Beata Nadzieja

«Jean a attrape la grippe et la balle» : ou caomment desambiguïser les mots sur l'exemple des verbes polonais

Annales Neophilologiarum nr 6, 107-116

2012

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

MUZEUM HISTORII POLSKI

ANNALES NEOPHILOLOGIARUM 6
Rok 2012

JĘZYKOZNAWSTWO

BEATA NADZIEJA[*]
Uniwersytet im. Adama Mickiewicza w Poznaniu

« JEAN A ATTRAPE LA GRIPPE ET LA BALLE » OU COMMENT DESAMBIGUÏSER LES MOTS SUR L'EXEMPLE DES VERBES POLONAIS

Le but du présent article est de présenter les résultats des recherches effectuées dans le cadre du projet « Technologie przetwarzania tekstu polskiego zorientowane na potrzeby bezpieczeństwa publicznego » (projet MNiSW R0002802)[1] dans le but de créer une description sémantique et syntaxique d'un échantillon remarquable de verbes polonais, et ceci de manière à la rendre « lisible » pour un logiciel, donc en essayant d'éviter toute ambiguïté possible.

1. L'ambiguïté – quelques généralités

Si l'on parle de la désambiguïsation, il faudrait d'abord rappeler brièvement la notion de l'ambiguïté. L'**ambiguïté** est la propriété d'un mot ou d'une suite de mots d'avoir plusieurs sens ou plusieurs analyses grammaticales possibles. La définition simple mais précise en est proposée par Catherine Fuchs qui dit qu'il y a l'ambiguïté « lorsqu'à une forme unique correspondent

---

[*] Beata Nadzieja od 1996 roku jest wykładowcą w Instytucie Językoznawstwa na Wydziale Neofilologii UAM. Obecnie przygotowuje rozprawę doktorską w Instytucie Filologii Romańskiej UAM.

[1] Cf. Z. Vetulani, J. Marciniak, T. Obrębski, G. Vetulani, A. Dąbrowski, M. Kubis, J. Osiński, J. Walkowska, P. Kubacki, K. Witalewski, *Zasoby językowe i technologie przetwarzania tekstu. POLINT-112-SMS jako przykład aplikacji z zakresu bezpieczeństwa publicznego*, Poznań 2010.

plusieurs significations »[2]. Les problèmes liés à l'ambiguïté ont été déja maintes fois discutés en linguistique car le phénomène est omniprésent dans chaque langue (rares sont les unités qui gardent le même sens dans tous leurs emplois). Bien que l'ambiguïté concerne tout type d'unités – à la fois les mots simples et les phrases, jusqu'aux textes entiers –, les lexicographes et les traductologues s'efforcent de la lever, surtout à cause des besoins du traitement automatique du langage qui exige un autre genre de dictionnaires.

L'ambiguïté peut se manifester :

– au niveau de la prononciation – les mots possédant différentes significations se ressemblent phonétiquement (p.ex. « *voire* » vs « *voir* » ou « *jumelles* » signifiant en français 'appareil optique' ou 'sœurs') ; il est question ici de l'ambiguïté lexicale – fruit de la polysémie ou de l'homonymie ;

– au niveau syntaxique de la phrase où telle ou autre segmentation détermine l'interprétation : par exemple la phrase française « *J'ai vu un homme avec des lunettes* » reçoit deux sens : 1) « c'est au moyen des lunettes que j'ai vu un homme » – où l'élément « *lunettes* » est attaché au verbe « *voir* », 2) « j'ai vu un homme qui portait des lunettes » où le même élément est attaché au complément d'objet direct « *un homme* ». Des situations pareilles apparaissent également en polonais. Les phrases comme « *Dziecko pogryzło szczenię* » sont difficiles à interpréter à cause de l'ordre libre des mots – deux lectures sont possibles : S-V-C et C-V-S ; ni la flexion des substantifs, ni le contexte (trop court ici) ne permettent de désambiguïser ces structures ;

– au niveau pragmatique où sans la reconnaissance de la situation de l'énonciation d'une phrase il est impossible de reconnaître la valeur de l'acte de langage – par exemple une phrase comme « *Il est tard* » peut signifier que le locuteur constate uniquement l'état de choses (l'acte de constatation) ou qu'il exprime, par exemple, sa volonté que son interlocuteur parte (l'acte de demander ou d'ordonner gentiment).

L'ambiguïté linguistique peut être de deux natures : volontaire et involontaire. La première échappe au contrôle des locuteurs, la deuxième est construite délibérément. La coexistence de plusieurs significations ou interprétations possibles, qui sont souvent considérées comme preuves de la richesse d'une langue donnée, constituent de toute évidence un véritable problème pour ceux qui cherchent à créer des logiciels pour le traitement automatique

[2] C. Fuchs, *Les ambiguïtés du français*, Paris 1996, p. 7.

du langage. Or de tels logiciels exigent l'existence de l'univocité entre formes et contenus (une transparence du langage est une condition *sine qua non* du bon fonctionnement des systèmes informatiques analysant le système naturel de l'homme). En effet, si un locuteur « humain » est capable de déchiffrer le sens à partir de la signification ambigüe d'une unité ou d'un énoncé concrets dans un contexte communicatif bien défini, ceci n'est pas le cas des ordinateurs[3].

2. Les démarches méthodologiques

Les travaux concernant la description des verbes en question se sont basés sur deux dictionnaires polonais : *Uniwersalny słownik języka polskiego* (USJP) de Stanisław Dubisz et Elżbieta Sobol[4] et *Słownik syntaktyczno-generatywny czasowników polskich* (SSGCP) de Kazimierz Polański[5]. Ils ont été choisis en guise de sources de définitions sémantiques des verbes polonais. La confrontation de ces deux sources a permis de construire un dictionnaire sémantico-syntaxique de plus de 2000 verbes polonais.

La sémantique de chaque verbe a été d'abord définie à partir des définitions fournies par deux dictionnaires de base. Ensuite, les chercheurs[6] ont pris en considération le schéma valenciel de chaque signification (établi à l'aide de SSGCP, les modifications et les corrections nécessaires ajoutées). L'idée d'attribuer un schéma particulier à chaque sens du mot vient de Tesnière[7] qui considère que tout verbe est doté d'une valence, c'est-à-dire est susceptible de régir un certain nombre d'actants. Ainsi définie, la valence va de pair avec le nombre et les traits sémantiques des arguments obligatoires auxquels il faut rajouter les éléments introducteurs de ces arguments. Le schéma valenciel, ainsi constitué et inclus dans un logiciel de traitement automatique des langues, pourrait donc constituer une information sur la signification exacte du verbe en question.

---

[3] Là, on voit bien la suprématie de la langue parlée sur l'écrit car à l'oral ce contexte communicatif est donné au moment de l'acte de la parole.

[4] S. Dubisz, E. Sobol, *Uniwersalny słownik języka polskiego*, Warszawa 2003.

[5] K. Polański, *Słownik syntaktyczno-generatywny czasowników polskich*, t. 1–4, Wrocław 1980–1990 ; t. 5, Kraków 1992.

[6] Le groupe se composait de quatre personnes : B. Nadzieja, A. Kaliska, A. Walkowska et W. Wojciechowska qui travaillaient sous la direction de Z. Vetulani et T. Obrębski.

[7] L. Tesnière, *Éléments de syntaxe structurale*, Paris 1959.

Il s'ensuit que l'analyse automatique ne saurait se passer de l'information valencielle[8].

Dans le corpus des verbes polonais soumis à l'analyse pour les besoins du projet il existe un certain nombre de prédicats verbaux qui possèdent plusieurs significations que l'on peut distinguer grâce à la valence. C'est le cas entre autres de :

CZEKAĆ[9] (ATTENDRE)

1. signification – « se tenir dans un lieu où quelqu'un ou quelque chose doit venir ou avoir lieu »

Nom.+Verbe+Prép NA + Acc « *Pies czekał na swojego pana.*[10] »
Nom.+Verbe+Compl.lieu « *Brat czekał na przystanku.* »
Nom.+Verbe+Compl.temps « *Piotr czekał pół godziny.* »
Nom.+Verbe+Prép NA + Acc+Compl.lieu « *Brat czekał na mnie na przystanku.* »
Nom.+Verbe+Prép NA + Acc+Compl.temps « *Pies długo czekał na swojego pana.* »
Nom.+Verbe+Compl.lieu+Compl.temps « *Piotr czekał na przystanku pół godziny.* »
Nom.+Verbe+Prép NA + Acc+Compl.lieu+Compl.temps « *Piotr czekał na autobus pół godziny na przystanku.* »

2. signification – « s'attendre à quelque chose, vouloir que quelque chose se produise »

Nom.+Verbe+Gen. « *Z utęsknieniem czekaliśmy świtu.* »

3. signification – « rester dans la même attitude, ne rien faire avant que cette chose ne se produise »

Nom.+Verbe+Compl.temps « *Sędzia czekał chwilę, a potem powtórzył pytanie.* »
Nom.+Verbe+Compl.manière « *Trzydzieści tysięcy robotników czekało z założonymi rękami.* »
Nom.+Verbe+Compl.temps+Compl.manière « *Sędzia czekał kilka minut bez słowa.* »

4. signification – « en parlant des choses – être promis, prévu pour quelqu'un »

Nom.+Verbe+Acc « *Czekały go wielkie wydatki.* »

---

[8] Ce travail typologique jouerait également un rôle important dans l'enseignement des langues où il faudrait sensibiliser les apprenants à apprendre les verbes toujours avec leur valence.

[9] La description des verbes – qui provient du dictionnaire de Polański – est une version simplifiée de la notation utilisée dans le projet.

[10] Les exemples cités viennent pour la plupart des cas du SSGCP.

3. Les indices et les rôles sémantiques

Les analyses des dictionnaires ont vite démontré que les verbes du type « *czekać* », syntaxiquement faciles à annoter par un logiciel, sont assez rares ; beaucoup plus souvent, différentes significations de la même forme verbale ont des schémas valenciels semblables. Cela veut dire que l'analyse devient plus compliquée et qu'une seule élaboration du schéma valenciel ne suffit pas à désambiguïser les verbes. Dans ces conditions, s'est imposée la nécessité d'une analyse et d'une description complémentaires, à savoir celles des arguments sélectionnés par le verbe. Dans son dictionnaire, Polański décrit les arguments – quoique de manière non systématique – à l'aide des indices sémantiques suivants :

[+Abstr] – abstrait	[Fl] – plante
[–Abstr] – concret	[Inf] – information
[+Anim] – animé	[Instit] – institution
[–Anim] – non-animé	[Instr] – instrument
[+Hum] – humain	[Liqu] – liquide
[–Hum] – non-humain	[Mach] – machine
[Coll] – collectif	[Mat] – matériel
[Elm] – force naturelle	[Pars] – partie

Les indices peuvent se combiner mutuellement. Par exemple [–Abstr/–Anim] définit la classe des objets concrets non-animés comme *pierre* ou *verre*, tandis que [+Hum/Pars] renvoie aux parties du corps humain (*main, tête, pied*). Il s'avère pourtant que la liste (ci-dessus) n'est pas suffisante pour exprimer précisément toutes les exigences syntaxiques des verbes[11]. C'est pourquoi, pour les besoins du projet, on a entrepris une démarche pour compléter les schémas valenciels, de même que les indices sémantiques proposés par Polański, consistant en l'ajout d'une autre information sur les rôles sémantiques.

Le *rôle sémantique* (notion introduite en linguistique par Charles J. Fillmore[12]) est la fonction sémantique du groupe nominal par rapport au

[11] Z. Vetulani, *Towards a Linguistically Motivated Ontology of Motion : Situation Based Synsets of Motion Verbs*, American Association for Artificial Intelligence (www.aaai.org) 2002.

[12] Ch. J. Fillmore, « The Case for Case », in *Universals in Linguistic Theory*, E. Bach, R. T. Harms (édit.), New York 1968.

prédicat verbal dans un schéma valenciel donné. Le même rôle sémantique peut avoir différentes réalisations dans la structure de la phrase, de même que la même position (dans la phrase) peut être occupée par des élements jouant différents rôles sémantiques. Il s'ensuit qu'une bonne (et complète) définition des arguments (à l'aide des rôles sémantiques justement) pourrait résoudre les problèmes de polysémie au niveau des verbes prédicatifs.

Ainsi, les rôles sémantiques deviennent un instrument utile dans la syntaxe à base sémantique : ils sont même « pertinents du point de vue du traitement automatique au niveau sémantique »[13]. En principe, l'ensemble des rôles employé dans la description des verbes en question est basé sur celui du projet de VerbNet – le plus large dictionnaire éléctronique de la langue anglaise (cf. Martha Palmer[14] ; les rôles sémantiques y sont appelés « thematic roles ») – et transformé – après quelques modifications – en un système opérationnel de rôles sémantiques suivants[15] :

Rôle	Définition	Exemples en polonais
Agent	Participe à l'action, effectue l'action de manière intentionnelle ou conformément aux lois de la nature ; initie ou contrôle le déroulement de l'action	Nom.+Verbe+Gen. « Pies bronił dzieci. » Nom. –> Agent [+Anim,–Hum]
Beneficiary (Benef)	N'effectue pas l'action mais il en tire profit ou perte ; ne participe pas à l'action	Nom.+Verbe+ Prép. NA+Acc. « Wszyscy głosowali na Jana Szeligę. » Prép. NA+Acc.–> Benef [+Hum]
Cause	Agent métaphorique ; le plus souvent l'objet, notion abstraite, mots de type « wszystko »	Nom.+Verbe+Compl.lieu abl[16]. « Droga szła w kierunku pól. » Nom. –> Cause [–Abstr,–Anim]
Experiencer (Exp)	Participant, conscient de quelque chose (état psychique, sensation)	Nom.+Verbe+Acc. « Jadły ich wszy i komary. » Acc. –> Exp [+Anim]
Patient	Participant de l'action, il éprouve l'action, l'état ou leurs effets	Nom.+Verbe+Acc. « Profesor leczył Piotra. » Acc. –> Patient [+Anim]
Recipient (Recip)	Destinataire dans l'action consistant à transmettre un objet, une propriété, une information ; aussi – appartenir à un groupe, constituer une partie de quelque chose	Nom.+Verbe+Prép. DO+Gen. « Napisz do mamy. Niech się nie martwi. » Prép. DO+Gen.–> Recip [+Hum]

[13] Z. Vetulani *et alli*, *Zasoby językowe i technologie przetwarzania tekstu...*, p. 183.

[14] M. Palmer, *Semlink : Linking PropBank, VerbNet and FrameNet, Proceedings of the Generative Lexicon Conference*, Sept. 2009, Pisa, Italy : GenLex-09.

[15] Ces modifications ont été effectuées par Z. Vetulani, T. Obrębski et B. Nadzieja.

[16] Les compléments de lieu peuvent recouvrir trois types de mouvement : *abl* (ablativ) – mouvement du côté d'une source, *adl* (adlativ) – mouvement vers une destination, *perl* (perlativ) – mouvement à travers un lieu.

Giver	Expéditeur dans l'action consistant à transmettre un objet, une propriété, une information	Nom.+Verbe+Prép. OD+Gen. « Dziewczynka chciała fortepian od ojca. » Prép. OD+Gen. –> Giver [+Hum]
Theme	Participant qui change sa position ou son état ou restant dans une position ou un état ; souvent un objet	Nom+Verbe+Acc+Compl.lieu adl « Brudną pościel dajesz do pralni ? » Acc –> Theme [–Abstr,–Anim]

Destination (Dest)	Objet, endroit ou temps vers lequel un mouvement, une action se déroulent	Nom+Verbe+Acc+Compl.lieu adl « Brudną pościel dajesz do pralni ? » Compl.lieu adl –> Dest [[endroit[17]]]
Location (Loc)	Endroit où l'action se déroule, où se trouve quelque chose	Nom.+Verbe+ŻE+Prép. W+Loc. « W swym przewodniku informuje, że warto zwiedzić to miasto. » Prép. W+Loc. –> Loc [Inf]
Source	Objet, endroit ou temps du côté duquel un mouvement, une action se déroulent ; aussi – cause, début de quelque chose	Nom.+Verbe+Gen.+Prép. Z+Gen. « Dziecku idzie krew z nosa. » Prép. Z+Gen. –> Source [+Anim,Pars]

Asset	Valeur (prix) dans les actions concernant le transfert de l'argent	Nom.+Verbe+Prép. ZA+Acc.+Nombre « Liczył 6 zł za kilogram truskawek. » Nombre –> Asset [[prix]]
Value	Valeur, aussi valeur mathématique ; tous les nombres et les expressions de mesure et leurs déterminants (*beaucoup, peu*)	Nom.+Verbe+Nombre+Acc. « Rodzina liczy pięć osób. » Nombre –> Value [[nombre, quantité]]

Instrument (Instr)	Objet, force utilisés comme instrument ; aussi la partie du corps qui effectue l'action	Nom.+Verbe+Acc.+Instr. « Mamili ludzi obietnicami. » Instr. –> Instr [–Abstr,–Anim] [+Abstr]
Material	Matériel, matière transformés en produit, utilisés lors de la production du produit ; aussi la nourriture	Nom.+Verbe+Acc.+Prép. Z+Gen. « Z wody, mąki i jajek otrzymamy jednolitą masę. » Prép. Z+Gen. –> Material [–Abstr,–Anim]
Predicate	Expression décrivant les particularités ou les caractéristiques des participants ou les relations entre eux	Nom.+Verbe+Acc.+Instr. « Ogród ten można nazwać parkiem. » Instr. –> Predicate [nom] [appellation]
Product	Produit, résultat de l'action	Nom.+Verbe+Acc. « Otwieram pracownię krawiecką. » Acc. –> Product [Instit]
Manner	Manière dont l'action est effectuée	Nom.+Verbe+CC manière « Wszyscy głosowali jawnie. » CC manière –> Manner [[manière, façon]]
Goal	But à atteindre	Nom.+Verbe+Prép. DO+Gen. « Jak zawsze szedł wprost do celu. » Prép. DO+Gen. –> Goal [+Abstr]
Action	Action, opération, acte	Nom.+Verbe+Acc.+Instr.+Prép. Z+Gen. « Psy ujadaniem obudziły podróżnego z lekkiego snu. » Prép. Z+Gen.–> Action [sommeil]
State	Etat dans lequel se trouvent les participants ; aussi les circonstances de l'action	Nom.+Verbe+CC « Piotr pada w pojedynku. » CC –> State [[circonstances]]

[17] La double parenthèse signifie que l'indice n'a pas été proposée par Polański dans SSGCP.

Time	Temps de l'action	Nom.+Verbe+Compl.temps « W 1453 roku pada Konstantynopol. » Compl.temps –> Time [[temps]]
Topic	Sujet dans les situations de communication	Nom.+Verbe+Prép. O+Loc. « Adaś pisze o swoim sukcesie. » Prép. O+Loc. –> Topic [[+Abstr]]

Les vingt-trois rôles paraissent être suffisants pour décrire la syntaxe et la valence d'un verbe quelconque. Pour illustrer les profits qui en découlent, nous proposons ci-dessous l'analyse du verbe *złapać* en polonais (fr. *attraper*) :

ZŁAPAĆ (ATTRAPER)

1. signification – « arriver à prendre, à saisir par la main ou par un outil »

Nom.+Verbe+Acc. « *Janek zręcznie złapał spadającą ze stołu książkę.* »

Nom.+Verbe+Acc.+Instr. « *Złapała wiadro obiema rękami, chlusnęła wodą na podłogę.* »

Nom.+Verbe+Acc.+Prép. ZA « *Złapał Wojtka za kark i jak szczeniaka wyrzucił na ulicę.* »

Nom.+Verbe+Acc.+Prép. ZA+Instr. « *Złapał Wojtka jedną ręką za kark i jak szczeniaka wyrzucił na ulicę.* »

2. signification – « saisir un homme ou un animal, le priver de la possibilité de bouger, l'arrêter »

Nom.+Verbe+Acc. « *Złapali Staszka i zaprowadzili go na komisariat.* »

Nom.+Verbe+Acc.+CC manière « *Złapał na wędkę sporego szczupaka.* »

3. signification – « métaph. en parlant d'un sentiment ou d'une sensation physique – apparaître soudainement »

Nom.+Verbe+Acc. « *Złapał go krwotok.* »

4. signification – « métaph. en parlant d'une force naturelle – apparaître soudainement, surprendre quelqu'un »

Nom.+Verbe+Acc. « *Złapała nas burza, zmokliśmy do suchej nitki.* »

Nom.+Verbe+Acc.+Compl.lieu « *Żeby tylko deszcz ich nie złapał w polu.* »

5. signification – « cour. tomber malade d'une maladie contagieuse »

Nom.+Verbe+Acc. « *Jacek złapał katar.* »

Puisque toutes les significations du verbe « *złapać »* actualisent le schéma bivalent dans lequel le premier argument est au Nominatif et le second à l'Accusatif, on n'arrive pas (dans l'analyse automatique) à désambiguïser la forme en question. Le recours aux rôles sémantiques peut venir ici en aide :

Signification	**1.**	**2.**	**3.**	**4.**	**5.**
Nominatif	Agent [+Hum]	Agent [+Anim]	Cause [+Abstr]	Agent [Elm][cataclysme]	Experiencer [+Hum]
Accusatif	Theme [–Abstr,–Anim][–Abstr]	Patient [+Anim]	Experiencer [+Hum]	Experiencer [+Hum]	Product [+Abstr]
Rôles accessoires	Instr –> Instr [Instr] Prép. ZA+Acc. –> Instr [–Abstr,–Anim][+Anim,Pars]	CC manière–> Manner [[manière]]		Compl.lieu –> Loc [[lieu]]	

Le tableau démontre que la même position dans la phrase peut être occupée par les arguments jouant différents rôles sémantiques conformément au sens du verbe (actualisé par ces arguments). Si l'on revient à l'intitulé du présent article, à savoir : « *Jean a attrapé la balle et la grippe »*, l'on voit que le même sujet *Jean* devient soit *Agent*, soit *Experiencer*, tandis que les deux arguments : *ballon* et *grippe* sont respectivement *Theme* et *Product*.

Il serait bien évidemment trop optimiste de considérer que cette analyse puisse recouvrir et expliquer toutes les nuances sémantiques et les ambiguïtés des verbes polonais. Ni l'éventail des rôles sémantiques, ni la description elle-même ne se veulent guère être exhaustifs. Néanmoins, grâce au fait que l'analyse est basée sur un échantillon important d'unité lexicales (plus de 2000 verbes), elle peut constituer un premier pas vers la construction d'un bon et efficace moyen de la désambiguïsation des verbes polonais.

Keywords: verb, arguments, thematic roles

"JOHN CAUGHT THE BALL" HOW TO CLARIFY THE MEANING OF WORDS: A STUDY OF POLISH VERBS

Summary

The aim of the present article is to present a solution to the problem of lexical ambiguity. The ambiguity of lexical items caused by homonymy or homophony does not create any problems for the addressee of messages due to the fact that they are placed within a particular communicational and situational context. Such a context, however, is absent in machine translation (MT). In order to avoid ambiguity of Polish verbs listed for the purposes of the project "Communication Safety", the specification of semantic roles describing the arguments of individual verbs has been drawn up. Their ambiguity was eliminated after determining the semantic roles of their arguments.

Translated by Sylwester Jaworski

"JEAN ZŁAPAŁ PIŁKĘ", CZYLI JAK DOPRECYZOWAĆ ZNACZENIE WYRAZÓW, NA PRZYKŁADZIE CZASOWNIKA POLSKIEGO

Streszczenie

Celem niniejszego artykułu jest przedstawienie zagadnienia wieloznaczności leksykalnej. Przyczyną wieloznaczności jednostek leksykalnych jest m.in. występowanie homonimii oraz homofonii. Wieloznaczność jednostek leksykalnych nie stwarza problemów odbiorcy wiadomości dzięki umiejscowieniu ich w danym kontekście komunikacyjnym i sytuacyjnym, jednakże tego rodzaju kontekstu brakuje w tłumaczeniu maszynowym. Aby uniknąć wieloznaczności polskich czasowników użytych na potrzeby projektu „Bezpieczeństwo komunikacyjne", sporządzono zestaw ról semantycznych opisujących argumenty pojedynczych czasowników. Dokładne ustalenie ról semantycznych argumentów ponad dwóch tysięcy polskich czasowników pozwoliło wykluczyć ich wieloznaczność.