*

1 pronom non traduit (pronom persan (1)) | Cause : erreur d'identification de l'antécédent ; 1 pronom mal traduit (pronom persan (2)) | Cause : mauvais placement au sein de la phrase

Pronom persan (1) : sujet nul « او » (personnel, sing.) | Antécédent : « دختر رنگپریده‌ای » → Tr. fr. : « une fille pâle effrayée » (SN, fém., sing.) | Place de l'antécédent :

la phrase précédent celle du pronom |Il faudrait remplacer « Ils » par « Elle » (pronome personnel, fém., sing.), car la locution verbale « prendre la main » se lie à « une fille pâle effrayée ».

Pronom persan (2) : sujet nul « آن‌ها » (démonstratif, plur.) | Antécédent : « دختر رنگ‌پریده و برادرش » → Tr. fr. : « une fille pâle effrayée et son frère » (2 SN, masc., plur.) | Place de l'antécédent : une partie (« une fille pâle effrayée ») se trouve dans la phrase précédent celle du pronom, alors que l'autre partie (« son frère ») se situe dans la même phrase que le pronom. | Pronom français (2) : « ils » (personnel, masc., plur.) | Il faudrait déplacer le pronom « ils » et le placer par la suite après « et » et avant « sont entrés », celui-ci étant le verbe auquel se lie ce pronom.

Les statistiques complètes concernant les pronoms bien traduits, mal traduits ou encore non traduits sont marquées dans le tableau ci-dessous.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتمال جامع علوم انسانی

Pronoms sujets de 3 ^e personne en corpus persan		و	آن	آنها	Total
Bien traduits		12 (dont 8 sujets nuls)/38 ≈ 32%	2 (sujets nuls)/3 ≈ 67%	7 (dont 1 sujet nul)/11 ≈ 64%	21 (dont 11 sujets nuls)/52 ≈ 40%
Mal traduits		24 (dont 12 sujets nuls)/38 ≈ 63%	1 (sujet nul)/3 ≈ 33%	3 (sujets nuls)/11 ≈ 27%	28 (dont 16 sujets nuls)/52 ≈ 54%
Non traduits		2 (dont 1 sujet nul)/38 ≈	0	1/11 ≈	3 (dont 1 sujet nul)/52

	5%		9%	≈ 6% (dont 4% corrects)
--	----	--	----	----------------------------------

Tableau 4 – Nombre et pourcentage de pronoms « bien traduits », « mal traduits », ou « non traduits » dans la langue cible

Conclusion

Le fait qu’au contraire du persan le français est une langue générée implique des problèmes pour la traduction automatique des pronoms sujets de 3^e personne depuis le persan vers le français. Dans l’intention d’étudier ces problèmes éventuels, nous avons tiré 50 extraits contenant lesdits pronoms d’*une collection d’écrits de Sadegh Hedayat*, afin que Google Translate, un traducteur automatique à base statistique, les traduise en français.

En conséquence, en vue d’évaluer la précision de la traduction Google et les conditions dans lesquelles Google offre une traduction correcte des pronoms sujets de 3^e personne (tout en tenant compte de l’impossibilité d’une analyse interphrastique de la part du système), également pour connaître les raisons de la différence du nombre de ces pronoms dans le texte-source et le texte-cible, nous avons fait une analyse commentée suivie de vérifications répétées. À travers cette analyse statistique et grammaticale, notre supposition concernant l’impact du pro-drop et de l’ellipse des pronoms persans sur la qualité de la traduction automatique a été attestée: nous avons constaté que Google a traduit ces pronoms avec 44% de précision (y compris des pronoms bien traduits (40%) et certains pronoms non traduits (4%)) ; en effet, il a offert une traduction correcte dans les cas où l’antécédent était dans la même phrase que le pronom, ou s’il se plaçait dans la phrase précédant celle du pronom, il était traduit

par un SN masculin. Quant aux raisons pour la différence numérale des dits pronoms dans le texte-source et le texte-cible, il faut les attribuer des fois à la nécessité de traduire certains pronoms persans omis en vertu de l'ellipse linguistique, des fois aux préférences algorithmiques de Google Translate.

Même s'il existe un nombre considérable de recherches comparatives au sujet de la traduction automatique des pronoms, celles-ci sont pour la plupart limitées aux langues voisines. La nouveauté dans notre recherche relève de notre effort pour étendre les études précédentes, dans la mesure du possible, à deux langues « éloignées », persane et française dans ce cas. C'est pour cette raison qu'une exigence pour mener davantage de recherches au sujet mentionné entre les langues éloignées se fait sentir. Enfin, nous concluons cette étude par une vieille citation de Martin Kay (1986), qui reste toujours valable aujourd'hui : « Il y a encore beaucoup à faire en traduction automatique des pronoms. »

Bibliographie

- Carl, M. (2003). Introduction à la traduction guidée par l'exemple (Traduction par analogie). In *Actes de la 10^{ème} conférence sur le Traitement Automatique des Langues Naturelles. (TALN)* (pp. 11-26). Récupéré de <https://www.aclweb.org/anthology/2003.jeptalnrecital-tutorial.1/>.
- Daas, A. (2013). La théorie du liage appliquée à la langue arabe moderne (la presse écrite) et à la langue française: analyse syntaxique comparative des réflexifs [thèse de doctorat, Université de Jijel], Dspace (p. 14). Récupéré de [La théorie du liage appliquée à la langue arabe moderne \(la presse écrite\) et à la langue française : analyse syntaxique comparative des réflexifs \(univ-jijel.dz\)](http://www.univ-jijel.dz:8080/jspui/handle/123456789/1000).

- Ducotterd, G. (2017). Contenu généré par l'utilisateur et traduction automatique: Évaluation comparative de deux systèmes: Google Translate vs. SYSTRAN [thèse de doctorat, Université de Genève]. Unige (p. 22). Récupéré de [Contenu généré par l'utilisateur et traduction automatique : Évaluation comparative de deux systèmes... | Archive ouverte UNIGE](https://archive-ouverte.unige.ch/unige:35617).
- Kay, M. (1986). Machine translation will not work. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, p. 286.
- Kerstens, J.G. & Ruys, E. (1994). *Generative syntaxis : een inleiding* (1^{ère} éd.). Noordhoff Uitgevers B.V., p. 218.
- Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press, p. 81.
- Loaiciga Sanchez, S. (2013). Résolution d'anaphores et traitement des pronoms en traduction automatique à base de règles. In *Traitement Automatique du Langage Naturel 2013 (TALN 2013)* (pp. 683-690). Récupéré de <https://archive-ouverte.unige.ch/unige:35617>.
- Rees, A. V. (2007). Les théories du liage appliquées aux données du français [mémoire de maîtrise, Université d'Utrecht]. Dspace (pp. 8-10). Récupéré de <https://dspace.library.uu.nl/handle/1874/23328>.
- Russo, L. (2010). La traduction automatique des pronoms clitiques. Quelle approche pour quels résultats ?. In *Actes de la 17^e Conférence sur le traitement automatique des langues naturelles (TALN 2010)*. Récupéré de http://www.iro.umontreal.ca/~felipe/TALN2010/Xml/Papers/taln_taln2010_submission_145.pdf.
- Russo, L. (2013). Pour une analyse linguistique comparative italien-français en vue d'une traduction automatique. Projets de recherche autour d'Its-2. *Publiforum*, (19). Récupéré de <http://publiforum.farum.it/index.php/publiforum/article/view/414>.
- Russo, L., Scherrer, Y., Goldman, J. P., Loaiciga Sanchez, S., Nerima, L., & Wehrli, É. (2011). Étude inter-langues de la distribution et des ambiguïtés syntaxiques des pronoms. In *Actes de la 18^e conférence sur le Traitement Automatique des Langues Naturelles (TALN)* (pp. 279-284). LIRMM. Récupéré de <https://archive-ouverte.unige.ch/unige:22823>.

- Russo, L., & Wehrli, É. (2011). La traduction automatique des séquences clitiques dans un traducteur à base de règles. *Traitement Automatique des Langues Naturelles*, 92. Récupéré de http://www.lirmm.fr/TALN2011/PDF_court/Russo_taln11_submission_165.pdf.
- Scherrer, Y., Russo, L., Goldman, J. P., Loaiciga Sanchez, S., Nerima, L., & Wehrli, É. (2011). La traduction automatique des pronoms: Problèmes et perspectives. In *Actes de la 18^e conférence sur le Traitement Automatique des Langues Naturelles (TALN)* (pp. 185-190). LIRMM. Récupéré de <https://archive-ouverte.unige.ch/unige:22824>.
- Weissenbacher, D., & Nazarenko, A. (2007). Identifier les pronoms anaphoriques et trouver leurs antécédents: l'intérêt de la classification bayésienne. In *Traitement Automatique des Langues Naturelles (TALN'07)* (pp.145-155). Récupéré de <https://hal.archives-ouvertes.fr/hal-00162093/>.
- بهارلو، محمد. (1372). *مجموعه‌ای از آثار صادق هدایت*. تهران: انتشارات طرح نو.

Sitographie

- Casswell, I. & Liang, B. (2020). Recent Advances in Google Translate. [Google AI Blog: Recent Advances in Google Translate \(googleblog.com\)](https://googleblog.com). Consulté le 26 novembre 2021.
- Fleck, F. (2019). Ellipse linguistique et ellipse narratologique. https://www.fabula.org/atelier.php?Ellipse_linguistique_et_ellipse_narratologique. Consulté le 5 janvier 2021.
- Je révise. (s.d.). L'antécédent d'un pronom relatif. <https://www.jerevise.fr/antecedent-pronom-relatif.html>. Consulté le 3 janvier 2021.
- La conjugaison. (s.d.). L'antécédent. <https://la-conjugaison.nouvelobs.com/regles/grammaire/l-antecedent-232.php>. Consulté le 3 janvier 2021.

Qwerty.wiki. (s.d.). Grammaire persane – Persian grammar. https://fr.qaz.wiki/wiki/Persian_grammar. Consulté le 2 janvier 2021.

Qwerty.wiki. (s.d.). Langue à sujet nul – Null-subject language. https://fr.qaz.wiki/wiki/Null-subject_language. Consulté le 2 janvier 2021.

