

## La phraséologie en langue, en dictionnaire et en TALN

Igor Mel'čuk, OLST, Université de Montréal

### 1 Introduction

Je n'ai pas besoin d'insister sur l'importance de la phraséologie pour la linguistique en général et pour le TALN en particulier. Cela serait essayer d'enfoncer une porte ouverte : sur ce point, la communauté linguistique est d'accord. Ce sur quoi on n'est pas tellement d'accord est le contenu exact de la notion « phraséologie », la façon dont les expressions phraséologiques doivent être décrites et, enfin, qu'est-ce qu'on peut faire avec ces expressions dans le traitement automatique de langues. Je vais me pencher sur ces trois points. Par conséquent, mon exposé sera divisé en quatre sections (sans compter cette Introduction). La section 2 propose la définition rigoureuse du concept 'phrasème' et offre une typologie exhaustive des phrasèmes, en établissant de ce fait les frontières de la phraséologie. La section 3 présente les fondements de description lexicographique des phrasèmes dans un Dictionnaire Explicatif et Combinatoire. C'est en fait la partie centrale de la présentation : je propose qu'une paire de bons dictionnaires est nécessaire et (presque) suffisante pour assurer – bien entendu, ensemble avec une grammaire – la traduction automatique de qualité acceptable. Bien entendu, une description systématique et formelle de la phraséologie sera bénéfique pour d'autres tâches du TALN (l'analyse et la génération des textes, entre autres). Dans la section 4, j'examine trois exemples difficiles de traduction automatique où les solutions viennent surtout du dictionnaire. Finalement, la section 5 résume les points les plus importants de la présentation.

La discussion est menée dans le cadre théorique Sens-Texte. Je suis obligé d'utiliser certaines notions et formalismes sans explication préalable. Pour des renseignements additionnels sur l'approche Sens-Texte, le lecteur est prié de s'adresser à Mel'čuk 1981, 1988a: 43-91, 1997, Kahane 2003a. Les termes techniques apparaissent en *Helvetica*.

### 2 La phraséologie en langue

Une expression phraséologique – on dit aussi « figée », « fixe », ou « idiomatique » – est un **syntagme**, expression constituée de plusieurs lexèmes (au moins deux) liés par des liens syntaxiques réguliers. Ainsi, l'expression *casser sa pipe*  $\approx$  'mourir' est syntaxiquement et morphologiquement organisée comme tous les syntagmes français du type « Verbe transitif—CO<sup>dir</sup> » : *casser*

*son jouet, aimer ses parents* ou *voir un tigre*. Même la séquence *casser sa pipe* peut signifier ‘casser sa pipe’ ! Ce qui rend cette expression spéciale est la présence du sens imprévisible : ‘mourir’. Une expression phraséologique, qui sera appelée dorénavant **phrasème**, est un syntagme phraséologisé, ou bien **non libre**. Il faut donc commencer par définir le syntagme **libre**.

### **Syntagme libre**

Un syntagme est **libre** si et seulement si [= ssi] chacun de ses composants  $L_i$  est sélectionné par le locuteur de façon **non contrainte** : strictement pour son sens, respectant les propriétés du  $L_i$  choisi, mais indépendamment des autres composants.

**Corollaire** : chacun de composants d’un syntagme libre peut être remplacé par n’importe quel synonyme sans affecter le sens du syntagme (en tenant compte, bien entendu, des propriétés particulières des synonymes substitués).

### **Syntagme non libre (= phraséologisé)**

Un syntagme est **non libre** si et seulement si [= ssi] au moins un de ses composants est sélectionné de façon **contrainte** : en fonction des autres composants.

Un syntagme non libre s’appelle **phrasème**.

**Corollaire** : aucun composant d’un phrasème ne peut être remplacé par n’importe quel synonyme (même si quelques synonymes peuvent être admis : *louper/manquer/rater le coche* ‘perdre une occasion de faire qqch de profitable’).

Un phrasème viole donc la **liberté de sélection** (de ses composants). Cette violation peut survenir soit dans le passage entre la R(eprésentation)Concept(uelle) et la RSém(antique) d’une expression linguistique, soit dans le passage entre la RSém et l’expression elle-même.

Dans le premier cas (où le locuteur choisit une RSém pour une RConcept de départ), on obtient des phrasèmes **pragmatiques**, ou **pragmatèmes** : dans une situation de communication langagière donnée, la sélection du sens pour exprimer la RConcept donnée est contrainte, de sorte qu’il faut utiliser un sens particulier et l’exprimer avec des moyens particuliers. Par exemple, soit le contenu informationnel ‘l’auteur de ce texte fait ressortir ce fragment de la citation en cause’, la situation d’énonciation étant la suivante : le locuteur veut mettre la remarque correspondante dans le texte imprimé. En français on écrit dans ce cas *C’est moi qui souligne*, en anglais *Italics/Emphasis mine* lit. ‘Italique/Emphase à moi’ et en allemand, *Hervorhebung des Autors* lit. ‘Rehaussement de l’auteur’. Par les gloses, on voit que les sens exprimés sont différents dans les

trois langues et qu'ils sont fixés dans chaque langue par la situation extralinguistique où le pragmatème en cause est utilisé.

Le deuxième cas (le locuteur choisit une expression langagière pour une RSém) donne des phrasèmes **sémantiques** : ici, le sens – c'est à dire, la RSém – est sélectionné librement, mais dans le syntagme qui l'exprime au moins un composant ne l'est pas. Les phrasèmes **sémantiques** incluent les **clichés**, les **collocations** et les **locutions**, qu'on distingue selon deux dimensions, **paradigmatique** et **syntagmatique** : le **degré de non-liberté** de sélection et la **compositionnalité**.

Pour caractériser de façon plus précise ces types de phrasèmes, nous avons besoin de la notion de compositionnalité de signes linguistiques. Un **signe linguistique**  $s$  est un triplet

$$s = \langle ' \sigma ' ; /s/ ; \Sigma \rangle, \text{ où :}$$

- '  $\sigma$  ' est le **signifié**, ou un contenu informationnel, le plus souvent un sens linguistique ;
- /s/ est le **signifiant**, ou un signal physique, le plus souvent une chaîne de phonèmes (ou de caractères – à l'écrit) ;
- $\Sigma$  est le **syntactique**, ou un ensemble d'informations spécifiant la cooccurrence de  $s$  avec d'autres signes.

Ainsi, le nom MONTAGNE est représenté en tant que signe comme suit :

$$\langle \text{'importante élévation du terrain'} ; /m\tilde{o}t\tilde{a}n/ ; \Sigma = \text{nom, commun, fém, ...} \rangle$$

Les signes simples se combinent en signes complexes par l'opération d'**union linguistique**  $\oplus$ . Cette opération est représentée par un ensemble de règles qui disent comment, dans une langue donnée, réunir les signes : les signifiés (en mettant les RSém des arguments dans les positions argumentales des prédicats, ...), les signifiants (en juxtaposant les chaînes de phonèmes et en appliquant les opérations linguistiques, ...) et les syntactiques (en retenant les données valables pour le signe complexe résultant).

### **Signe linguistique complexe compositionnel**

|| Un signe linguistique complexe **AB** est **compositionnel** ssi  $\mathbf{AB} = \mathbf{A} \oplus \mathbf{B}$ .

Cela signifie que, pour le signe  $\mathbf{AB} = \langle \text{'AB'} ; /AB/ ; \Sigma_{AB} \rangle$ , le signifié 'AB' = 'A'  $\oplus$  'B', le signifiant /AB/ = /A/  $\oplus$  /B/ et le syntactique  $\Sigma_{AB} = \Sigma_A \oplus \Sigma_B$ .

De cette définition, il découle que la compositionnalité est une notion absolue, qui ne connaît pas de degrés : un signe complexe est compositionnel ou pas. La propriété d'être compositionnel

concerne les trois composants du signe de façon indépendante ; dans ce qui suit nous n'allons considérer que la compositionnalité du signifié, soit la **compositionnalité sémantique**.

Un syntagme libre est nécessairement compositionnel : c'est grâce à cette propriété que la communication langagière est possible. En fait, maîtriser la langue **L** veut dire avoir dans son cerveau tous les signes simples de **L** et toutes les règles de l'opération  $\oplus$  pour **L**.

La sélection de lexies se fait sur l'axe paradigmatique de la langue, alors que leur combinaison se passe sur l'axe syntagmatique. Ainsi notre caractérisation des phrasèmes est exhaustive.

Donnons maintenant les définitions des trois types majeurs de phrasèmes sémantiques.

### **Cliché**

Un phrasème sémantique est un **cliché** ssi 1) aucun de ses composants n'est sélectionné librement (tous les composants sont pris en bloc, comme un tout ; le cliché est donc complètement contraint) et 2) il est compositionnel.

**Exemples** : *Ce qu'il fallait démontrer* (et non pas, par ex., *\*Ce qu'il était nécessaire de prouver*) ; *Autrement dit* ; *C'est moi qui te/vous le dis !* ; *dans huit jours* (plutôt que *dans une semaine*) ; *Dans le doute, abstiens-toi*.

### **Collocation**

Un phrasème sémantique est une **collocation** ssi 1) un de ses composants est sélectionné librement et l'autre non, étant choisi en fonction du premier (la collocation est donc mi-contrainte) et 2) il est compositionnel.

**Exemples** : *AGILE comme un singe*, *AIMER à la folie*, *HOMMAGE vibrant*, *IGNORANCE crasse*, *JACASSER comme une pie*, ... ; *attacher la CEINTURE DE SÉCURITÉ*, *établir <libeller> un CHÈQUE*, *tenir sa PAROLE*, *battre (le) PAVILLON*, *décerner un PRIX*, *exercer une PROFESSION*, ...

Une collocation est constituée d'une **base**, lexie choisie librement par le locuteur (en petites majuscules dans nos exemples), et d'un **collocatif**, choisi en fonction de la base. Elle est sémantiquement compositionnelle, car son sens est divisible en deux parties telles que la première correspond à la base et la deuxième est exprimée par le collocatif.

### **Locution**

Un phrasème sémantique est une **locution** ssi 1) aucun de ses composants n'est sélectionné librement (il est donc complètement contraint) et 2) il est non compositionnel.

Une locution est indiquée par des coins surélevés : ' ... '.

**Exemples** : 'par acquit de conscience' 'pour en être absolument sûr', 'Un ange passe' 'Un silence soudain s'établit dans une compagnie', 'nager entre deux eaux' 'se maintenir entre les opinions

opposées', 's'en mordre la langue' 'regretter d'avoir dit quelque chose', 'vache à lait' 'personne dont on tire profit', 'aller au charbon' 'devoir accomplir un travail désagréable', 'rendre son tablier' 'quitter un poste', ...

Une locution est caractérisable par le degré de sa **transparence** : la propriété de son sens d'inclure, dans un certain degré, les sens de ses composants. Trois cas de figure sont à distinguer ici ; pour les caractériser, nous allons introduire la notion de pivot sémantique.

### Pivot sémantique (d'une sens)

Posons qu'un sens 'S' peut être divisé en deux parties, 's<sub>1</sub>' et 's<sub>2</sub>' ('S' = 's<sub>1</sub>' ⊕ 's<sub>2</sub>').

La partie 's<sub>1</sub>' du sens 'S' est appelé le **pivot sémantique** de 'S' ssi l'autre partie 's<sub>2</sub>' est un prédicat dont 's<sub>1</sub>' est l'argument : 'S' = 's<sub>2</sub>'('s<sub>1</sub>').

Soulignons que le pivot sémantique d'un sens 'S' est logiquement différent du **composant communicativement dominant** de 'S'. Ainsi, pour le sens 'réussir un examen', le pivot sémantique est 'examen', alors que le composant dominant est 'réussir'. Cependant, dans beaucoup de cas, les deux peuvent coïncider. Le pivot sémantique sera identifié dans nos exemples par **les ha-**  
**chures**.

- La **quasi-locution** : son sens inclut les sens de tous ses composants mais aussi un sens additionnel « imprévisible » qui est le pivot sémantique ; c'est la locution la plus transparente.

**Exemples** : 'donner le sein à Y' '**nourrir le bébé Y** de façon à mettre le mamelon d'un sein dans la bouche de Y', 'tache solaire' '**région sur la surface** du soleil ... perçue comme une tache', 'pâte dentifrice' '**substance fabriquée** sous forme de pâte destinée à nettoyer les dents', ...

- La **semi-locution** : son sens inclut le sens seulement d'un de ses composants, mais non pas dans le rôle de pivot sémantique ; c'est la locution mi-transparente.

**Exemples** : 'fruits de mer' '**animaux** de mer **comestibles** qui ne sont pas des poissons', 'bain de foule' '**acte** qui consiste à se mêler à la foule', 'eau lourde' '**composé chimique** à base d'eau où les deux atomes d'hydrogène sont remplacés par deux atomes de deutérium', 'prendre l'eau' '**[les chaussures]** être défectueux de façon à laisser passer l'eau', 'casser les oreilles [à Y]' '**[sons continus déplaisants]** causer des sensations déplaisantes dans les oreilles de Y', ...

- La **locution complète** : son sens n'inclut le sens d'aucun de ses composants ; c'est la locution la moins transparente, qui parfois n'est pas transparente du tout.

**Exemples** : 'river son clou [à Y]' 'remettre Y à sa place, en le réduisant au silence', 'faire table rase' 'décider de ne tenir aucun compte de ce qui existait auparavant', 'au fur et à mesure'

[*que Y*] ‘en même temps que Y et successivement’, ‘*hors pair*’ ‘exceptionnel’, ‘*en tenue* (*costume, habit*) d’Adam/d’Ève’ ‘complètement nu/e’, ‘*s’envoyer en l’air*’ ‘éprouver le plaisir sexuel’, ...

Une typologie exhaustive de phrasèmes (les types majeurs) est présentée dans la Figure 1 ci-dessous.

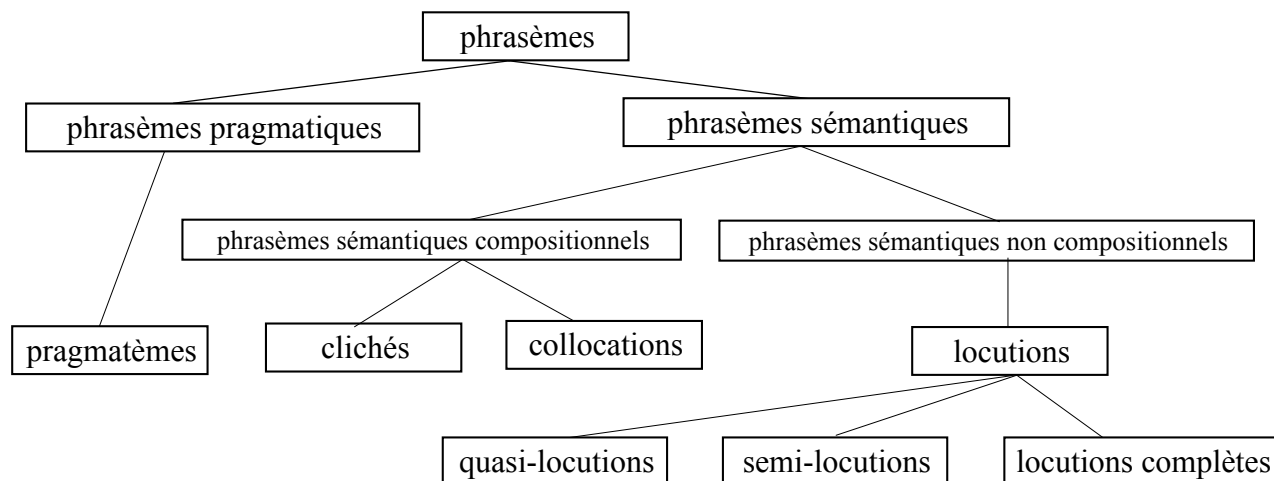


Figure 1. Typologie des phrasèmes

### 3 La phraséologie en dictionnaire

Le dictionnaire considéré ici est le *Dictionnaire Explicatif et Combinatoire* [= DEC] ; ses principes, sa structure et ses notions de base sont pris pour acquises (voir Mel’čuk 1988b, Mel’čuk *et al.* 1984-1999, Mel’čuk *et al.* 1995, Mel’čuk 2006).

Une locution est une **unité lexicale**, ou une **lexie**, tout comme un lexème. Les locutions sont donc décrites dans le DEC de la même façon que les lexèmes : chacune possède son propre article de dictionnaire, ayant la même structure que l’article d’un lexème, avec quand même une différence importante : puisqu’une locution est un syntagme, elle est munie de l’arbre de sa S(structure)Synt(axique)P(rofonde). Par exemple:

‘S’ENVOYER EN L’AIR’ ‘éprouver un plaisir sexuel’, locution verbale

SSyntP : ENVOYER<sub>REFL</sub>-II→EN-II→AIR<sub>DEF,SG</sub>

Le nombre de locutions (= phrasèmes non compositionnels) dans une langue est entre 10 000 et 20 000 (ainsi, le dictionnaire Rey & Chanterau 1993 en contient à peu près 9 000) ; elles peuvent être tout simplement stockées comme telles dans le dictionnaire, dont elles ne gonflent pas le volume de façon trop visible.

Par contre, les phrasèmes compositionnels, en particulier les collocations, ne sont pas des lexies : ils n'ont pas d'articles de dictionnaires propres et sont décrits à l'intérieur des articles de leurs bases. Par exemple, *applaudir à tout rompre* ne doit pas apparaître comme une entrée séparée, mais sous APPLAUDIR ; *tirer un chèque*, non plus : on le met sous CHÈQUE ; et ainsi de suite. Le nombre de collocations dans les langues de type SAE (*Standard Average European*) est très élevé : au moins dix fois le nombre de lexèmes. Leur description lexicographique exige donc un appareil formel qui permette leur présentation systématique, élégante et se prêtant facilement au traitement automatique. Un tel appareil, ce sont les **fonctions lexicales** [= FL] .

Il n'est pas bien entendu possible de présenter ici la notion de FL ni d'en offrir un tour d'horizon assez détaillé (voir Mel'čuk 1974: 78-109, 1982, 1996, 2003a, b, Kahane 2003b, Kahane & Polguère 2001, Wanner (ed.) 1996). Je vais me limiter à donner ici quelques exemples de FL syntagmatiques standard et non standard, pour démontrer ensuite comment les FL peuvent être mises à profit dans le TALN.

#### FL syntagmatiques standard

##### • FL verbales

###### – Verbes supports

	RESPONSABILITÉ [d'avoir fait Y]	SOINS	ACCUSATION	AIDE
Oper <sub>1</sub>	<i>porter</i> [ART ~]	<i>donner</i> [ART ~s à N <sub>Y</sub> ]	<i>lancer</i> [ART ~ contre N <sub>Y</sub> ]	<i>venir</i> [à l'~(→de N <sub>Y</sub> )]
Func <sub>1</sub>	~ <i>incombe</i> [à N <sub>X</sub> ]	~s <i>s'adressent</i> [à N <sub>Y</sub> ]	————	~ <i>parvient</i> [à N <sub>Y</sub> de N <sub>X</sub> ]
Labor <sub>12</sub>	————	<i>entourer</i> [N <sub>Y</sub> de ~s]	<i>mettre</i> [N <sub>Y</sub> en ~]	<i>venir</i> [en ~ à N <sub>Y</sub> ]

###### – Verbes de réalisation

	PRIX [récompense]	MÉDECIN	PIÈGE	ASPHALTE
Real <sub>2</sub>	<i>obtenir</i> [ART ~]	<i>voir</i> [ART ~]	<i>tomber</i> [dans ART ~]	————
Fact <sub>2</sub>	~ <i>va</i> [à N <sub>Y</sub> ]	~ <i>voit</i> [N <sub>Y</sub> ]	~ <i>prendre</i> [N <sub>Y</sub> ]	~ <i>couvre</i> [N <sub>Y</sub> ]
Labreal <sub>12</sub>	<i>récompenser</i> [N <sub>Y</sub> de ART ~]	————	<i>prendre</i> [N <sub>Y</sub> au ~]	<i>couvrir</i> [N <sub>Y</sub> d'~]

##### • FL adjectivales (intensificateurs et atténuateurs)

	TREMPÉ	BOURRÉ 'soûl'	RESPIRER	RÔLE	RIRE
Magn	~ <i>comme une soupe</i>	<i>raide</i> ~	~ <i>à pleins poumons</i>	~ <i>crucial</i>	~ <i>à se rouler par terre</i>

	BLESSÉ <sub>N</sub>	DIFFÉRENCE	SALAIRE	REGARDER	SOMMEIL	ÂGE
AntiMagn	~ <i>léger</i>	~ <i>ténue</i>	~ <i>maigre</i>	~ <i>du coin de l'œil</i>	~ <i>de chat</i>	<i>bas</i> ~

### FL syntagmatiques non standard

CAFÉ ‘boisson chaude...’

sans ajout de produit laitier : noir | postposé

BOIRE [l'alcool]

un verre d'un trait : //faire cul sec

PRIER

Autorités pertinentes

vous P. de Y-er : Merci [de V('Y')<sub>INF</sub>] [sur un panneau]

## 4 La phraséologie en TALN

Les pragmatèmes, les clichés et les locutions doivent être listés dans le dictionnaire, et nous avons montré, comment.

Ce sont les collocations qui représentent le plus gros problème dans le traitement automatique, vu leur nombre et leur variété. Je pense que la seule solution raisonnable de ce problème est l'usage des fonctions lexicales.

Les FL peuvent être utilisées en TALN – surtout dans la traduction automatique et la génération de textes – des deux façons suivantes.

- Les FL assurent la sélection lexicale correcte dans les cas de type *grave maladie* ~ russe *tjažělaja bolezn'* lit. ‘lourde maladie’, *mettre* [N<sub>V</sub>] *en danger* ~ russe *podvergat'* [N<sub>V</sub>] *opasnosti<sub>DAT</sub>* lit. ‘soumettre [N<sub>V</sub>] au danger’ ou *prendre la fuite* ~ russe *obratit'sja v begstvo* lit. ‘se transformer en fuite’. De telles équivalences « exotiques » sont facilement couvertes par des paires de dictionnaires du type DEC.
- Les FL sont sous-jacentes au paraphrasage du niveau syntaxique profond, ce qui permet de construire la SSyntP acceptable pour la phrase de sortie dans la situation de distorsion entre le vocabulaire de la phrase et sa SSyntP. Par exemple, soit la phrase anglaise (1a) et ses traductions russe et française (1b-c) :

(1) a. anglais *She competes internationally.*

b. russe *Ona učastvuet v meždunarodnyx sorevnovanijax.*

c. français *Elle participe à des compétitions internationales.*

Le verbe signifiant ‘[to] compete’ (dans le sens visé) n'existe ni en russe ni en français. Cependant, un verbe V peut toujours être paraphrasé par le nom déverbatif et son verbe support :  $V \Leftrightarrow S_0(V) \leftarrow \text{II} - 0 \text{per}_1(S_0(V))$ , ce qui donne *compete*  $\Leftrightarrow$  *participate in competition(s)* ; *competition* a des équivalents russe et français directs.

Pour le système universel de paraphrasage SyntP, voir Mel'čuk 1974: 141-176, 1988c, 1992 and 2004, Mel'čuk & Wanner 2006, and Milićević 2007 ; l'utilisation des FL dans la génération



de textes est décrite dans Iordanskaja *et al.* 1996 et Lareau & Wanner. Le système de paraphrasage pour le russe a été implémenté et testé dans une série d'expériences : Apresjan & Cinman 1998 and 2002.

Je vais maintenant présenter trois exemples de traduction qui sont difficiles à cause des collocations impliquées, pour montrer comment les FL permettent d'assurer de bons résultats.

**NB** : Pour être sûr de mes traductions, je ne considère que la traduction vers le russe, en fournissant des gloses détaillées.

**Exemple 1** : le verbe anglais STRIKE 'frapper'

Soit la phrase anglaise (2) :

(2) *The book thief struck again* lit. 'Le voleur de livres a frappé de nouveau'.

Sa traduction russe, la plus littérale dont je suis capable, est

(3) *Knižnyj vor snova soveršil kražu* lit. 'Le voleur de livres de nouveau a commis un vol'.

Pas question de traduire le verbe STRIKE 'frapper' par quelque chose comme UDARJAT' 'frapper' : le résultat serait absolument incompréhensible. Il faut traduire par la collocation *soveršit' kražu* 'commettre un vol'. Mais comment établir l'équivalence *strike* = *soveršit' kražu* ? D'autant plus que dans de nombreux contextes STRIKE a d'autres équivalences :

*The hurricane struck the island again* lit. 'L'ouragan a frappé l'île de nouveau'. =

*Uragan snova obrušilsja na ostrov* lit. 'L'ouragan de nouveau s'est abattu sur l'île'.

*The bullet struck him in the shoulder* lit. 'La balle l'a frappé dans l'épaule'. =

*Pulja popala emu v plečo ...* lit. 'La balle l'a atteint à l'épaule ...'

*A suicide bomber struck in the market* lit. 'Un terroriste kamikaze a frappé au marché'. =

*Terrorist-smertnik podorval sebja na rynke* lit. 'Un terroriste kamikaze a fait exploser soi au marché'.

Dans le cas sous considération la réponse vient immédiatement : tous ces emplois de STRIKE sont des valeurs de la FL  $\text{Fact}_\theta$  :  $\text{Fact}_\theta(L)$  = 'faire ce que (la dénotation de) L est censé faire d'après sa nature'. Dans le DEC russe, on doit avoir:

$\text{Fact}_\theta(\text{VOR 'voleur'})$  : *krast'* 'voler', *soveršat' kražu* 'commettre un vol'

$\text{Fact}_\theta(\text{URAGAN 'ouragan'})$  : *obrušit'sja* 's'abattre'

$\text{Fact}_\theta(\text{PULJA 'balle'})$  : *popast'* 'atteindre'

$\text{Fact}_\theta(\text{TERRORIST-SMERTNIK 'terroriste kamikaze'})$  : *podorvat' sebja* 'se faire exploser'

Dans le DEC anglais, on indiquera la même chose pour les emplois de STRIKE en cause :  $\text{Fact}_\theta(\text{THIEF 'voleur'}) = \textit{strike} 'frapper', etc.$

**Exemple 2** : le verbe polonais OBOWIĄZYWAC ‘obliger’

Considérons le texte d’un panneau vu dans le hall d’un immeuble à Varsovie :

(4) polonais

*Meszkanców budynku obowiązuje cisza nocna*  
 habitant-PL.ACC immeuble-SG.GÉN obliger-PRÉS.3SG silence-SG.NOM nocturne  
 lit. ‘Le silence nocturne oblige les habitants de l’immeuble’.

Le sens en est immédiatement clair : on prie les habitants de ne pas faire de bruit pendant la nuit.

La traduction russe est par conséquent aussi immédiatement claire :

(5) russe

*Žiteli doma objazany noč’ju sobljodat’ tišinu*  
 habitant-PL.NOM immeuble-SG.GÉN sont.obligés la.nuit observer silence-SG.ACC  
 ‘Les habitants de l’immeuble sont obligés d’observer la nuit le silence’.

On peut obtenir cette traduction à partir de (4) en n’utilisant que des informations dictionnairiques en ce qui concerne les phrasèmes. Nous y voyons deux difficultés : la traduction du verbe OBOWIĄZYWAC ‘obliger’ et de l’expression *cisza nocna* ‘silence nocturne’.

La première difficulté est liée au régime particulier du verbe polonais, que ne possède pas son équivalent russe, tout à fait comme son équivalent français : *\*tišina objazyvaet... \**‘le silence oblige...’. Le DEC polonais doit avoir, pour OBOWIĄZYWAC ‘obliger’ l’indication suivante :

$$X \text{ obowiązuje } Y\text{-a} \Leftrightarrow Y \text{ est obligé de } \text{Rea}_1(X)\text{-II} \rightarrow X$$

Cette indication assure la bonne construction de la phrase<sup>1</sup>. Dans un DEC russe, on trouve, sous TIŠINA, que  $\text{Rea}_1(\text{TIŠINA ‘silence’}) = \text{sobljudat’ ‘observer’}$ , ce qui permet de construire la partie initiale de la phrase russe : *Žiteli doma objazany sobljodat’ tišinu*.

La deuxième difficulté relève du fait qu’en russe, le nom TIŠINA ‘silence’ a le collocatif *noč’naja* ‘nocturne’ qui correspond exactement à l’adjectif polonais *nocna* ‘nocturne’ ; cependant, on ne peut pas dire *\*sobljudat’ noč’nuju tišinu !* Pour pallier cela, il suffit d’indiquer, dans le sous-article de  $\text{Rea}_1(\text{TIŠINA ‘silence’})$ , que le modificateur temporel ou locatif du nom doit être transféré vers le verbe SOBLJUDAT’ :

$$*\text{SOBLJUDAT’-II} \rightarrow \text{TIŠINA-ATTR} \rightarrow \Xi_{(\langle\text{temp}\rangle, \langle\text{loc}\rangle)} \Rightarrow \Xi_{(\langle\text{temp}\rangle, \langle\text{loc}\rangle)} \leftarrow \text{ATTR-SOBLJUDAT’-II} \rightarrow \text{TIŠINA}$$

(Cette règle particulière correspond à la règle générale de paraphrasage syntaxique profond : Mel’čuk 1992: 50, Règle n° 19.) Le résultat est *sobljudat’ noč’ju tišinu*. Une bonne traduction a ainsi été obtenue.

**Exemple 3** : le nom français APPOINT ‘montant exact payé par X à Y pour Z et qui n’exige pas de donner l’argent en retour’

Écriteau dans un autobus français :

(6) *Merci de faire l’appoint.*

Sa traduction russe n’est pas évidente, mais en voici une variante possible :

(7) *Platite za proezd bez sdači* lit. ‘Payez pour le trajet sans monnaie rendue’  
[la traduction a été proposée par M. Glovinskaja].

Considérons pas à pas comment une telle équivalence peut être produite par une paire de bons dictionnaires.

- *Merci de Y* est décrit dans un DEC français comme une FL non standard sous PRIER :

autorités pertinentes  
vous prient de Y-er : Merci [de V(‘Y’)INF [sur un écriteau ou panneau]

Dans un DEC russe, on trouve sous PROSIT’ ‘prier’ la même FL non standard :

autorités pertinentes  
vous prient de Y-er : V(‘Y’)IMPÉR.2PL [sur un écriteau ou panneau]

- *Faire* est décrit dans un DEC français comme  $\text{Rea}l_1$  de APPOINT :

$\text{Rea}l_1(\text{APPOINT})$  : faire [l’~]

- APPOINT est traduit comme PLATA BEZ SDAČI lit. ‘paiement sans monnaie rendue’.

- $\text{Rea}l_1(\text{PLATA ‘montant’})$  : //platit’ ‘payer’

Les quatre pas indiqués permettent d’obtenir *Platite bez sdači* ‘Payez sans monnaie rendue’. Mais le russe exige ici une précision : *platit’ za čto?* ‘payer pour quoi?’ – *za proezd* ‘pour le trajet’. Cette précision doit venir de l’information concernant la situation d’énonciation : si l’écriteau est placé dans un moyen de transport publique, on ajoute *za proezd* ; si c’est près d’un guichet d’une billetterie, *za bilet* ‘pour le billet’ est de mise ; si c’est à la caisse d’une cantine, cela sera *za edu* ‘pour la nourriture’.

Il existe aussi une voie plus simple d’établir l’équivalence en cause, notamment décrire *faire l’appoint* comme une FL non standard de PAYER :

de sorte que Y ne doit pas  
rendre la monnaie à X : //faire l’appoint

La FL non standard correspondante en russe serait :

PLATIT’ ‘payer’

de sorte que Y ne doit pas  
rendre la monnaie à X : *bez sdači* ‘sans monnaie rendue’

L'équivalence s'obtient immédiatement. J'ai voulu quand même indiquer les pistes multiples qui peuvent mener vers le même résultat.

## 5 Conclusions

Voici les cinq points les plus importants de cette communication :

1. Les phrasèmes constituent une partie importante du lexique d'une langue. Ils doivent donc être présentés dans un dictionnaire formel (du type DEC) de façon détaillée et systématique.
2. La clé pour la production automatique du texte d'une qualité relativement élevée est le dictionnaire du type DEC.
3. Ce dictionnaire doit réserver une place d'honneur aux collocations décrites en termes de fonctions lexicales.
4. Les fonctions lexicales doivent être exploitées sous deux angles : pour la sélection lexicale et pour le paraphrasage syntaxique profond.
5. Le système de paraphrasage doit faire partie de tout système réussi de TALN.

## Remerciements

Le texte de la présente communication a été lu et critiqué par L. Iordanskaja, Ph. Langlais, S. Mille, A. Polguère et L. Wanner, que je remercie de tout mon cœur. J'ai fait de mon mieux pour prendre en charge leurs remarques et suggestions ; tous les défauts et erreurs survivants ne sont imputables qu'à moi-même.

## Références

- Apresjan, Jurij & Cinman, Leonid (1998) Perifrazirovanie na komp'jutere [Paraphrasing with Computer]. *Semiotika i informatika*, 36, 177-202.
- Apresjan, Jurij & Cinman, Leonid (2002) Formal'naja model' perefrazirovanija predloženij dlja sistem pererabotki tekstov na estestvennyx jazykax [A Formal Model of Sentence Paraphrasing for Natural Language Text Processing Systems]. *Russkij jazyk v naučnom osveščenii*, No. 2 [= 4], 102-146.
- Grossmann, Francis & Tutin, Agnès (eds.) (2003) *Les collocations : analyse et traitement*. [= *Travaux et recherches en linguistique appliquée, Série E : Lexicologie et lexicographie*, n° 1]. Amsterdam: De Werelt.
- Iordanskaja, Lidija, Kim, Myunghee & Polguère, Alain (1996) Some Procedural Problems in the Implementation of Lexical Functions for Text Generation. In: Wanner (ed.) 1996: 279-297.
- Kahane, Sylvain (2003a) The Meaning-Text Theory. In: Agel, Vilmos, Eichinger, Ludwig, Eroms, Hans-Werner, Hellwig, Peter, Heringer, Hans Jürgen, & Lobin, Henning (eds.) *De-*

- pendency and Valency. An International Handbook of Contemporary Research*, vol. 1. Berlin—New York: W. de Gruyter, 546-570.
- Kahane, Sylvain (2003b) Sur le lien entre la définition lexicographique et les fonctions lexicales: une blessure profonde dans le DEC. In: Grossman & Tutin (eds.) 2003: 61-73.
- Kahane, Sylvain & Polguère, Alain (2001) Formal Foundations of Lexical Functions. In: *Proceedings of "COLLOCATION: Computational Extraction, Analysis and Exploitation", 39th Annual Meeting and 10<sup>th</sup> Conference of the European Chapter of the Association for Computational Linguistics*, Toulouse, 8-15.
- Lareau, François & Wanner, Leo (2007) Towards a Generic Multilingual Dependency Grammar for Text Generation. In: Tracy Holloway King & Emily M. Bender (eds.), *Proceedings of the Grammar Engineering Across Frameworks 2007 Workshop. CSLI Publications*, Stanford, CA: Stanford University, <http://csli-publications.stanford.edu>.
- Mel'čuk, Igor (1974) *Opyt teorii lingvističeskix modelej Smysl - Tekst* [Outline of a Theory Meaning-Text Type Linguistic Models]. Moskva: Nauka. [Reprint: 1999.]
- Mel'čuk, Igor (1981) Meaning-Text Models: A Recent Trend in Soviet Linguistics. *Annual Review of Anthropology*, 10, 27-62.
- Mel'čuk, Igor (1988a) *Dependency Syntax: Theory and Practice*. Albany, N.Y.: State University of New York Press.
- Mel'čuk, Igor (1988b) Semantic Description of Lexical Units in an *Explanatory Combinatorial Dictionary*: Basic Principles and Heuristic Criteria. *International Journal of Lexicography*, 1: 3, 165-188.
- Mel'čuk, Igor (1988c) Paraphrase et lexique dans la théorie linguistique Sens-Texte. In: Bès, Gabriel & Fuchs, Catherine (eds), 1988, *Lexique et paraphrase [= Lexique 6]*. Lille: Presses Universitaires de Lille, 13-54.
- Mel'čuk, Igor (1992) 1992 Paraphrase et lexique: Vingt ans après. In: Mel'čuk *et al.* 1992: 9-58.
- Mel'čuk, Igor (1996) Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In: Wanner, Leo (ed.), *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins, 37-102.
- Mel'čuk, Igor (1997) *Vers une linguistique Sens-Texte. Leçon inaugurale*. Paris: Collège de France.
- Mel'čuk, Igor (2003a) Collocations dans le dictionnaire. In: Szende, Thomas (ed.), *Les écarts culturels dans les dictionnaires bilingues*. Paris: Honoré Champion, 19-64.
- Mel'čuk, Igor (2003b) Les collocations : définition, rôle et utilité. In: Grossmann & Tutin (eds.) 2003: 23-31.
- Mel'čuk, Igor (2004) Verbes supports sans peine. *Lingvisticae Investigationes*, 27: 2, 203-217
- Mel'čuk, Igor (2006) Explanatory Combinatorial Dictionary. In: G. Sica, ed., *Open Problems in Linguistics and Lexicography*, 2006, Monza (Italy): Polimetrica Publisher, 225-355. Voir aussi <http://www.polimetrica.com/?p=productsList&sWord=lexicography>
- Mel'čuk, Igor, Clas, André & Polguère, Alain (1995) *Introduction à la lexicologie explicative et combinatoire*. Bruxelles : Duculot.

- Mel'čuk, Igor & Polguère, Alain (1991) Aspects of the Implementation of the Meaning-Text Model for English Text Generation. In: N. Lancashire (ed.), *Research in Humanities 1*, Oxford: Clarendon Press, 204-215.
- Mel'čuk, Igor & Polguère, Alain (2007) *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Bruxelles: De Boeck.
- Mel'čuk, Igor & Wanner, Alain (2006) Syntactic Mismatches in Machine Translation. *Machine Translation*, 21: 81-138.
- Mel'čuk, Igor, et al. (1984, 1988, 1992, 1999) *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques I-IV*. Montréal: Les Presses de l'Université de Montréal.
- Milićević, Jasmina (2007) *La paraphrase. Modélisation de la paraphrase langagière*. Bern etc.: Peter Lang.
- Rey, Alain & Chanterreau, Sophie (1993) *Dictionnaire des expressions et locutions*. Paris: Dictionnaires LE ROBERT.
- Wanner, Leo (ed.) (1996) *Lexical Functions in Lexicography and Natural Language Processing*. Amsterdam/Philadelphia: Benjamins.

## Notes

<sup>1</sup> Voici encore trois exemples du même type pour le verbe polonais OBOWIĄZYWAC 'obliger' (je remercie G. Sekunda qui me les a fournis) :

- (i) *Czy sądu w tej sprawie nie obowiązuje tajemnica lekarska?*  
lit. 'Est-ce que le secret médical n'oblige pas le tribunal dans cette affaire ?' =  
'Est-ce que le tribunal n'est pas obligé, dans cette affaire, de **garder** le secret médical ?'
- (ii) *Kodeks obowiązuje także osoby zatrudnione na podstawie umów prawa prywatnego*  
lit. 'Le code oblige également les personnes employées dans le cadre de contrats de droit privé'. =  
'Les personnes employées dans le cadre de contrats de droit privé sont également obligées de **respecter** le code'.
- (iii) *Uczestników obowiązuje noszenie na czapkach białych pokrowców*  
lit. 'Le port de couverture blanche sur leur casquette oblige les participants'. =  
'Les participants sont obligés de **faire** le port d'une couverture (⇒ **porter** une couverture) blanche sur leur casquette'.