Un modèle linguistique fonctionnel : le modèle Sens-Texte

Igor Mel'čuk Observatoire de linguistique Sens-Texte Université de Montréal, Montréal, Canada igor.melcuk@umontreal.ca

1 Modèles dans la science

Comment sait-on ce qui s'est passé quelques secondes après le Big Bang? Personne n'était là, et aucun dispositif d'observation ne pouvait exister au moment de la création de l'univers. Cependant, nous savons beaucoup de choses sur l'univers de ces temps révolus, et ce, grâce à la construction des modèles cosmologiques abstraits – systèmes d'équations, qui, s'appuyant sur les lois physiques connues et la logique, relient entre eux des phénomènes observables qui ont eu lieu des milliards d'années après le Big Bang. À partir de ces modèles, les scientifiques peuvent tirer des conclusions sur l'état de l'univers aux différents stages de son existence, y compris les états précédant l'apparition de l'Homme.

Ce n'est qu'un exemple parmi des milliers et des milliers. Chaque fois qu'un chercheur n'est pas capable d'observer directement la structure interne de l'objet ou du phénomène étudié, il recourt à un modèle. Ce n'est pas une exagération que de dire qu'une science « dure », c'est, avant tout, la construction de modèles. On le sait pertinemment au moins depuis Galilée. « Il n'y a pas de savant qui ne pense continuellement par modèles – même s'il ne l'avoue ni aux autres, ni à lui-même » (Auger 1965: 4).

La linguistique, dont l'objet est la langue humaine, est dans la même position que la cosmologie. La langue, un système de règles extrêmement complexes, est encodée dans le cerveau de chaque locuteur et, de ce fait, elle est inaccessible à l'observation directe : les linguistes ne peuvent pas ouvrir les crânes ou pénétrer à volonté le cerveau avec leurs électrodes. La seule solution est alors le recours aux modèles. La Grammaire Générative de Chomsky a solidement implanté l'idée de modélisation en linguistique. En même temps, une grande contribution a été apportée par le travail d'envergure mené dans les diverses branches de la linguistique computationnelle. À présent, nous pouvons considérer comme acquis le principe de modélisation linguistique. Néanmoins, il nous faut préciser le concept même de modèle linguistique et établir les types de modèles les plus prometteurs. Le terme modèle est ambigu ; nous avons besoin d'une définition rigoureuse.

Soit une entité **E** (un objet ou un système d'objets); **E** fonctionne dans ce sens qu'elle reçoit des entrées (*inputs*) observables et produit les sorties (*outputs*) correspondantes, également observables. Le chercheur s'intéresse au fonctionnement de **E** (plutôt qu'à sa structure interne, qui n'est de toute façon pas observable). Pour décrire **E**, il construit un modèle fonctionnel de **E** – c'est-à-dire, **M**(**E**).

M(E) est un modèle fonctionnel de **E**, si et seulement si **M(E)** est un système d'expressions symboliques créé par le chercheur pour décrire le fonctionnement de **E**, c'est-à-dire que **M(E)** associe aux entrées données les mêmes sorties que **E**.

Le modèle **M** est *fonctionnel* dans les deux sens du terme : 1) **M** cherche à représenter le fonctionnement de l'entité étudiée et 2) il le fait en utilisant des fonctions (au sens mathématique), de sorte que **M** lui-même est en fait une fonction très complexe.

2 Le modèle Sens-Texte

Le modèle Sens-Texte sera caractérisé en trois temps : les postulats de la théorie sous-jacente, les représentations linguistiques qu'il utilise, et des exemples de quelques règles du modèle.

2.1 Les postulats sous-jacents au modèle

L'élaboration du modèle Sens-Texte a été entreprise il y a 50 ans à Moscou – par A. Zholkovsky et le présent auteur, avec quelques collègues, avant tout, Ju. Apresjan. Une esquisse assez détaillée du modèle Sens-Texte peut être trouvée dans Mel'čuk 2012. La théorie linguistique sous-jacente à ce modèle, ou la théorie linguistique Sens-Texte, est basée sur les trois postulats suivants.

Postulat 1

Une langue est un système de règles qui décrit la correspondance entre un ensemble infini de sens et un ensemble infini de textes.

Les sens sont représentés par des objets formels appelés les représentations sémantiques [RSém], les textes l'étant également par des objets formels appelés les représentations phonétiques [RPhon]. Le postulat 1 peut être réécrit symboliquement comme (1):

(1) {RSém_i}
$$\angle$$
langue® {RPhon_i} | $i \neq j$, $0 < i, j \le \infty$

D'un point de vue logique, la correspondance Sens-Texte (c'est-à-dire, « ∠langue® ») est bidirectionnelle, car elle représente une équivalence (entre sens et textes). Mais, dans une perspective linguistique, elle doit être étudiée et décrite dans la direction Sens ® Texte : la langue naturelle sert avant tout à parler, ce qui veut dire « à passer d'un sens à un texte correspondant ». La synthèse linguistique, ou la production de textes à partir de sens donnés, est plus importante pour la linguistique que l'analyse, ou la compréhension des textes. L'orientation de la recherche et de la description linguistiques − des sens vers les textes − donne une priorité absolue à l'étude de la synonymie, en particulier à la paraphrase linguistique (Miliéevié 2007).

Postulat 2

La correspondance Sens-Texte en (1) doit être décrite par un dispositif logique, ou un système de règles, qui constitue un modèle linguistique fonctionnel : un modèle Sens-Texte.

Un modèle Sens-Texte prend des sens, ou des RSém, comme entrées, et produit en sorties des textes, ou des RPhon, de la même façon que le font les locuteurs natifs. C'est dans ce sens qu'un modèle Sens-Texte est une fonction mathématique : $\mathbf{f}(RS\acute{e}m) = \{RPhon_j\}$. Appliquée à une RSém, cette fonction – ou bien ce modèle – produit l'ensemble de toutes RPhon (quasi-)synonymes qui correspondent à cette RSém selon le jugement des locuteurs.

La correspondance Sens-Texte est multi-multivoque. D'une part, une RSém peut correspondre à un nombre astronomique de RPhon (plusieurs millions), car la synonymie est incroyablement riche ; d'autre part, une RPhon peut aussi exprimer beaucoup de RSém, l'ambigüité étant également répandue. La correspondance Sens-Texte est trop complexe pour être décrite en bloc – cela nous oblige à la découper en des correspondances moins compliquées. De ce fait, nous avons besoin du Postulat 3.

Postulat 3

Pour décrire la correspondance Sens-Texte sont introduits deux niveaux de représentation intermédiaires : une représentation syntaxique, qui concerne la phrase, et une représentation morphologique, qui vise le mot-forme.

Comme résultat, un modèle Sens-Texte a l'organisation suivante :

$$\begin{array}{c} \text{(2) } \{RS\acute{e}m_i\} \Leftrightarrow \{RSynt_k\} \Leftrightarrow \{RMorph_i\} \Leftrightarrow \{RPhon_i\} \\ \text{s\'emantique} \quad \text{syntaxe} \quad \text{morphologie} + \\ \text{phonologie} \end{array}$$

Les expressions en gras sont les noms des modules majeurs du modèle.

Les représentations de tous les niveaux, sauf le niveau sémantique, sont subdivisées en deux sous-niveaux : le sous-niveau profond, qui est orienté vers le sens, et le sous-niveau de surface, orienté vers le texte. En comptant le niveau final de sortie – le niveau phonique/graphique, nous obtenons un ensemble de sept représentations :

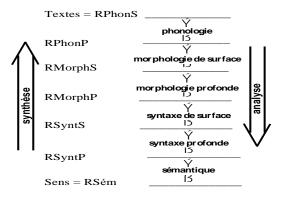


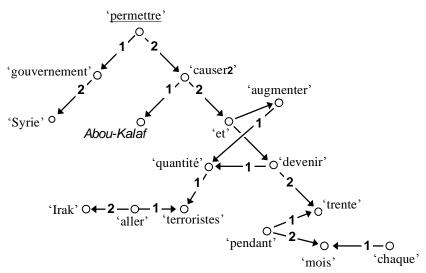
Figure 1. La structure du modèle Sens-Texte.

2.2 Représentations linguistiques dans le modèle Sens-Texte

NB: Pour faire simple, je n'utiliserai ici que les représentations graphiques des phrases et des autres expressions linguistiques.

Pour rendre les idées de base du modèle Sens-Texte plus claires, je vais donner des exemples de représentations linguistiques (drastiquement simplifiées) de tous les niveaux.

(3) Une structure sémantique de départ (implémentée par les phrases en (8))



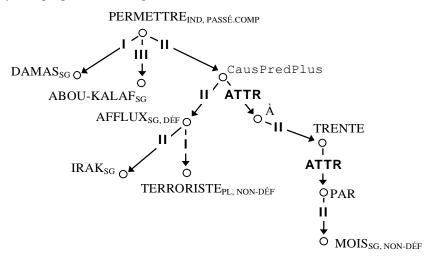
- 1. 'Causer2' exprime la causation agentive, c'est-à-dire volontaire.
 - 2. La lecture littérale de cette structure sémantique : 'Le gouvernement de la Syrie a permis qu'Abou-Kalaf cause2 que la quantité de terroristes qui vont à l'Irak augmente et devienne trente chaque mois'.

C'est une des quatre composantes d'une RSém – une structure sémantique [SSém], la structure de base d'une RSém ; formellement, c'est un réseau dont les nœuds sont étiquetés de sémantèmes (sens de lexies désambigüisées) et les arcs, de numéros appelés à distinguer les arguments d'un même prédicat.

Les trois autres composantes d'une RSém – la structure sémantico-communicative, la structure rhétorique et la structure référentielle – ne sont pas montrées. Dans ce qui suit, je ne présenterai que les structures de base de chaque représentation.

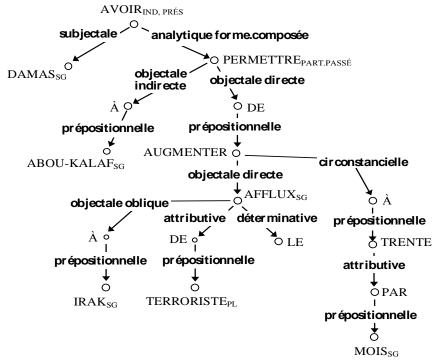
La SSém en (3) peut être verbalisée par un nombre fort élevé de phrases, desquelles seules trois sont montrées en (8).

(4) Structure syntaxique profonde de la phrase (8a)



Une SSyntP est un arbre de dépendances ; ses nœuds sont étiquetés de lexies pleines de la langue à l'étude (y compris les locutions et les fonctions lexicales – voir 3.3), et ses branches sont étiquetées des noms des 12 relations syntaxiques profondes. (Les lexies structurales, c'est-à-dire « grammaticales », ne sont pas représentées dans cette structure.)

(5) Structure syntaxique de surface de la phrase (8a)



La structure syntaxique de surface d'une phrase est aussi un arbre de dépendances ; mais ses nœuds sont étiquetés de tous les lexèmes de la phrase (y compris les pronoms et les mots structuraux), et ses branches sont étiquetées au moyen des noms des relations syntaxiques de surface. (Ces relations sont spécifiques à chaque langue ; leur nombre est autour de 50 à 60 relations par langue.)

(6) Structure morphologique profonde de la phrase (8a)

Les symboles "|," "||" et "|||" représentent les pauses de longueur différente : ce sont des éléments de la structure morphologo-prosodique profonde.

$DAMAS_{SG}$	$AVOIR_{IND,PR\acute{E}S,3,SG}$	PERMETTRE _{PART_PASSÉ}	À	ABOU-KALAF $_{SG}$ \parallel
DE	AUGMENTER _{INF}			
$LE_{MASC, SG}$	$AFFUX_{SG}$	DE TERRORISTE $_{PL}$	EN	IRAK _{SG}
À TRENTE	PAR	MOIS _{sc}		

La structure morphologique profonde d'une phrase est une chaîne de lexèmes (de cette phrase) munis de toutes les valeurs de leurs catégories flexionnelles, c'est-à-dire des grammèmes.

(7) Structure morphologique de surface de la phrase (8a)

Le symbole "
" représente l'opération d'union linguistique, qui met ensemble des signes linguistiques selon les règles de la langue et les propriétés de ces signes.

La structure morphologique de surface d'une phrase est une chaîne de groupes de morphèmes, chacun correspondant à un mot-forme ; cette chaîne est envoyée à l'entrée du module morphologique de surface, qui construit la phrase correspondante, en l'occurrence la phrase en (8a).

- (8) a. Damas a permis à Abou-Kalaf d'augmenter l'afflux de terroristes en Irak à trente par mois.
 - b. Le gouvernement syrien a laissé Abou-Kalaf accroître la quantité de terroristes pénétrant en Irak jusqu'à trente chaque mois.
 - c. Abou-Kalaf a eu le feu vert des dirigeants syriens pour faire monter le nombre des militants d'Al-Qaeda qui affluent en Irak à trente par mois.

2.3 Règles linguistiques dans un modèle Sens-Texte

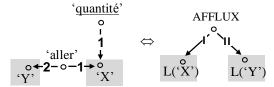
Trois classes majeures de règles linguistiques seront illustrées: règles sémantiques, syntaxiques et morphologiques. Ces règles décrivent la « traduction » d'une représentation linguistique du niveau n en une représentation du niveau n+1. Les règles données ci-dessous sont utilisées dans les transitions entre la représentation sémantique en (3) et la phrase (8a); tout comme les représentations linguistiques, les règles sont simplifiées.

2.3.1 Règles Sens-Texte sémantiques

Il est impossible de donner des exemples de chaque type de règles sémantiques ; je me limiterai à trois règles.

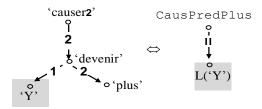
« L('a') » signifie 'lexie qui exprime le sens 'a'; le grisé marque le contexte de la règle – les éléments qui ne sont pas affectés par la règle, mais dont la présence est nécessaire pour que la règle puisse s'appliquer.

Règle sémantique lexémique individuelle (≈ un article de dictionnaire)



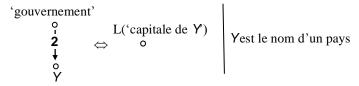
Le sens 'quantité de X qui vont à Y' peut être exprimé, entre autres, par le nom AFFLUX [de L(X) à L(Y)].

Règle sémantique lexico-fonctionnelle individuelle



Le sens 'causer2 que Y devienne plus grand [que Y a été]' peut produire la fonction lexicale complexe (voir 3.2) CausPredPlus, dont la valeur est spécifiée pour son actant syntaxique profond II dans le dictionnaire (augmenter la quantité, aggraver le conflit, élargir l'horizon, etc.)

Règle sémantique métonymique générale



Le sens 'gouvernement du pays Y' peut être exprimé par le nom de la capitale de Y ('le gouvernement de la Russie' \Leftrightarrow MOSCOU, 'le gouvernement des États Unis' \Leftrightarrow WASHINGTON, etc.).

2.3.2 Règles Sens-Texte syntaxiques

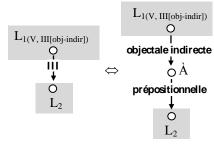
Pour les règles syntaxiques, il faut illustrer séparément les règles syntaxiques profondes et les règles syntaxiques de surface ; je donnerai deux règles de chaque type.

Règles syntaxiques profondes (syntaxe profonde)

Introduction de la valeur d'une FL



Introduction du complément d'objet indirect



Exemple : $envoyer_{L_1}$ une carte à $Alain_{L_2}$

Règles syntaxiques de surface (syntaxe de surface)

Le symbole "+" spécifie l'ordonnancement linéaire, et "..." indique un « trou » possible entre deux éléments de la proposition.

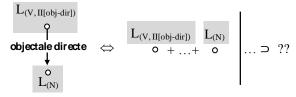
Placement du participe passé dans une forme composée



L'ensemble d' $ADV_{(type\ a)}$ doit être donné par des règles à part : les adverbes DÉJÀ, TOUJOURS, etc., pas d'adverbe de manière en **-ment** (mais : *a effectivement obtenu*), certains parenthétiques, etc.

Exemple: Ayant, comme on sait, nettoyé les écuries d'Augias, ...

Placement du complément d'objet direct



Ici également, on a besoin de règles supplémentaires qui précisent les éléments lexicaux pouvant apparaître entre le verbe transitif et le syntagme du CO^{dir}; j'indique par « ?? » mon ignorance à ce sujet.

Exemples: augmenter l'afflux; As-tu reçu hier ma lettre?

Sur la syntaxe de surface, voir Mel'čuk 1988 et 2009.

2.3.3 Règles Sens-Texte morphologiques

Pour les règles morphologiques, il faut également illustrer séparément les règles morphologiques profondes et les règles morphologiques de surface ; deux règles de chaque type seront données.

Règles morphologiques profondes (morphologie profonde)

```
PL \Leftrightarrow {PL} (terroriste + \emptyset \sim terroriste + s)
PL \Leftrightarrow \mathbf{A}^{\mathbf{al} \Rightarrow \mathbf{aux}} (apophonie du type <math>id\acute{e}al \sim id\acute{e}aux)
```

Règles morphologiques de surface (morphologie de surface)

```
{PL} \Leftrightarrow -s | non après -au, -eau, -eu et non après -s, -z, -x
-x | après -au, -eau, -eu
— | après -s, -z, -x
\mathbf{A}^{al} \Rightarrow_{\mathbf{aux}} \Leftrightarrow /al/\Rightarrow/o/
```

3 Modélisation des deux phénomènes linguistiques centraux : les choix lexicaux paradigmatiques et syntagmatiques

On sait depuis F. de Saussure et R. Jakobson que l'activité langagière du Locuteur se déroule selon deux axes : l'axe paradigmatique, où le Locuteur effectue la sélection des unités linguistiques à utiliser, et l'axe syntagmatique, où les unités sélectionnées sont combinées pour produire des énoncés. Je vais considérer ici un aspect particulier de l'activité paradigmatique : les choix lexicaux ; quant à l'activité syntagmatique, nous parlerons de la combinatoire lexicale restreinte. Ces deux problèmes sont réputés être des pierres d'achoppement pour toute théorie linguistique.

3.1 Choix lexicaux paradigmatiques

Les choix lexicaux paradigmatiques sont de deux types majeurs : les choix libres, quand une lexie L est sélectionnée directement et seulement pour son sens ; et les choix restreints, quand une lexie L est sélectionnée au lieu d'une autre lexie L' comme une fonction de L' (L remplace L' dans le texte). Les choix libres sont modélisés par les décompositions sémantiques, et les choix restreints, par les dérivations sémantiques.

3.1.1 Décompositions sémantiques

Les phrases (9a) et (9b) sont synonymes :

- (9) **a**. Jean est sûr que Marie est à Nancy. ≡
 - b. Jean ne doute pas que Marie est à Nancy.

Cela signifie au moins trois choses: 1) ces phrases sont mutuellement substituables dans le texte salva significatione; 2) leurs négations sont aussi synonymes: Jean n'est pas sûr que $P \equiv Jean$ doute que P; 3) les deux phrases ne sont pas factives et, par conséquent, peuvent être continuées par ... mais ce n'est pas vrai. Quelles informations sur les lexèmes $S\hat{U}R$ et DOUTER doivent être stockées dans le cerveau du Locuteur pour qu'il soit capable de les manipuler comme il le fait ? Nous ne pouvons pas le savoir de façon certaine ; mais nous pouvons proposer un modèle plausible.

Suivant A. Zholkovsky, A. Bogusławski et A. Wierzbicka, nous proposons que le sens de ces lexèmes (comme tous sens lexicaux en général, c'est-à-dire tous les sémantèmes – à quelques exceptions près) est constitué de sens plus simples ; ou bien, en d'autres termes, que les sens – lexicaux et autres – sont décomposables. Ces sens plus simples sont, à leur tour, décomposables en des sens encore plus simples et ainsi de suite, jusqu'à ce qu'on atteigne les primitifs sémantiques, ou les atomes de sens ; à ce sujet, voir l'œuvre abondante de Wierzbicka. Considérons les phrases suivantes (le dièse « * » indique une suite inacceptable ; le verbe CROIRE [que P] est pris au sens de 'avoir la croyance « P »') :

- (10) **a**. Jean croit que Marie est arrivée, mais il n'en est pas sûr.
 - **b**. Jean est sûr que Marie est arrivée, *mais il ne le croit pas.
 - **c**. Jean croit que Marie est arrivée, "mais il en doute.
 - **d**. Jean est sûr que Marie est arrivée. ≡ Jean ne doute pas que Marie est arrivée.
 - e. Jean n'est pas sûr que Marie soit arrivée.≅ Jean doute Marie soit arrivée.

Pour qu'un Locuteur (ou un ordinateur) puisse traiter correctement les expressions en (10), en établissant leur acceptabilité et synonymie, il est suffisant de représenter les sens considérés comme suit (les crochets « []] » marquent le présupposé²):

- (11) a. X est sûr que P: '¶Ayant la croyance « P a lieu », ¶X n'est pas disposé à admettre que P n'a pas lieu'.
 - **b.** X doute que P: '[N'ayant pas la croyance « P a lieu »,], X est disposé à admettre que P n'a pas lieu'.

Avec ces définitions lexicographiques, nous obtenons pour les phrases en (10) les décompositions sémantiques suivantes :

- (12) **a**. 'Jean a la croyance « Marie est arrivée », mais [ayant la croyance « Marie est arrivée »,]] il est disposé à admettre que Marie n'est pas arrivée'.
 - b. '[Ayant la croyance « Marie est arrivée »,] Jean n'est pas disposé à admettre que Marie n'est pas arrivée,
 #mais il n'a pas la croyance « Marie est arrivée » (la contradiction logique est marquée en gras).
 - c. 'Jean a la croyance « Marie est arrivée », "mais [n'ayant pas la croyance « Marie est arrivée»,]] il est disposé à admettre que Marie n'est pas arrivée' (une autre contradiction).
 - d. '[Ayant la croyance « Marie est arrivée »,] Jean n'est pas disposé à admettre que Marie n'est pas arrivée'. ≡ '[Ayant la croyance « Marie est arrivée » [cette composante sémantique est le résultat d'une double négation],] Jean n'est pas disposé à admettre que Marie n'est pas arrivée'.
 - e. '[Ayant la croyance « Marie est arrivée »,] Jean est disposé à admettre que Marie n'est pas arrivée'. ≅ '[N'ayant pas la croyance « Marie est arrivée »,] Jean est disposé à admettre que Marie n'est pas arrivée'.

En (10e), les deux phrases sont quasi synonymes, mais pas absolument synonymes comme les phrases en (10d): le seconde phrase en (10e) exprime un plus bas niveau de certitude que la première. C'est cela que montrent les décompositions en (12e): les deux phrases affirment que Jean est disposé à admettre que Marie n'est pas arrivée'; cependant, la première présuppose chez Jean la croyance « Marie est arrivée », alors que la seconde ne porte pas ce présupposé.

Comme on le voit, les décompositions proposées permettent une description formelle et cohérente des données en (10). Les décompositions sémantiques constituent un outil descriptif de toute première importance dans le cadre du modèle Sens-Texte.

3.1.2 Dérivation sémantique

Si un Locuteur veut parler de la personne qui est au volant d'une voiture, il a besoin du lexème CONDUCTEUR. Mais la personne qui « conduit » une locomotive est appelée un MÉCANICIEN, celui « conduisant » un avion est un PILOTE, et le « conducteur » d'une diligence est un COCHER. De façon similaire, une personne qui s'occupe d'une ferme est un FERMIER, celui qui s'occupe d'un magasin est un COMMERÇANT, et le propriétaire d'une auberge est un AUBERGISTE. D'autre part, le « client » d'un restaurant est un CLIENT, celui d'un hôpital, un PATIENT, celui d'une prison est un PRISONNIER, celui d'un théâtre, un SPECTATEUR, celui d'une école, un ÉLÈVE, etc. On voit ici des liens sémantiques réguliers CONDUCTEUR : VOITURE = MÉCANICIEN : LOCOMOTIVE = PILOTE : AVION = COCHER : DILIGENCE. Ces liens rappellent les liens de dérivation, sauf que, dans ce cas, les membres des paires analysées ne manifestent aucune similarité formelle (morphologique). Cependant, il est clair que, pour la modélisation de production de textes, les relations lexicales de ce type doivent être spécifiées explicitement. Cela peut être fait au moyen de fonctions lexicales, introduites en 3.3 ; ce sont des fonctions lexicales paradigmatiques.

3.2 Les choix lexicaux syntagmatiques

La cooccurrence lexicale restreinte constitue un problème de taille pour une description lexicographique. Deux exemples simples suffiront ici :

- 1) On TIRE une conclusion, mais on FAIT une déduction et MÈNE un raisonnement; on PREND la fuite et une décision mais on DONNE un conseil et l'assaut; on DISPOSE d'un droit, mais on A une diarrhée et PIQUE un somme; etc.
- 2) On est con COMME UN PANIER, mais malheureux COMME LES PIERRES; on aime à LA FOLIE, croit DUR COMME FER et conseille VIVEMENT; quelque chose peut être cassé EN MILLE MIETTES, on refuse et on s'arrête NET, il pleut COMME VACHE QUI PISSE; etc.

Une étude poussée des expressions de ce type montre que :

Dans la grande majorité des cas, la cooccurrence lexicale restreinte (c'est-à-dire, synchroniquement arbitraire) se manifeste dans l'expression d'un nombre limité de sens très abstraits, « presque grammaticaux ».

Ainsi, dans le premier exemple ci-dessus, ce sens est ≈ 'faire', et dans le second, il est ≈ 'très/de façon intense'.

Le fait crucial est qu'un sens du type considéré correspond à une fonction (mathématique) : la lexie qui contrôle la cooccurrence (= pour laquelle les cooccurrents appropriés doivent être sélectionnés) est son argument, tandis que

les cooccurrents eux-mêmes constituent la valeur, de sorte qu'on a $\mathbf{f}(L) = \{L_1, L_2, ..., L_n\}$. Pour les exemples ci-dessous, deux fonctions sont proposées - ce sont des fonctions lexicales standard simples (voir la section suivante):

```
Le verbe support Oper<sub>1</sub>
                                                         L'intensificateur Magn
Oper<sub>1</sub>(erreur)
                         = faire [ART \sim]
                                                         Magn(majorit\acute{e}) = grande / antépos\acute{e} < \acute{e}crasante / postpos\acute{e}
Oper<sub>1</sub>(assistance)
                         = fournir [ART ~]
                                                         Magn(assistance) = généreuse / postposé
Oper<sub>1</sub>(rage)
                         = être [en ~]
                                                         Magn(rage)
                                                                               = aveugle
                                                         Magn(con)
                                                                               = comme un panier
Oper<sub>1</sub>(conclusion)
                        = tirer [ART \sim]
                                                         Magn(pleuvoir) = comme vache qui pisse
Oper<sub>1</sub>(activité)
                         = déployer [ART ~]
```

Les expressions décrites par ces fonctions lexicales sont des collocations : la base de la collocation – la lexie sélectionnée par le Locuteur pour son sens indépendamment des autres lexies - correspond à l'argument de la fonction ; et le collocatif, sélectionné en fonction de la base, est un des éléments de sa valeur. (Sur les collocations, voir Mel'čuk 1998, 2003 et 2015: 293sqq.)

3.3 Fonctions lexicales

Trois propriétés des fonctions lexicales standard simples s'avèrent cruciales :

- Les fonctions lexicales standard simples ne sont pas nombreuses : entre 50 et 60.
- Les fonctions lexicales sont universelles, dans le sens où elles sont valables pour toutes les langues du monde.
- · Les fonctions lexicales conviennent également pour la description des choix lexicaux restreints autant paradigmatiques que syntagmatiques. Autrement dit, elles assurent une description homogène et systématique des dérivations sémantiques et des collocations.

Plusieurs exemples de fonctions lexicales rendront cette notion plus claire. (Sur les fonctions lexicales, voir Wanner, ed. 1996 et Mel'čuk 2015: 155 *sqq*.)

= chute

Fonctions lexicales paradigmatique (= dérivations sémantiques)

```
1. Nom d'action/de propriété : S<sub>0</sub>
```

```
S<sub>0</sub>(accepter)
                        = acceptation
                                                        S_0(tomber)
S<sub>0</sub>(capable)
                        = capacité
                                                        S_0(vain)
                                                                              = vanité
2. Nom de patient : S<sub>2</sub>
                                                                              = cible
S_2(crime)
                        = victime
                                                        S_2(tirer)
S2(vendre)
                        = marchandise
                                                        S<sub>2</sub>(parler [à])
                                                                              = interlocuteur
3. Adjectif de possibilité active : Able<sub>1</sub>
Able_1(dissoudre)
                        = soluble
                                                        Able_1(guerre)
                                                                              = belliqueux
Able_1(nuire)
                        = nocif
                                                        Able_1(vomir)
                                                                              = nauséeux
```

Fonctions lexicales syntagmatiques (= collocations)

1. Adjectif d'évaluation positive : Bon

```
Bon(compliment)
                    = bien tourné
                                               Bon(analyse)
                                                                 = lucide
Bon(se sentir)
                    = frais et dispos
                                               Bon(temps)
                                                                 = beau / antéposé
```

2. Verbes supports: Oper, Func, et Labor,

```
Oper<sub>1</sub>(enquête)
                         = mener [ART \sim]
                                                          Oper<sub>2</sub>(examen)
                                                                                   = passer [ART \sim]
Func_1(accent)
                        = tombe [sur N_X]
                                                          Func_2(raid)
                                                                                   = vise [N_Y]
                                                          Func_0(feu)
                                                                                   = brûle
Func_0(calme_N)
                         = règne
Labor_{12}(h\acute{e}ritage) = recevoir [N_Y en ~]
                                                          Labor<sub>32</sub>(h\acute{e}ritage) = laisser[N_Y en ~]
```

3. Verbes de réalisation : Real_i, Fact_i et Labreal_{ij}

```
Real_1(faim)
                           = assouvir [ART \sim]
                                                               Real_1(traitement) = répondre [à ART ~]
Fact<sub>1</sub>(rivière)
                                                               Fact<sub>2</sub>(bombe)
                                                                                          = frappe [N_Y]
                           = se jette [dans N_X]
                           = est à l'affiche, passe, se joue
Fact<sub>0</sub>(film)
Labreal<sub>12</sub>(pi\grave{e}ge) = prendre [N<sub>Y</sub> au ~]
                                                               Labreal<sub>21</sub>(pi\grave{e}ge) = tomber [dans ART ~]
```

4. Préposition de localisation : Locin

```
Loc_{in}(liste) = sur [ART \sim]  Loc_{in}(fin) = \grave{a} [ART \sim]  Loc_{in}(vacances) = en [\sim]  Loc_{in}(pass\acute{e}) = dans [ART \sim]
```

Faute de temps, je ne peux pas parler ici de fonctions lexicales non standard, de fonctions lexicales composée et de configurations de fonctions lexicales.

3.4 Corrélations entre le sens et les collocatifs d'une lexie

La description proposée des sens lexicaux et de la cooccurrence lexicale restreinte permet, assez souvent, de préciser nos définitions lexicographiques. Prenons, par exemple, le nom APPLAUDISSEMENTS. Le Petit Robert le définit comme 'battement des mains en signe d'approbation, d'admiration ou d'enthousiasme'. Cette définition semble correcte à première vue, mais elle ne tient pas compte des collocatifs de ce nom : tonnerre d'applaudissements ; Magn assourdissants, à tout rompre, nourris vs. AntiMagn clairsemés, quelques, rares, timides ; Son : les applaudissements fusent, sonnent, retentissent. Ces cooccurrents témoignent de deux choses :

- Premièrement, les applaudissements constituent, avant tout, un événement sonore.
- Deuxièmement, les applaudissements sont graduables : la force et la fréquence de battements des mains sont proportionnelles au degré d'approbation, d'admiration ou d'enthousiasme.

Par conséquent, la définition du nom APPLAUDISSEMENTS et celle du verbe APPLAUDIR doivent être corrigées :

applaudissements de X à Y: 'son produit par le fait que X applaudit Y'.

X applaudit Y : '□ bat des mains en signe d'approbation, d'admiration ou d'enthousiasme de X pour Y, la force et la fréquence de ces battements étant proportionnelles au degré d'approbation, d'admiration ou d'enthousiasme'.

De cette façon, les définitions lexicographiques et les collocations se confortent mutuellement (Iordanskaja & Polguère 2005, Iordanskaja 2007). Cette corrélation est vitale pour la description de la langue, un système « où tout se tient » [F. de Saussure].

3.5 Le Dictionnaire explicatif et combinatoire

Pour qu'une lexie L puisse être utilisée correctement dans tous les contextes possibles, le locuteur doit avoir dans son cerveau une masse importante d'informations sur L. Dans un modèle linguistique, toutes ces informations doivent être consignées dans un dictionnaire d'un type particulier : le *Dictionnaire explicatif et combinatoire* [DEC]. Ce dictionnaire est explicatif dans le sens où il est ancré dans le sémantisme des lexies : chaque L y reçoit une décomposition sémantique rigoureuse (= « explication » de L). Il est combinatoire dans le sens où il spécifie, pour chaque L, la totalité de sa cooccurrence restreinte – syntaxique et lexicale. Au sein du modèle Sens-Texte, le DEC joue un rôle de premier plan : l'entrée de la lexie vedette L contient toutes les données sémantiques et toutes les données combinatoires sur L, les deux types de données exploités par les règles du modèle Sens-Texte. Il s'ensuit que ce modèle est lexical de façon essentielle.

Le DEC a été décrit en détail à plusieurs reprises (Mel'čuk 2006, Mel'čuk *et al.* 1984–1999, Mel'čuk *et al.* 1995, Mel'čuk & Polguère 2007, Mel'čuk 2013: 259 *sqq.*, Mel'čuk & Milićević 2014: vol. 1: 269 *sqq.*); on peut donc se limiter ici à indiquer ses propriétés définitoires.

Propriétés générales du DEC

- Le DEC est un lexique théorique élaboré dans le cadre d'une théorie linguistique globale et formalisée : la théorie Sens-Texte.
- Le DEC est un lexique formalisé, écrit en termes de plusieurs métalangages lexicographiques.
- Le DEC est complet (≈ exhaustif) au niveau de chaque entrée.

Propriétés spécifiques du DEC

- Le DEC est un dictionnaire actif, fournissant toutes les informations lexicographiques en direction Sens ® Texte.
- Le DEC est un dictionnaire sémantique : la définition (= RSém) de la lexie vedette L est sous-jacente à l'entrée de L et y détermine les autres informations.
- Le DEC couvre systématiquement toute la cooccurrence lexicale restreinte de L en termes de fonctions lexicales.
- Le DEC traite les lexèmes et les locutions de la même façon : toutes ces unités lexicales constituent des vedettes des entrées correspondantes.

11

• Le DEC prévoit une seule entrée pour chaque lexie ; les lexies reliées entre elles par la relation de polysémie sont réunies dans une super-entrée, appelée le vocable.

4 Conclusion: La valeur des modèles fonctionnels en linguistique

Un modèle fonctionnel d'une langue naturelle a une grande utilité pratique au moins dans trois domaines technologiques et sociaux :

- Traitement informatique des langues naturelles.
- Enseignement et apprentissage des langues et de la linguistique (voir, par exemple, Polguère 2008).
- Compilation des livres de référence sur la langue, tels que des dictionnaires de tous les types, des grammaires pédagogiques et des manuels.

Le caractère formel du modèle Sens-Texte et son orientation (« Comment telle ou telle idée est-elle exprimée dans telle ou telle langue ? ») sont d'une valeur spéciale sous cet angle.

Cependant, l'impact théorique du modèle Sens-Texte apparaît encore plus important. Le progrès scientifique jusqu'à présent touchait surtout les aspects physiques de l'univers : matière et énergie. Depuis que le premier homininé a commencé à parler et, de ce fait, est devenu *Homo sapiens*, l'humanité a développé de nouveaux moyens de transport (y compris des vaisseaux spatiaux), a augmenté notre force physique de plusieurs ordres de magnitude (il suffit de penser à la bombe H), a renforcé nos organes de perception (microscopes électroniques et radiotélescopes), a élargi nos possibilités de communication (les média électroniques et l'Internet), etc. Nous avons pénétré l'atome et les profondeurs de l'univers ; nous savons beaucoup sur les origines de notre monde et la structure de nos gènes. Mais nous n'avons pas encore fait un progrès comparable dans le domaine de traitement d'information par le cerveau humain. Ce problème, qui est central à la vie en générale et surtout à la vie humaine, reste toujours enveloppé de mystère. Nous n'en savons pas assez sur le fonctionnement de notre cerveau ; le renforcement de nos capacités mentales est toujours une des tâches centrales de la science moderne. Face aux défis du 21^{me} siècle, l'humanité a un besoin aigu de modèles fiables du processus de la pensée humaine, du raisonnement logique et, pourquoi pas, des émotions humaines. Il semble que cet objectif soit très bien compris par la communauté scientifique internationale et qu'un nombre toujours croissant de scientifiques soit d'accord avec un tel programme.

Néanmoins, de façon surprenante, on a tendance à oublier – ou à ignorer ? – ce fait vital :

La seule clé fiable de la pensée humaine est la langue.

Sans bien comprendre comment la langue fonctionne dans notre psyché nous ne pourrons pas bien comprendre le traitement des informations mentales par le cerveau. C'est pourquoi je crois que les modèles fonctionnels de langue en général, et le modèle Sens-Texte en particulier, ont acquis aujourd'hui une signification spéciale. La linguistique doit enfin occuper une place d'honneur parmi les sciences « dures »; et les modèles fonctionnels, qui incarnent l'approche typiquement scientifique des phénomènes complexes, y apporteront une contribution précieuse.

Remerciements chaleureux

Merci infiniment à mes cobayes de laboratoire, mes trois premiers relecteurs : Rémi Camus, Lidija Iordanskaja et Svetlana Krylosova. Leurs remarques et suggestions m'ont aidé à améliorer mon texte.

Notes

¹ Une expression qui contient ou régit une subordonnée complétive est dite factive si et seulement si elle implique la vérité de cette subordonnée. Par exemple, dans la phrase *Jean sait que Marie est à Nancy*, la proposition principale *Jean sait* implique la vérité de la subordonnée *Marie est à Nancy*, car on ne peut pas faire suivre la phrase par ...mais ce n'est pas vrai: Jean sait que Marie est à Nancy, mais [#]ce n'est pas vrai.

² Le présupposé est la partie 'σ'' d'un sens 'σ' qui reste affirmée quand 'σ' est nié, c'est-à-dire une partie de sens inaccessible à la négation. Par exemple, dans la phrase *Jean, qui est paresseux, est venu*, le sens 'Jean est paresseux' constitue un présupposé, car la phrase *Jean, qui est paresseux, n'est pas venu* continue à affirmer que Jean est paresseux.

Références

- Puisqu'il est impossible de présenter ici une bibliographie relativement complète ou même représentative, je me permets de recourir à l'autopromotion et de n'indiquer que quelques-unes des publications pertinentes.
- Auger, P. 1965. Les modèles dans la science. Diogène, n° 52, 3-15.
- Iordanskaja, L. 2007. Lexicographic Definition and Lexical Cooccurrence: Presuppositions as a 'No-Go' Zone for the Meaning of Modifiers. In: Gerdes, K., Reuther, T. & Wanner, L., eds, *Meaning-Text Theory 2007 [MTT'07]*, München/Wien: Wiener Slawistischer Almanach, 209–218.
- Iordanskaja, L. & Polguère, A. 2005. Hooking Up Syntagmatic Lexical Functions to Lexicographic Definitions. In: Apresjan, Ju. & Iomdin, L., eds, *East-West Encounter, Second International Conference on Meaning-Text Theory [MTT'05]*, Moskva: Jazyki slavjanskoj kul'tury, 170–186.
- Mel'c]uk, I. 1981. Meaning-Text Models: A Recent Trend in Soviet Linguistics. *Annual Review of Anthropology*, v. 10, 27–62.
- Mel'čuk, I. 1988. Dependency Syntax: Theory and Practice. Albany (N.Y.): The SUNY Press.
- Mel'čuk, I. 1992 2000. Cours de morphologie générale, vol. 1 5. Montréal: Les Presses de l'Univ. de Montréal.
- Mel'čuk, I. 1995. Phrasemes in Language and Phraseology in Linguistics. In: Everaert, M., van der Linden, E.-J., Schenk, A. & Schreuder, R., eds, *Idioms. Structural and Psychological Perspectives*, Hillsdale (N.J.) Hove: Lawrence Erlbaum, 167–232.
- Mel'čuk, I., 1996. A Tool for the Description of Lexical Relations in the Lexicon. In: Wanner, ed. 1996: 37–102.
- Mel'čuk, I. 1998. Collocations and Lexical Functions. In: Cowie, A., ed., *Phraseology. Theory, Analysis, and Applications*, Oxford: Clarendon Press, 23–53.
- Mel'c]uk, I. 2003. Collocations dans le dictionnaire. In: Szende, Th., réd., *Les écarts culturels dans les dictionnaires bilingues*, Paris: Honoré Champion, 19–64.
- Mel'c]uk, I. 2006. Explanatory Combinatorial Dictionary. In: Sica, G., ed., *Open Problems in Linguistics and Lexicography*. Monza (Italy): Polimetrica, 225–355.
- Mel'c]uk, I. 2007. Lexical Functions. In: Burger, H., Dobrovol'skij, D., Kühn, P. & Norrick, N., eds, *Phraseology*. *An International Handbook of Contemporary Research*, Berlin/New York: de Gruyter, 119–131.
- Mel'c]uk, I. 2009. Dependency in Natural Language. In: Polguère, A. & Mel'c]uk, I., eds, *Dependency in Linguistic Description*, Amsterdam/Philadelphia: John Benjamins, 1–110.
- Mel'cluk, I. 2012. Semantics: From Meaning to Text. Amsterdam/Philadelphia: John Benjamins.
- Mel'cluk, I. 2013. Semantics: From Meaning to Text. Vol. 2. Amsterdam/Philadelphia: John Benjamins.
- Mel'c]uk, I. 2015. Semantics: From Meaning to Text. Vol. 3. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, I. et al. 1984-1999. *Dictionnaire explicatif et combinatoire du français contemporain : Recherches lexico-sémantiques I IV*. Montréal: Les Presses de l'Université de Montréal.
- Mel'čuk, I., Clas, A. & A. Polguère, A. 1995. *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Mel'čuk, I. & Milićević, J. 2014. *Introduction à la linguistique*. Vol. 1 3. Paris: Hermann.
- 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.
- Milićević, J. 2007. La paraphrase. Modélisation de la paraphrase langagière. Bern: P. Lang.
- Wanner, L., ed. 1996. *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins.