

**19, 2013**

**Ricerche Dottorali in Francesistica - Nos voix/voies pour la  
Recherche**

---

**Lorenza RUSSO**

**Pour une analyse linguistique comparative italien-français en vue d'une  
traduction automatique. Projets de recherche autour d'Its-2**

---

**Per citare l'articolo:**

<https://www.publifarum.farum.it/index.php/publifarum/article/view/414>

---

Rivista Publifarum

[publifarum.farum.it](http://publifarum.farum.it)

---

Documento accessibile online:

<https://www.publifarum.farum.it/index.php/publifarum/article/view/414/673>

Documento generato automaticamente 25-08-2020

---

# Pour une analyse linguistique comparative italien-français en vue d'une traduction automatique. Projets de recherche autour d'Its-2

Lorenza RUSSO

---

## Indice

[Abstract](#)

[1. Introduction](#)

[2. Le projet de recherche](#)

[3. Le projet de thèse](#)

[4. Conclusion](#)

[Bibliographie](#)

---

## Abstract

In this article we present two different but complementary research projects. The first one is a university project which aims to develop and improve a multilingual rule-based machine translation called Its-2. We will then explain in more detail our work in developing lexicons and grammar for this project. Our second project, that is, our thesis project, is focused on closely-related language machine translation. Our aim is not only to understand how to improve the machine translation between languages from the same family and, more specifically, between Italian and French,

but also to emphasize the importance of linguistic information to obtain grammatically correct translations. We will then explain our point of view by discussing the results of some tests performed on Its-2.

## 1. Introduction

La recherche que nous présentons dans cet article s'étale sur deux projets complémentaires: un projet de recherche de département et le projet de thèse. En ce qui concerne le premier projet, il vise au développement d'un traducteur automatique multilingue à base de règles syntaxiques, appelé Its-2 (WEHRLI ET AL., 2009). Dans le cadre de ce projet, nous sommes chargée de l'implémentation des lexiques mono- et bilingue italien-français, de la création de règles syntaxiques pour l'analyse du français et de l'italien tout comme pour la traduction de l'italien vers le français et vice versa. Le projet de thèse sur la traduction automatique entre langues proches, <sup>1</sup> vise à éclaircir les tenants et les aboutissants de ce type de traduction, afin de mieux comprendre comment améliorer la traduction entre langues qui appartiennent à la même famille et, dans notre cas spécifique, entre l'italien et le français.

Nos buts principaux sont donc au nombre de deux : i) pousser les recherches en traduction automatique entre langues proches, en particulier pour la paire de langues italien-français et ii) souligner l'importance d'une information linguistique fine pour l'obtention de traductions automatiques grammaticalement correctes.

Dans la prochaine section, nous décrirons plus en détail notre travail de recherche au sein du Département de Linguistique de l'Université de Genève, puis nous présenterons notre projet de thèse et les recherches menées pour ce projet. Finalement, la dernière section est dédiée aux conclusions et aux futures recherches.

## 2. Le projet de recherche

Le travail de recherche en traitement automatique du langage naturel que nous menons depuis 2007 dans le cadre du projet MulTra - Multilingual Translation (WEHRLI ET NERIMA, 2008), porte sur l'implémentation du système de traduction automatique à base de règles Its-2. Its-2 est un système multilingue qui bénéficie

d'une analyse syntaxique profonde.<sup>2</sup> Cette dernière est effectuée par Fips (WEHRLI ET NERIMA, 2009), analyseur syntaxique qui utilise des grammaires basées sur le modèle Government and Binding (CHOMSKY, 1995) et spécifiques à chaque langue traitée par son système.<sup>3</sup>

La stratégie de traduction d'Its-2 s'articule sur trois moments principaux: l'analyse syntaxique, le transfert lexical et syntaxique et la génération morphosyntaxique (RUSSO ET WEHRLI, 2011a). Lors de la première phase, c'est-à-dire l'analyse syntaxique profonde et complète de la phrase source, Fips donne des informations sur la nature morphosyntaxique des mots qui la composent et il crée la structure arborescente de la phrase source. Sur la base de ces informations, Its-2 trouve la tête lexicale de la structure source et cherche la correspondance lexicale de cette tête dans la base de données bilingue. D'après les informations morphosyntaxiques contenues dans cette dernière, Its-2 projette une structure arborescente cible. Après la tête lexicale, ce sont les sous-arbres gauches et droits qui sont pris en considération par le système. Cela constitue la phase de transfert lexical et syntaxique. La structure arborescente cible une fois obtenue, la traduction s'achève, lors de la troisième phase, par la génération et la réorganisation morphosyntaxique de la phrase cible, pour qu'elle soit grammaticalement correcte.

Notre travail de recherche dans le cadre de ce projet porte sur plusieurs aspects : au plan lexical, tout d'abord, mais aussi au plan syntaxique et de transfert. Du point de vue lexical, nous améliorons la couverture lexicale monolingue et bilingue du traducteur Its-2 tout comme de l'analyseur syntaxique.<sup>4</sup> Les lexiques monolingues et bilingues contiennent de nombreuses informations morphologiques et syntaxiques nécessaires pour que le système soit capable de correctement reconnaître, analyser et traduire une langue donnée. L'importance de ces informations ou, plus en général, de la couverture de ces lexiques est liée aux performances du système : on risque d'avoir une analyse incomplète ou erronée de la phrase source et donc une traduction incomplète ou incorrecte si le système doit traiter un mot inconnu. Les informations contenues dans les lexiques sont des informations morphologiques sur la flexion d'un mot donné, sur son genre, nombre et personne, sur sa représentation phonétique ainsi que sur ses variantes, d'un côté. D'autres informations telles que le type et le nombre d'arguments du lexème

de base et leurs fonctions grammaticales et thématiques sont également présentes dans la base de données lexicale (RUSSO, 2012).

Au niveau syntaxique, il est nécessaire d'implémenter les analyses syntaxiques de l'analyseur Fips en créant des règles d'attachement de constituants adjacents. Pour la plupart, ces règles sont spécifiques à la langue traitée et elles sont décrites dans un pseudo-formalisme qui se base sur la forme générale d'une règle de réécriture, mais qui est exprimée sous la forme de fonctions booléennes, où a et b représentent le premier et le deuxième constituant d'une opération de combinaison. Dans l'exemple (1), on peut voir la règle pour l'attachement d'un déterminant à un substantif.

- (1)DP + NP
  - b.IsType(CommonNoun)
  - a.AgreeWith(b,{gender, number})

Selon ce formalisme, a et b indiquent les deux constituants adjacents de l'opération de combinaison, qui dans la règle en (1) sont le déterminant (D) et le substantif (N). De plus, des conditions sont associées à chaque règle : la condition "Agree with", par exemple, impose un accord entre les deux constituants, en indiquant que le constituant pris en question, qui dans notre exemple est le a, c'est-à-dire le déterminant, concorde en genre et nombre avec l'autre constituant, qui dans notre cas est le b, le substantif, auquel il s'attache.

Finalement, pour améliorer la traduction automatique de Its-2, nous devons créer des règles de transfert qui permettent au système de traduction de passer de l'arbre syntaxique source, créé par Fips, à un arbre syntaxique cible, grammaticalement et syntaxiquement correct par rapport aux règles de la langue cible. Ces règles de transfert deviennent nécessaires dans le cas de différences syntaxiques entre les paires de langues que le système traduit. <sup>5</sup> Pour ce qui est de la paire de langues qui nous intéresse ici, c'est-à-dire l'italien et le français, il est nécessaire de créer des règles de transfert pour gérer des différences syntaxiques. Parmi celles-ci se trouvent, entre autres, les différences liées à la position des pronoms clitiques (position proclitique - dite aussi préverbale - ou enclitique - dite aussi postverbale).

Pour pouvoir avoir une traduction automatique correcte des pronoms clitiques,

plusieurs règles de transfert et de génération sont nécessaires. Tout d'abord, il faut que le système puisse traduire un pronom clitique enclitique en italien par un pronom clitique proclitique en français, comme par exemple dans les phrases infinitives subordonnées («Vorrei sostituirlo» - «J'aimerais le remplacer»), ou dans les structures gérondives («Mi addormento leggendo una favola» - «Je m'endors en te lisant une fable»). Pour ce qui est de la traduction des séquences clitiques, considérons la phrase «Gianni glielo ha già detto» («Gianni le lui a déjà dit»). Les règles de transfert et de génération sont utiles au système pour transformer le pronom proclitique datif en italien «gli» en un pronom enclitique datif en français, «lui», ainsi que pour transformer le pronom proclitique accusatif en italien «lo» en un pronom enclitique accusatif en français «le». De plus, les règles de transfert seront indispensables au système pour réorganiser les deux pronoms de la séquence clitique selon les règles du français, c'est-à-dire en générant l'ordre [ - CLacc - CLdat ] plutôt que l'ordre [ - CLdat - CLacc ], ce dernier étant propre à l'italien. <sup>6</sup>

Tournons-nous maintenant vers notre deuxième projet de recherche, c'est-à-dire vers notre travail de thèse.

## **3. Le projet de thèse**

### **3.1. Les motivations de ce travail**

Comme mentionné plus haut, notre travail de thèse se focalise sur la traduction automatique entre langues proches, et plus précisément entre l'italien et le français.

À l'heure actuelle, les recherches en traduction automatique entre langues proches n'ont porté que très marginalement sur la paire de langues italien-français, préférant plutôt développer des systèmes pour d'autres langues européennes, telles que l'espagnol et le portugais (ARMENTANO OLLER ET AL., 2006), ou sur des langues dites «minoritaires», telles que le catalan (CANALS-MAROTE ET AL., 2001) et l'occitan (ARMENTANO OLLER ET FORCADA, 2006). Un autre courant de recherche a pris en considération les langues slaves, notamment le serbe et le slovaque (HAJIČ ET AL., 2000), mais on peut compter aussi des travaux sur le turc

(TANTUG ET AL., 2007) et le tatar de Crimée (ALTINTAS ET CICEKLI, 2002) ainsi que sur le gaélique d'Irlande et le gaélique écossais (SCANNEL, 2006). De plus, il existe des recherches en traduction automatique sur des langues scandinaves, telles que le suédois, le norvégien et le danois (VOLK, 2008).

L'hypothèse générale sur laquelle se base la plupart des traducteurs automatiques pour des paires de langues proches est qu'une analyse syntaxique de surface suffirait à l'obtention de traductions automatiques compréhensibles, vu les caractéristiques morphosyntaxiques que les langues proches partagent. L'intérêt des chercheurs en traduction automatique entre langues proches est donc principalement dû au fait que la création et l'implémentation de tels systèmes sembleraient plus rapides en comparaison au développement d'un traducteur automatique pour langues éloignées. Ainsi, si on traduit de l'italien en allemand – c'est-à-dire entre deux langues éloignées –, il faut que le traducteur automatique soit capable de tenir compte d'ajustements syntaxiques de la phrase dans la langue cible (2a). Le même problème ne se pose pas dans le cas de langues voisines, comme l'italien et le français, qui partagent une partie des structures syntaxiques (2b).

- (2)a. Oggi ho mangiato una mela.  
Heute ich habe gegessen einen Apfel.  
Heute habe ich einen Apfel gegessen.
- (2)b. Oggi ho mangiato una mela.  
Aujourd'hui ai mangé une pomme.  
Aujourd'hui j'ai mangé une pomme.

Le comportement syntaxique similaire entre des langues proches présente aussi un deuxième avantage pour la traduction automatique : le partage des ambiguïtés lexicales et syntaxiques. Les langues proches ayant un comportement syntaxique similaire, il y a de bonnes probabilités que les cas d'ambiguïtés soient facilement transposables de la langue source à la langue cible (3a). Le même phénomène considéré du point de vue des langues éloignées pose, par contre, plus de problèmes, le traducteur automatique nécessitant plus d'informations linguistiques pour résoudre l'ambiguïté (3b). <sup>7</sup>

- (3)a. Je parle aux étudiants du collège.

[TP [DP Je] parle [VP [PP aux [DP [NP étudiants]]][PP du [DP [NP collège]]]]  
[TP [DP Je] parle [VP [PP aux [DP [NP étudiants [PP du [DP [NPcollège]]]]]]]]  
Parlo agli studenti del collegio.

[TP [DP ] Parlo [VP [PP agli [DP [NP studenti]]][PP del [DP [NP collegio]]]]

[TP [DP ] Parlo [VP [PP agli [DP [NP studenti [PP del [DP [NPcollegio]]]]]]]]

- (3)b. Je parle aux étudiants du collège.

I speak to the students of the college.

I speak to the students about the college.

Au vu de ces données, on serait tenté d'affirmer que la traduction automatique entre langues proches peut amener à de meilleurs résultats que celle entre langues éloignées vu les caractéristiques linguistiques partagées par les langues voisines. Mais le problème est bien plus complexe. Pour s'en rendre compte, il suffit de considérer que toutes ambiguïtés ne sont pas transposables d'une langue à une autre. C'est le cas, par exemple, d'une phrase telle que «Ho conosciuto la ragazza che ha sposato Gianni», dans laquelle la phrase relative («che ha sposato Gianni») peut être traduite de deux façons différentes en français, selon que l'on analyse «che» comme sujet de la relative – «J'ai connu la fille qui a épousé Gianni» – ou comme son objet direct – «J'ai connu la fille que Gianni a épousée». En outre, si on tient compte d'autres aspects linguistiques qui différencient les langues proches, telles que les pronoms clitiques («Rideva raccontandomi una bugia» – «Elle riait en me racontant un mensonge»), on comprend facilement que de telles différences ne peuvent que très difficilement être traitées par un traducteur automatique qui effectue une analyse superficielle de la phrase source.

Compte tenu de ces problèmes, notre travail de thèse se donne comme but de répondre à deux questions en particulier : i) est-ce qu'on obtient de meilleurs résultats si on traduit entre langues proches qu'entre langues éloignées ? ; ii) une analyse syntaxique de surface suffit-elle pour obtenir des traductions automatiques entre langues proches de bonne qualité ?

### **3.2. Les évaluations**

Pour répondre à ces questions, nous avons mené deux évaluations différentes : une première, afin d'examiner l'hypothèse selon laquelle traduire automatiquement

entre langues proches peut mener à de meilleurs résultats que de traduire entre langues éloignées; et une deuxième pour vérifier si une analyse syntaxique superficielle de la phrase suffit pour l'obtention de résultats de bonne qualité.

Pour notre première évaluation, nous avons choisi de tester notre système Its-2 sur un corpus multilingue parallèle constitué de vingt communiqués de presse de l'Administration Nationale Suisse, pour un total d'environ 5.500 mots par corpus et par sens de traduction. Il s'agit de communiqués de presse datant de 2010 dans les trois langues nationales suisses - l'allemand, le français et l'italien - et dans leur traduction en anglais. <sup>8</sup> Le choix de ce corpus a été motivé par son caractère multilingue, qui nous a permis de tester Its-2 sur plusieurs paires de langues. En particulier, six couples de langues ont été pris en considération : français-italien ainsi qu'italien-français ; français-anglais ainsi qu'anglais-français ; allemand-français et anglais-italien. Notre objectif principal a été de comparer la qualité des traductions entre langues proches (français-italien) à celle des traductions entre langues éloignées (français-anglais, allemand-français, anglais-italien) afin d'examiner l'impact de la proximité ou de la distance linguistique sur la qualité des traductions proposées.

L'évaluation des traductions proposées par Its-2 s'est effectuée à l'aide de la mesure BLEU (PAPINENI ET AL. 2002). Cette dernière permet, en effet, de comparer automatiquement la traduction proposée par le système de traduction automatique à une traduction de référence, effectuée dans la plupart des cas par un traducteur humain. Cette comparaison automatique se fait sur la base de n-grams de traduction, c'est-à-dire sur des fragments de textes de longueur variable. En particulier, on compare les n-grams de la traduction automatique aux n-grams de la référence de traduction et on compte le nombre de correspondances. Plus il y a de n-grams qui coïncident, plus la traduction proposée est considérée comme correcte.

Les résultats que nous avons obtenus montrent que globalement Its-2 atteint de meilleurs résultats pour les langues proches (0.28 points BLEU en total) alors que pour des langues éloignées il atteint des résultats moins positifs (0.16 points BLEU en total). Ces derniers sont plus bas pour les paires de langues allemand-français et anglais-italien. Seulement la paire anglais-français, quant à elle, atteint de meilleurs résultats dans les deux sens de traduction (0.28 points BLEU en total). Le

fait que globalement cette paire de langues obtient de meilleurs résultats par rapport à la paire de langues français-italien ne contredit qu'à première vue l'idée que la traduction automatique entre langues proches mène à de meilleurs résultats par rapport à celle entre langues éloignées. On ne peut pas négliger, en effet, que dans le cas spécifique du français et de l'anglais les efforts d'implémentation ont été plus importants. Cela est dû tout simplement au fait que ces deux langues ont été les premières à être développées dans Its-2. Compte tenu du fait que le même type d'effort a été investi dans l'implémentation des paires de langues prises en considération pour cette évaluation de traduction - exception faite pour la paire français-anglais -, la différence entre les scores qu'Its-2 atteint pour la paire de langues proches et pour les paires de langues éloignées ne peut être due qu'à la distance linguistique entre ces couples de langues et à la nécessité, par conséquent, d'une phase de transfert plus détaillée pour les langues éloignées.

La deuxième évaluation que nous avons menée vise, quant à elle, à examiner si une analyse syntaxique superficielle de la phrase suffit à obtenir de résultats corrects. Comme déjà mentionné plus haut, même si les langues proches partagent de nombreuses caractéristiques morphosyntaxiques, les différences ne manquent pas. Pour cette deuxième évaluation, nous avons décidé de nous focaliser sur le phénomène des pronoms clitiques, compte tenu du fait que les recherches actuelles en traduction automatique ne se focalisent pas directement sur la traduction de ce phénomène linguistique, préférant plutôt aborder le problème de la traduction automatique entre langues proches d'une façon plus générale. Nous avons donc mené deux études exploratoires sur la traduction automatique des pronoms clitiques entre langues proches. En effet, dans d'autres traducteurs automatiques entre langues proches, le phénomène des pronoms clitiques n'est pris en compte que rarement et souvent dans la partie concernant les problèmes à résoudre. <sup>9</sup> Pour notre travail, par contre, il s'agit du problème principal.

Afin d'avoir une vision d'ensemble qui n'ignore pas la diversité dans les deux principales approches en traduction automatique - c'est-à-dire l'approche linguistique et l'approche statistique <sup>10</sup> -, nous avons choisi pour cette évaluation de tester deux systèmes structurellement différents : le premier système, Its-2, en tant que représentant des systèmes de traduction automatique basés sur des règles linguistiques ; et le deuxième, Google Translate, disponible en ligne, en tant

que représentant des systèmes à base statistique. En particulier, la stratégie de traduction effectuée par Google Translate se révèle bien différente de celle d'Its-2 puisque Google Translate se base sur une approche statistique qui exploite des techniques d'apprentissage statistique appliquées au modèle de traduction [11](#).

Pour notre évaluation, nous avons créé nous-mêmes deux corpus bilingues, un en italien et un en français, en faisant en sorte de contrôler la structure syntaxique, le cas du clitique et sa position dans la phrase, notre but étant d'éviter l'interaction possible entre différents phénomènes linguistiques (RUSSO, 2011). Les corpus comptent un total d'environ 300 phrases pour chaque sens de traduction, des clitiques accusatifs et des clitiques datifs ainsi que trois structures syntaxiques : les structures transitives non réfléchies à l'indicatif, les structures infinitives subordonnées et les structures présentant un gérondif. Lors de l'évaluation manuelle des traductions, nous avons considéré une traduction comme correcte seulement dans le cas où le clitique était présent dans la phrase cible, généré correctement dans son cas, genre et nombre ainsi qu'attaché à son réel nœud d'attachement dans la langue cible.

Les résultats ont montré une nette différence dans la qualité des traductions des systèmes étudiés : Its-2, le traducteur automatique bénéficiant de règles linguistiques atteint de meilleurs taux de traductions correctes que Google Translate, le système à base statistique, et cela, indépendamment de la structure syntaxique de la phrase ou du type de clitique (RUSSO, 2011). Cela nous amène à confirmer la nécessité d'une analyse syntaxique profonde pour un système de traduction automatique, même dans le cas de langues proches – comme l'italien et le français –, notamment en présence de phénomènes syntaxiques très pointus et complexes comme celui de la cliticisation des pronoms.

## **4. Conclusion**

Dans cet article, nous avons illustré les deux projets de recherche dans lesquels nous sommes impliquée : un projet de recherche de département, portant sur l'implémentation du traducteur automatique Its-2 et notre projet de thèse, portant sur la traduction automatique entre l'italien et le français.

Ce deuxième projet en particulier prend ses début d'une constatation: l'italien et

le français ne font que marginalement l'objet de recherches en traduction automatique, bien que tout un courant des recherches en traitement automatique du langage ait porté, depuis les années '90, à la création et à l'implémentation de nombreux systèmes pour les langues proches, tels que l'espagnol et le portugais, l'espagnol et le catalan ou encore le serbe et le slovaque.

Compte tenu des caractéristiques morphosyntaxiques partagées entre les langues proches et aussi de l'hypothèse de base en traduction automatique entre langues proches - à savoir qu'il suffit d'effectuer une analyse syntaxique de surface pour traduire entre langues proches - nous avons mené deux évaluations : une première visant à comprendre si effectivement la traduction entre langues proches est plus rapide que la traduction entre langues éloignées ; et une deuxième visant à examiner si une analyse syntaxique superficielle de la phrase suffit à obtenir de résultats corrects.

Les résultats obtenus nous permettent de conclure que la traduction automatique entre langues proches présente des avantages non seulement en termes d'efforts d'implémentation mais aussi en termes de résultats par rapport à une traduction automatique entre langues éloignées. Cependant, une analyse syntaxique profonde se révèle indispensable si on veut obtenir des traductions automatiques grammaticalement correctes, même pour les langues proches et notamment en présence de phénomènes linguistiques très pointus.

Par la suite, les résultats obtenus par Its-2 seront comparés à ceux d'autres traducteurs automatiques, et notamment, à ceux des traducteurs conçus pour des paires de langues proches, et cela, pour appuyer avec plus de force nos résultats. De plus, nous comptons élargir les évaluations sur les pronoms clitiques à d'autres types de textes et à d'autres corpus plus conséquents : les résultats d'une recherche sur la fréquence des pronoms dans les textes (RUSSO ET AL., 2011c) montrent bien que les pronoms clitiques se trouvent dans différents styles de textes (littéraire, journalistique, etc.) et qu'ils sont utilisés d'une façon plus consistante en italien par rapport au français. Cela est en partie motivé par des raisons linguistiques et en partie par des raisons de style. En ce qui concerne les raisons linguistiques, il y a plus de verbes pronominaux en italien qu'en français. Pour ce qui est des raisons de style, au lieu du passif classique en français, on utilise plutôt des tournures verbales pronominales en italien <sup>12</sup>. Cela constitue une

nouvelle et intéressante voie de recherche dont il faudrait tenir compte à l'avenir.

## Bibliographie

K. ALTINTAS ET I. CICEKLI, « A Machine Translation System Between a Pair of Closely Related Languages », dans Proceedings of the 17th International Symposium on Computer and Information Sciences (ISCIS 2002), Orlando, Florida, Crc Press, 2002, p. 192-196.

C. ARMENTANO OLLER, R. C. CARRASCO, A. CORBÌ-BELLOT, M. FORCADA, M. GINESTRI-ROSELL, S. ORTIZ-ROJAS, S., J. PÉREZ-ORTIZ, G. RAMIREZ-SANCHEZ, F. SANCHEZ-MARTINEZ ET M. SCALCO, « Open-Source Portuguese-Spanish Machine Translation », dans Computational Processing of the Portuguese Language, Proceedings of the 7th International Workshop on Computational Processing of Written and Spoken Portuguese (PROPOR), 2006, p. 50-59.

C. ARMENTANO OLLER ET M. FORCADA, « Open-Source Machine Translation between Small Languages: Catalan and Aranese Occitan », dans Strategies for developing Machine Translation for Minority Languages (5th SALTMIL workshop on Minority Languages) (organized in conjunction with LREC 2006), 2006, p. 51-54.

R. CANALS-MAROTE, A. ESTEVE-GUILLÉN, A. GARRIDO-ALENDIA, A. GUARDIOLA-SAVALL, A. ITURRASPE-BELLVER, S. MONTSERRAT-BUENDIA, S. ORTIZ-ROJAS, H. PASTOR-PINA, P. PÉREZ-ANTÒN ET M. FORCADA, « The Spanish-Catalan Machine Translation System InterNOSTRUM », dans Proceedings of MT Summit VIII : Machine Translation in the Information Age, 2001, p. 73-76.

CHOMSKY, The Minimalist Program, Cambridge, Mass., MIT Press, 1995.

L. DABÈNE ET C. DEGACHE, « Comprendre les langues voisines », *Ela - Études de linguistique appliquée. Revue de Didactologie des langues-cultures*, n. 104, Didier Erudition, Paris, 1996.

A. GARRIDO-ALENDIA, P. GILABERT-ZARCO, J. PÉREZ-ANTÒN, A. PERTUSA-IBANEZ, G. RAMIREZ-SANCHEZ, F. SANCHEZ-MARTINEZ, M. SCALCO ET M. FORCADA, « Shallow Parsing for Portuguese-Spanish Machine Translation », dans Language technology for Portuguese shallow processing tools and resources, Lisbon, Portugal, 2003.

J. HAJIČ, J. HRIC ET V. KUBOŇ, « Machine translation of Very Close Languages »,

- Applied Natural Language Conferences, dans Proceedings of the Conference on Applied Natural Language Processing, Seattle, Washington, 2000.
- L. NERIMA, V. SERETAN ET E. WEHRLI, « Le problème des collocations en TAL », dans Nouveaux cahiers de linguistique française, n. 27 (2006), Genève, 2006, p. 95-115.
- H. NEY, « One Decade of Statistical Machine Translation : 1996-2005 », dans Proceedings of MT Summit X, 2005, p. 12-17.
- K. PAPINENI, S. ROUKOS, T. WARD ET W.-J. ZHU, « BLEU : a Method for Automatic Evaluation of Machine Translation », dans Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, 2002, p. 311-318.
- L. RUSSO, « La traduction automatique entre langues proches : les pronoms clitiques en italien et en français », dans G. BERNARD BARBEAU, C. GAGNÉ ET G. LEBLANC (éd.), Actes des XXIV Journées de Linguistique (JDL), CIRAL, Québec, Canada, 2011, p. 141-153.
- L. RUSSO, « Les lexiques bilingues pour les traducteurs automatiques ou l'apport de la lexicographie computationnelle », dans G. DOTOLI, C. BOCCUZZI, M. LO NOSTRO (éd.) Le dictionnaire bilingue. Tradition et innovation, collection "Biblioteca della Ricerca", section "Linguistica", Fasano, Schena Editore - Paris, Éditions Alain Baudry et C.ie., 2012, p.197-212.
- L. RUSSO ET E. WEHRLI, «Traduction automatique et aide terminologique: le traducteur de mots en contexte TWiC et le traducteur de phrases Its-2», dans C. VALLINI, A. DE MEO ET V. CARUSO (éd.) Traduttori e traduzioni, Napoli, Liguori, 2011a, p. 301-310.
- L. RUSSO ET E. WEHRLI, « La traduction automatique des séquences clitiques dans un traducteur à base de règles », dans Actes du 18e Colloque International TALN 2011 (Traitement automatique du Langage Naturel), Université de Montpellier (France), organisé par l'association ATALA (Association pour le Traitement Automatique des Langues), 2011b.
- L. RUSSO. Y. SCHERRER, J-P. GOLDMAN, S. LOÁICIGA, L. NERIMA ET E. WEHRLI, « Étude inter-langues de la distribution et des ambiguïtés syntaxiques des pronoms », dans Actes du 18e Colloque International TALN 2011 (Traitement automatique du Langage Naturel), Université de Montpellier (France), 2011c.

K. SCANNEL, « Machine Translation for Closely Related Language Pairs », dans Proceedings of the Workshop "Strategies for developing Machine Translation for Minority Languages", LREC, Genova, Italie, 2006, p. 103-107.

A. TANTUG, E. ADALI ET K. OFLAZER, « Machine Translation between Turkic Languages », dans Proceedings of the ACL 2007 Demo and Poster Sessions, Prague, 2007, p. 189-192.

M. VOLK, « The Automatic Translation of Film Subtitles. A Machine Translation Success Story? », dans Festschrift for Anna Sagvall Hein, Uppsala, 2008.

E. WEHRLI ET L. NERIMA, « Traduction multilingue : le projet MulTra », dans Actes du 15<sup>e</sup> Colloque International TALN 2008 (Traitement automatique du langage naturel), Avignon, France, 2008.

E. WEHRLI ET L. NERIMA, « L'analyseur syntaxique Fips », *Journée thématique ATALA « Quel analyseur syntaxique pour le français ? »*, Université Paris Diderot - Paris 7, 2009. Article disponible à la page: <http://alpage.inria.fr/iwpt09/atala/fips.pdf>

E. WEHRLI, L. NERIMA ET Y. SCHERRER, « Deep Linguistic Multilingual Translation and Bilingual Dictionaries », dans Proceedings of the Fourth Workshop on Statistical Machine Translation, Athens, 2009, p. 90-94.

---

## Note

[↑ 1](#) On définit deux langues comme proches quand elles sont typologiquement et génétiquement apparentées (DABÈNE ET DEGACHE 1996).

[↑ 2](#) Lors d'une analyse syntaxique profonde le système identifie les constituants d'une phrase - tels que les groupes nominaux, verbaux, etc. - et il crée un arbre syntaxique complet de la phrase tout en spécifiant les fonctions de ses constituants ou leurs structures internes. Par contre, une analyse syntaxique de surface (appelée aussi shallow parsing en anglais) ne spécifie ni les fonctions des constituants dans la phrase, ni leurs structures internes.

[↑ 3](#) Actuellement, les langues traitées par le système sont l'allemand, l'anglais, le français, l'italien et l'espagnol. D'autres langues, telles que le japonais, le roumain, le russe et le grec sont en phase d'implémentation.

↑ 4 Remarquons que des lexiques de collocations sont aussi présents. Pour plus d'informations sur ces lexiques, nous renvoyons le lecteur au travail de (RUSSO 2012) ainsi qu'au travail de (NERIMA ET AL. 2006).

↑ 5 À l'heure actuelle, le système traduit les paires de langues suivantes: anglais-français, allemand-français, italien-français et espagnol-français. D'autres paires de langues, telles qu'allemand-anglais, anglais-italien, allemand-italien et japonais-anglais, sont en cours d'implémentation.

↑ 6 Pour plus d'information sur l'analyse et la traduction automatique des séquences clitiques, voir (RUSSO ET WEHRLI 2011b).

↑ 7 Les analyses syntaxiques de l'exemple (3a) sont les analyses de Fips. En particulier, les étiquettes syntaxiques sont en anglais: par exemple, NP est un Noun Phrase (syntagme nominal), VP un Verb Phrase (syntagme verbal) et PP un Prepositional Phrase (syntagme prépositionnel). Ces deux analyses syntaxiques de la phrase source en français mènent à deux traductions différentes en anglais (3b) alors que la traduction en italien garde l'ambiguïté (3a).

↑ 8 Ces textes sont disponibles à l'adresse : <http://www.news.admin.ch>.

↑ 9 Ce dernier est le cas, par exemple, de deux systèmes de traduction automatique proposés par l'Universitat d'Alicant et par l'Universitat Politècnica de Catalunya, à savoir, le système interNOSTRUM (CANALS-MAROTE ET AL. 2001) et le système Traductor Universia (GARRIDO-ALENDA ET AL. 2003), fusionnés dans la plate-forme open-source Apertium et adoptés aussi pour d'autres langues officielles de l'Espagne.

↑ 10 L'approche statistique ou stochastique se base plutôt sur des calculs mathématiques. En général, à la place de règles linguistiques, les traducteurs statistiques utilisent des modèles probabilistes qui cherchent à identifier laquelle des phrases dans la langue cible a le plus de probabilités d'être la traduction d'une phrase donnée dans la langue source. En d'autres mots, étant donné un fragment source S pour lequel il faut générer sa traduction dans un fragment cible C, le système analyse tout d'abord la distribution des traductions de S observées dans les corpus bilingues, pour ensuite générer le segment cible C sur la base du plus haut pourcentage de probabilité calculé avec le théorème de Bayes. Pour plus d'informations sur les systèmes de traduction basés sur des approches statistiques, nous renvoyons le lecteur à consulter (NEY 2005).

↑ 11 Relevons que, puisqu'il s'agit d'un système commercialisé, les informations relatives à la stratégie de traduction effectuée par Google Translate sont très peu nombreuses. Les quelques informations présentées ici pour Google Translate sont disponibles à la page [http://www.google.com/intl/fr/help/faq\\_translation.html](http://www.google.com/intl/fr/help/faq_translation.html).

↑ 12 Pour plus d'informations à ce propos, nous renvoyons le lecteur à l'article de (RUSSO ET AL. 2011c).