

ZYGMENT VETULANI, GRAŻYNA VETULANI

Université Adam Mickiewicz, Poznań

vetulani@amu.edu.pl, gravet@amu.edu.pl

SYNONYMIE ET GRANULARITÉ DANS LES BASES LEXICALES DU TYPE WORDNET

Abstract. Zygmunt Vetulani, Grażyna Vetulani, *Synonymie et granularité dans les bases lexicales du type WordNet* [Synonymy and granularity in the WordNet-like lexical data bases], *Studia Romanica Posnaniensia*, Adam Mickiewicz University Press, Poznań, vol. XLII/1: 2015, pp. 113-127. ISBN 978-83-232-2854-7. ISSN 0137-2475. eISSN 2084-4158. DOI: 10.14746/strop.2015.421.008

In this paper we present a contribution to the transformation of PolNet, a Polish wordnet developed at the Adam Mickiewicz University in Poznań, into a Lexicon Grammar of Polish. The current step consists in including verb-noun collocations and relations linking the verbal synsets to noun synsets. We focus on the concept of synonymy for two kinds of predicative entities: verbs and verb-noun collocations and on synset granularity problems that emerged at this stage of the project. This work was sponsored by the Polish National Program for Humanities (grant 0022/FNiTP/H11/80/2011).

Keywords: wordnet, synonymy, granularity, valency structure, lexicon grammar.

1. INTRODUCTION

Grâce à la recherche étymologique nous connaissons les lois de l'évolution sémantique des mots et nous savons que leur sens change avec l'évolution du monde réel (changement de référent ou de connaissance sur le référent) ou selon les circonstances d'emploi (contexte, emprunts à l'intérieur du système, figures rhétoriques, etc.).

L'évolution sociale, culturelle et technologique fait que de nouveaux concepts font apparaître de nouveaux termes, mais les circonstances exactes de ces apparitions restent souvent inconnues, à l'exception de certaines créations *ad hoc* comme ce fut le cas du mot *robot*, inventé en 1920 par l'écrivain tchèque Joseph Čapek pour son frère Karel, pour désigner une espèce d'« homme artificiel », héros de la fameuse pièce de théâtre *R.U.R.* (« Rossum's Universal Robots »). La signification initiale du mot s'est effacée, plus d'un demi-siècle plus tard, au profit de celle utilisée actuellement.

Si le phénomène d'apparition de mots nouveaux est évident, le processus de formation de sens nouveaux pour les mots fonctionnant déjà dans le système reste souvent obscur. Sans aucun doute, les transformations sémantiques se déroulent de manière autonome dans chaque langue, bien que, parfois, elles puissent aussi interférer. Dans tous les cas, elles conduisent au phénomène de la *polysémie* (identifiée

par certains linguistes avec l'*homonymie*). Le polonais en fournit un exemple intéressant avec le mot *zamek* qui correspond en français aux signifiés de : 'château', 'serrure', 'fermeture éclair', 'culasse' ou 'trappe'¹. Remarquons que toutes ces significations se rattachent au sens général de 'fermer', mais en français, l'idée générale de fermeture ne se manifeste pas dans une forme linguistique (un signifiant).

Beaucoup de chercheurs (dont Polański, 1995 : 408) opposent la *polysémie* où les différentes significations d'un mot ont une source étymologique commune, à l'*homonymie* où il est difficile de trouver un tel lien. Par exemple, est homonymique en polonais le mot *bal* signifiant 'soirée dansante' (un calque du français) ou 'tronc d'arbre'. En français, on peut illustrer le même phénomène par *élan* qui signifie une sorte de mouvement (pol. *rozpęd*) ou un animal (pol. *łoś*).

Si du point de vue de la recherche diachronique l'opposition polysémie-homonymie est importante, il n'en n'est pas de même quand il s'agit de la création de modèles de langues utiles pour les applications technologiques où l'approche synchronique prévaut. Dans ce deuxième cas, l'opposition entre les termes de *polysémie* et *homonymie* peut être ignorée.

Un autre phénomène commun pour les langues, et essentiel à la compréhension et au traitement automatique du langage, est la *synonymie*. Par *synonymie* nous entendons l'expression du même contenu par deux (ou plusieurs) formes linguistiques différentes (Polański, 1995 : 533)². De même que la polysémie, la synonymie résulte de l'évolution linguistique et des phénomènes d'interférence (l'évolution et l'interférence se croisent souvent). Les emprunts lexicaux ajoutent des significations nouvelles ou apparentées à celles existant déjà dans le système. On observe alors soit l'éloignement soit le rapprochement sémantique ou pragmatique entre les formes. Par exemple, l'usage a abouti au rapprochement des termes *pomme de terre* et *patate* en français et, de manière similaire, *ziemniak* et *kartofel* en polonais. Pour l'anglais, la conquête normande de l'Angleterre a eu pour conséquence la création des sociolectes dérivés du vieux français normand d'une part et du vieil anglo-saxon (*Old English*) de l'autre. Les paires comme *people-folk*, *liberty-freedom* sont devenus synonymes, identiques sur le plan sémantique et pragmatique.

2. TRAVAUX ANTÉRIEURS

Notre recherche sur la base lexicale PolNet (Polish WordNet)³, qui constitue le cadre des recherches présentées dans cet article, s'inspire des travaux de G.A. Miller

¹ Stratégie défensive utilisée en hockey sur glace.

² Nous allons nous limiter à la *synonymie* des mots bien que l'idée-même soit parfois appliquée à d'autres formes (Polański, *ibidem*). Plus précisément, nous allons appliquer la synonymie aux mots désambiguïsés (pol. *slowosens*, ang. *word-meaning pair*).

³ Travaux initiés à l'Université Adam Mickiewicz de Poznań au début des années 2000 sous la direction de Z. Vetulani.

(1990) sur Princeton WordNet (PWN), ainsi que de ceux de P. Vossen dans le cadre du projet EuroWordNet. Lors du développement de PolNet, nous nous sommes aussi basés sur les travaux de M. Gross (1994) liés au lexique-grammaire, sur ceux de K. Polański (1980-1992 ; *Syntaktyczno-generatywny słownik czasowników polskich*), ainsi que sur les projets FrameNet (Fillmore *et al.*, 2002) et VerbNet (Palmer, 2009), tous les deux proches du concept de *lexique-grammaire*.

3. SYNONYMIE

L'importance de la synonymie est déjà attestée au VII^e siècle avant notre ère comme en témoignent les fragments en écriture cunéiforme d'un dictionnaire assyrien de synonymes⁴. Parmi les linguistes il n'y a pas de consensus concernant la nature de la synonymie, mis à part le consensus sur l'idée générale que la signification des mots synonymiques est proche ou identique. Le manque de critères clairs et effectifs pour identifier ce phénomène est un problème majeur, en particulier pour la création de bases de données sémantico-lexicales.

Pour reconnaître la synonymie, le plus souvent, on propose comme test le remplacement d'un mot par un autre pour vérifier si, après une telle opération, le sens de l'énoncé reste le même. Gottfried Leibnitz (d'après Vossen, 2002) a défini un critère très rigoureux :

Two expressions are synonyms if the substitution of one for the other never change the truth value of a sentence in which the substitution is made⁵.

Cette définition accorde à la synonymie des propriétés algébriques très utiles (la synonymie devient une relation *d'équivalence*, c'est-à-dire elle est *réflexive*, *symétrique* et *transitive*). En même temps, il s'avère que la synonymie (au sens de Leibnitz) est rarement observée dans la langue (les exemples se trouvent le plus souvent parmi des mots importés des sociolectes ou d'autres langues).

Étant donné que la pratique quotidienne du langage n'impose pas de contraintes aussi dures que celles proposées par Leibnitz, dans l'objectif de créer des modèles linguistiques à l'usage informatique visant la simulation de la compétence linguistique humaine, G.A. Miller et Ch. Fellbaum (Miller *et al.*, 1990) ont proposé un critère plus faible :

Two expressions are synonyms in a linguistic context *C* if the substitution of one for the other in *C* does not alter the truth value⁶.

⁴ Article « Synonym ». Dans : *Wikipedia*, <http://en.wikipedia.org/wiki/Synonym>, consulté le 20.12.2014.

⁵ Deux expressions sont synonymiques si la substitution de l'une par l'autre ne change jamais la valeur de vérité (*truth value*) de la phrase à laquelle l'opération de substitution a été appliquée.

⁶ Deux expressions sont synonymiques dans un contexte linguistique *C* si la substitution de l'une par l'autre dans le contexte *C* ne change pas la valeur de vérité (*truth value*).

Cela signifie qu'un seul contexte *C* suffit pour identifier la synonymie (Vossen, 2002 : 18). Cette définition laisse un champ d'interprétation ouvert en fonction du sens attribué au terme « contexte linguistique *C* »⁷.

À son tour, dans le projet EuroWordNet, Vossen appelle *équivalence sémantique* sa version de synonymie⁸. Deux éléments sémantiquement équivalents doivent avoir la même extension (« synonyms denote the same range of entities »), mais en même temps, ils ne doivent pas être nécessairement identiques en ce qui concerne le registre, le dialecte, le style ou la morphosyntaxe.

Vossen propose également des tests pour vérifier l'extension des concepts auxquels correspondent les mots examinés. Ce sont des schémas de phrases dans lesquelles les variables sont remplacées par des mots-candidats. Si les mots réussissent le test cela signifie leur substitutivité sémantique (extensionnelle), mais pas nécessairement leur équivalence pragmatique. Les deux tests qui suivent⁹ ont été publiés dans le rapport final de l'EuroWordNet (Vossen, 2002)¹⁰.

Test 1. Synonymie des noms communs

On substitue les noms examinés (désambiguïsés) à X et à Y dans (a) et (b) :

(a) Si qch/qq est X alors qch/qq est aussi Y.

(b) Si qch/qq est Y alors qch/qq est aussi X.

Le test de synonymie est réussi si (a) et (b) sont vrais.

Conditions : X et Y sont des noms au singulier ou au pluriel¹¹.

Exemple :

(a) Si c'est une pomme de terre alors c'est aussi une patate.

(b) Si c'est une patate alors c'est aussi une pomme de terre.

Résultat : *patate* et *pomme de terre* sont des synonymes.

Test 2. Synonymie des verbes

On substitue les verbes examinés (desambiguïsés) à X et à Y dans (a) et (b) :

(a) Si qch/qq X alors qch/qq Y.

(b) Si qch/qq Y alors qch/qq X.

Le test de synonymie est réussi si (a) et (b) sont vrais.

Conditions : X et Y sont des verbes à la 3^e personne du singulier, au présent.

⁷ Pour éviter la confusion, certains auteurs (dont Miller et Fellbaum, 1990) préfèrent parler de « ressemblance sémantique » (*semantic similarity*) au lieu de synonymie (voir aussi Vossen, 2002 : 18). Il est possible d'envisager des contraintes sur le contexte *C* qui impliqueraient la ressemblance pragmatique des synonymes (registre, etc.).

⁸ Appliquée dans l'Euro WordNet.

⁹ Dans le cadre du projet EuroWordNet des tests semblables pour d'autres relations sémantiques ont été également proposés.

¹⁰ En ce qui concerne le critère (dérivé) servant à conclure que les mots ne sont pas sémantiquement équivalents, Vossen propose de trouver une relation sémantique (autre que l'identité) qui les relie (p. ex. *person* et *police force* sont liés par la relation : membre–groupe).

¹¹ Pour le polonais, il faut respecter en outre le cas grammatical (ici, l'instrumental).

Exemple :

(a) Si qq marche alors qq se déplace à pied.

(b) Si qq se déplace à pied alors qq marche.

Résultat : *marcher* et *se déplacer à pied* sont des synonymes.

Il existe encore des réalisations de l'idée de synonymie appliquées dans les systèmes ontologiques autres que celles présentées ci-dessus (voir à ce propos p. ex. Z. Vetulani (2004) sur la synonymie des verbes de mouvement où l'auteur a proposé de considérer la similarité entre les situations auxquelles se rapportent les verbes – les situations étant représentées par les structures formelles du type attribut-valeur).

4. PRINCETON WORDNET

Dans les années 80, G.A. Miller et ses collaborateurs du Laboratoire des Sciences Cognitives de l'Université de Princeton ont lancé une nouvelle méthode de description sémantique du lexique qui s'est avérée utile surtout dans la recherche d'information sur Internet. L'idée centrale de cette méthode était de présenter le lexique sous forme d'une base de données avec l'utilisation de relations de synonymie et d'hyponymie. La base, nommée WordNet (PWN)¹², est composée de classes de synonymes appelées *synsets* qui forment une hiérarchie régie par la relation de l'hyponymie (et la relation contraire de l'hyponymie)¹³. D'autres relations sémantiques (méronymie, antonymie, etc.) y sont représentées aussi. Le point faible de l'approche de Miller est le critère de synonymie (ci-dessus) qui – seul – ne suffit pas pour former les *synsets* car il ne garantit pas la transitivité quand le contexte *C* change. Pour remédier à ce défaut il faut renforcer le critère initial de la synonymie en imposant la référence à un même contexte *C*.

En réalité, pour bien déterminer les éléments d'un *synset* il est également utile de compléter sa description par une définition (le plus souvent informelle) rédigée sous la forme de glose.

5. WORDNET EN TANT QU'ONTOLOGIE – EXIGENCES CONCERNANT LA SYNONYMIE

Le succès du système PWN est dû en partie au fait qu'il peut être considéré comme une ontologie car tous les éléments d'un *synset* représentent le même con-

¹² On dit souvent Princeton WordNet (PWN) pour bien distinguer le système initial par rapport à une multitude de systèmes similaires qui fonctionnent pour un grand nombre de langues particulières.

¹³ Plus précisément, le champ de la relation d'hyponymie est constitué de tous les *synsets* et la relation même est induite par la relation linguistique (qui porte le même nom d'*hyponymie*) entre les sens des mots. Il en est de même pour l'hyponymie.

cept et la relation hyperonymie/hyponymie correspond à celle de généralisation/spécialisation des concepts. Les systèmes du type WordNet encodent de manière directe la relation qui lie les concepts aux mots qui les représentent dans la langue. Ce trait est très utile pour les applications informatiques dans le domaine de l'intelligence artificielle impliquant le traitement du langage naturel. Il est conforme à la fameuse formule prononcée par Gruber (1933) : “an ontology is a formal, explicit specification of a shared conceptualization” [une ontologie est une spécification formelle et explicite de la conceptualisation partagée (par l'ensemble des utilisateurs de la langue)]¹⁴.

6. SYNONYMIE DANS LE SYSTÈME POLNET

6.1. POLNET

Le projet du WordNet polonais « PolNet – Polish WordNet » a été lancé à l'Université Adam Mickiewicz de Poznań par Z. Vetulani en 2006. Il a été directement inspiré par le Princeton WordNet et constitue une réponse aux besoins d'une ontologie générale correspondant à la conceptualisation naturelle et commune de l'ensemble des utilisateurs de la langue polonaise. Ce type d'ontologie est utile, et parfois indispensable, pour la création de modèles formels de la compétence linguistique et par la suite, pour l'implémentation de systèmes d'intelligence artificielle possédant cette compétence.

Les systèmes ressemblant au WordNet ont un avantage sur les ontologies traditionnelles car ils rendent compte de manière explicite des relations entre les mots de la langue et les concepts de l'ontologie. De plus, ces relations sont présentées de manière très simple (appartenance du mot à la classe des mots qui représentent le concept en question). Pour assurer la compatibilité entre la conceptualisation partagée par l'ensemble des utilisateurs de la langue polonaise et l'ontologie PolNet, on a décidé de construire le système selon le modèle « merge », c'est-à-dire indépendamment de tout autre wordnet (en particulier du PrincetonWordNet). Il existe des ontologies qui sont développées selon la méthode « expand » consistant à traduire PWN, qui est plus rapide et moins coûteuse, mais, en même temps, elle risque de recopier la conceptualisation socio-culturelle propre à l'anglais (implicite dans PWN).

¹⁴ À noter qu'une relation identique existe entre les objets du langage formel de logique et le modèle algébrique construit dans la démonstration du théorème de Gödel qui dit que *tout système logique non-contradictoire possède un modèle*, c'est-à-dire un univers dans lequel tous les théorèmes de ce système sont vrais. Henkin a montré comment construire un tel modèle à partir de « briques » qui sont des classes de synonymes du système logique en question. Le fait que la synonymie des termes (du langage formel en considération) soit une relation d'équivalence (c'est-à-dire réflexive, symétrique et transitive) constitue la base du raisonnement de Henkin. La construction elle-même ressemble à celle des wordnets. Pour que l'analogie soit complète il faut donc que la synonymie, qui est à la base de formation des synsets, soit, elle aussi, une relation d'équivalence (Z. Vetulani, 2012).

L'algorithme utilisé pour la création du PolNet (pour les substantifs) a été publié en 2007 (Z. Vetulani *et al.*, 2007). Il a été appliqué dans l'environnement VisDic, et ensuite, dans sa version multi-utilisateurs DebVisDic (Pala *et al.*, 2007 ; DebVisDic a été conçu pour développer et maintenir les systèmes du type WordNet créés par le groupe de Karel Pala de l'Université Masaryk de Brno en République Tchèque). La première distribution du PolNet (version 1.0), sous la forme d'une licence CC, a eu lieu lors du colloque LTC (décembre 2011) et peu après, au Global WordNet Conference (janvier 2012), ainsi que sur la page web www.ltc.amu.edu.pl. Actuellement, elle est accessible aussi en utilisant Meta-Share.

La partie essentielle de la première phase du projet PolNet (jusqu'à 2010) concernait les synsets des noms communs, ainsi que les relations entre les synsets induites par les relations sémantiques existant entre les mots (pol. *slowosens*). La sélection initiale du matériel lexical (en premier temps, un peu plus de 10 000 mots) ciblait le lexique général attesté dans une des versions initiales du corpus IPI PAN (Przepiórkowski, 2004) du polonais qui a servi de base pour extraire les listes de fréquence des mots¹⁵.

Le problème essentiel consistant à distinguer les significations différentes d'un mot a pu être résolu, dans la plupart des cas, grâce aux dictionnaires accessibles en ligne (*Uniwersalny słownik języka polskiego – USJP*) et *Słownik języka polskiego* (SJP, M. Szymczak). L'algorithme de création et de description des synsets (Z. Vetulani *et al.*, 2007) utilisait les tests de synonymie (voir ci-dessus). Étant donné que les tests sont de nature extensionnelle et ignorent les différences entre les variantes pragmatiques (parmi lesquelles la substitutivité ne se pratique jamais ou presque), pour atteindre le niveau de granulation respectant le style, le registre, etc., on a eu recours aux définitions informelles sous la forme de gloses (dans la plupart des cas, définitions de dictionnaires adaptées à nos besoins).

L'organisation classique des wordnets des noms est basée sur la hiérarchie induite par la relation hyperonymie/hyponymie. Au sommet de cette hiérarchie se trouvent les notions les plus générales, au bas de la hiérarchie – les notions plus particulières. Cette structuration a pour but de rendre possible l'héritage de propriétés – mécanisme souvent utilisé dans la représentation des connaissances et dans l'inférence logique. L'idée est analogue à la méthode de classification utilisée par Linné dans *Systema Naturae* [Les systèmes de la Nature] au XVII^e siècle.

¹⁵ Ayant en vue à ce stade l'extension du système à d'autres parties du discours (surtout les verbes), nous avons compilé une liste de 761 noms communs qui ont servi à Polański (1990) de descripteurs sémantiques pour caractériser les exigences sémantico-syntaxiques des prédicats verbaux (Z. Vetulani, 2003). Nous avons également pris en compte le vocabulaire propre au domaine de la sécurité publique (sécurité des stades de football), ce qui a été motivé par les besoins de l'application informatique (POLINT-112-SMS) choisie pour tester la validité du PolNet dans le cadre du développement des systèmes IA avec compétence linguistique (Z. Vetulani *et al.*, 2010).

L'extension de PolNet aux concepts prédicatifs conduit à l'introduction dans le système des relations entre les synsets verbaux et les synsets nominaux. Parmi ces relations il y a celles qui associent les concepts (représentés par les synsets) aux positions argumentales ouvertes par le prédicat. Ces relations sont appelées *rôles sémantiques*. Étant donnée la nature extensionnelle des relations hiérarchiques d'hyponymie/hyperonymie, les tests proposés pour les noms (Test 1) suffisent en principe pour constituer les synsets. Pour les verbes et les autres mots prédicatifs nous admettons que, à part le Test 2 (ci-dessus), il est nécessaire de prendre en considération les valeurs des rôles sémantiques concernés.

6.2. STRUCTURE DE VALENCE

Le fait de relier les *synsets* verbaux aux synsets nominaux par des relations afin de mettre en évidence les conditions de connectivité sémantique et syntaxique des arguments sélectionnés par les prédicats de nature grammaticale différente (verbes ou noms) est un trait typique du *lexique-grammaire* (Z.Vetulani, G.Vetulani, 2013). Cette approche impose à la synonymie des prédicats de fortes contraintes de granularité¹⁶ par la prise en compte des *structures de valence*.

Par *structure de valence* nous entendons une structure de données contenant l'information sur toutes les contraintes morpho-syntaxiques et sémantiques des éléments qui occupent les positions d'argument (nombre, genre, cas grammatical, etc.).

Dans le PolNet, nous exigeons que les synonymes aient la même structure de valence (y compris les mêmes valeurs des rôles sémantiques associées aux arguments). Ainsi, la structure de valence constitue un indice de la signification.

Formellement, les rôles sémantiques sont des fonctions (au sens mathématique, c'est-à-dire qu'elles attribuent une valeur à un (ou plusieurs) argument(s) dans la matrice syntaxique attribuée au synset. Les fonctions prennent leurs valeurs dans une ontologie formée de synsets (ontologie PolNet). Elles indiquent les rôles joués par les acteurs dans la situation à laquelle se rapporte la phrase (Agent, Patient, Bénéficiaire, ...). Le fait de garder la structure de valence avec les synsets (comme une partie de la description) a des avantages du point de vue des applications informatiques (traitement automatique du langage, parsing, communication homme-machine, etc.).

7. GRANULARITÉ

La granularité des synsets verbaux dépend directement de la définition de la synonymie des verbes. La granularité fine correspond à une différenciation minutieuse entre les significations proches, ce qui résulte en un grand nombre de synsets. Par la suite, le système devient utile pour la représentation des connaissances.

¹⁶ Par *granularité* nous entendons la taille relative d'un synset (voir la section suivante).

Avec l'apparition dans PolNet de synsets prédicatifs verbaux la question concernant le type d'information distinctive pour les synsets a pris une importance majeure. Le critère d'identité de structure de valence pour les synonymes verbaux fait de la valence une marque distinctive pour un synset.

Tableau 1. Représentation en DEBVisDic (simplifiée) d'un synset verbal (pol. *pomóc*, fr. *aider*) et de son synonyme en forme de collocation verbo-nominale (pol. *udzielić pomocy*, fr. *apporter de l'aide*) :

POS : v ID: 3441

Synonyms: {pomóc:1, pomagać:1, udzielić pomocy:1, udzielać pomocy:1} ('aider')

Definition: « wziąć (brać) udział w pracy jakiejś osoby (zwykle razem z nią), aby ułatwić jej tę pracę » ('participer au travail de qq pour lui apporter un soutien')

Frame : Agent(N)_Benef(D)

Frame : Agent(N)_Benef(D) Action(pre(w)+NA(L))

Frame : Agent(N)_Benef(D) Manner

Frame : Agent(N)_Benef(D) Action(pre(w)+NA(L)) Manner

Usage : Agent(N)_Benef(D) /*Pomagam jej (Je l'aide)*/

Usage : Agent(N)_Benef(D) Action('w'+NA(L)) ; *Pomagam jej w odrabianiu lekcji. /Je l'aide pour faire ses devoirs/*

Usage : Agent(N)_Benef(D) Manner; *Chętnie udzielam jej pomocy / Je l'aide volontiers/*

Usage : Agent(N)_Benef(D) Manner Action('w'+NA(L)); *Chętnie udzielam jej pomocy w lekcjach. /Je l'aide volontiers pour faire ses devoirs/*

Semantic_role : [Agent] {człowiek:1, istota ludzka:1, ... } /({homme, être humain,...})/

Semantic_role : [Benef] {człowiek:1, istota ludzka:1, ... } /({homme, être humain,...})/

Semantic_role : [Action] {czynność:1} %({activité})

Semantic_role : [Manner] {ADVERB_FEATURE_QUALITY} /{un concept de l'ontologie générale/

Dans l'exemple ci-dessus, les lignes contenant le schéma prédicat–argument (marquées « frame ») sont compatibles entre elles. Nous parlons de la compatibilité des schémas prédicat–arguments dans la structure de valence quand pour deux lignes « frame » quelconques l'une est sous-ligne de l'autre, ou bien il existe une troisième ligne dont les deux sont des sous-lignes ; p. ex. :

'Agent(N)_Benef(D) Manner'

et

'Agent(N)_Benef(D) Action('w'+NA(L))'

sont des sous-schémas pour

'Agent(N)_Benef(D) Action('w'+NA(L)) Manner'.

Voici les phrases qui en sont une illustration (dans l'ordre) :

*Robert chętnie udziela pomocy Barbarze (...)/Robert aide volontiers Barbara (...)/,
Robert (...) udziela pomocy Barbarze w lekcjach /Robert aide (...) Barbara pour faire ses devoirs/,
Robert chętnie udziela pomocy Barbarze w lekcjach /Robert aide volontiers Barbara pour faire ses devoirs/.*

Dans ce qui suit, nous allons prendre en considération uniquement les structures de valence dans lesquelles tous les schémas prédicat–arguments sont compatibles. De telles structures seront appelées *simples*. Les tableaux qui suivent présentent des fragments du code (simplifiés) avec la structure de valence pour trois sens choisis du verbe *szanować* (fr. *respecter*). C'est un exemple de cas où le verbe polysémique peut apparaître dans des structures variées, incompatibles entre elles : les sens sont différenciables grâce aux structures de valence. Les différences vont se manifester dans les schémas prédicat–arguments et dans les valeurs des rôles sémantiques. On observe ce phénomène lorsqu'on compare la valeur du rôle Agent pour *szanować* des tableaux 2 et 3. D'autre part, la différence dans la description des synsets présentés dans les tableaux 3 et 4 se manifeste au niveau de la valence, en particulier par les descriptions informelles présentées dans les lignes marquées Définition. Des différences de ce genre, exprimées sous forme descriptive, peuvent parfois être partiellement exprimées par la relation d'hyperonymie. Par exemple, un des signifiés du mot *szanować* ('*szanować:3*') a pour hyperonyme un des signifiés du mot *dbać*, alors qu'aucun des sens du mot *dbać* n'est hyperonyme pour *szanować:2*.

Tableau 2. Fragment de la description du mot *szanować:1*

Synonyms: {szanować:1, poważać:1}
 Frame : Agent(N)_Benef(Acc)
 Frame : Agent(N)_Benef(Acc) Cause(za + Acc)
 Definition : Stosunek jednej osoby do drugiej, polegający na uznaniu jej wysokiego statusu
 Semantic_role : [Agent]{człowiek:1, istota ludzka:1, ...}
 Semantic_role : [Benef]{człowiek:1, istota ludzka:1, ...}
 Semantic_role : [Cause]{motyw:1, pobudka:3, powód:1, przyczyna:1, racja:3, tytuł:6, wzgląd:2, źródło:3}
 Usage : Agent(N)_Benef(Acc); *Adam szanuje swoich rodziców.*
 Usage : Agent(N)_Benef(Acc) Cause('za'+ Acc); *Adam szanuje Sonię za mądrość.*

Tableau 3. Fragment de la description de *szanować:2*

Synonyms : {szanować:2}
 Frame : Agent(N)_Object(Acc)
 Frame : Agent(N)_Object(Acc) Cause('za'+Acc)
 Definition : Stosunek osoby lub instytucji do zjawiska lub zdarzenia, połączony z uznaniem znaczenia bądź wagi tego zjawiska lub zdarzenia
 Semantic_role : [Agent]{człowiek:1, istota ludzka:1, ...}
 Semantic_role : [Agent]{instytucja:1, placówka:2, ośrodek:1}
 Semantic_role : [Object]{koncepcja:1, pomysł:1, myśl:2, rozwiązanie:5, koncept:1, patent:4, wymysł:2}
 Semantic_role : [Object]{zdarzenie:1, wydarzenie:1, historia:4, }
 Semantic_role : [[Object]{zwyczaj:1, praktyka:4, obyczaj:1, tradycja:3}
 Semantic_role : [Cause]{ motyw:1, pobudka:3, powód:1, przyczyna:1, racja:3, tytuł:6, wzgląd:2, źródło:3}
 Usage : Agent(N)_Object(Acc) ; *Unia Europejska szanuje postanowienia traktatu w Schengen.*

Usage : Agent(N)_Object(Acc) ; *Gość szanuje zwyczaje panujące w danym domu.*

Usage : Agent(N)_Object(Acc) Cause('za'+Acc) ; *Gość szanuje zwyczaje panujące w danym domu za ich piękno.*

Tableau 4. Fragment de la description de *szanować*:3

Synonyms: {szanować:3}

Frame : Agent(N)_Object(Acc)

Definition : Dbać, ochraniać przed zniszczeniem, uszczerbkiem

Semantic_role : [Agent]{człowiek:1, istota ludzka:1,...}

Semantic_role : [Agent]{instytucja:1, placówka:2, ośrodek:1}

Semantic_role : [Agent]{kraina:5, kraj:1, państwo:4, ziemia:8}

Semantic_role : [Object]{1stOrderEntity:1}

Usage: Agent(N)_Object(Acc) ; *Adam szanuje swój samochód.*

Usage: Agent(N)_Object(Acc) ; *Każdy kraj szanuje swoje zabytki.*

La granularité établie au niveau correspondant aux différences exprimables à l'aide d'une définition est utile pour le traitement automatique, tant dans l'analyse que dans la génération du texte. Malheureusement, l'utilisation stricte de ce principe est difficile car les exigences morpho-syntaxiques des mots prédicatifs ne sont pas nécessairement invariantes par rapport à la relation de synonymie traditionnelle. Pour le polonais, nous observons ce phénomène au moment de la prise en compte, à côté des prédicats simples (verbes), des prédicats composés (collocations verbo-nominales).

L'existence de collocations verbo-nominales, parfois synonymiques des formes simples, répond aux besoins d'expressivité et de souplesse de la langue, ceci par la contribution sémantique du verbe supportant le mot prédicatif. Ce verbe modifie le sens de la collocation en apportant l'information sur le registre, l'aspect, etc. D'un autre côté, dans beaucoup de cas, les différences initiales entre les prédicats verbaux et les collocations qui leur correspondent ont tendance à disparaître. Par conséquent, ils peuvent être considérés comme synonymiques. Mais, en même temps, il arrive souvent en polonais que le verbe prédicatif et la collocation correspondante n'aient pas la même structure de valence, ce qui se manifeste par des différences au niveau des propriétés morpho-syntaxiques des arguments (changement de cas de déclinaison, usage de préposition, etc.).

8. STRUCTURES DE VALENCE OU CLASSES D'OBJETS ?

L'application de la conception de la structure de valence (en tant qu'indice formel de la signification du prédicat et, en même temps, d'un synset) et comparable à celle de *classes d'objets* de G. Gross dans la description des prédicats de la langue française. G. Gross définit la notion de classe d'objets comme : « ensemble de substantifs, sémantiquement homogènes, qui détermine une rupture d'interprétation

d'un prédicat donné, en délimitant un emploi spécifique » (2008 : 11). En même temps, G. Gross considère que « les classes d'objets ne sont pas des concepts sémantiques abstraits mais des entités construites sur des bases syntaxiques et déterminées par la signification des prédicats » (*ibidem*). Autrement dit, c'est un outil de description, dérivé de la signification du prédicat, ce qui présuppose que les significations des prédicats sont données *a priori*.

Une classe d'objets est un ensemble de mots dont chacun peut remplir la position argumentale pour un sens donné du prédicat. Dans des cas particuliers, il peut s'avérer que certaines classes d'objets ne se laissent pas caractériser par les synsets (p. ex. quand les significations distinguées par G. Gross ne correspondent pas aux significations distinguées en polonais selon la structure de valence (décrite plus haut)). Indépendamment des différences entre l'approche de G. Gross et ce que nous proposons, les deux approches sont conformes au principe de distinguer les significations en fonction de différences de position des arguments : « autant d'emplois différents (des prédicats) qu'il y a de classes (d'objets) différentes en position argumentale » (G. Gross, 2002 : 96).

9. CHANGEMENT DE STRUCTURE DE VALENCE

L'application stricte du principe que tout synset est caractérisé par l'information syntaxico-sémantique sous la forme de structure de valence commune fait que dans certains cas la collocation paraphrasante n'appartient pas au même synset que le verbe simple (bien qu'elle ait le même sens, définition et attribution des rôles sémantiques). Les exemples *upoważnić – dać pełnomocnictwo* (fr. *autoriser – donner l'autorisation*) permettent d'observer ce phénomène :

Piotr upoważnił adwokata (Acc).

Piotr dał pełnomocnictwo adwokatowi (Dat).

Les deux phrases ont le même contenu et sont identiques sur le plan pragmatique. La seule différence est d'ordre syntaxique et concerne le prédicat (simple vs composé), de même que la flexion de l'argument qui joue le rôle du Bénéficiaire (remplacement de l'accusatif par le datif). Parfois, la différence est marquée par la présence d'une préposition : *szanować – mieć szacunek* (*respecter – avoir du respect*) où l'argument de la collocation est introduit par la préposition *dla* suivie du génitif.

Piotr szanuje ojca (Acc).

Piotr ma szacunek dla ojca (prép(*dla*)+Gen).

En définitive, la solution que nous proposons pour le problème de transformations consiste en introduction des relations entre les synsets apparentés et décrivant le type de transformations nécessaires. Ainsi, le synset du verbe *upoważnić* est lié

à celui de la collocation *dać pełnomocnictwo* par la relation `Trans_Case_Beneficient_Acc_Dat` qui informe sur la nécessité du changement de cas.

En fin de compte, la solution admise dans le système PolNet permet de :

- obtenir la granulation des synsets verbaux qui correspond aux structures de valence simples,

- faire en sorte que deux synsets différents, dont l'un contient le verbe simple et l'autre la collocation paraphrasante, ayant des structures de valence semblables¹⁷ (non identiques) soient reliés par la relation transformationnelle qui décrit les différences.

10. PERSPECTIVES

Les travaux sur les verbes simples et les collocations verbo-nominales¹⁸ ont ouvert une nouvelle phase de transformation du PolNet initial en un système de type lexique-grammaire. Ils nous ont amenés à porter plus particulièrement attention aux problèmes de granulation qui sont essentiels dans la construction de systèmes du type WordNet. C'est d'autant plus important que les grammaires et dictionnaires traditionnels ne sont pas conçus en fonction des besoins de la linguistique informatique et du génie linguistique. Les travaux sur le lexique-grammaire du polonais sous la forme de PolNet sont en cours et seront poursuivis, notamment pour ce qui concerne les aspects quantitatifs et l'extension à d'autres parties du discours.

RÉFÉRENCES

- FILLMORE, Charles J., BAKER, Collin F. & SATO, Hiroaki (2002) : « The FrameNet Database and Software Tools ». Dans : *Proceedings of the Third International Conference on Language Resources and Evaluation. Vol. IV*, Las Palmas : LREC.
- GROSS, Gaston (2008) : « Les classes d'objets ». *Lalies* 28, Presses de l'ENS, Paris : Éditions Rue d'Ulm, 111-165 (<https://halshs.archives-ouvertes.fr/halshs-00410784>; consulté le 1.01.2015).
- GROSS, Gaston (2002) : « Recherches théoriques et enseignement des langues ». Dans : Alicja KACPRZAK (éd.), *Points communs : linguistique, traductologie, glottodidactique*, Łódź : Wydawnictwo Biblioteka, 88-101.
- GROSS, Maurice (1994) : « Constructing Lexicon-Grammars ». Dans : Beryl T. Sue ATKINS & Antonio ZAMPOLLI (éds.), *Computational Approaches to the Lexicon*, Oxford, UK : Oxford University Press, 213-263.
- GRUBER, Thomas R. (1993) : « A translation approach to portable ontology specifications ». *Knowledge Acquisition*, vol. 5(2) : 199-220.

¹⁷ Nous disons que deux structures de valence sont *semblables* quand elles ont les lignes 'frame' ainsi que l'attribution des rôles sémantiques respectivement identiques.

¹⁸ Recherche initiale dont le résultat a été l'acquisition et la description d'une partie essentielle des collocations verbo-nominales du polonais ; recherches en cours (G. Vetulani, 2000, 2012).

- MILLER, George A., BECKWITH, Richard, FELLBAUM, Christiane, GROSS, Derek & MILLER, Katherine (1990) : « WordNet: An online lexical database ». *Int. J. Lexicograph.* 3, 4 : 235-244.
- PALA, Karel, HORÁK, Aleš, RAMBOUSEK, Adam, VETULANI, Zygmunt, KONIECZKA, Paweł, MARCINIAK, Jacek, OBRĘBSKI, Tomasz, RZEPECKI, Paweł & WALKOWSKA, Justyna (2007) : « DEB Platform tools for effective development of WordNets in application to PolNet ». Dans : Zygmunt VETULANI (éd.), *Proceedings of the 3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, October 5-7, 2005, Poznań, Poland*, Poznań : Wydawnictwo Poznańskie, 514-518.
- PALMER, Martha (2009) : « Semlink: Linking PropBank, VerbNet and FrameNet ». Dans : *Proceedings of the Generative Lexicon Conference*, Sept. 2009, Pisa, Italy: GenLex.
- POLAŃSKI, Kazimierz (éd.) (1980-1992) : *Słownik syntaktyczno-generatywny czasowników polskich*, vol. I-IV, Wrocław : Ossolineum, 1980-1990 ; vol. V, Kraków (1992) : Instytut Języka Polskiego PAN.
- POLAŃSKI, Kazimierz (éd.) (1995) : *Encyklopedia językoznawstwa ogólnego*. Wrocław-Warszawa-Kraków : Ossolineum.
- PRZEPIÓRKOWSKI, Adam (2004) : *The IPI PAN Corpus. Preliminary Version*. Warszawa : IPI PAN.
- VETULANI, Grażyna (2000) : *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych*. Poznań : Wydawnictwo Naukowe UAM.
- VETULANI, Grażyna (2012) : *Kolokacje werbo-nominalne jako samodzielne jednostki języka. Syntaktyczny słownik kolokacji werbo-nominalnych języka polskiego na potrzeby zastosowań informatycznych. Część I*. Poznań : Wydawnictwo Naukowe UAM.
- VETULANI, Zygmunt (2003) : « Linguistically Motivated Ontological Systems ». Dans : Nagib CALLAOS, William LESSO, Klaus-Dieter SCHEWE & Elsayed ATLAM (éds.), *Proceedings of the 7th World Multiconference on Systemics, Cybernetics and Informatics, July 27-30, 2003, Orlando, Florida, USA*, vol. XII (*Information Systems, Technologies and Applications: II*), Int. Inst. of Informatics and Systemics, 395-400.
- VETULANI, Zygmunt (2004) : « Towards a Linguistically Motivated Ontology of Motion: Situation Based Synsets of Motion Verbs ». Dans : Valerie BARR & Zdravko MARKOV (éds.), *Proceedings of the Seventheens International Florida Artificial Intelligence Research Society Conference (FLAIRS-04)*, AAAI Press (2004), Menlo Park, California, 813-817.
- VETULANI, Zygmunt (2012) : « Wordnet and Gödel's Completeness Theorem ». *Studies in Logic, Grammar and Rhetoric* 27 (40) : 213-221.
- VETULANI, Zygmunt & MARCINIAK, Jacek (2011) : « Natural Language Based Communication between Human Users and the Emergency Center: POLINT-112-SMS ». Dans : Zygmunt VETULANI (éd.), *Human Language Technology. Challenges for Computer Science and Linguistics. LTC 2009. Revised Selected Papers*, LNAI 6562, Berlin-Heidelberg : Springer-Verlag, 303-314.
- VETULANI, Zygmunt, MARCINIAK, Jacek, OBRĘBSKI, Tomasz, VETULANI, Grażyna, DĄBROWSKI, Adam, KUBIS, Marek, OSIŃSKI, Jędrzej, WALKOWSKA, Justyna, KUBACKI, Piotr & WITALEWSKI, Krzysztof (2010) : *Zasoby językowe i technologie przetwarzania tekstu. POLINT-112-SMS jako przykład aplikacji z zakresu bezpieczeństwa publicznego (in Polish) [Language resources and text processing technologies. POLINT-112-SMS as example of homeland security oriented application]*. Poznań : Wydawnictwo Naukowe UAM.
- VETULANI, Zygmunt & VETULANI, Grażyna (2013) : « Through Wordnet to Lexicon Grammar ». Dans : Fryni KAKOYIANNI-DOA (éd.), *Penser le lexique-grammaire : perspectives actuelles*, Paris : Éditions Honoré Champion, 531-545.

- VETULANI, Zygmunt, WALKOWSKA, Justyna, OBREBSKI, Tomasz, KONIECZKA, Paweł, RZEPECKI Paweł & MARCINIAK, Jacek (2007) : « PolNet – Polish WordNet project algorithm ». Dans : Zygmunt VETULANI (éd.), *Proceedings of the 3rd Language and Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics, October 5-7, 2007, Poznań, Poland*, Poznań : Wydawnictwo Poznańskie, 172-176.
- VOSSEN, Piek, BLOKSMA, Laura, RODRIGUEZ, Horacio, CLIMENT, Salvador, CALZOLARI, Nicoletta & PETERS, Wim (1998) : « The EuroWordNet Base Concepts and Top Ontology », Version 2, Final, January 22, 1998 (Euro WordNet project report) (<http://www.vossen.info/docs/1998/D017.pdf> ; consulté le 22.04.2011). (also: Vossen, P. (et al.) (2002): *EuroWordNet General Document, Version 3. Final, July 1*, <http://www.vossen.info/docs/2002/EWNGeneral.pdf>; consulté le 1.01.2015).