

La théorie du procédé constructif

Cette théorie, dont certains aspects ont été exposés au **Coling 94**, et qui bénéficie d'un nouvel intérêt en raison du système automatique de traduction, au stade expérimental, projeté par N.T.T. (Laboratoires des Sciences de la Communication) (sens J/E), est issue des travaux d'une grande école linguistique japonaise. En cours de formalisation avancée, grâce au Pr. Masahiro Miyazaki, elle s'appuie sur un lointain cousinage avec la Grammaire Générale de Port-Royal (1660) et plus récemment avec la Grammaire Générative de Chomski.

Un des fondements essentiels de cette théorie réside dans la théorie du sens, notamment au plan du comportement de l'être humain, vis-à-vis d'un autre, quand il s'agit de transmettre une signification. Selon cette théorie, trois processus interagissent : celui du Monde des Objets du discours, celui de la Reconnaissance de ce monde par son interlocuteur, et celui de leur Expression. Une relation dynamique s'établit entre ces trois pôles pour construire un sens intelligible par deux êtres en communication (voir schéma 1).

L'émetteur communique, de façon abstraite ou concrète, son rapport au monde des objets - il les identifie et y attache sa propre image - qu'il conceptualise. Il recourt à des règles linguistiques pour transmettre ce qu'il a *reconnu*.

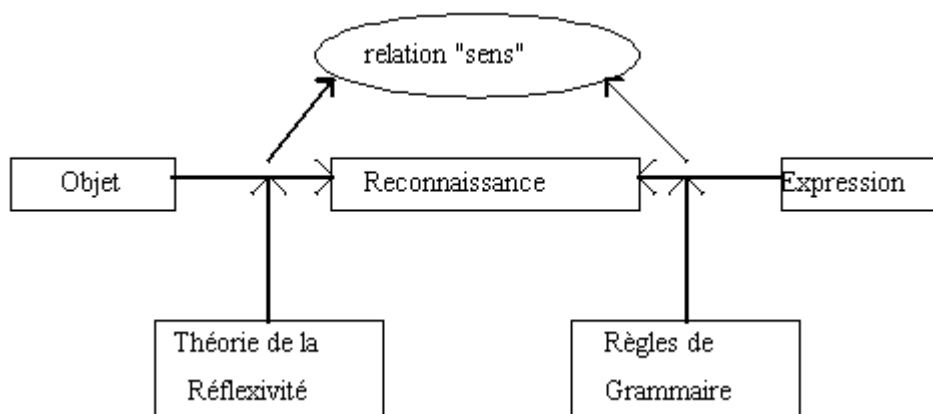


Schéma 1 : théorie du Procédé Constructif (Motori Tokieda)

Les objets du monde extérieur recouvrent trois notions : la *Substance*, c'est-à-dire les concepts de base, les *Attributs* (couleur, taille, possesseur, action effectuée sur l'objet) et les *Relations* entre les objets en question. L'émetteur, lui, constitue son identité, en se projetant dans la représentation des choses. L'expression finale est donc issue de la combinaison entre reconnaissance objective et représentation subjective.

Pour la langue japonaise, langue agglutinante, l'expression finale englobe une partie objective représentée par les noms, les verbes et les adjectifs, et une partie subjective représentée par des "joshis", c'est-à-dire des suffixes. Donc, le japonais accumule les mots et n'indique l'inflexion

subjective qu'à l'aide de ces suffixes ajoutés au verbe ; alors que l'anglais (et le français), langues inflectives, sont construites d'une manière diamétralement différente et affectent aux signes linguistiques des catégories nominales ou verbales.

Selon les concepteurs de cette théorie, ce transcodage d'un système de représentation linguistique à un autre est nécessaire pour aller du japonais vers une autre langue indo-européenne. La T.C.P. repose elle-même sur une méthode de traduction dite multi-niveaux (voir schéma 2). Ici, les connaissances du traducteur en matière de catégories subjectives sont mises en relation avec le caractère subjectif de la phrase à traduire. Ensuite, le caractère objectif de cette phrase est rendu objectivement en anglais, en respectant les niveaux idiomatique, sémantique-valence verbale et standard. Enfin, les catégories subjectives servent à former une phrase anglaise correcte, à partir de la structure objective de la phrase.

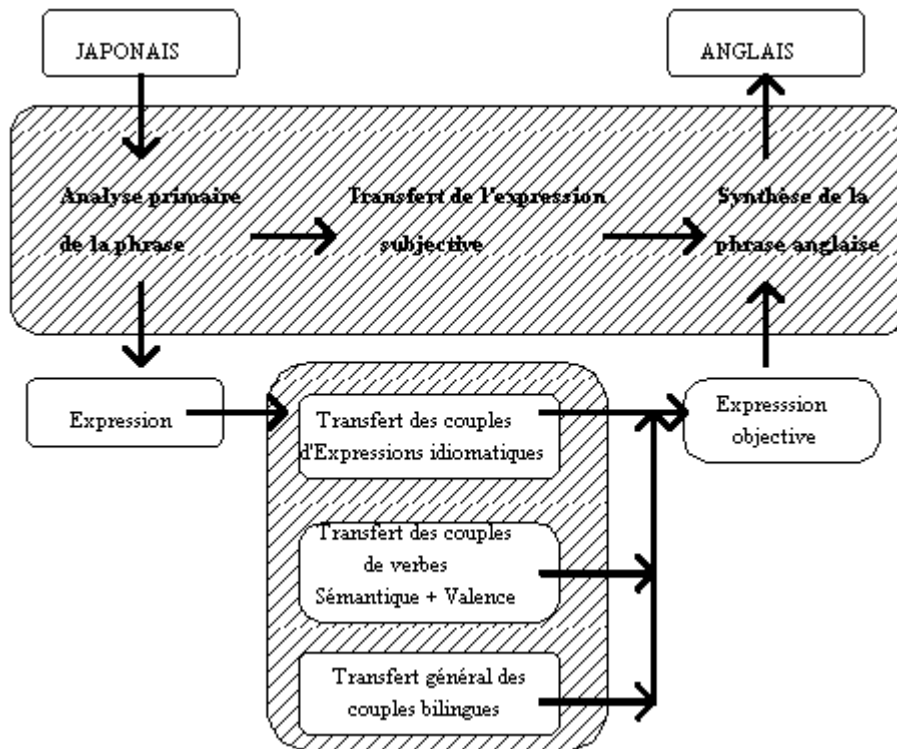


Schéma 2 : méthode de traduction dite multi-niveaux

Le traducteur conçu par N.T.T., ALT-J/E est doté d'un **dictionnaire sémantique** de 100 000 mots communs, 200 000 noms propres, 70 000 termes techniques et 30 000 autres mots, tous sont classés en 3000 catégories sémantiques, ventilés en 80 domaines et regroupés par style. Il est également pourvu d'un **dictionnaire de transfert** de 16 000 structures verbales dans le sens japonais/anglais, ainsi que d'un **dictionnaire de règles syntaxiques et sémantiques** adaptées au domaine d'application requis. Celui-ci permet de prendre en compte le *contexte* au niveau d'un paragraphe, de traduire des expressions idiomatiques, de réécrire automatiquement des phrases japonaises en phrases standard, de traduire spécifiquement les enclitiques japonais, de traiter les appositions, de générer des nombres et de comptabiliser les noms en anglais, de générer des pronoms possessifs en anglais, enfin d'ordonner correctement les mots pour l'écriture des adresses.

On passerait d'un taux de cohérence et de fidélité de 30 %, (avec les meilleurs systèmes japonais/anglais du marché) à quelque 80 % avec cette procédure.

Notons, pour clore, que lors du **Coling 94**, le traitement automatique des langues naturelles a représenté une part importante des communications. En traduction automatique, en particulier, plusieurs méthodologies ont fait l'objet d'exposés :

- la GBMT ou traduction automatique basée sur la grammaire (TABR)
- la EBMT ou traduction automatique fondée sur l'exemple (TABE)
- les méthodes statistiques, chères à IBM

Il a été affirmé qu'une certaine association des 3 types de méthodes pouvait être bénéfique. Les langues dont il a été le plus question, furent, dans l'ordre, le japonais, l'anglais, le coréen, le chinois, suivis de loin par l'allemand, l'espagnol, le français et l'indonésien. Il est à noter que le Japon garde un net avantage dans les échanges scientifiques avec la France (1 chercheur français présent à la conférence, pour 6 chercheurs japonais présents lors d'une conférence en France).

Au cours de la visite des laboratoires d'ATR, diverses théories ont été exposées, notamment la rapidité de traduction grâce aux ordinateurs à structure parallèle ou massivement parallèle, l'utilisation du contexte pour la résolution des ambiguïtés syntaxiques, la génération automatique de grammaire à partir d'un texte arbitraire, le concept de *white board*, exposé par Christian Boitet (GETA-IMAG).