APPROCHE ANALYTIQUE DES SYSTÈMES DE TRADUCTION

Elena Carmen ANTAL
Grup Scolar de Industrializarea Lemnului Piteşti

Résumé *: L'article suivant, dédié à l'analyse des différents types de systèmes de traduction, est le fruit d'un stage de recherche de quatre mois à l'université de Nice Sophia Antipolis.*
Pour commencer, on fera un parallèle entre la TA et la TAO. On va passer en revue les débuts de la traduction automatique et ensuite on va faire référence au renouveau qui est survenu dans ce domaine pendant les années '80.
Les chapitres suivants seront consacrés aux différentes approches :

- *l'approche directe ;*
- *l'approche indirecte à structure de pivot ;*
- *l'approche indirecte à structure de transfert.*

La dernière partie sera dédiée à des types de traduction dont on parle moins et qui sont, par la suite, moins connues :

- *la traduction à base de connaissances*
- *la traduction basée sur l'exemple*

Mots-clés*: traduction automatique, traduction à base de connaissance, traduction basée sur l'exemple*

Introduction

La TA et la TAO sont issues de la demande de plus en plus croissante enregistrée sur le marché de la traduction. L'homme a ressenti la nécessité d'avoir un outil pour traduire plus rapidement. C'est pour cela que l'on a fait appel à l'informatique et l'on a crée des systèmes de traduction automatique.

TA versus TAO

La TAO est un ensemble logiciel qui permet la transduction d'un texte d'une langue source en une ou plusieurs langues (s) cible(s).

Les champs d'application de la TA sont : l'information économique, la veille technologique, l'information scientifique et technique, la veille technologique, l'information économique.

La TAO fait intervenir le traducteur humain, qui use de la TA comme d'un instrument pour mieux faire son travail.

Si l'on compare la TA et la traduction humaine d'un point de vue économique et quantitatif, la TA est préférable, mais d'un point de vue qualitatif, la traduction humaine donne de meilleurs résultats.

La "préhistoire" de la TA

En remontant dans le passé, on retrouve en **1629**, le *Dictionnaire multilingue* de Descartes, qui contient des mots traduits dans plusieurs langues, ayant comme nœud commun le latin.

En **1933**, G. Artsrouni publie un *Dictionnaire électromécanique.* L'auteur met en œuvre des mécanismes (un dispositif électromécanique) pour arriver à l'équivalent du mot dans la langue cible.

L'auteur prévoit une étape de pré édition, où les mots sont à la forme de base (l'infinitif pour les verbes), ensuite l'une de post édition : contenant les mots bruts. On introduit le facteur humain dans le processus: c'est lui qui doit faire l'accord, rétablir la flexion etc.

BDD-A5464 © 2006 Universitatea din Pitești
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 02:40:41 UTC)

Les débuts de la TA se situent vers **1947**, avec la création d'un *Dictionnaire automatisé* (Booth et Britten) à Princetown.

En **1949**, le Mémorandum de Warren Weaver (Georgetown), spécialiste en cryptographie, montre que le passage d'une langue à l'autre est un problème d'encodage et de décodage on crée un premier prototype, pour la traduction du russe vers l'anglais, sous la direction de L. Dorstert, disposant de 250 mots et de 6 règles. C'était un système direct, de langue à langue.

C'est en **1951** que le premier article sur la TA apparaît, signé par Bar Hillel, fondateur du courant pragmatique. (Massachusetts Institute of Technology).

Une année plus tard, en **1952** se déroule le premier colloque sur la TA (MIT). En **1953** est créé un prototype de traduction russe anglais : le système GAT (Georgetown Automatic Translation), avec l'entreprise IBN. Cette année-là sont traduits les premiers textes, en chimie, en disposant seulement de dix règles et cent mots.

En**1959**, le rapport Bar Hillel montre que, pour mieux traduire, il faut développer les recherches en linguistique.

Pendant la Conférence de Londres, un autre moment important dans l'histoire de la TA (**1961**), on énonce les principes des systèmes de deuxième génération. Selon ces principes, les systèmes de traduction doivent fonctionner comme des binômes, faisant d'abord l'analyse de la LS et ensuite la génération dans la langue cible.

Mais en **1966**, le rapport ALPAC montre qu'il n'y avait pas de résultat concret. On insiste sur les limites des systèmes directs, inspirés de l'expérience de Georgetown. Ce rapport négatif « tue » les recherches en TAO, faute de ressources financières.

Il suit, donc, la période entre **1969 et 1980**, dénommée métaphoriquement "la traversée du désert", on peut noter, pourtant, des parutions marquantes marquantes dans le domaine.

Ainsi, après la création du CETA (Centre d'Etudes en Traduction Automatique), on enregistre le premier succès en matière : le système Systran, dû à Peter Thoma et issu de la collaboration entre les Américains et les Russes en vue du déroulement d'un programme spatial. Le système Systran, créé pour la traduction de documents, à la NASA, est basé sur la traduction des mots. Il reste pourtant au même niveau que le système de Georgetown (GAT).

Après la création de TITUS, en **1973**, il est à noter le succès historique du système TAUM METEO, créé au Canada en **1976.** L'un des initiateurs du système, Alain Colmérauer, avait mis au point un système d'interrogation en langage naturel. Ce système, nommé « Q système » stockait les connaissances de façon à pouvoir l'interroger. Il rencontre Richard Kitredge et ils ont l'idée de traduire les bulletins météo, ceux-ci ayant un vocabulaire restreint et des structures semblables. On fait d'abord l'analyse des phrases (mots et structure de la phrase) et ensuite la génération. Le système permet le passage d'une langue à l'autre, à l'aide d'une structure de transfert.

Le système ARIANE, mis au point par B.Vauquois en 1978, est conçu en trois étapes : l'analyse de la langue source, le transfert des structures et du lexique et enfin l'équivalence entre les deux langues. La même année est crée le système ROBRA.

En 1980, paraît le système SUSY conçu par des Allemands en collaboration avec le centre de Grenoble en trois étapes : analyse, transfert et génération.

Le renouveau des années '80 est marqué par un très grand nombre de systèmes de traduction : MicroCAT, et MacroCAT, Metal, MOPTRANS, *D.L.T*, ROSETTA, Alps, SPANAM, TITRAN, Atlas, *L.M.T.*, LOGOS, EUROTRA, Translator, Globalink, Pivot, HiCats . L'analyse syntaxique est au centre de ces systèmes. Ils sont conçus pour

BDD-A5464 © 2006 Universitatea din Pitești
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 02:40:41 UTC)

aider l'homme dans la traduction et non pas pour se substituer à lui.

Les différents types de traduction

A des degrés divers, et selon leur compétence et leur entraînement à les utiliser, les traducteurs humains disposent d'une large palette de logiciels capables de standardiser, homogénéiser et de faciliter les traductions en ligne ou hors ligne. Ceux-ci se distinguent en :

- logiciels de traduction automatique
- logiciels de traduction assistée parmi lesquels :
  - les logiciels multifonction à mémoire de traduction
  - les outils de localisation
  - les dictionnaires électroniques
  - systèmes de gestion de terminologie
  - les outils de rédaction en langue étrangère assistés

Selon la part prise par le traducteur humain à l'ensemble des opérations de traduction, on a pris l'habitude de distinguer entre Traduction humaine assistée par ordinateur (HAMT) et traduction automatique enrichie par traducteur humain (MAHT).

Selon les périodes où ils ont été conçus, leurs fondements linguistiques et informatiques diffèrent.

Les premiers furent ceux basés sur **la transformation**. On passe d'une langue à une autre, c'est-à-dire que chacun des outils (analyseurs morphologiques, lexicaux, syntaxiques et sémantiques) est fondé sur **une approche contrastive** (description d'une langue par rapport à l'autre) (décennie 70).

Les deuxièmes furent ceux fondés sur **des connaissances linguistiques** (décennies 80), c'est-à-dire que les analyseurs comportaient des représentations abstraites (modèles), des phrases de la langue L1 et de la langue L2.

Les troisièmes introduisent une langue intermédiaire appelée **Pivot** ou **interlangue**, vers laquelle convergeraient toutes les descriptions des langues source et cible.

Les quatrièmes firent triompher la vogue **des mémoires de traduction**, c'est-à-dire de systèmes de traduction fondés sur des exemples (textes déjà traduits) à partir desquels on pouvait dégager des similitudes totales ou partielles d'énoncés à traduire. Enfin, en rapport avec la conviction que la mise au point de systèmes de traduction issus de descriptions exhaustives des langues n'était pas pour la prochaine étape, furent conçus des systèmes dits " probabilistes " impliquant un inventaire statistique des textes à traduire pour établir des règles de traduction adaptée et mettre en oeuvre des modèles d'analyse et de génération de textes en rapport avec ces règles.

BDD-A5464 © 2006 Universitatea din Pitești
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 02:40:41 UTC)

Types d'approches

Les systèmes de traduction emploient deux types d'approche :

1. *l'approche directe* : on traduit directement d'une langue dans l'autre ;
2. *les approches indirectes*, où l'on traduit indirectement :
   - approche à *structure de pivot* : de la langue source on va vers un langage pivot et on génère dans la langue cible.
   - approche à *structure de transfert* : on génère des structures à l'aide desquelles on va faire une traduction dans la langue cible.

L'approche directe

Les principes de ce type d'approche visent l'analyse et la génération contextuelle basée sur le mot, à l'aide d'un dictionnaire ou d'une grammaire.

Mais ce type d'approche pose certains **problèmes**

- Il existe un couplage entre l'analyse et la génération, ce qui fait que pour quatre langues, il faut avoir seize systèmes de traduction.
- Les données ne sont pas réutilisables, à cause du manque d'homogénéité.
- Il y a une faible couverture syntaxique.
- Le traitement sémantique n'est pas présent.
- On se heurte à des problèmes de maintenance (si l'on change une règle d'un côté, il peut intervenir des problèmes de l'autre côté).

Dans ce type d'approche, le même système fait en même temps, l'analyse et la génération. L'analyse est refaite chaque fois que l'on ajoute une nouvelle langue. S'il y a un système de codification pour traduire d'une langue LN 1 dans une langue LN 2, il n'est pas bon pour traduire dans une langue LN 3.

L'approche indirecte

Le principe de ce système vise l'analyse d'un énoncé en LS en interlingua (pour la structure profonde), pour le gérer ensuite en langue cible.

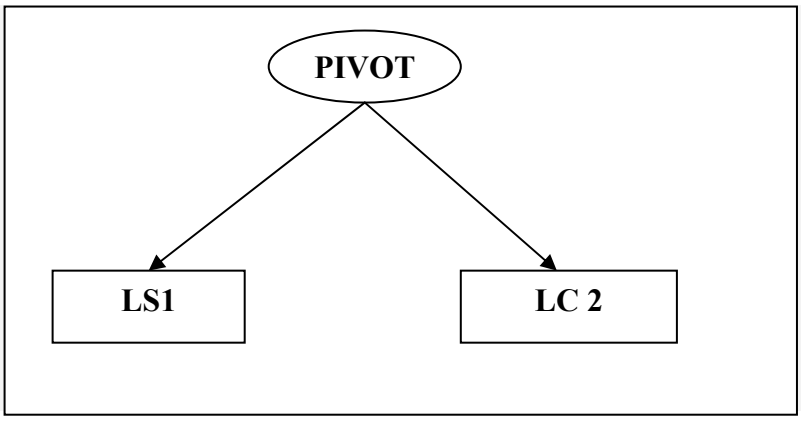
Les avantages de cette démarche sont la séparation de l'analyse et génération et la modularité (la réutilisation des ressources).

Mais il y a aussi certains problèmes avec la définition exacte de *l'interlingua*, pour laquelle on avait proposé le langage logique, l'aymara, l'espéranto.

Le groupe CETA (Centre d'Etudes de Traduction Automatique) avait proposé de passer, pour faire la traduction, par une structure profonde (notion de Chomsky), sémantique, qui serait identique pour toutes les langues.

BDD-A5464 © 2006 Universitatea din Pitești
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 02:40:41 UTC)

Schéma de l'approche indirecte à structure de pivot



L'approche indirecte à structure de transfert

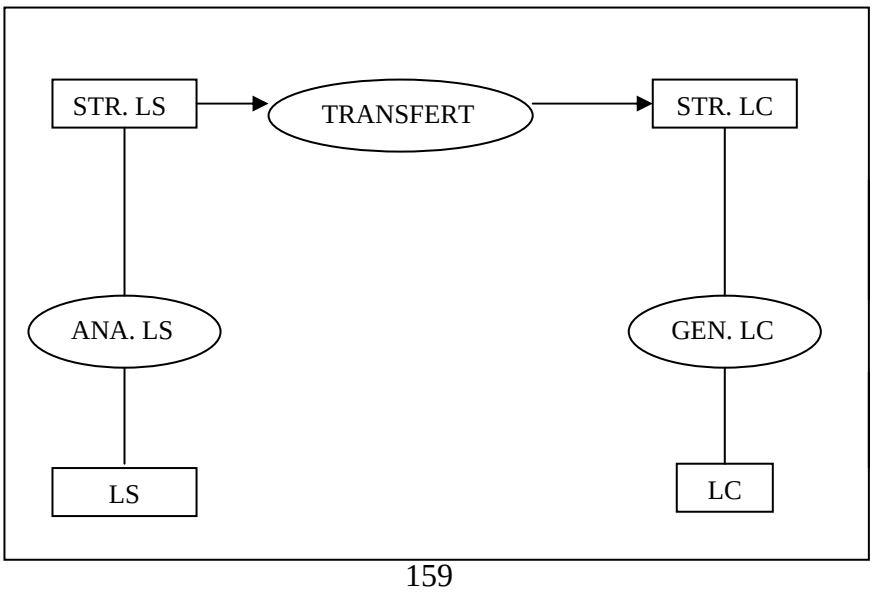
Principes :

Il faut faire la représentation structurale de l'information en langue source, le transfert structural et lexical et ensuite la représentation structurale en langue cible, puis la génération en langue cible. Le transfert est double. Il y a éventuellement la modification des structures et ensuite le transfert des mots.

Cette approche a comme avantage d'être une solution de compromis par rapport à l'approche indirecte à structure de pivot.

Mais le problème c'est qu'il y a une approche principalement formelle, car la sémantique ne joue qu'un faible rôle.

SCHEMA



BDD-A5464 © 2006 Universitatea din Pitești
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 02:40:41 UTC)

Définition:

Méthode de traduction automatique fondée sur la prise en compte du sens, à l'aide des méthodes de représentation des connaissances empruntées à l'intelligence artificielle.

Il s'agit d'une théorie interprétative de la traduction.

Aspects linguistiques

Dès 1958, Bar Hillel avait vu qu'il y avait un problème dans les situations d'ambiguïté.

Par exemple, pour les mots: "glace", "coq", "aimer", manger", "donner" on peut avoir plusieurs sens entre lesquels on peut choisir en fonction du contexte, car le mot acquiert le sens en contexte. S'il s'agit de l'ordinateur, il faut lui donner les outils pour qu'il puisse choisir lui aussi.

Aspects intralinguistiques

Après avoir réfléchi sur les aspects linguistiques qui se rapportent à une seule langue, il faut prendre en compte les aspects intralinguistiques, qui se rapportent aux différences des deux langues à traduire.

La traduction basée sur l'exemple

Dans ce système de traduction, on compare des textes et leur traduction et on extrait des éléments linguistiques de cette comparaison.

Le projet Eurolangues est une façon de valoriser le projet Eurotra. On compare les traductions et on trouve des exemples de traduction de sa base. Le système propose des variantes de traduction et c'est au traducteur d'en choisir.

On passe ainsi à des systèmes d'approche indirecte et à un traitement syntaxique. Au début, il s'agissait d'une approche directe et d'un traitement lexical.

Conclusion

On remarque donc, dans l'évolution de la TAO, deux tendances. Il s'agit, d'abord, du passage de l'approche directe à l'approche indirecte et ensuite d'un traitement lexical à un traitement syntaxique, même sémantique. Dans les deux cas, la sémantique n'est pas suffisamment prise en compte.

Bibliographie

Courtin, J, *Un metalangage pour l'analyse morphologique*, CNRS, 1969
Delavenay, É., *La machine à traduire*, PUF, 1963
Fisher, R., *Dictionnaire anglais-français des nouvelles technologies*, Eyrolles, 1987
Mounin, G., *La machine à traduire : histoire des problèmes linguistiques*, Mouton, 1964
Vauquois, B., *La Traduction automatique à Grenoble* Dunod - Association Jean-Favard pour le développement de la linguistique quantitative, 1975

BDD-A5464 © 2006 Universitatea din Pitești
Provided by Diacronia.ro for IP 216.73.217.142 (2026-11-15 02:40:41 UTC)