



You have downloaded a document from
RE-BUŚ
repository of the University of Silesia in Katowice

Title: Role et description du contexte dans la traduction automatique des textes : approche orientee objets

Author: Beata Śmigielska

Citation style: Śmigielska Beata. (2011). Role et description du contexte dans la traduction automatique des textes : approche orientee objets. "Romanica Cracoviensia" ([T.] 11 (2011), s. 422-432).



Uznanie autorstwa - Bez utworów zależnych Polska - Ta licencja zezwala na rozpowszechnianie, przedstawianie i wykonywanie utworu zarówno w celach komercyjnych i niekomercyjnych, pod warunkiem zachowania go w oryginalnej postaci (nie tworzenia utworów zależnych).



UNIwersYTET ŚLĄSKI
W KATOWICACH



Biblioteka
Uniwersytetu Śląskiego



Ministerstwo Nauki
i Szkolnictwa Wyższego

Beata Śmigielska

Université de Silésie
de Katowice

RÔLE ET DESCRIPTION DU
CONTEXTE DANS LA TRA-
DUCTION AUTOMATIQUE
DES TEXTES – APPROCHE
ORIENTÉE OBJETS

Cet article illustre au moyen d'un exemple le rôle et la façon de description du contexte dans la traduction automatique des textes.

D'abord, nous allons montrer les traductions des phrases fournies par l'un des systèmes de traduction les plus populaires dont l'accès libre est donné par Google. Cela nous permettra d'analyser et de mettre en rapport l'influence aussi bien des contextes proches que des contextes lointains et leur grande importance dans le processus de la bonne traduction de la langue faite par la machine.

Ensuite, nous présenterons, à travers des emplois du mot *temps*, la manière dont on peut décrire les contextes en question dans le cadre de la méthode orientée objets créée par W. Banyś.

Tous les linguistes lexicographes partagent l'opinion que la polysémie des mots est le phénomène qui pose les problèmes les plus ardues à la description de la langue naturelle. On se rend très bien compte du fait que si c'est l'homme qui fait la traduction du texte en n'importe quelle langue, c'est lui qui résout tous les problèmes qui résultent du caractère polysémique des mots et, qui plus est, il le fait le plus souvent de manière efficace. Il a la capacité de choisir le bon équivalent dans la langue cible et il le fait plus ou moins consciemment. Comment donc décrire la langue de sorte que l'on puisse obtenir les mêmes résultats de traduction, à cette différence près que ce serait à la machine de réaliser cette tâche difficile ?

Prenons tout d'abord quelques exemples de phrases traduites du français vers le polonais par le système de traduction proposé par Google et essayons de voir, d'un côté, s'il fera des fautes de traduction résultant de la polysémie des mots et, de l'autre côté, à quel point ces traductions seront correctes.

Avant de passer à la traduction automatique des phrases faite par le système mentionné, analysons les acceptions du substantif français *temps*, qui est polysémique, dégagées par le *Grand dictionnaire français-polonais* :

TEMPS m.

1. Czas

p.ex. : *Le temps passe vite.*

2. Czas, czasy, epoka
p.ex. : *Les temps sont bien changés.*
3. Czas, pora, okres
p.ex. : *Le temps des vacances a déjà terminé.*
4. Czas (sport)
p.ex. : *Le sportif a amélioré son temps.*
5. Czas (geografia, astronomia)
p.ex. : *Le temps atomique est encore plus précis que la rotation de la terre.*
6. Pogoda, czas
p.ex. : *Aujourd'hui le temps est ensoleillé.*
7. Czas (gramatyka)
p.ex. : *Les temps composés se conjuguent avec un verbe auxiliaire.*
8. Takt, suw (mechanika)
p.ex. : *C'est un moteur à deux temps.*
9. Takt, rytm (muzyka)
p.ex. : *C'est un fort temps.*

Comme nous le voyons, les auteurs du *Grand dictionnaire français-polonais (GDFP)* énumèrent neuf acceptions du substantif français *temps*. Par contre, p.ex. *Le Grand Robert de la langue française (GRLF)*, dégage encore plus d'acceptions pour la même entrée. Et d'après ce que dit W. Banyś (2005 : 59) à propos du nombre des sens des mots polysémiques dans le contexte de la discussion sur le nombre de sens d'un mot et de la traduction « il y a autant de sens différents d'un mot dans la langue source que de ses traductions différentes dans la langue d'arrivée ... ». On voit donc facilement que le mot *temps* est généralement représenté en polonais par quatre types de traductions différentes, telles que *czas*, *pogoda*, *takt*, *suw*. Remarquons à ce propos que les mots *epoka* et *okres* ont leurs équivalents beaucoup plus appropriés en français : pour *epoka* – *une époque*, pour *okres* – *une période* et ils peuvent fonctionner dans les contextes correspondants comme synonymes plus ou moins proches pour le mot *czas*.

En nous limitant à l'un des sens du mot *temps*, à savoir *pogoda*, testons le système de traduction automatique de Google du point de vue de sa capacité de retrouver dans sa base de données lexicales, telle quelle est, l'équivalent correct de ce mot en polonais. Le premier test consistera donc à proposer au système douze phrases simples, assez typiques pour le domaine de la météorologie contenant le mot *temps* employé au sens de *pogoda* (les phrases en français seront choisies au hasard et leur version en polonais sera présentée en italique).

P.ex. :

Tableau 1

1. Le temps est splendide.	<i>Pogoda jest wspaniała.</i>
2. Le temps est magnifique.	<i>Pogoda jest piękna.</i>
3. Le temps est chaud.	<i>Jest gorąco.</i>
4. Le temps est sec.	<i>Jest sucho.</i>
5. Le temps est froid.	<i>?x Pogoda jest zimno.</i>
6. Le temps est couvert.	<i>?x Czas jest pokryte.</i>
7. Le temps est pluvieux.	<i>Pogoda jest deszczowa.</i>

Tableau 2

8. Le temps se dégrade.	<i>Pogoda się pogarsza.</i>
9. Le temps se gâte.	?x <i>Pogoda jest zepsute.</i>
10. Le temps se refroidit.	?x <i>Pogoda stygnie.</i>
11. Le temps se met au beau.	?x <i>Pogoda jest słoneczna.</i>
12. Le temps s'améliore.	<i>Pogoda jest coraz lepsza.</i>

Dans le Tableau 1 nous avons quelques phrases de structure *temps+est+adjectif qualificatif* tandis que dans le Tableau 2 le *temps* est suivi d'un verbe au présent de l'indicatif, suivant le schéma : *temps+verbe au présent de l'indicatif*. On voit bien que le système de traduction automatique en question a généré dans ces cas-ci les phrases correspondantes en polonais plus au moins correctes sauf qu'il a commis quatre fautes de nature syntaxique (5, 6, 9 – ce n'est pas par ailleurs un point fort du système) et de nature sémantique (5, 6, 9, 10, 11). Ce qui est certainement à remarquer, c'est le bon choix du mot *pogoda* parmi d'autres sens possibles du mot en question dans presque tous les cas présentés. Là où il est possible de ne pas employer ce mot explicitement dans la phrase, le système l'a omis correctement (3, 4). Pour le moment il n'y a qu'une erreur de traduction qui résulte du caractère polysémique du mot analysé, à savoir le système a choisi *czas* au lieu de *pogoda* dans l'exemple numéro 6.

Nous continuerons donc le test avec d'autres exemples de phrases simples ayant les mêmes structures que les précédentes, mais en y ajoutant quelques informations supplémentaires, typiques pour les prévisions météorologiques (adverbes de temps, de lieu, de manière) qui devraient aider le système, comme elles aident souvent l'homme, à retrouver le bon équivalent dans la langue cible.

P.ex. :

Tableau 3

A.	13. Le temps est couvert.	?x <i>Czas jest pokryte.</i>
	14. Le temps est couvert de nuages.	?x <i>Pogoda jest pochmurno.</i>
B.	15. Le temps est incertain.	?x <i>Czas jest niepewna.</i>
	16. Le temps est incertain aujourd'hui.	?x <i>Czas jest teraz niepewna.</i>
C.	17. Le temps est lourd.	?x <i>Czas jest ciężki.</i>
	18. Le temps est lourd aujourd'hui.	?x <i>Dziś pogoda jest ciężki.</i>
D.	19. Le temps est triste.	?x <i>Czas jest smutne.</i>
	20. Le temps est triste aujourd'hui.	?x <i>Teraz jest smutne.</i>
	21. Le temps est triste dans le Nord.	?x <i>Pogoda jest ponury na północy.</i>
E.	22. Le temps varie.	?x <i>Czas jest różny.</i>
	23. Le temps varie très vite.	<i>Pogoda zmienia się bardzo szybko.</i>
F.	24. Le temps se brouille.	?x <i>Czas zaciera.</i>
	25. Le temps se brouille très vite.	?x <i>Czas zaciera się bardzo szybko.</i>

G.

26. Le temps se maintient.	?x <i>Pogoda w ładowniach.</i>
27. Le temps se maintient humide.	?x <i>Wilgotnej pogody trwa.</i>

Dans le Tableau 3, parmi quinze phrases, il n'y en a qu'une seule (23) qui est traduite correctement. D'autres exemples prouvent que l'adjonction des éléments adverbiaux influence, dans une certaine mesure, la traduction au niveau sémantique. Il faut souligner que dans les cas où nous avons ajouté des adverbes aux phrases, le choix de l'équivalent est souvent bien fait par la machine. Toutes les phrases contiennent cependant des fautes de nature syntaxique, sauf 23.

Le pas suivant dans nos analyses concernera les structures de phrases de type : *il fait beau temps, il fait mauvais temps* ou *il fait +un+(adj)+temps+(adj)*, qui sont d'ailleurs très représentatives pour le domaine étudié et, de ce point de vue, elles devraient être facilement reconnues par le traducteur automatique.

P.ex. :

Tableau 4

28. Il fait un temps horrible.	<i>Pogoda jest okropna.</i>
29. Il fait un temps insupportable.	<i>Pogoda jest nie do zniesienia.</i>
30. Il fait un temps ensoleillé.	<i>Pogoda jest słoneczna.</i>
31. Il fait un temps maussade.	<i>Pogoda jest ponura.</i>
32. Il fait un temps glacial.	?x <i>Pogoda jest zimno.</i>
33. Il fait un temps de chien.	?x <i>Pies pogody.</i>
34. Il fait mauvais temps.	?x <i>Pogoda jest zła.</i>
35. Il fait mauvais temps ce week-end sur Paris.	?x <i>To jest zła pogoda w ten weekend w Paryżu.</i>
36. Il fait un mauvais temps automnal.	?x <i>Miał słabe jesienna pogoda.</i>
37. Il fait beau temps.	?x <i>Pogoda jest dobra.</i>
38. Il fait beau temps à Paris.	?x <i>To miło, pogoda w Paryżu.</i>
39. Il fait un beau temps printanier.	?x <i>To sprawia, że piękna pogoda wiosną.</i>

On voit donc que le système a bien traduit le mot *temps* dans les phrases ci-dessus, à ceci près que seulement les quatre : 28, 29, 30, 31 sont acceptables de tous les points de vue. Les phrases qui contiennent des éléments que l'on peut traiter dans une certaine manière en tant qu'expressions soit plus ou moins figées soit, au moins, représentatives pour le domaine étudié ont conduit le système à choisir, au niveau du sens, l'équivalent polonais *pogoda* et non pas *czas*, ce qui est le cas le plus fréquent dans les phrases du type précédent (Tableau 3). On pourrait soupçonner que dans tous nos exemples analysés aussi bien la construction impersonnelle *il fait* que la co-présence de certains éléments p.ex. de certains adjectifs qualificatifs de type : *printanier, automnal, ensoleillé, maussade, glacial, insupportable, de chien, horrible* activent probablement la bonne traduction du mot en permettant de discriminer ses différents emplois (cf. Śmigielska 2007).

Et encore quelques exemples intéressants :

P.ex. :

Tableau 5

40. On a annoncé le mauvais temps pour dimanche prochain.	?x Zapowiedział on w złym czasie dla następnej niedzieli.
41. Marie a annoncé le mauvais temps pour dimanche prochain	?x Mary ogłoszenia w złym czasie dla następnej niedzieli.
42. La station météo a annoncé le mauvais temps pour dimanche prochain.	?x Stacja meteorologiczna powiedziała, że pogoda w najbliższą niedzielę.
43. Le journaliste a annoncé le mauvais temps pour dimanche prochain.	?x Dziennikarz zapowiada pogody w najbliższą niedzielę.
44. Le vent de l'est, les températures basses – tout cela annonce le mauvais temps pour dimanche prochain.	?x Wschodnie wiatry i niskie temperatury – wszystko w niewłaściwym czasie delegowania do następnej niedzieli.
45. J'ai regardé la météo hier soir. On a annoncé le mauvais temps pour dimanche prochain. Il pleuvra, il fera un vent violent sur la majeure partie du pays.	?x Patrzyłem na pogodę w nocy. Zapowiedział on w złym czasie dla następnej niedzieli. Deszcz pada, to pozwoli gwałtownego wiatru na większości kraju.

Les phrases du Tableau 5 nous permettent de tester si le système est capable de faire la traduction adéquate de l'expression *annoncer le mauvais temps* (*zapowiadać złą pogodę*). Cinq premières phrases (40, 41, 42, 43, 44) sont des phrases simples, mais dans chacune apparaissent des éléments différents en position du sujet. Là où en position du sujet nous avons *station météo* et *journaliste*, donc les entités qui sont censées *annoncer le temps* p.ex. dans les médias où le temps est généralement annoncé, le traducteur automatique choisit la forme de *pogoda*. Même si *Marie*, p.ex. travaillant en tant que journaliste, pourrait le faire aussi, le système « n'en sait rien » et ne réagit pas convenablement. La dernière phrase est mise dans le contexte météorologique plus large. De son côté droit et gauche il y a des informations lexicales telles que *météo*, *pleuvra*, *fera un vent* qui, même si elles auraient dû le faire, n'indiquent pas tout de même au système le bon chemin à prendre.

Ces exemples et ces réflexions nous amènent à constater que le système de Google est encore loin d'être efficace en traduction automatique des textes et il fait beaucoup de fautes aussi bien au niveau syntaxique qu'au niveau sémantique. Il ne fait certainement pas recours dans sa base de données lexicales à de telles informations, parce qu'il n'en dispose probablement pas, qui permettent de discriminer les sens du mot avec la précision nécessaire à la reconnaissance ou à la génération de phrases correctes. Le système, que nous avons testé, c'est un système de traduction qui s'appuie sur les techniques de « l'apprentissage statistique ». Le modèle de traduction ainsi construit est fondé sur un grand nombre de textes bilingues ou multilingues parallèles, traduits toujours par l'homme à partir d'une langue de départ vers une ou des langues d'arrivée.

Au cours des analyses de tous les exemples cités ci-dessus, on observe que dans le processus de traduction le système prend souvent, mais malheureusement pas toujours, en considération les contextes proches où apparaît le mot à traduire, ce qui est le plus important dans le cas des mots de caractère polysémique. D'où vient que dans certains

cas il est capable de prévoir la traduction correcte dans la langue cible (ici en polonais). Par contre les contextes plus lointains ne sont pas pris en compte dans ce processus, ce qui faciliterait sans aucun doute, si c'était le cas, la tâche.

Que le phénomène de la polysémie pose, comme nous l'avons remarqué au début, le plus de problèmes à la description précise de la langue naturelle est un fait reconnu par tous ceux qui ont travaillé sur différentes méthodes permettant de désambigüiser les sens des mots dans le cadre de traduction automatique. Dans ce qui suit nous nous concentrerons sur l'approche orientée objets élaborée par W. Banyś (2001a,b, 2005) qui propose quelques solutions intéressantes et efficaces à ce type de problèmes. La méthode, comme beaucoup d'autres, souligne qu'il faut étudier toujours les sens des mots en fonction de leurs emplois concrets dans les textes. Pour indiquer donc tous les sens des unités lexicales de manière précise, il est nécessaire très souvent de se référer aux contextes, plus ou moins proches, où elles apparaissent. Comment donc décrire le substantif français *temps*, vu naturellement son caractère polysémique, d'après les principes de l'approche mentionnée de sorte que sa traduction automatique vers le polonais soit adéquate?

Commençons, dans cette présentation très brève, par le principe que chaque élément (objet) de la langue se caractérise par la co-présence d'attributs et d'opérations d'un certain type qui l'accompagnent dans la phrase. C'est une sorte de collocations qui reposent sur un rapprochement particulièrement fréquent de deux éléments du lexique. Vu que la machine, n'étant pas encore à l'heure actuelle capable de raisonner toute seule, a besoin d'une base de données lexicales précises sur lesquelles elle peut travailler, c'est l'homme qui doit introduire, d'une manière ou d'une autre, sur une fiche de description établie, toutes les données sur le comportement linguistique du mot analysé du point de vue, d'un côté, des opérations que son référent effectue et qui sont effectuées sur lui et, de l'autre côté, des attributs qui se réfèrent à sa structure.

Voici un échantillon de description contrastive français-polonais du mot *temps* (partie sémantique) :

Tableau 6

<i>FR</i>	<i>PL</i>
TEMPS	POGODA
[Classe d'objets:] <i>phénomènes naturels</i>	[Klasa obiektowa:] <i>zjawiska naturalne</i>
[Définition:] <i>état du climat, de l'atmosphère en un endroit</i>	[Definicja:] <i>stan atmosfery ziemskiej w danym miejscu i czasie</i>
[Synonymes:] <i>climat</i>	[Synonimy:] <i>klimat</i>
[Super classe2:] <i>événements</i>	[Superklasa2:] <i>zdarzenia</i>
[Super classe1:] <i>phénomènes</i>	[Superklasa1:] <i>zjawiska</i>
[Domaine:] <i>météorologie</i>	[Dziedzina:] <i>meteorologia</i>
[Attributs:]	[Atrybuty:]
<i>beau temps</i>	<i>ładna pogoda</i>
<i>mauvais temps</i>	<i>brzydka pogoda</i>
<i>temps automnal</i>	<i>jesienna pogoda</i>
<i>temps chaud</i>	<i>ciepła pogoda</i>

<i>temps froid</i>	<i>zimna pogoda</i>
<i>temps incertain</i>	<i>niewyraźna pogoda</i>
<i>temps magnifique</i>	<i>cudowna pogoda</i>
<i>temps maussade</i>	<i>ponura pogoda</i>
<i>temps pluvieux</i>	<i>deszczowa pogoda</i>
<i>temps tempéré</i>	<i>umiarkowana pogoda</i>
<i>etc.</i>	<i>etc.</i>
[Opérateurs:]	[Operatory:]
<i>le temps change</i>	<i>pogoda się zmienia</i>
<i>le temps s'adoucit</i>	<i>pogoda się ociepla</i>
<i>le temps s'éclaircit</i>	<i>pogoda się przejaśnia</i>
<i>le temps se gâte</i>	<i>pogoda się psuje</i>
<i>le temps se maintient</i>	<i>pogoda się utrzymuje</i>
<i>le temps se rafraîchit</i>	<i>pogoda się ochładza</i>
<i>etc.</i>	<i>etc.</i>
<i>il fait un + adj. + temps</i>	<i>jest + przym. + pogoda</i>
<i>il fait un temps + adj.</i>	<i>jest + przym. + pogoda</i>
<i>annoncer le temps</i>	<i>zapowiedzieć pogodę</i>
<i>attendre le temps</i>	<i>czekać na pogodę</i>
<i>parler du temps</i>	<i>rozmawiać o pogodzie</i>
<i>penser au temps</i>	<i>mysleć o pogodzie</i>
<i>prévoir le temps</i>	<i>przewidzieć pogodę</i>
<i>etc.</i>	<i>etc.</i>
[Extensions :]	[Ekstensje :]
<i>il fait beau temps</i>	<i>jest ładna pogoda</i>
<i>il fait mauvais temps</i>	<i>jest brzydka pogoda</i>
<i>le temps à ne pas mettre le chien dehors.</i>	<i>pogoda, że psa nie wyrzucisz</i>
<i>le temps à ne pas mettre le nez dehors.</i>	<i>pogoda, że nosa nie wysuniesz na zewnątrz</i>
<i>le temps de chien</i>	<i>pogoda pod psem</i>
<i>etc.</i>	<i>etc.</i>
[Frames :]	[Kadr :]
<i>temps</i>	<i>pogoda</i>

La description du type ci-dessus du mot polysémique *temps* tient compte de tous les éléments (opérateurs et attributs) qui entrent en contact avec lui dans les phrases réelles produites par les locuteurs de la langue et qui constituent son contexte d'emploi le plus proche. Et, de ce point de vue, il est naturel que c'est, entre autres, grâce à ses co-occurrences qu'il soit possible d'arriver à expliciter tout sens du mot. Il ne faut pas oublier que puisque, dans cette optique, la description des sens des mots est faite toujours en termes de leurs emplois concrets, tous les mots de la langue se caractérisant par un certain nombre d'opérateurs et d'attributs, qui leur sont propres, peuvent être regroupés suivant ce critère en classes de mots appelée *classe d'objets* (Gross 1994, 1997, 2008 ; Le Pesant 1994).

On obtient ainsi une description du monde, composé (de classes) d'objets du monde, tel qu'il est vu par la langue et dans la langue, ce qui fait que la base de données lexicales de ce type, compilant des informations sur le comportement linguistique des éléments de la langue en fonction aussi des connaissances sur le monde, devient en même temps un thésaurus. (Banyś 2001a : 22).

Dans l'approche orientée objets on indique donc à chaque fois les ensembles des caractéristiques opérationnelles et attributives des mots et, ce qui en résulte, on les répartit en classes d'objets correspondantes. Parmi tous les paramètres importants de la description, il y en a un qui nous paraît indispensable pour la traduction, et surtout celle qui est automatique : c'est la nécessité de préciser le cadre où un mot donnée « réalise » son sens particulier. Toutes les informations de nature sémantique qui servent à caractériser de manière précise une classe d'objets sont organisées par le *frame* (*cadre*) qui « représente typiquement un état de choses statique présentant des caractéristiques typiques des objets typiques qu'il renferme » (Banyś 2000 : 45). Le cadre étant une situation, un contexte plus large, il donne des indices précis quant au sens des mots (cf. Minsky 1975, Schank, Abelson 1977). C'est une information essentielle et très souvent c'est elle qui décide, en dernière instance, de la bonne traduction des textes comportant des mots polysémiques. Il arrive que seuls les opérateurs et les attributs, donc le contexte immédiat des mots, permettent de les désambiguïser, mais il y a des cas où on est obligé de chercher plus loin et de faire recours au contexte plus large qui nous indique en effet de quel cadre il s'agit et, en conséquence, de quel sens du mot il est question.

Analysons p.ex. des emplois ambigus du mot *temps* du type :

1. *C'est un temps triste.*
2. *C'est un mauvais temps.*

Ces deux phrases, même si, au premier coup d'œil, faciles à traduire, peuvent s'avérer difficiles sans faire recours au contexte ou à la situation où elles ont été prononcées ou écrites. Il y a donc deux possibilités de les interpréter en polonais :

Ad 1. Pogoda jest smutna./To jest smutny czas.

Ad 2. Jest brzydka pogoda./To jest zły czas.

Afin de faire la désambiguïstation des sens du mot en question, il faut donc spécifier les cadres où il est employé. Cette spécification se fait par le listing du plus grand nombre d'éléments de la langue qui sont apparentés les uns avec les autres par le sens et qui créent un réseau sémantique. Pour définir le cadre où le mot *temps* est employé au sens de *pogoda*, il suffit d'énumérer les mots qui apparaissent le plus souvent dans son voisinage quand on parle du temps, pris au sens météorologique. Cette démarche peut s'avérer minutieuse et exigeant beaucoup d'efforts et de temps, si on le fait manuellement mot par mot, sauf si l'on se sert de dictionnaires électroniques déjà existants qui disposent d'une base de données lexicales riches en informations sur différents types de relations entre les mots, telles que: analogies, synonymes, antonymes, collocations. Et c'est grâce à ce type de dictionnaires que contiennent p.ex. *Antidote*, *Sensagent*, *Cordial*, et beaucoup d'autres, que ce travail peut se faire semi automatiquement. Dans le cadre donc de *temps-pogoda* on prendrait sûrement en considération, dans un premier temps, des mots clés pour le domaine, tels que p.ex. : *vent*, *pluie*, *ciel*, *soleil*, *neige*, *arc-en-ciel*, *nuage*, *brume*, *température*, etc. Le principe est que si l'on a plus d'éléments qui sont listés dans le cadre, on a plus de chances d'obtenir la traduction adéquate d'un mot polysémique.

Analysons de ce point de vue quatre fragments de textes tirés d'Internet qui contiennent les phrases ambiguës du type dont nous avons parlé plus haut :

Dimanche 30 mai, s'est déroulé le CSI jeunesse. Nous attendions le beau, mais c'est un temps triste qui nous accompagna toute la journée, avec un léger passage pluvieux.....Ce qui n'a pas dérangé les athlètes !!

(<http://athlevald.over-blog.com/article-csi-jeunesse-30-05-2010-51551602.html>, 29.10.2010)

Les éléments : *le beau, passages pluvieux* nous renvoient au sens *pogoda* du mot *temps*.

On nous a habitués à la logique, au temps de l'horloge et du calendrier, c'est pratique pour aller prendre le train, souhaiter les anniversaires, mais c'est faux. C'est un temps triste. Au fond on ne sait jamais l'heure qu'il est.

(Jean SULIVAN, <http://www.biblioconcept.com/citations/temps.htm>, 29.10.2010)

Dans ce fragment il n'y a pas d'éléments appartenant au cadre *pogoda* donc il n'est pas possible de traduire le mot *temps* par ce terme. Pour trouver le bon équivalent du mot, le système devrait chercher dans d'autres cadres proposés.

J'avais envie de faire 1:01 mais dans ma tête je savais que j'allais faire 1:02 », a commenté Manaudou, qui n'a pas caché la déception que lui procure ce deuxième mauvais temps en deux compétitions, après l'Open de Paris. « C'est difficile parce que c'est un mauvais temps, je n'aime pas nager pour faire ça. Même si ça me sert, je ne vois pas trop l'intérêt. Ça me déprime un peu quand je vois le temps, ou que je n'arrive pas à partir par rapport aux autres... »

(<http://www.rfi.fr/sportfr/afp/001/spo/newsmlmmd.18a3b92756c211542f03ea35e0cca6ee.731.asp>, 29.10.2010)

Le texte ci-dessus ne contient non plus aucun élément sémantiquement lié au domaine météorologique. La traduction *pogoda* devrait donc être bloquée dans cette situation-là.

...tu travailles sous la pluie? J'ai remarqué que beaucoup de gens se plaignent de la pluie et ils disent que c'est un mauvais temps, pour moi, c'est l'Espoir et L'Espérance, la nature a besoin de bons arrosages venus du ciel...

(<http://kriscounette.centerblog.net/6487169-135-DOS-D-ANE>, 29.10.2010)

La *pluie* et *ciel* indiquent immédiatement l'équivalent polonais *pogoda*.

Nous venons de voir que la méthode orientée objets permet de décrire les différents emplois des mots grâce à quelques principes importants que nous n'avons présentés dans ce texte que de manière très générale. Aussi bien le regroupement des mots en classes d'objets que le listing des co-occurrences des mots y appartenant, situés dans un cadre donné, bien spécifié, sont de bons outils qui servent à lever les ambiguïtés et à éviter une grande partie des erreurs de traduction. Mais, même si la description des mots suivant l'approche présentée est tellement précise et constitue un système cohérent, il existe quand même des cas douteux qui sont ambigus et difficiles à traduire pour l'homme et ils le sont d'autant plus pour la machine. P.ex. :

La pluie tombait si serrée qu'elles ne regrettaient presque pas de ne pas aller voir leur oncle Alfred. Elles parlaient de jouer au loto, lorsqu'elles virent le dindon traverser la cour en courant. Il se mit à l'abri sous le hangar, secoua ses plumes mouillées et essuya son grand cou...

– *C'est un mauvais temps pour les dindons, fit observer Delphine, et pour les autres bêtes aussi. Heureusement, ça ne dure jamais longtemps. Mais s'il pleuvait pendant quarante jours et quarantenuits ?*

(D'après Marcel Aymé, *Les contes bleus du Chat perché*)

Il pleut et c'est un mauvais temps pour les mouvements. Des troupes passent, retournant en train vers Lille. Hier et aujourd'hui, on a visité les caves pour y loger des hommes en cas de bombardement.

(<http://home.nordnet.fr/~jdujardin/haubourdin/g16m09.html>, 29.10.2010)

Pour rendre compte de ce type de phrases, nous avons à prendre en considération des éléments plus subjectifs que précédemment et nous appuyer sur une interprétation encore davantage préférentielle qu'absolue, s'il peut y en avoir une d'une manière générale dans ce type de cas.

Or, dans les phrases en question, on a affaire à une construction contenant la préposition *pour* : *c'est un mauvais temps pour...*, qui semble suggérer, à la différence de *il fait mauvais temps* où *mauvais temps* constitue une expression tendant à être figée (remarquons à ce propos aussi la difficulté des expressions du type *?x il fait mauvais temps pour...*, avec 0 occurrences sur le web), que ce n'est pas tellement *le temps – pogoda* qui est visé, *le temps – pogoda* semblant être généralement une donnée objective, mais *le temps – czas*, qui peut être considéré comme ayant une différente valeur pour différents sujets du point de vue de différents buts visés et ceci malgré l'apparition dans les deux phrases des actualisateurs standard de l'interprétation météorologique du mot *temps* du type : *La pluie tombait (...). Il se mit à l'abri sous le hangar, secoua ses plumes mouillées et essuya son grand cou* ou *Il pleut et c'est un mauvais temps pour les mouvements* (dans cette dernière phrase l'apparition de la conjonction *et* semble renforcer, sans exclure totalement l'autre interprétation, davantage l'interprétation non météorologique du mot *temps*, puisque, du point de vue de la pertinence des énoncés, la répétition de la même ou presque la même information dans les deux éléments dans une construction conjonctive a une valeur plutôt limitée).

Autant dire que les approches basées sur une analyse de type statistique des textes parallèles tels que le module de Google devraient être complétées par des analyses de type probabiliste, préférentiel, comme p.ex. une approche orientée objets et c'est la jonction des deux qui semble fournir un outil de traduction assistée par la machine de plus en plus efficace et adéquat.

BIBLIOGRAPHIE

- BANYŚ Wiesław, 2000, *Système de si en français moderne. Esquisse d'une approche cognitive*, Katowice : Wyd. Uniwersytetu Śląskiego.
- BANYŚ Wiesław, 2001a, Bases de données lexicales électroniques – une approche orientée objets, Partie I: Questions de modularité, *Neophilologica* 15, 7–28.
- BANYŚ Wiesław, 2001b, Bases de données lexicales électroniques – une approche orientée objets, Partie II: Questions de description, *Neophilologica* 15, 206–248.
- BANYŚ Wiesław, 2005, Désambiguïsation des sens des mots et représentation lexicale du monde, *Neophilologica* 17, 57–76.
- GROSS Gaston, 1994, Classes d'objets et description des verbes, *Langages* 115, 15–30.
- GROSS Gaston, 1997, Un outil sémantique pour la traduction automatique : les classes d'objets, *Indo-French Symposium on NLP, Hyderabad*, 21–26.
- GROSS Gaston, 2008, Les classes d'objets, *Lalie* 28, 113–165.
- LE PESANT Denis, 1994, Les compléments nominaux du verbe lire, une illustration de la notion de « classe d'objets », *Langages* 115, 31–46.

- MINSKY Marvin, 1975, A Framework for Representing Knowledge, (in :) *The Psychology of Computer Vision*, P.H. Winston (ed.), New York : McGraw-Hill.
- SCHANK Roger, ABELSON Robert, 1977, *Scripts, Plans, Goals and Understanding*, Hillsdale : Lawrence Erlbaum.
- ŚMIGIELSKA Beata, 2007, Remarques sur la traduction automatique et le contexte, *Neophilologica* 19, 253–267.

Summary

The role and description of context in automatic translation of texts – the object oriented approach

The article is devoted to the description of intra-sentential and extra-sentential context in automatic translation based on the object oriented approach proposed by W. Banyś.

In the process of analyzing and describing a given word in terms of the number of its meanings, in order to solve one of the main problems concerning the automatic translation of texts, which is the phenomenon of polysemy, a linguist comes across different kinds of difficulties. It is often difficult to determine the appropriate meaning of a given word if the analysis is limited to just one sentence in which the word appears. Very often it is necessary to seek for more information outside the sentence, in a much broader context, which is strictly connected with the notion of frames and scripts supplying us with essential data necessary to reveal a correct translation of a given word in the target language.

The author of the article presents on the chosen examples how to precisely describe context in the object oriented approach in order for the automatic translation to be accurate.

Streszczenie

Rola i opis kontekstu w tłumaczeniu automatycznym tekstów – ujęcie zorientowane obiektowo

Artykuł jest poświęcony opisowi roli bliższego i dalszego kontekstu w tłumaczeniu automatycznym tekstów w ujęciu zorientowanym obiektowo autorstwa W. Banyśa.

Podczas analizy i opisu danego słowa pod względem liczby posiadanych znaczeń, w celu rozwiązania jednego z głównych problemów w automatycznym tłumaczeniu tekstów, którym jest zjawisko polisemii, bardzo często można natrafiać na różnego typu trudności. Ograniczając się w swej analizie jedynie do zdania, w którym wystąpiło dane polisemiczne słowo, często niezwykle trudno określić jego właściwe znaczenie. W takich przypadkach należy znaleźć potrzebne informacje poza zdaniem, sięgając do szerszego kontekstu, który ściśle wiąże się z pojęciem kadru i skryptu dostarczającego niezbędnych wskazówek, koniecznych do dobrego tłumaczenia na język docelowy.

Opierając się na wybranych przykładach, autorka przedstawia, jak precyzyjnie opisać kontekst w ujęciu zorientowanym obiektowo, aby tłumaczenie tekstów wykonane przez maszynę było poprawne.