

**LES MÉCANISMES DISCURSIFS DU SENS ET DE SIGNIFICATION;  
DE COMPRÉHENSION ET D'INTERPRÉTATION DANS LA  
PERSPECTIVE D'ÉTUDES SUR LA TRADUCTION AUTOMATIQUE /  
SENSE AND SIGNIFICATION DISCOURSE MECHANISMS; ON  
COMPREHENSION AND INTERPRETATION IN THE  
PERSPECTIVE OF STUDIES IN AUTOMATIC TRANSLATION /  
MECANISMELE DISCURSIVE ALE SENSULUI ȘI SEMNIFICAȚIEI.  
DESPRE ÎNȚELEGERE ȘI INTERPRETARE DIN PERSPECTIVA  
STUDIILOR REFERITOARE LA TRADUCEREA AUTOMATĂ<sup>1</sup>**

*Abstract:* This text presents the reflections on the meaning and understanding in the context of studies on machine/automatic translation. How to define and describe the understanding? Indeed, research on machine translation demonstrate the problem in the theoretical approach to the concepts of "meaning" and "sense", "understanding" and "interpretation", which is a serious impediment in transferring linguistic data to mathematical language algorithms.

*Key words:* understanding, interpretation, meaning, signification, automatic/machine translation, discourse analysis, Linguistic discourse analysis

*Résumé:* Ce texte fait état des réflexions sur le sens et la signification en langues dans la perspective d'études sur la traduction automatique. Comment définir et décrire la compréhension et l'interprétation face à de nouveaux défis technologiques, telle est la question de départ. En effet, les recherches sur la traduction automatique démontrent les lacunes dans l'approche théorique des notions de « signification », de « sens », de « compréhension » et d'« interprétation », ce qui constitue un véritable écueil dans le transfert des données linguistiques vers la langue mathématique des algorithmes.

*Mots-clés:* compréhension, interprétation, sens, signification, traduction automatique, discours, linguistique discursive.

En partant du constat de Georges Kleiber (1999 :5) concernant le sens à savoir « Le sens résiste à tout, sauf à la ... compréhension », nous nous proposons, dans le cadre de cette étude, de réfléchir sur cette compréhension en langue justement. Comment la définir et surtout la décrire face à de nouveaux défis technologiques, telle est la question de départ. En effet, les recherches sur la traduction automatique démontrent les lacunes dans l'approche théorique des notions de « signification », de « sens », de « compréhension » et d'« interprétation », ce qui constitue un véritable écueil dans le transfert des données linguistiques vers la langue mathématique des algorithmes.

#### **Linguistique vs ingénierie**

Le fait est que la problématique des traductions est devenue l'un des problèmes linguistiques majeurs de ces dernières années et dans la perspective de ces études changent

---

<sup>1</sup> Joanna Jereczek-Lipińska, Université de Gdańsk, Pologne, filjj@ug.edu.pl.

la recherche en sémantique et même plus globalement en linguistique. En sachant aussi que les recherches portant sur le sens dépassent aujourd'hui plus que jamais avant le champ disciplinaire de la linguistique et devient ainsi un terrain d'investigations inter- et pluridisciplinaires.

Commençons par un bref rappel terminologique et une courte revue historique des toutes premières versions des traducteurs automatiques. Retracer leur évolution permet de constater leur fiabilité potentielle et réelle justement et les points faibles. Ensuite l'on s'interrogera si le sens a de la signification dans la perspective d'études portant sur la traduction automatique et ici l'on parcourra les théories linguistiques du sens et de la signification. Deuxièmement l'on essaiera de faire correspondre les différentes conceptions du sens et de la signification par rapport aux versions successives des traducteurs automatiques.

Ensuite, l'objet sera d'évoquer les nombreux problèmes et défis qui résistent dans l'art de la traduction automatique. Suivra la signification et le sens dans la perception d'études de la linguistique discursive. Et enfin, l'on terminera avec les défis et perspectives dans l'évolution à venir du phénomène de traduction automatique.

#### **Méthodologies de recherches**

Cet article s'inscrit dans la perspective d'études interdisciplinaires portant sur la linguistique discursive, sur l'ingénierie informatique et sur la traduction. L'objet est ici de démontrer la perception des relations entre les phénomènes complexes de sens et de signification, de compréhension et d'interprétation d'un côté et de ce combat permanent dans la création du traducteur automatique idéal voire optimal. Parcourons d'abord quelques notions principales inévitables, en essayant de mettre de l'ordre dans ce champ terminologique. Il s'agira d'abord de distinguer les notions de linguistique informatique, d'informatique (ingénierie) linguistique, de traitement automatique des langues, d'algorithme, de traduction automatique voire informatique et enfin, la notion de traduction assistée par ordinateur. Il importe de distinguer ici en fonction de l'apport humain dans le processus de traduction, les différents genres de traduction, à commencer par la traduction assistée par ordinateur, en passant par la traduction assistée par l'homme et enfin en arrivant à la traduction totalement automatique.

Ainsi, la linguistique informatique est un courant de linguistique qui s'occupe de la langue en partant des modèles mathématiques ce qui débouche par exemple sur des analyseurs morphologiques ou syntaxiques.

L'informatique voire ingénierie linguistique par contre sera l'utilisation des ordinateurs à des fins linguistiques, en d'autres termes c'est une science qui utilise les ordinateurs pour analyser les structures de la langue et pour construire des modèles d'analyse linguistique dans le but de créer des traducteurs automatiques et les dictionnaires électroniques.

Les chercheurs français (Rastier, Cavazza, Abeillé, 1994) distinguent encore au sein de cette ingénierie linguistique un courant d'études qu'ils appellent le traitement automatique des langues.

La traduction automatique consiste à utiliser des algorithmes dans la traduction de textes d'une langue à l'autre. Et l'algorithme est ici envisagé comme une suite cohérente et

bien définie d'actions nécessaires à la réalisation d'une tâche. C'est une méthode de calcul et d'analyse de données en l'occurrence linguistiques.

La traduction assistée par ordinateur (de l'ang. *computer-assisted translation* – CAT) est une forme de traduction assistée par ordinateur grâce à l'utilisation de logiciel CAT.

Dans ce qui suit nous nous proposons de distinguer les différents genres de traduction en fonction de l'apport du facteur humain dans l'action de traduire. L'on y distingue trois possibilités :

- la traduction assistée par ordinateur ; - c'est la traduction entièrement réalisée par le traducteur qui s'aide seulement de différents logiciels ou programmes d'ordinateur pour rendre le processus plus rapide et efficace, comme p. ex. de dictionnaires électroniques, logiciel CAT ou autres ;
- la traduction assistée par homme ; - c'est une traduction réalisée par machine et révisée et revue par l'homme, avant, pendant ou après le processus ;
- la traduction automatique – qui ne prévoit donc pas d'intervention humaine à aucun moment du processus, le texte de base est soumis à la machine qui ensuite génère un texte d'arrivée, c'est le plus grand défi qui reste à relever.

#### **Esquisse historique**

Les chercheurs du domaine situent les débuts des recherches portant sur la traduction automatique pendant la période de la deuxième guerre mondiale aux Etats-Unis où c'était la période du développement important des ordinateurs et de l'informatique. Durant les années 50-60 apparaissent les tout premiers logiciels composés de dictionnaires bilingues dont l'objectif était donc de traduire les unités lexicales sans avoir à tenir compte de la structure syntaxique. Ils étaient surtout connus en Grande Bretagne, URSS, en France, au Canada, au Japon et en Pologne.

Les années 60-70 c'est la période de nombreuses tentatives, pas toujours très réussies, de la création des dictionnaires des structures syntaxiques. D'où les recherches portant sur les différentes méthodes d'analyse de la syntaxe. Leurs résultats ont abouti aux versions de dictionnaires élargis à ces programmes d'analyse et de synthèse automatiques de la syntaxe. Seulement vu la capacité de mémoire opérationnel et la puissance, les premiers logiciels étaient destinés à traiter les textes techniques. Les principaux obstacles à l'époque étaient dus au fait que la description cohérente, théorique et systématique des langues avec la description des mécanismes de la sémantique et de la grammaire formelle y compris les structures syntaxiques n'avait encore vu le jour. C'est d'ailleurs à ce stade-ci qu'on en a ressenti un grand besoin.

L'on a tenté de créer des logiciels dont l'objectif était de trouver les équivalents de deux types à savoir, les équivalents sémantiques d'un côté et de signification de l'autre. Les chercheurs se sont concentrés ensuite sur la recherche de la méthode de la représentation formelle du contenu sémantique des phrases en partant de l'idée qu'il existe des sens universels. Mais on s'est vite rendu compte de l'illusion de ces recherches ne fût-ce que par les problèmes dus aux ambiguïtés constitutives des langues. Plus loin on cherchait, plus les problèmes d'ordre sémantique apparaissaient et dérangeaient la création du logiciel de traduction automatique optimal. Dans les années 70-80 on peut situer la période du

développement de la linguistique informatique qui a désormais subdivisé les recherches dans ce domaine en plusieurs étapes successives notamment la création :

- de la base mémorielle lexicographique (les dictionnaires) ;
- des logiciels de traitement de synthèse et d'analyse de la phrase d'une langue donnée ;
- des logiciels de traitement sémantique de textes ;
- des logiciels de traitement de textes spécialisés qui analysaient et le contenu sémantique et syntaxique en sachant que la caractéristique des textes spécialisés est l'emploi de la structure stabilisée, fort conventionnelle et d'un contenu sémantique limité et prévisible.

A cette époque, les chercheurs IBM ont publié les résultats des expériences portant sur le système Candide qui utilisait les méthodes d'analyses statistiques. Et parallèlement le groupe de chercheurs japonais a commencé les études sur les méthodes d'analyse de corpus considérés comme des ensembles de textes et des exemples de traduction.

En résumé, nous pouvons diviser les recherches sur le logiciel de traduction en fonction de la méthode utilisée en étapes suivantes : la méthode symbolique valable pour les premières versions de traducteurs et la méthode statistique utilisant les grands ensembles de textes (les corpus). Dans les années 90 ont commencé les études sur le traducteur vocal qui englobait la reconnaissance et la synthèse vocale basée sur ces méthodes en même temps.

#### **Les théories linguistiques du sens et de la signification**

Polanski définit l'explication des expressions comme la recherche des expressions équivalentes, mais plus analytiques. La signification serait donc considérée comme la relation reliant deux arguments et donc un prédicat composé de deux arguments et comme tel constitue des sujets d'études des chercheurs de différentes disciplines : de philosophie, de psychologie, de logique, de sociologique et sociolinguistique, d'informatique et de sémantique et de linguistique en général. Kleiber (1999) estime que nous pouvons distinguer au sein de ces études, deux catégories les théories extensionnelles de signification et d'interprétation ainsi que les théories intentionnelles de signification et de sens.

Ainsi, aurait-on à l'intérieur de cette première catégorie les théories suivantes :

- intentionnelle – qui part du point de vue de l'émetteur ;
- associationniste – qui s'oriente en fonction du point de vue du récepteur ;
- référentielle – qui présuppose la thèse de vérité universelle ;
- interactive – qui envisage les points de vue de l'émetteur et du récepteur.

La deuxième catégorie par contre se subdivise en théories suivantes :

- structurelle et formelle (le structuralisme, le fonctionnalisme) où donc le mot est au centre des études ;
- théorie de la grammaire générative et le distributionnalisme, le transformationnalisme ainsi que le modèle opérationnel de la syntaxe où l'étude se concentre sur la phrase ;
- la théorie de la sémantique logique, du texte, du cognitivisme et de la pragmatique – le texte constitue le point de départ des études.

Considérons à présent le sens du point de vue strictement sémantique quand elle apparaît comme discipline étudiant les relations entre les mots et les objets qu'ils

représentent. Les recherches sur les unités minimales de sens ont abouti à une théorie selon laquelle le sens se déduit de la somme des significations des composantes. Le sens était donc perçu comme un attribut de la forme et le processus de l'interprétation consistait à décoder le code contenu dans la langue.

#### **Les conceptions différentes du sens et de la signification et les versions des logiciels de traduction qui y correspondaient**

De ces différentes théories de signification et d'interprétation découlaient les différents modèles de logiciels de traduction, en voici un relevé.

Les tout premières versions du traducteur automatique étaient construites autour d'un dictionnaire bilingue où le sens était situé dans les unités lexicales. Ensuite sont venus les logiciels avec des méthodes d'analyse syntaxique basés sur des grammaires formelles et qui considéraient donc la phrase comme porteuse de sens. Les logiciels suivants contenaient l'analyse et synthèse automatiques de la syntaxe du texte spécialisé qui s'inspiraient des théories considérant le texte comme unité minimale de sens.

#### **Les difficultés et défis dans l'art de la traduction automatique**

Nous allons nous interroger sur ce qui constitue du point de vue sémantique les difficultés voire plutôt des défis dans le processus efficace de la traduction. Ainsi, le premier obstacle qui s'impose ici est la synonymie, même s'il n'y a pas de synonymes parfaits (par ex. entendre/écouter ne veut pas dire la même chose), leur existence complique la tâche du traducteur automatique. En sachant qu'il y a le choix, tout choix étant porteur de signification, les difficultés apparaissent d'emblée. Car choisir ici n'est jamais innocent. Choisir un mot plutôt qu'un autre qui lui est proche c'est s'exprimer en toute conscience et en assumer la responsabilité. Vient ensuite le phénomène d'homonymie, encore plus tordu et complexe, car sous la même graphie (p.ex. le verbe *gagner*) se cachent plusieurs mots donc significations donc de sens et donc aussi d'interprétations possibles. Le discours figé pose autant si ce n'est plus de difficultés dans les logiciels de traduction, sa complexité consiste aux problèmes de repérage d'abord et ensuite dans la recherche des équivalents qui n'existent souvent même pas. Comme le sens de l'expression ne correspond jamais à la somme des significations des composantes, il est extrêmement compliqué de les traduire pour un traducteur humain. Le phénomène à ne pas négliger est l'évolution des langues. La langue change et son évolution prépare souvent des surprises. La polysémie et l'ambiguïté provoquent également des cas de malentendus dans la communication entre locuteurs de la même langue, entre forcément locuteurs venant des cultures différentes, quoi dire de ce travail soumis à la machine. Viennent ensuite le sous-entendu et l'implicite voire même l'ironie qui se lisent derrière les mots, les phrases, le texte, au coup par coup et ne sont donc pas donnés comme tel dans le texte, en d'autres termes ils ne sont pas inscrits dans le code que l'on pourrait ensuite seulement décoder. Tout comme dans le cas des effets de sens où même celui qui prononce ou écrit le propos n'est pas capable au préalable de détecter voire éviter tous les effets de sens possibles et qui apparaîtront ou pas au récepteur. Les problèmes de syntaxe et les différences dans les structures syntaxiques entre les langues font l'objet de nombreuses analyses pour pouvoir les soumettre à la machine également sans avoir de succès spectaculaires. Viennent

enfin toutes sortes d'ambiguïté et d'ambivalence aussi bien au niveau lexical, comme illustre cet exemple :

*Il a trois passions : la culture, la culture physique et la culture des orchidées*

Ou au niveau morphologique (la façon de marquer le pluriel : *ils semblent normaux, les filles jouent, ils dorment, ils dansent*), tant au niveau syntaxique : *Quel auteur cite ce conférencier ?, la photo de Paul* ou encore *Attention au chien ! Attention aux enfants !*

Au niveau phonétique (dans le cas des traductions vocales) et enfin au niveau pragmatique le plus complexe des précédents : *C'est chaud !* –plusieurs interprétations possibles – *Mange, tant que c'est chaud !* ou *Touche pas car c'est chaud* ou autres.

### **Le sens et la signification dans la perspective de recherches en linguistique discursive**

Réfléchissons d'abord quelle est la motivation qui a entamé les études de linguistes discursivistes. Premièrement, le fait que les phrases pouvaient contenir les éléments qu'on ne pouvait pas interpréter au niveau de ces phrases. Deuxièmement, l'interprétation d'un propos donné ne se limite pas à la somme des significations des phrases qui le constituent. L'argument suivant portait sur le fait que le sens échappe à toutes les définitions fournies par les chercheurs des autres courants linguistiques. Pour la première fois, il devient non pas un objet d'études en soi mais un instrument d'analyse. Ce sens est considéré comme un phénomène dynamique et non pas statique et on accepte désormais qu'il soit imprévisible, on ne l'inscrit pas dans le code, comme étant une donnée préconstruite et donnée, comment interpréter pourtant les propos tout simple *la peur du policier, le train de Paris*. La linguistique discursive perçoit pour la première fois dans l'histoire de cette recherche linguistique le sens tel qu'il est donc comme phénomène hétérogène et complexe avec une structure irrégulière. On lui admet son élasticité, son caractère évolutif voire changeant et que l'on puisse le concevoir comme résultat des négociations et donc l'action en cours des locuteurs. Le sens est différentiel et unique, c'est un fait réel à un moment donné. Il y a de nombreux facteurs qui déterminent le sens et la signification :

- la situation voire le contexte ;
- la polyphonie ;
- la focalisation envisagée notamment par Nølke (2001).

Ainsi, la recherche du sens présuppose non pas le décodage, mais l'activité de l'émetteur et du récepteur car le sens naît dans le discours, dans l'interdiscours, il y a donc du sens et non pas le sens : *Il est 20h ; Tout à fait Thierry*. Il faudrait donc non pas un corpus lexicographique, un corpus de phrases, de textes mais plutôt un corpus d'usages donc de discours.

Ces constats ont permis de faire avancer de manière significative les recherches sur la traduction automatique car ont éliminé un certain nombre d'illusions sur ce sens qui pouvait facilement se soumettre à l'algorithme.

La notion de signification dans la traduction automatique

La recherche linguistique distingue trois types de significations :

- la signification locutionnelle qui présuppose la relation entre la langue et la signification lexicographique des signes. Ce type de signification ne dépend pas du contexte ;

- la signification logique qui se base sur la logique du propos et est également indépendante du contexte ;
- la signification contextuelle qui présuppose l'adaptation de la formule à la convention d'usage et de situation, elle dépend du contexte.

Toute traduction devrait prendre en considération les trois significations en sachant que la dernière pose aux traducteurs le plus de difficultés.

En 1997 Kay a envisagé un exemple suivant comme relevant de ce troisième type de signification : *Où voulez-vous que je me mette ?* Si l'on traduit de manière directe ce propos on aurait *Gdzie Pan/Pani chce / chcecie abym się umieścił ?* Ainsi, selon le contexte aurait-on :

- *Gdzie Pan/Pani chce (chcecie) abym usiadł? (Où je dois m'asseoir?)*
- *Gdzie mam zaparkować? (Où je dois me garer?)*
- *Gdzie mam się podpisać? (Où je dois signer ?)*

Ceci présuppose donc un choix entre au moins les propositions ci-dessus que la machine devrait trancher. On sait bien que ce n'est pas réel.

Il s'ensuit que la sémantique de l'époque des recherches en linguistique discursive respecte les principes suivants:

- la sémantique n'a de sens que si elle se rapporte à la réalité discursive,
- les recherches commencent par un usage pour aller vers le modèle théorique et la description qui en découle ;
- la sémantique est ouverte au cognitivisme ;
- la sémantique perçoit le sens comme étant un phénomène dynamique qui naît dans une négociation et au cours d'une interaction.

Les perspectives de recherches sur la traduction automatique – en guise de conclusion. Les recherches sur la traduction automatique sont condamnées à suivre les études discursives où la langue est perçue comme une base de données qui sont un ensemble de discours donc ensemble de réalités linguistiques classées en corpus. Ces différents discours avec leurs paramètres de situation, la description de la structure et des marques morphosyntaxiques et sémantique du texte composés avec de l'algorithme déboucheraient sur l'analyse et le repérage de l'équivalent et donc le processus de traduction efficace.

Les recherches en traduction automatique dépendent également de l'avancée au sein d'autres disciplines surtout qu'aujourd'hui c'est inter-, pluri- et transdisciplinarité qui dictent la loi.

L'on peut prévoir que les recherches sur la traduction automatique vont se concentrer sur deux axes à savoir sur :

- l'intelligence artificielle étant définie comme la base de données dynamiques, riche et élargie, actualisée en permanence et représentative de la complexité constitutive des langues ;
- l'algorithme actif et actualisé selon le besoin.

Le sens, la signification, la compréhension et l'interprétation démontrent qu'il est aujourd'hui possible de trouver les relations d'équivalence seulement au niveau du discours en acceptant l'approximation des réponses trouvées.

### ***Bibliographie***

- Abeillé, Anne, 2002, *Une grammaire électronique du français*, CNRS Editions, Paris.
- Adam, Jean-Michel, 1990, *Éléments de linguistique textuelle*, Editions Mardaga, Paris.
- Blanchard Sylvie, Korach Dominique, 1995, *Vocabulaire*, Editions Nathan, Paris.
- Carré, René, Jean-François Dégremont, Maurice Gross, Jean-Marie Pierrel et Gérard Sabah, 1991, *Langage humain et machine*, Presses du CNRS, Paris.
- Cavazza, Marc, Rastier, François, Abeille Jacques, 1994, *Sémantique pour l'analyse. De la linguistique à l'informatique*, Elsevier Masson, Issy les Moulineaux
- Dubois, Jean, 1969, *Grammaire structurale du français : la phrase et les transformations*, Paris : Larousse.
- Gross, Maurice, 1997, « Synonymie, morphologie dérivationnelle et transformations », in *Langue* vol. 31, n° 128, pp. 72-90, Larousse, Paris, <http://cat.inist.fr/?aModele=afficheN&cpsidt=2457286>
- Kay Paul & Brent Berlin, 1997, "Science ≠ Imperialism: There are non-trivial constraints on color categorization", in *Brain and Behavioral Sciences*, <http://www1.icsi.berkeley.edu/~kay/>
- Kleiber, Georges, 1999, *Problèmes de sémantique. La polysémie en questions*, Ed. du Septentrion, Lille.
- Labesse, Henri, 1985, « Un analyseur syntaxique du français », in *Revue québécoise de linguistique* (Numéro thématique : Linguistique et informatique), vol. 14, n° 2, pp. 103-117.
- Nølke, Henning, 2001, *Le regard du locuteur 2 : Pour une linguistique des traces énonciatives*, Paris, Éditions Kimé.
- Polański Kazimierz, Jurkowski Marian, 1993, *Encyklopedia językoznawstwa ogólnego*, Ossolineum, Warszawa.
- Rastier François, 1997, *Meaning and Textuality*, University of Toronto Press.
- Wehri, Éric, 1997, *L'analyse syntaxique des langues naturelles : problèmes et méthodes*, Masson : Paris.
- <http://theses.ulaval.ca/archimede/fichiers/22111/ch04.html#d0e890>

