

La théorie Sens ↔ Texte : histoire, théorie et applications

Françoise Bacquellaine
franba@letras.up.pt ; shirleybac@gmail.com
Faculdade de Letras da Universidade do Porto

Le titre de cette communication ressemble étrangement au titre du colloque. J'ai découvert dernièrement la théorie Sens-Texte dans deux contextes différents. Le premier est l'enseignement du Français Langue Étrangère (FLE), le second est la néologie. Cette approche originale de l'objet complexe qu'est toute langue naturelle a éveillé ma curiosité. Or, quel meilleur moyen pour découvrir une théorie que d'en retracer la genèse, d'en définir les fondements et d'en expérimenter les applications ?

C'est cette découverte récente que je voudrais partager avec vous aujourd'hui, car cette théorie marginale, longtemps ignorée par la communauté internationale de linguistes, suscite de plus en plus d'intérêt depuis une dizaine d'années, comme le soulignait Jasmina Milićević (2006 : 1-2). Ce succès tardif est dû à la concurrence des théories génératives-transformationnelles et de l'approche dominante, plutôt sémasiologique qu'onomasiologique, qui procède plutôt selon la perspective du récepteur que selon celle de l'émetteur. L'atout de la théorie Sens-Texte est sans aucun doute son large champ d'application couvrant plusieurs domaines de recherche en plein essor à l'époque actuelle : non seulement le traitement automatique des langues naturelles (TALN) et ses nombreux sous-domaines, mais aussi l'enseignement/apprentissage des langues, la lexicologie et la lexicographie.

Comme le titre l'indique, mon exposé se divise en trois parties : l'histoire de la genèse et du développement de la théorie Sens-Texte ; une présentation succincte de ses fondements : le concept de « langue », les trois postulats sur lesquels elle repose, la définition du modèle général et d'un modèle particulier,

celui de Mel'čuk ; pour terminer, je présenterai un exemple d'application à la lexicographie et à l'enseignement/apprentissage des langues.

1. Histoire

Au début de l'ère informatique, la traduction automatique a connu un essor remarquable aussi bien à l'Est qu'à l'Ouest. Les scientifiques étaient persuadés que l'on pouvait modéliser la langue comme tout autre phénomène naturel. Or, la tâche s'est révélée bien plus ardue que prévu. Arnold *et al.* (1994: 15) ont comparé l'évolution de la traduction automatique aux montagnes russes. Après la montée vertigineuse, due à l'enthousiasme du monde scientifique et à l'abondance de fonds attribués à la recherche dans les années 1950, on assiste à la dégringolade hallucinante au début des années 1960 et à une remontée en pente douce depuis le début des années 1970. La dégringolade s'explique par les maigres résultats malgré les efforts déployés, dont fait état le fameux rapport ALPAC (Automatic Language Processing Advisory Committee), commandé en 1964 par l'Académie des Sciences états-unienne et publié en 1966. Les investisseurs se tournent vers d'autres projets, mais les recherches reprennent timidement dans les années 1970, notamment au Canada. En effet, la traduction automatique et l'informatique en général sont capables de résoudre de nombreux problèmes dans ce pays bilingue, où la traduction occupe une place importante parmi les services, notamment publics. Les investissements consentis ont produit des outils performants tels que le système de traduction automatique METEO ou le Grand dictionnaire terminologique, deux exemples d'outils canadiens qui rendent des services inestimables aux traducteurs.

C'est dans les années 1960, dans un contexte de désenchantement face aux insuffisances de la grammaire générative-transformationnelle, qu'est née la théorie Sens-Texte à Moscou. Igor Mel'čuk est l'un de ses pères fondateurs et sa leçon inaugurale au Collège de France en 1997 présente les résultats de 30 ans de recherche pour mettre au point un modèle qui, selon lui, constitue « [l]e meilleur moyen d'[accéder] à la compréhension du fonctionnement de la langue », celle-ci étant elle-même « [l]e meilleur moyen d'[accéder] à la compréhension du fonctionnement de la pensée » (1997 : 37). Nous allons donc le suivre dans ce parcours.

Au début des années 1960, Mel'čuk découvre les *fonctions lexicales*, qu'il met au point avec l'aide d'Alexander Žolkovskij pour améliorer les performances de la traduction automatique en termes de *collocations*, que Mel'čuk définit comme des « cooccurrences lexicales restreintes » par opposition aux

« cooccurrences lexicales libres ». Il s'agit donc d'expressions phraséologiques telles que « tirar uma fotografia » en portugais qui correspond à « prendre une photo » en français. Ces fonctions « permettent d'établir les correspondances interlinguistiques entre cooccurrents contextuels de façon directe et logique » (1997 : note 12, p.39).

Juri Asprejan les rejoint pour former l'École sémantique de Moscou en 1965, année où Mel'čuk et Žolkovskij publient, en russe, les fondements de la Théorie Sens-Texte. Une présentation plus détaillée paraîtra deux ans plus tard et sera bientôt traduite en anglais puis en français en 1970. Mais l'équipe est dissoute car Mel'čuk est contraint à l'exil en 1976. Il continue ses recherches au Canada, où il s'installe en 1977.

A la tête d'une nouvelle équipe constituée de 4 à 6 membres, il publie 4 volumes du *Dictionnaire explicatif et combinatoire du français contemporain* (DEC) aux Presses de l'Université de Montréal : *Recherches lexico-sémantiques I* en 1984, *II* en 1988, *III* en 1992 et *IV* en 2000. Parallèlement, en 1995, il publie, avec André Clas et Alain Polguère, un ouvrage théorique – *Introduction à la lexicologie explicative et combinatoire* – qui « se veut une initiation systématique à la lexicologie moderne », comme l'indique l'éditeur De Boeck Université. Ces deux ouvrages, lexicographique et lexicologique, ont servi de tremplin à la théorie Sens-Texte parmi les linguistes francophones.

L'année 1997 est marquée par deux faits importants. D'une part, l'Observatoire de Linguistique Sens-Texte est fondé à l'Université de Montréal, d'autre part, Mel'čuk occupe la chaire internationale du Collège de France au cours de l'année universitaire 1997-1998.

Depuis 2003, une conférence internationale sur la théorie Sens-Texte est organisée tous les deux ans : à Paris, d'abord – ce qui constitue traditionnellement une sorte de consécration au sein de la Francophonie – puis à Moscou en 2005, à Klagenfurt, en Autriche, en 2007 et la prochaine doit se tenir à Montréal en juin 2009.

Le lexique actif du français, publié par Mel'čuk et Polguère en 2007, est beaucoup moins ambitieux que le DEC, mais il offre à la fois un exemple d'application lexicographique et pédagogique qui permettra d'illustrer cette théorie dans la troisième partie de mon exposé.

2. Théorie

Il est difficile de donner une idée claire de ce qu'est la théorie Sens-Texte en quelques minutes. Je vais donc me limiter aux traits principaux en commençant

par le point de départ, c'est-à-dire la conception de la langue sur laquelle repose cette théorie extrêmement complexe. Pour Mel'čuk et les co-fondateurs de la théorie Sens ↔ Texte, la langue est un mécanisme ou un système de règles qui permet au locuteur de faire deux choses : « parler » et « comprendre ».

« Parler », cela signifie que, grâce à la langue, l'émetteur peut « **faire correspondre à un sens**¹ qu'il veut exprimer **tous les textes** de sa langue qui, d'après lui, peuvent véhiculer ce sens **et choisir** celui qui passe le mieux dans les circonstances concrètes d'un acte langagier donné » ; d'autre part, « comprendre la parole », cela signifie que, grâce à la langue, le récepteur peut « **faire correspondre à un texte** qu'il perçoit **tous les sens** que, d'après lui, ce texte peut véhiculer **et choisir** celui qui passe le mieux dans les circonstances concrètes d'un acte langagier donné » (Mel'čuk, 1997 : 1).

Actuellement, l'ordinateur est capable de réaliser la première opération selon l'ensemble fini de données, de règles lexicales et grammaticales produites et encodées selon un modèle, qu'il soit génératif-transformationnel ou Sens ↔ Texte. L'obstacle principal au fonctionnement des moteurs de traduction automatique est justement le problème du choix qui suppose une initiative dont l'ordinateur est incapable (voir à ce sujet Alan Melby, 1995).

Outre cette définition du concept de *langue*, trois postulats complètent les fondements de cette théorie. Selon le premier, « [l]a langue est un système fini de règles qui spécifie une CORRESPONDANCE² multi-multivoque entre l'ensemble infini dénombrable de sens et un ensemble infini dénombrable de textes » (Mel'čuk, 1997 : 4). On peut représenter cette correspondance (1) selon la formule :

$$\{\text{RSém}_i\} \text{ langue}; \langle == \rangle; \{\text{RPhon}_j\} \mid 0 < i, j \in \infty \text{ (idem : 5)}$$

La partie à gauche du trait vertical représente la correspondance mathématique entre les représentations sémantiques, c'est-à-dire les « objets symboliques formels » représentant les « sens », et les représentations phoniques, c'est-à-dire les « objets symboliques formels » représentant les « textes » (*ibidem*). Les sens et les textes prennent forme grâce à la « langue », au système linguistique quel qu'il soit.

¹ C'est moi qui souligne.

² C'est l'auteur qui souligne.

À droite du trait vertical se trouvent les conditions auxquelles est soumise la transposition du *sens* en *texte* et vice versa. Si les règles constituent un ensemble fini, les *sens* et les *textes* sont en nombre infini ($0 < i, j < \infty$), mais on peut en dénombrer une partie, par exemple dans un *corpus*. C'est pourquoi cet ensemble est à la fois infini et dénombrable, ce qui permet à Mel'čuk de le qualifier de « discret » au sens mathématique et informatique du terme.

Selon le deuxième postulat, « [l]a correspondance (1) doit être décrite par un DISPOSITIF LOGIQUE, qui constitue un modèle fonctionnel de la langue de type Sens-Texte ; il doit être élaboré et présenté dans la direction Sens → Texte » (*ibidem*). Si la langue permet la correspondance dans les deux sens, la théorie Sens-Texte se limite à l'approche onomasiologique, au parcours dans le sens de la synthèse ou de la production de la parole et ne se préoccupe pas du parcours inverse, sémasiologique, dans le sens de l'analyse ou de la compréhension de la parole. Le modèle Sens-Texte est doublement fonctionnel : d'une part, il permet de représenter, de « modéliser » le comportement ou le fonctionnement de la langue ; d'autre part, il est organisé comme une fonction au sens mathématique du terme.

Quant au troisième et dernier postulat, il pose que

[d]ans la description de la correspondance (1), deux NIVEAUX INTERMÉDIAIRES de représentation des énoncés sont nécessaires pour mettre en lumière les faits linguistiques pertinents : la représentation SYNTAXIQUE [= RSynt], qui correspond aux régularités spécifiques à la PHRASE, et la représentation MORPHOLOGIQUE [= RMorph], qui correspond aux régularités spécifiques au MOT (*idem* : 6).

Comme tout modèle linguistique, un modèle Sens-Texte est une entité créée pour représenter l'entité étudiée ; c'est un système abstrait d'expressions symboliques, de règles ; c'est un modèle fonctionnel permettant de représenter, de modéliser le comportement observable d'un objet dont la structure interne est inobservable, ce que Mel'čuk appelle une « boîte noire » (*idem* : 4). Un tel modèle est facilement « falsifiable (au sens de Popper) » (*ibidem*), c'est-à-dire qu'un seul contre-exemple suffit à en démontrer l'inadéquation.

La particularité du modèle Sens-Texte réside dans plusieurs de ses caractéristiques que je résume ici. Il est discret et doublement fonctionnel. L'approche est onomasiologique. Il est aussi global et intégral puisqu'il tient compte de toutes les composantes de la langue et tend donc à la présenter « comme un tout indivisible » (*idem* : 7). Il est « équatif » ou « traductif » dans

la mesure où il fait correspondre à chaque représentation sémantique toutes les représentations phoniques qui peuvent l'exprimer dans une langue donnée et qui permettent donc au locuteur de « traduire » un *sens* donné en un *texte* qui l'exprime (*idem* : 6-7). Enfin, ce modèle se fonde sur la synonymie des énoncés ou « paraphrasage » puisque, à un sens donné, correspondent plusieurs textes synonymes (*idem* : 7).

Il s'agit là du tronc commun. Le modèle particulier que Mel'čuk a présenté au Collège de France il y a une dizaine d'années a bien sûr évolué et nous en verrons une application récente pour terminer cet exposé. La caractéristique « la plus générale et la plus saillante » de ce modèle, est, comme il le dit lui-même, « son caractère stratificationnel³ et modulaire » (*idem* : 21). Cette « stratification » découle du troisième postulat qui introduit des niveaux intermédiaires entre la représentation sémantique et la représentation phonique. Mel'čuk la schématise ainsi (*idem* : 8):

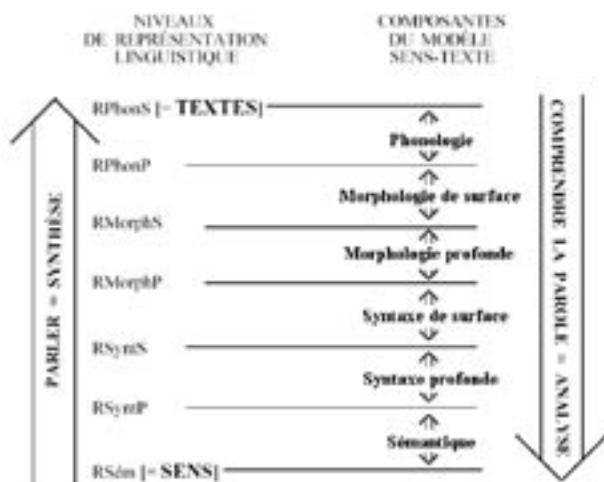


Figure 1 – Modèle Sens-Texte

La flèche de gauche indique le parcours onomasiologique de la synthèse et celle de droite, le parcours sémasiologique de l'analyse. Dans la colonne de gauche se trouvent les niveaux de représentation linguistique : sémantique,

³ L'expression « stratifié » employée par Polguère (1998) me semble plus correcte.

syntactique, morphologique et phonologique. Seul le niveau de représentation sémantique n'est pas dédoublé. Tous les autres se divisent en représentations profonde et de surface, la représentation de surface phonique étant le texte lui-même, écrit ou oral. Tous ces niveaux sont purement linguistiques, alors que le niveau de la représentation sémantique profonde sortirait du contexte exclusivement linguistique.

Chaque représentation est constituée d'une structure de base et de plusieurs autres structures qui se superposent à la première et la précisent. Les structures sont des objets formels. Par exemple, la structure de base d'une représentation sémantique est sa structure sémantique, « qui reflète le sens propositionnel, ou objectif, de l'énoncé représenté » (*idem* : 9), mais elle comporte également une structure sémantico-communicative, exprimant les oppositions telles que thème/rhème ou emphatisé/neutre, et une structure rhétorique, reflétant les intentions du locuteurs, telles que l'ironie, le pathétique, les registres de langues, etc.

Dans la colonne de droite, figurent les composantes, ou « modules », du modèle Sens-Texte de Mel'čuk. Un module est « un ensemble de règles qui assurent la correspondance entre les représentations de deux niveaux adjacents » (*idem* : 15).

Par souci de concision, je ne m'étendrai pas sur les modules et les structures, dont on trouvera des exemples dans la bibliographie. Disons simplement que le principe directeur du modèle Sens-Texte est de séparer le plus possible la représentation et la manipulation des phénomènes linguistiques qui sont de nature différente. Contrairement à ceux qui cherchent à tout prix à « construire une représentation unifiée de la phrase qui exprime SIMULTANÉMENT les liens hiérarchiques, anaphoriques et linéaires entre les éléments lexicaux », le mot d'ordre du modèle Sens-Texte est « distinguer et séparer » : « chaque aspect identifiable d'un fait à décrire reçoit une représentation autonome, pour établir ensuite des règles de correspondance entre ces diverses représentations » (*idem* : 21).

3. Applications

Les modèles Sens-Texte peuvent s'appliquer non seulement au traitement automatique des langues naturelles, notamment à la traduction automatique, à la génération de textes et aux systèmes d'interrogation de bases de données en langue naturelle, mais aussi à l'enseignement/apprentissage des langues maternelles et étrangères et à la création d'ouvrages de référence tels que les dictionnaires de collocations, de régimes ou de synonymes, les répertoires terminologiques, les grammaires de consultation ou les manuels. L'exemple du

Lexique actif du français de Mel'čuk et Polguère, publié en 2007, nous permettra de conclure cette brève présentation de la théorie Sens-Texte sur deux types d'application : lexicographique et pédagogique.

Le LAF présente quatre caractéristiques : c'est un dictionnaire « spécialisé » (Mel'čuk et Polguère, 2007 : 14) « qui aide à s'exprimer » (*idem* : 15), c'est un « échantillon » (*idem* : 17) de dictionnaire « formalisé » (*idem* : 16).

Il ne s'agit pas d'un dictionnaire spécialisé au sens habituel du terme, c'est-à-dire une nomenclature ou un répertoire terminologique. Le LAF est un dictionnaire spécialisé car il se focalise sur la description en profondeur de deux phénomènes particuliers : les collocations et les dérivations sémantiques (*cf. infra*).

Contrairement à la plupart des dictionnaires qui servent à comprendre et sont donc un outil d'analyse selon le parcours Texte à Sens, le LAF aide à s'exprimer. C'est un « outil permettant à son utilisateur d'encoder linguistiquement sa pensée » (*idem* : 15) selon le parcours Sens à Texte.

Le LAF est un dictionnaire expérimental qui « ne couvre qu'un sous-ensemble (...) du lexique français » (*idem* : 17). Il s'agit donc d'un échantillon de dictionnaire destiné à être complété. Son but est à la fois lexicologique et pédagogique. Il vise à « proposer une nouvelle façon de décrire le lexique » et à « servir de modèle de référence pour l'élaboration d'une pédagogie de l'enseignement des connaissances lexicales » (*ibidem*).

Enfin, le LAF est un dictionnaire formalisé car l'usage de nombreux formalismes permet de présenter une grande quantité d'informations sous une forme compacte. Cette caractéristique présente toutefois l'inconvénient que l'utilisateur « doit d'abord se familiariser avec la structure et les formalismes de ce dictionnaire pour en tirer le meilleur parti possible » (*idem* : 16). C'est pourquoi la première partie de ce dictionnaire original constitue une sorte de mode d'emploi qui doit d'abord être digéré.

Le LAF est un dictionnaire spécialisé qui ne contient que « des mots français qui contrôlent un ensemble significatif de dérivations sémantiques ou de collocations » (*idem* : 14). Ces *vocables* sont polysémiques comme « coup de balai » ou monosémiques comme « abeille ». Un vocable polysémique se compose de plusieurs *lexies* homonymes dont les sens sont apparentés de façon plus ou moins nette. Par exemple, le vocable « coup de balai » correspond à deux *lexies*. La première signifie au sens propre « nettoyage » et la deuxième, au sens figuré, « élimination ».

« Nettoyage » et « élimination » sont des *étiquettes sémantiques* correspondant à chaque lexie du vocable « coup de balai ». Les étiquettes sémantiques des lexies du LAF sont classées selon une *hiérarchie* – la « hiérarchie des étiquettes sémantiques » (HES)⁴. La HES est donc propre au français. Elle est fondée non seulement sur la relation hyperonyme/hyponymes mais aussi sur la synonymie et d'autres « liens qui relèvent de la dérivation sémantique » (*idem* : 43). Il ne s'agit pas seulement d'établir des relations entre les noms, mais aussi entre toutes les parties du discours. Les auteurs rejettent la segmentation « en **plusieurs**⁵ sous-hiérarchies (...) correspondant [chacune] à une partie du discours », ils considèrent que « la classification en parties du discours est avant tout grammaticale et non sémantique » et qu'« [e]lle ne doit donc pas conduire à une segmentation de la hiérarchie sémantique du dictionnaire » (*idem* : 44). Cette hiérarchie est « relativement 'plate' », car elle se développe plutôt en largeur qu'en profondeur, et elle « n'est pas strictement arborescente » puisque « certaines de ses étiquettes possèdent plus d'une mère ».

Les lexies décrites dans le LAF sont aussi organisées par *champs sémantiques*. Un champ sémantique est « un ensemble de lexies qui se regroupent naturellement dans l'esprit du locuteur car leurs définitions partagent une composante de sens particulière, qui joue un rôle relativement important dans ces définitions » (*idem* : 45). On trouvera par exemple « féliciter », « élogieux » et « exposé » parmi des dizaines d'autres lexies dans le champ sémantique de la communication.

Le LAF ne contient pas de définition au sens traditionnel du terme. Il est construit selon deux axes de description lexicographique : les *dérivations sémantiques* et les *collocations*. Ces deux axes reposent sur des liens lexicaux orientés, paradigmatiques dans le cas des dérivations sémantiques et syntagmatiques dans le cas des collocations. En effet,

[u]ne dérivation sémantique est une relation particulière entre deux lexies : une lexie de départ [par exemple *coup de soleil* ou *galop*] et une lexie « sémantiquement construite » à partir de celle-ci [par exemple *brûlure* ou *course*]. Il s'agit donc d'un lien lexical orienté : de la lexie d'origine vers la lexie dérivée. Ce type de lien est appelé **lien paradigmatique**⁶ (*idem* : 22).

⁴ Disponible sur le site de l'Observatoire de linguistique Sens-Texte : <http://olst.ling.umontreal.ca/>.

⁵ Ce sont eux qui soulignent.

⁶ Ce sont les auteurs qui soulignent.

Par contre,

Une collocation est une expression linguistique constituée de deux lexies, dont l'une – la base – contrôle l'emploi de l'autre – le collocatif. Une collocation n'est donc pas une relation. Cependant, une collocation présuppose une relation particulière de contrôle fonctionnel entre les deux lexies qui la constituent : la base contrôle le choix du collocatif. Ce contrôle est également un lien orienté : de la base du collocatif vers le collocatif. Ce type de lien lexical est appelé *lien syntagmatique* (*ibidem*).

Comme il s'agit dans les deux cas de liens lexicaux orientés, l'équipe de Mel'čuk a modélisé ces deux phénomènes au moyen d'un même outil descriptif : les *fonctions lexicales* correspondant à un ensemble restreint de patrons de liens lexicaux dérivationnels et collocationnels identifiés en 1995 dans l'*Introduction à la lexicologie explicative et combinatoire*. Comme l'explique l'auteur, « [c]haque lien lexical standard est modélisé comme une fonction mathématique qui s'applique aux lexies pour retourner la liste de leurs dérivés sémantiques ou de leurs collocatifs » (*idem* : 23). En langage mathématique, cela donne : $f(x) = y$, où « f » représente le lien lexical et « y » toutes les expressions possibles du lien lexical pour la lexie « x ». Par exemple, la fonction *Instrument* qui associe à la lexie-argument « coup de balai » la valeur du dérivé sémantique « balai » s'écrit : « *Instrument* (coup de balai) = balai » et la fonction *Donné de façon peu soignée* qui associe à la même lexie-argument la valeur des collocatifs « petit, rapide » s'écrit : « *Donné de façon peu soignée* (coup de balai) = petit, rapide ».

La macrostructure du LAF se présente comme un dictionnaire standard. Il comporte des entrées de vocables, un sommaire de l'entrée dans le cas des vocables polysémiques et un ou plusieurs articles décrivant chacun une lexie.

Quant à la microstructure, le mot-vedette est directement suivi de ses caractéristiques grammaticales : partie du discours, genre, etc., mais aussi, s'il y a lieu, des indications concernant le registre, les contraintes morphologiques (invar., pas de pl.,...) ou syntaxiques (emploi adverbial, ...). Jusqu'ici, rien de bien original puisque le *Petit Robert* fournit le même genre de données. Viennent ensuite – et c'est là que le LAF se distingue des autres – l'étiquette sémantique (« nettoyage » et « élimination » pour « coup de balai »), la formule actancielle dans le cas des prédicats sémantiques qui possèdent des actants (« Coup de balai donné par l'individu X dans le lieu Y »), les synonymes (« balayage »), les dérivations sémantiques (hyperonyme : « nettoyage ») et les collocations (soigné : « bon », « grand ») ainsi que les informations nécessaires à leur emploi (prépositions, emploi ou omission de l'article, position par rapport à la lexie, ...).

Finalement, le LAF propose, comme le *Petit Robert*, un ou plusieurs exemples et, le cas échéant, la liste des locutions incluant le mot-vedette. Par exemple, à l'entrée du mot-vedette « balai », on trouve comme locutions « coup de balai », « balai d'essuie-glace », « balai mécanique », « comme un balai », « manche à balai », « balai à chiottes » et « Du balai ! ». L'article du mot-vedette « coup de balai » contient toutes les rubriques sauf la liste de locutions et il permettra d'illustrer la microstructure (*idem* : 168-169) :

COUP DE BALAI, locution nominale, masc

I NETTOYAGE : *Il a donné un coup de balai dans sa chambre.*

II ÉLIMINATION : *Il faudrait un sérieux coup de balai dans l'administration publique.*

I NETTOYAGE

Coup de balai donné par l'individu X dans le lieu Y [Prép_{loc} N]

☞ balayage

Génér. nettoyage **Instrument avec lequel X donne un C.D.B.** balai **Donné de façon soignée** bon, grand | antépos **Donné de façon peu soignée** petit, rapide | antépos [X] **donner un C.D.B.** donner, passer [ART ~ Prép_{loc} N_Y]

- *Il suffit d'un dernier coup de balai pour finir de nettoyer ta chambre.*

II ÉLIMINATION

Coup de balai donné par la personne X [= de N, A_{poss}] visant les individus ou l'entité sociale Y [= contre N (**ex.** *contre les politiciens corrompus*), Prép_{loc} N (**ex.** dans les milieux politiques)]

☞ élimination, épuration, nettoyage ☞ ☞⁷ congédiement

Important grand, important, sérieux ! antépos [X] **donner un C.D.B.** donner, effectuer, opérer [Art ~ contre/ Prép_{loc} N_Y]

Cet acte pourrait mener à un sérieux coup de balai dans les milieux de la lutte antiterroriste. La nouvelle administration effectue un grand coup de balai à la tête de la fonction publique.

⁷ « Introduit une liste de lexies qui, sans être des synonymes du mot-vedette, sont tout de même perçues comme en étant sémantiquement très proches » (*idem* : 74)

Le LAF est donc conçu comme un outil de recherche d'informations permettant, d'une part, de découvrir des familles sémantiques de lexies, soit à partir de la HES disponible en ligne, soit à partir des étiquettes sémantiques ou des champs sémantiques consignés dans les annexes 1 et 2 du LAF (*idem* : 51-52), et, d'autre part, d'« exprimer linguistiquement ses idées » selon le parcours onomasiologique : le locuteur qui se demande comment on appelle « l'espace de comptoir dans une gare où on achète les billets » trouvera la lexie « guichet » dans l'article du vocable « gare » (*idem* : 48-51).

Mais ce dictionnaire a aussi été conçu comme un outil pédagogique favorisant le parcours du réseau lexical du français plutôt que la simple consultation pour permettre aux apprenants de prendre conscience des relations paradigmatiques et syntagmatiques existant au sein de ce réseau extrêmement complexe. Les auteurs proposent trois types d'activités : exploration du lexique, production linguistique et lexicographie.

L'exploration de la structure du lexique peut se faire par l'étude de la polysémie, trop souvent absente des cours de langues. La structure polysémique d'un vocable donné peut être comparée, d'une part, à celle d'un dérivé nominal, verbal, adjectival ou adverbial, d'autre part, à celle de vocables dont certaines lexies sont classées sous la même étiquette sémantique. Ce type d'activité permet de mieux comprendre le « développement, souvent anarchique et imprévisible, de la polysémie », d'identifier des « patrons de polysémie », c'est-à-dire des « familles de relations sémantiques » récurrentes (métaphores, métonymies, rétrécissement ou élargissement du sens, etc.), et de prendre conscience de l'importance du régime et des collocations propres à chaque lexie, si l'on veut parler correctement, c'est-à-dire faire vivre les mots (*idem* : 57-58). L'élaboration de mini-lexiques thématiques à partir des étiquettes et des champs sémantiques est une autre façon d'explorer la structure du lexique. Il ne s'agit pas d'élaborer de simples listes de lexies, mais bien de les analyser pour transformer chaque liste en réseau lexical structuré selon divers critères tels que « les types de signification, les parties du discours, les proximités sémantiques, etc. » (*idem* : 56-57)

Le LAF permet d'accéder aux dérivés sémantiques et aux collocations pour produire un texte à partir d'un sens, c'est-à-dire exprimer linguistiquement ses idées (*cf. supra*). Il permet aussi d'explorer le potentiel de paraphrasage, car « [l]a langue est (...) plus qu'un outil pour exprimer nos pensées : c'est un outil pour exprimer nos pensées de multiples façons » (*idem* : 58). Les règles de paraphrasage « mettent en jeu des procédés se situant au confluent du lexique et de la grammaire » (*idem* : 59) et ce type d'exercices présente donc un double intérêt pour le cours de langue.

Enfin, le LAF constitue le point de départ de nombreuses activités lexicographiques. Le LAF comporte des erreurs et les apprenants peuvent les corriger en utilisant diverses ressources traditionnelles et électroniques (*corpus*, informateurs, dictionnaires). Ils peuvent aussi le compléter en introduisant des quasi-synonymes, des synonymes dans d'autres registres de langue/variétés régionales, voire en créant les entrées de vocables féminins pour les vocables masculins décrits, ou vice versa, ce qui peut révéler des différences de structuration polysémique, comme entre « taureau » et « vache » ou « tigre » et « tigresse ». Ils peuvent enfin rédiger les définitions analytiques de lexies sémantiquement apparentées à partir de leur étiquette sémantique, de leur structure prédicative et de leurs éventuelles affinités sémantiques décrites par les liens paradigmatiques et syntagmatiques. Cette liste n'est pas exhaustive et les auteurs invitent les enseignants et les apprenants à faire preuve de créativité pour exploiter leur outil (*idem* : 59-61).

4. Conclusion

En somme, le LAF est un outil déroutant lorsqu'on l'ouvre pour la première fois, mais après avoir lu le mode d'emploi et navigué sur le site de l'OLST, il apparaît comme un outil innovant qui permet en effet le parcours du sens au texte et la réalisation d'activités pédagogiques sortant des sentiers battus. Toutefois, outre l'inconvénient que constituent ses nombreux formalismes, il est regrettable que les exemples soient forgés et manquent donc parfois de naturel.

Quant aux perspectives d'avenir de la linguistique Sens-Texte, Mel'čuk lui attribue une utilité pratique dans le TALN, l'enseignement/apprentissage des langues étrangères et maternelles et la création d'ouvrages de référence. Il lui attribue aussi un intérêt théorique « encore plus important », celui de développer « de bons modèles de la pensée humaine » afin de répondre au « besoin aigu d'un SUPERCERVEAU⁸, c'est-à-dire [de] machines capables de penser à une échelle que l'Homme seul ne saurait atteindre », face au problème crucial du XXI^e siècle, à savoir « le manque de ressources naturelles pour une population qui s'accroît » (Mel'čuk, 1997 : 36). Certes, « on ne pourra jamais comprendre comment notre cerveau traite les informations sans savoir comment la langue y fonctionne », certes, « les modèles fonctionnels de la langue [tels que le modèle Sens-Texte] (...) peuvent contribuer de façon essentielle à la compréhension que nous avons

⁸ C'est l'auteur qui souligne.

(...) de notre pensée » (*idem* : 37), certes, « la meilleure approche existant à ce jour pour mener à bien le traitement automatique des langues » est sans doute la théorie Sens-Texte (Polguère, 1998 : 9), mais, comme Melby, je doute que la machine parviennent jamais à *penser*, à prendre des décisions, des initiatives, à créer, à innover. Mel'čuk reconnaît lui-même que l'ensemble des sens et celui des textes sont tous deux infinis. Il admet que son modèle est synchronique et ne tient pas compte de la variation géographique et sociale. Or, la langue est un objet dynamique. Des néologismes lexicaux et syntaxiques sont créés tous les jours et de manière différente selon les régions, voire les individus. Il semble donc utopique d'élaborer un programme informatique qui aurait tout prévu, selon un ensemble fini de règles, consignées dans un dictionnaire et une grammaire. La machine ne fera jamais que ce pour quoi elle a été programmée.

Et quand bien même on y parviendrait, cela est-il vraiment souhaitable ? Comme le fait remarquer Melby avec humour en citant Douglas Robinson, pour conclure son article de 1995, un ordinateur doté de la capacité de faire des choix résultant de sa propre volonté, des choix éthiques dont il serait seul responsable, peut refuser un beau jour de faire ce qu'on lui demande : « Douglas Robinson (1992) puts it well. He asks whether a machine translation system that can equal the work of a human might not 'wake up some morning feeling more like watching a Charlie Chaplin movie than translating a weather report or a business letter' » (1995: 16).

J'espère avoir éveillé la curiosité de ceux qui, comme moi, découvrent seulement maintenant, 40 ans après sa genèse, la théorie Sens-Texte, son modèle fonctionnel et ses applications.

Références

- ALPAC 1966. *Language and machines: Computers in Translation and Linguistics*. A Report by the Automatic Language Processing Advisory Committee. Washington, D.C.: Division of Behavioral Sciences, National Academy of Sciences, National Research Council.
- Arnold, D.J.; Balkan, L.; Meijer, S.; Lee, R. 1994. *Machine Translation: an Introductory Guide*. Londres : NCC Blackwell.
- De Boeck Université, page consultée le 20 octobre 2008 sur le World Wide Web : <http://universite.deboeck.com/livre/?GCOI=28011100391790&fa=details> .
- Melby, A.K. 1995. *Why can't a computer translate more like a person ?* Article en ligne

- téléchargé le 30 novembre 2006 du World Wide Web : <http://www.mt-archive.info/Melby-1995.pdf> .
- Mel'čuk, I. 1976. *Das Wort. Zwischen Inhalt und Ausdruck*. München : Wilhelm Fink Verlag.
- Mel'čuk, I.; Arbatchewsky-Jumarie, N.; Elnitsky, L.; Iordanskaja, L., Lessard, A. 1984. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I*. Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk, I; Pertsov, N. 1987. *Surface Syntax of English. A formal model within the Meaning-Text Framework*. Amsterdam : John Benjamins Publishing Company.
- Mel'čuk, I.; Arbatchewsky-Jumarie, N.; Dagenais, L.; Elnitsky, L.; Iordanskaja, L.; Lefebvre, M.-N.; Mantha, S. 1988. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques II*. Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk, I.; Arbatchewsky-Jumarie, N.; Iordanskaja, L.; Mantha, S. 1992. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques III*. Montréal: Les Presses de l'Université de Montréal.
- Mel'čuk, I.; Clas, A.; Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Mel'čuk, I. 1997. *Vers une linguistique Sens-Texte*. Leçon inaugurale au Collège de France téléchargée le 07 juin 2008 du World Wide Web : <http://www.olst.umontreal.ca/pdf/MelcukColldeFr.pdf> .
- Mel'čuk, I.; Arbatchewsky-Jumarie, N.; Iordanskaja, L.; Mantha, S.; Polguère A. 1999. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques IV*. Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk, I. 2001. *Semantics and the Lexicon in Modern Linguistics*. Article téléchargé le 07 juin 2008 du World Wide Web : <http://www.olst.umontreal.ca/pdf/MelcukMTTIntro.pdf> .
- Mel'čuk, I.; Polguère, A. 2007. *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Bruxelles : De Boeck & Larcier.
- Milićević, J. 2006. A short Guide to the Meaning-Text Linguistic Theory. *Journal of Koralex*. 8 : 187-233. Téléchargé le 07 juin 2008 du World Wide Web : <http://www.olst.umontreal.ca/pdf/IntroMTTJM.pdf> .
- Polguère, A. 1998a. La Théorie Sens-Texte. *Dialangue*. 8-9 : 9-30. Téléchargé le 07 juin 2008 du World Wide Web : <http://www.olst.umontreal.ca/pdf/PolgIntroTST.pdf> .
- Polguère, A. 1998b. Pour un modèle stratifié de la lexicalisation en génération de texte. *Traitement Automatique des Langues (T.A.L.)*. 39 (2) : 57-76. Téléchargé le 07 juin 2008 du World Wide Web : <http://www.olst.umontreal.ca/FrEng/LexAPol.pdf> .
- Université de Montréal 2009 (dernière mise à jour). *Observatoire de linguistique Sens-Texte*. Dernière consultation le 08 avril 2009 sur le World Wide Web : <http://olst.ling.umontreal.ca/> .

Valente, R. 2002. *La « Lexicologie explicative et combinatoire » dans le traitement des unités lexicales spécialisées*. Thèse de doctorat, Université de Montréal, Faculté des arts et des sciences, Département de linguistique et de traduction. Téléchargée le 07 juin 2008 du World Wide Web : www.theses.umontreal.ca/theses/nouv/valente_r-s/these.pdf.