

Problèmes d’acceptabilité des occurrences relevées dans les corpus informatisés : exemple des prédicats nominaux désignant une propriété ou un trait de caractère¹

Acceptability problems of the occurrences noted
in the corpora – on the example of nominal predicates
designating properties and character traits

Grażyna Vetulani

Adam Mickiewicz University, Poznań, Poland

gravet@amu.edu.pl

ORCID : 0000-0002-2138-3704

Abstract

In this paper, we present the problems encountered during our research in the field of nominal predication in Polish, in particular concerning nouns referring to properties and character traits. In a previous work on predicates referring to various activities and behaviors, carried out with the use of electronic corpora (the National Corpus of Polish Language, as well as electronic sources available on the Internet), we obtained a dictionary of verb-noun collocations with high lexical coverage comparing to traditional dictionaries (not based on electronic corpus research). Using the same corpus data and methodology as before, we conducted a study of the class of predicative nouns referring to properties and character traits. However, the results obtained following the application of this method proved to be unsatisfactory. In this article, we consider the reasons for this in order to optimize research methods later in our work.

Keywords: corpus linguistics, nominal predicates, feature names, construction of dictionaries

¹ Cet article est en partie un compte-rendu des travaux financés par le Ministère de la Recherche et de l’Enseignement en Pologne réalisés entre le 1.02.2012 et le 30.04.2015 sous ma direction [Programme de Recherche : NPRH, Nr 0022/FNiTP/H11/80/2011, intitulé *Rozbudowa zasobów cyfrowych języka polskiego w zakresie słowników walencyjnych w kierunku leksykonu-gramatyki zorientowana na potrzeby zastosowań informatycznych w humanistyce (Lexicon-grammar oriented extension of Polish digital valency dictionaries for computer applications in humanities)*].

1. INTRODUCTION

Bien que les recherches linguistiques d'aujourd'hui utilisent de plus en plus les corpus électroniques, et qu'elles exploitent également les méthodes d'analyse informatisées, les résultats obtenus à la suite de leur application peuvent s'avérer peu satisfaisants. Dans cet article, nous voulons rendre compte des problèmes rencontrés lors des travaux sur les prédicats nominaux du polonais et plus particulièrement sur ceux qui renvoient aux propriétés et aux traits de caractère (les éléments réunis dans la Classe II, Vetulani, 2000). Dans la première monographie consacrée aux noms prédicatifs du polonais (Vetulani, 2000), analysés selon les principes du *lexique-grammaire* (cf. les travaux antérieurs du L.A.D.L. de Paris-VII dirigés par Maurice Gross), nous avons distingué cinq classes de prédicats, selon le type d'unités comprises dans chacune d'entre elles. Ainsi, la Classe I contient les noms désignant les différents comportements, activités et opérations ; dans la Classe II se trouvent les noms renvoyant aux traits de caractère et aux propriétés ; la Classe III réunit les noms de maladies et la Classe IV ceux qui désignent les professions. La Classe V diffère des précédentes, car elle comprend les prédicats qui entrent dans les constructions avec un *verbe support événementiel*.

Les travaux antérieurs, de nature comparative entre le français et le polonais, concernaient les prédicats nominaux désignant tout type d'activité et de comportement (Classe I, Vetulani, 2000). Ils montraient l'efficacité de l'utilisation des corpus électroniques et des systèmes de recherche automatique (Vetulani, 2012). Lorsque par la suite nous avons examiné les éléments de la Classe II, nous nous attendions donc à de bons résultats. Or, les occurrences relevées jusqu'à présent paraissent moins prometteuses, d'une part par rapport aux résultats concernant la langue française pour le même type d'unités, et de l'autre, par rapport à nos attentes – justifiées par la bonne expérience de l'étape antérieure.

Dans cet article, nous allons chercher les causes des insuffisances mentionnées, malgré l'exploration de plusieurs corpus informatisés et l'utilisation des logiciels d'extraction performants, puis nous proposons des solutions appropriées afin d'obtenir des données de qualité supérieure. Ces données serviront à enrichir le dictionnaire des noms prédicatifs du polonais, réunissant déjà les éléments de la Classe I accompagnés de la description sémantico-syntaxique codée et destinée aux applications informatiques (Vetulani, 2012). Dans cet article nous indiquons aussi quelques futurs axes de recherche en lien avec les besoins lexicographiques actuels.

2. L'OBJET DE L'ÉTUDE

2.1. L'APPROCHE DITE *LEXIQUE-GRAMMAIRE*

Les recherches sur les substantifs désignant les propriétés et les traits de caractère en polonais s'inscrivent dans un projet de longue haleine remontant aux années 90, au moment où nous avons entrepris nos démarches concernant la description des prédicats nominaux. Nous avons pris comme point de départ plusieurs travaux antérieurs, fructueux surtout pour le français (Giry-Schneider, 1978 ; Danlos, 1980 ; Vivès, 1983 ; Gross & Vivès, 1986 ; Gross, G., 1987)², mais aussi pour d'autres langues, telles que le grec, le coréen, l'arabe, le malgache, etc. (à cette époque, ces travaux s'effectuaient au L.A.D.L.³ de l'Université Paris-VII sous la direction de Maurice Gross). Dès le début, nous avons décidé d'utiliser les mêmes principes méthodologiques que ceux appliqués au français, connus sous le terme du *lexique-grammaire*, tout en ayant pleinement conscience du fait qu'il faudrait surmonter certains obstacles liés à la terminologie car, dans notre cas, il s'agissait du polonais, langue ayant de longue date sa propre tradition grammaticale et son propre appareil notionnel. Il nous semblait utile de pouvoir appliquer à ce système une même méthodologie : d'un côté pour vérifier l'efficacité de la méthode d'analyse face à une langue slave, et de l'autre, pour créer un premier dictionnaire sémantico-syntaxique des prédicats nominaux du polonais.

Les méthodes descriptives du *lexique-grammaire* préconisent "une interdépendance systématique de deux composantes de la langue qui sont apparemment de nature totalement différente : celle des règles formelles de la grammaire et celle des entrées lexicales porteuses du sens" (Kakoyianni-Doa, 2014, p. 15). Par conséquent, en décrivant le sens d'un élément, il faut prendre en considération l'organisation des éléments au niveau des structures syntaxiques dans lesquelles l'élément en question apparaît dans l'usage et d'en rendre compte, d'une manière ou d'une autre, dans le dictionnaire. Kakoyianni-Doa ajoute ensuite que "la découverte des unités d'une langue et de leurs propriétés est soumise à une exigence de catégorisation. Comme les résultats de la description doivent être suffisamment formels pour permettre une application au traitement automatique des langues, ils prennent la forme de tableaux à double entrée – les "tables" ou "matrices", qui présentent les entrées lexicales en regard avec leurs propriétés syntaxico-sémantiques" (Kakoyianni-Doa, 2014, p. 15).⁴ Cela veut dire que pour chaque usage de l'élément donné, on présente dans les tables les cadres syntaxiques admis.

² Ces ouvrages ont suivi *Méthodes en Syntaxe* de Maurice Gross (Gross, 1975).

³ Laboratoire d'Automatique Documentaire et Linguistique.

⁴ Les chercheurs du L.A.D.L. représentent les éléments linguistiques dans une sorte de dictionnaires hautement formalisés, appelées *tables syntaxiques*. Guillet (1991, p. 120) dit : « Chaque ligne de la matrice est une phrase élémentaire (un « sens » dans la tradition lexicographique). »

2.2. LES PRÉDICATS NOMINAUX EN POLONAIS

Dans un premier temps, après avoir pris certaines distances par rapport à la tradition grammaticale polonaise (qui faisait analyser en premier lieu le verbe comme centre logique et sémantique de la phrase), nous avons emprunté la définition du prédicat à la logique mathématique et admis qu'est prédicat tout élément qui décrit une propriété du sujet ou une relation entre le sujet et les compléments, c'est-à-dire entre les arguments (Carnap, 1934 ; Grzegorzczuk, 1969 ; Gross, M., 1981). Ainsi, il était naturel d'admettre qu'un prédicat se réalise dans le discours sous la forme des mots appartenant à différentes catégories grammaticales, y compris à la catégorie des noms (Vetulani, 2000). En ce qui concerne ces derniers, peuvent être reconnues comme prédictives uniquement les formes dans leur usage abstrait car ce sont elles qui remplissent les conditions dont il est question. Mais la description des prédicats nominaux ne peut pas se limiter à la constatation qu'il s'agit des abstraits par opposition aux concrets (cf. en pol. *telefon* dans l'expression *wykonać telefon do kogoś* où le nom *telefon* veut dire 'appel' et *telefon* en tant qu'appareil, ou en fr. *action* dans *mener une action* et *action* comme 'part de capital de société'). La description exhaustive d'un prédicat nominal exige que l'on prenne en compte ses contextes d'emploi afin de découvrir tous ses sens (presque chaque forme prédictive est polysémique). Définir le ou les sens d'un prédicat implique alors l'observation attentive des éléments qui l'accompagnent dans la structure. En premier lieu, il s'agit du verbe car le prédicat nominal, ne possédant la possibilité d'exprimer ni la personne, ni le temps, ni l'aspect, est lié à un *verbe support* en formant avec lui un syntagme. D'ailleurs, c'est une raison pour laquelle, au fur et à mesure de nos études sur les prédicats nominaux du polonais, il nous a fallu reconnaître une sorte de tour prédictif (*collocation*) à structure *verbe support + nom prédictif*, d'autant plus qu'au niveau de chaque expression de ce type, on observe une forte lexicalisation entre les éléments constitutifs (voir à ce sujet Vetulani, 2012). L'analyse de la dépendance réciproque du prédicat nominal et du verbe support permet, semble-t-il, d'y voir un signe composé (du moins en polonais car le plus souvent ce sont des expressions conventionnelles dans lesquelles le choix du verbe n'est pas d'emblée prévisible). Dans la plupart des cas, le prédicat nominal sélectionne un verbe support approprié [cf. *pełnić funkcję* (fr. *remplir une fonction*), *wszcząć alarm* (fr. *sonner/donner l'alarme*), *wymierzyć karę* (fr. *infliger une punition*), etc.].

À part le verbe support (ou les verbes supports car une forme prédictive en admet souvent plusieurs) qu'il faut présenter explicitement dans le dictionnaire à côté du nom, il est nécessaire de prendre en compte d'autres éléments cooccurrents et les définir avec rigueur. Il s'agit notamment de fournir les informations sur les contraintes grammaticales pour un sens donné du prédicat qui concernent le nombre, la nature et la détermination des arguments, mais aussi la façon dont ils sont introduits dans la phrase (avec ou sans prépositions). On doit aussi rendre compte de la présence dans la phrase des modificateurs – si la structure l'exige (cf. *être d'une grande bonté* et **être*

de bonté). En ce qui concerne le polonais, outre les traits mentionnés, on doit ajouter ceux qui concernent la flexion casuelle des prédicats et des arguments, ainsi que préciser l'aspect (*perfectif* ou *imperfectif*) pour chaque verbe support.⁵ La méthode de description détaillée des contraintes grammaticales pour chaque emploi de la forme prédicative permet de lever sa polysémie.

2.2.1. LES RÉSULTATS OBTENUS (CLASSE I)

Après avoir adopté la méthode d'analyse du lexique-grammaire, nous avons recensé et décrit sous la forme codée et destinée à des applications ultérieures plus de 14600 formes prédicatives (*collocations verbo-nominales*) du polonais se trouvant actuellement sur un CD joint à la monographie (Vetulani, 2012).⁶ Ce dictionnaire comprend les collocations relevées dans le corpus informatisé⁷ parmi lesquelles, à côté des expressions connues, il y a celles qui sont moins conventionnelles. Nous les avons gardées toutes pour témoigner à la fois de la créativité et du dynamisme du système polonais dans le domaine de la prédication nominale. Comme nous l'avons dit plus haut, les prédicats entrant dans les collocations réunies représentaient la Classe I (Vetulani, 2000), peu homogène encore (mais une classification plus détaillée est possible selon les mêmes principes) dans laquelle se trouvent les substantifs désignant des activités et comportements variés comme : *aprobata* (fr. *approbation*), *bitwa* (fr. *bataille*), *bląd* (fr. *faute*), *ucieczka* (fr. *fuite*) et ainsi de suite. Chaque collocation est accompagnée d'un schéma (d'un code) étant une illustration formelle de son sens et correspondant à la structure d'emploi retrouvée dans le corpus. Tous les emplois des prédicats se résument en un schéma général : *collocation verbo-nominale/verbe support (cas grammatical du nom prédicatif)/(MOD)/N1Prép (cas grammatical)/N2Prép (cas grammatical) /...*⁸.

Sans entrer dans les détails concernant la construction des entrées, nous présentons ci-dessous un exemple tiré du dictionnaire, à savoir le nom prédicatif *nadzór* (en fr. *surveillance*) formant avec des verbes supports appropriés différentes collocations :

⁵ Indépendamment de ce qu'on sait à propos des *variantes aspectuelles* d'un *verbe support neutre (standard)* qui ont été analysées pour le français (Vivès, 1983) et que l'on peut retenir aussi dans l'analyse du polonais, il s'agit ici de l'aspect grammatical, une catégorie propre aux langues slaves qui diffère de ce qu'on définit comme *mode d'action*. En polonais, comme dans d'autres langues de la même famille, un verbe est d'emblée (au niveau de l'infinitif) soit *perfectif*, soit *imperfectif*.

⁶ En même temps, nous reconnaissons que la description comprise dans le dictionnaire, bien qu'elle suive les principes méthodologiques du lexique-grammaire, est moins rigoureuse par rapport à celle que l'on retrouve dans les tables syntaxiques élaborées au L.A.D.L.

⁷ Narodowy Korpus Języka Polskiego (NKJP).

⁸ En polonais, on distingue 7 cas pour la flexion nominale : *mianownik (Nominatif)*, *dopelniacz (Génitif)*, *celownik (Datif)*, *biernik (Accusatif)*, *miejsownik (Ablatif)*, *narzędnik (Instrumental)*, *wolacz (Vocatif)*. Les symboles *N1*, *N2*,... renvoient aux arguments du prédicat qui s'introduisent dans la phrase directement ou par une préposition (*Prép*).

nadzór, m

być pod nadzorem/być pod(N)/N1(D);MOD,
mieć nadzór/mieć(B)/N1nad(N),
obejmować nadzorem/obejmować(N)/N1(B),
obejmować nadzór/obejmować(B)/N1nad(N),
objąć nadzorem/objąć(N)/N1(B),
objąć nadzór/objąć(B)/N1nad(N),
pełnić nadzór/pełnić(B)/N1nad(N),
podejmować nadzór/podejmować(B)/N1nad(N),
podjąć nadzór/podjąć(B)/N1nad(N),
podlegać nadzorowi/podlegać(C)/N1(D),
posiadać nadzór/posiadać(B)/N1nad(N),
pozostać pod nadzorem/pozostać pod(N)/N1(D),
pozostawać pod nadzorem/pozostawać pod(N)/N1(D),
prowadzić nadzór/prowadzić(B)/N1nad(N),
rozpościerać nadzór/rozpościerać(B)/N1nad(N),
roztaczać nadzór/roztaczać(B)/N1nad(N),
roztoczyć nadzór/roztoczyć(B)/N1nad(N),
sprawować nadzór/sprawować(B)/N1nad(N),
wykonywać nadzór/wykonywać(B)/N1nad(N),
wypełniać nadzór/wypełniać(B)/N1nad(N),
zastosować nadzór/zastosować(B)/N1nad(N),
znajdować się pod nadzorem/znajdować się pod(n)/N1(D),

Pour chaque collocation, le dictionnaire fournit plusieurs exemples d'emploi en contexte relevés dans le corpus exploité. Nous ne pouvons les citer ici, faute de place (cf. Vetulani, 2012, pp. 150-152).

2.2.2. LES RECHERCHES EN COURS SUR LES PRÉDICATS DÉSIGNANT LES PROPRIÉTÉS ET LES TRAITS DE CARACTÈRE (CLASSE II)

Ayant confirmé l'efficacité du *lexique-grammaire* dans l'analyse d'une langue slave, nous avons décidé, avec l'intention d'enrichir notre dictionnaire, d'entreprendre une étude systématique des prédicats de la Classe II (Vetulani, 2000), à savoir des noms désignant les propriétés et les traits de caractère. Ces travaux sont actuellement en cours. En ce qui concerne la nature grammaticale de ce type de prédicats, comme tous les autres prédicats, ils constituent le centre logique et sémantique de la phrase, mais le plus souvent ils ne sélectionnent qu'un argument, peu importe sa nature : humaine ou concrète ; cf. *głupota Jana* (fr. *la stupidité de Jean*), *złośliwość Jana* (fr. *la méchanceté de Jean*), *złośliwość charakteru* (fr. *la méchanceté du caractère*), *złośliwość uwagi* (fr. *la méchanceté de la remarque*), *dobro serca* (fr. *la bonté*

du coeur), *chropowatość powierzchni* (fr. *rugosité de la surface*), etc. Du point de vue sémantique, ces prédicats sont proches des adjectifs (aussi prédicatifs) ; cf. *głupota Jana/Jan jest głupi* (fr. *la stupidité de Jean/Jean est stupide*), *złośliwość Jana/Jan jest złośliwy* (fr. *la méchanceté de Jean/Jean est méchant*), etc.⁹

Dans la langue polonaise, parmi les substantifs désignant les propriétés et les traits de caractère se trouvent des formes dérivées des adjectifs et terminées en *-ość* (*aktywność, bezpretensjonalność, bezbronność*) ou en *-oć* (*dobroć*), mais aussi en *-izm, -yzm* (*egoizm, autentyzm, autokrytycyzm*), en *-cja, -sja* (*arogancja, agresja*), en *-ctwo, -stwo* (*cwaniactwo, dyletanctwo*) et en *-ga* (*odwaga, powaga*)¹⁰. On y rencontre encore des dérivés des substantifs comme *czasownikowość* (en fr. 'propre au verbe') ou d'autres, moins réguliers, du type : *animusz* (fr. *entrain*), *chłód* (fr. *froidueur*), *donkiszoteria* (fr. *don-quichottisme*), *dynamika* (fr. *dynamique*), etc. Parfois, on y voit des formes parallèles du point de vue sémantique comme p. ex. *pedanteria/pedantyzm* (fr. *pédanterie*) ou *konserwatywność/konserwatyzm* (fr. *conservatisme*).

Les difficultés liées au recensement et à l'analyse des prédicats de la Classe II ne sont pas liées seulement à leur variation formelle. Il s'agit encore de la taille de l'ensemble, de la polysémie des mots et de la qualité des corpus d'analyse.

3. LES RAPPORTS ENTRE LA QUALITÉ DES CORPUS ET LA QUALITÉ DU DICTIONNAIRE

3.1. L'INÉGALITÉ ENTRE LES CORPUS FRANÇAIS ET POLONAIS LORS DE LA PREMIÈRE ÉTAPE DES TRAVAUX SUR LES PRÉDICATS NOMINAUX

En raison de la taille du système linguistique, de son caractère ouvert et dynamique, ainsi que de sa variation interne, il est clair que seuls des corpus de bonne qualité¹¹ permettent d'arriver à des résultats fiables. C'est d'autant plus important

⁹ Cette corrélation est tout à fait naturelle. Dans chaque système linguistique, on observe une variation grammaticale dans l'expression d'une même idée. Cela ne veut pas dire que dans chaque langue tout prédicat s'actualise au niveau du discours à travers toutes les catégories grammaticales. Parfois, on observe des réalisations parallèles en français et en polonais, comme dans les cas de *pragnąć/pragnienie/pragnący* et *désirer/désir/désireux* (corrélation entre un verbe, un nom et un adjectif prédicatifs), parfois non, p. ex. *nadzieja* en polonais n'a ni son correspondant verbal, ni adjectival (c'est un *prédicat autonome*), tandis que le français possède à la fois *espoir* (un prédicat nominal) et *espérer* (un prédicat verbal).

¹⁰ Traduction française des éléments cités, conformément à l'ordre : *activité, simplicité, vulnérabilité, bonté, égoïsme, authenticité, autocritique, arrogance, agression, ruse, dilettantisme, courage, gravité*.

¹¹ Rappelons ici comment définit *le corpus* J. Sinclair : « Un corpus est une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage » (Habert, Nazarenko & Salem, 1997, p. 4).

dans le cas des études comparatives qui exigent l'existence de corpus de taille et de qualité comparables.

En ce qui concerne les prédicats nominaux en français, nous avons souligné à une autre occasion (Vetulani, 2018, pp. 344-345) que leur recensement « s'effectuait dès le début dans de très bonnes conditions. Déjà dans les années 70 a été créée une base de données de textes français (*FRANTEXT*) qui comportait des textes littéraires et philosophiques, mais aussi scientifiques et techniques afin de fournir des exemples pour le *Tre sor de la Langue Franaise*. Une fois le dictionnaire termin , la base *FRANTEXT* a continu     voluer : d'abord mise   disposition dans les ann es 80 sous forme d'un CD (*Discotext*), ensuite mise sur le web (en 1998) ». Elle est r guli rement enrichie et disponible en ligne (<https://www.frantext.fr/>). Aujourd'hui, elle est la seule   proposer des recherches sur des textes allant des ann es 1180   2020, avec un fonds contemporain tr s riche. *FRANTEXT* correspond de plein droit   ce qu'on consid re comme *corpus national* car :

- avec un volume de 262 millions de mots (ao t 2021) il est le plus complet des corpus annot s du franais ;
- il constitue une collection de donn es linguistiques organis es selon les crit res explicites pour obtenir un  chantillon repr sentatif de la langue franaise ;
- il refl te une grande diversit  linguistique ;
- il continue d' tre actualis  et maintenu ;
- en particulier, il est dot  de fonctionnalit s permettant de charger, de cr er, de visualiser et d' diter des corpus de travail pour des applications linguistiques et informatiques.

  la m me  poque, apparaissaient en France d'autres collections de textes  lectroniques, de m me que des syst mes de recherche informatis s efficaces.¹²

 tant donn  le manque de corpus et de syst mes comparables pour le polonais au d but de nos recherches sur les pr dicats nominaux, nous avons utilis  comme source pour nos observations en premier lieu un dictionnaire traditionnel (Szymczak, 1987). Ce corpus sur papier, normatif et limit , nous a permis de rassembler les premi res donn es : il a fourni plus de 8000 formes pr dicatives, de m me qu'un certain nombre d'informations sur leur usage. Certes, la lecture de la totalit  des explications comprises dans le dictionnaire  tait p nible et surtout lente, mais en fin de compte elle nous a permis de relever un certain nombre d'informations pertinentes li es   l'usage grammatical des  l ments analys s. Ces informations  taient n cessaires pour attribuer   chaque pr dicat un ou plusieurs mod les correspondant   son fonctionnement syntaxique.  videmment, le manque de compatibilit  entre les corpus analys s pour le franais et le polonais, au d but de nos recherches, a eu un impact sur la forme du

¹² P. ex. tr s t t, a  t  cr e pour le franais le syst me *INTEX*, un analyseur lexical constitu  des dictionnaires  lectroniques et des programmes d'analyse informatis s que nous avons vu fonctionner au L.A.D.L. (Silberztein, 1993).

dictionnaire. Par rapport aux tables syntaxiques du L.A.D.L. (très riches en informations concernant les structures syntaxico-sémantiques des prédicats français) le format que nous avons proposé était modeste, mais conforme aux principes théoriques admis.

3.2. LE PROGRÈS LIÉ À L'EXTRACTION AUTOMATISÉE DES DONNÉES

Au moment où ont été disponibles les premiers corpus électroniques de la langue polonaise (fin des années 90 ; cf. Przepiórkowski, 2004), de même que les systèmes informatiques d'extraction des données, nous y avons vu l'occasion d'accélérer les travaux sur la prédication nominale, ainsi que d'élargir notre dictionnaire.¹³ Il est vrai que les premières applications des méthodes d'analyse informatisée n'ont pas apporté de résultats très probants, surtout en raison de la qualité du corpus initial qui, à l'époque, n'avait pas encore tous les attributs de *corpus national*. Au fur et à mesure le corpus s'est développé pour devenir une base textuelle sérieuse (Przepiórkowski, Bańko, Górski & Tomaszczyk, 2012). On le connaît aujourd'hui sous le nom de *Narodowy Korpus Języka Polskiego* (NKJP). Nos travaux sur les prédicats nominaux du polonais allaient de pair avec l'évolution des ressources électroniques pour cette langue. Grâce à des méthodes d'analyse novatrices, nous avons créé un dictionnaire nouveau (pour les mêmes unités de la Classe I), enrichi par rapport au premier, c.-à-d. un dictionnaire de collocations verbo-nominales (Vetulani, 2012).

Comme nous l'avons déjà dit, la même approche empirique, basée sur les corpus électroniques et l'extraction informatisée des données, a été appliquée dans l'analyse des prédicats nominaux polonais désignant les propriétés et les traits de caractère. Les travaux s'effectuaient selon les mêmes règles, connues de l'étape précédente de nos recherches, consistant en :

- a) l'extraction automatique des concordances du corpus NKJP,
- b) le traitement automatique des concordances pour obtenir une liste de collocations candidates à l'analyse ultérieure,
- c) la lecture des contextes d'emploi par les lexicographes qui renaient ou rejetaient les constructions relevées.

Après avoir surmonté les problèmes liés à la polysémie des formes, comme ceux qu'on observe dans le prédicat *konsekwencja* qui se lie soit au verbe *wykazać* avec lequel il forme l'expression *wykazać konsekwencję w...* (voulant dire 'être conséquent'), soit au verbe *wyciągnąć* (*wyciągnąć konsekwencje* correspond à *tirer des conséquences de qch*), nous n'avons retenu que les formes employées au sens de 'qq, q.ch possède un trait'. Pour ces formes nous avons proposé le même format de dictionnaire, afin de garder la structure des codes définie au départ. La seule nou-

¹³ Les principaux travaux s'effectuaient dans les années 2007-2009 dans le cadre du Projet de Recherche financé par le MNiSW (NPRH, Nr 0022/FNiTP/H11/80/2011).

veauté du format par rapport à celui qui a été défini dans (Vetulani, 2012) consiste à distinguer deux types de structures d'emploi pour un nom prédicatif. Le groupe I comprend les structures dans lesquelles le nom est employé au *nominatif* (jouant la fonction du sujet dans la phrase – selon l'approche traditionnelle) et le groupe II toutes les autres.

Nous présentons ici un exemple tiré du dictionnaire actualisé : le nom *bezczynność* (fr. *inactivité*) :

bezczynność, ż

I

bezczynność gryzie/bezczynność gryzie/N0 (B),
bezczynność cechuje/bezczynność cechuje/N0 (B),

II

zachowywać bezczynność/zachowywać (B),
wykazać bezczynność/wykazać (B),
wykazywać bezczynność/wykazywać (B),
usychać w bezczynności/usychać w (Ms),
dopuścić się bezczynności/dopuścić się (D),
gnić w bezczynności/gnić w (Ms),
pozostawać w stanie bezczynności/pozostawać w stanie (D),
trwać w bezczynności/trwać w (Ms),

3.3. LES TENTATIVES POUR REMÉDIER AUX PROBLÈMES D'ACQUISITION DES DONNÉES

Malgré l'utilisation d'un corpus solide (NKJP) et l'application des systèmes performants d'extraction des données (Vetulani, Obrębski & Vetulani, 2006, 2007, 2008), lors de la recherche des contextes d'emploi pour les prédicats désignant les propriétés et les traits de caractère, nous n'avons pas encore obtenu de résultats satisfaisants. Notre déception découle avant tout du fait qu'un grand nombre de formes n'ont pas du tout apparu dans le corpus, du moins au sens qui nous intéressait (p. ex. *awaryjność, barokowość, barwność, bezdźwięczność, bezideowość, familiarność, fałszywość, filantropijność, humanitarność, kompetencyjność, niepodatność, niepodważalność, sensowność, seryjność, sękatność, siność, tonalność, trzpiotowatość, układność, umysłowość, uporczywość, urazowość, urokliwość, warowność, wczesność, wiatropylność* et beaucoup d'autres).

Pour remédier à cet état de choses nous avons entrepris les démarches de vérification consistant en l'utilisation des mêmes systèmes d'extraction sur d'autres corpus disponibles sur Internet. Ces essais n'ont pourtant pas amélioré les résultats. Les corpus soumis à l'analyse ont en plus dévoilé beaucoup de constructions à verbes inten-

sifs du registre relâché, parfois vulgaires, parfois peu acceptables, que le dictionnaire traditionnel ne notait pas.¹⁴

Nous présentons ci-dessous plusieurs contextes authentiques :

to co jest tu pisane nie musi **powalać komicznością**. Ma raczej, Fakt, że moje budowle **nie powalają epickością**, ale jednak..., Może moje pytanie jest głupie i **śmierdzi amatorstwem**, niektóre tytuły tak **dają po oczach japońskością**, że aż mnie nosi, język **owrzodził kaśliwością** uwag, **zajeżdża** i kabotynizmem, **dulszczyzną** czy jak zwał tak zwał, „Mój eliksir” może **zajeżdża** nieco **melodramatycznością**, ale,

Nous allons donc en conséquence poursuivre les travaux sur les prédicats nominaux désignant les propriétés et les traits de caractère.

4. CONCLUSION

Notre analyse a certes montré qu'à l'heure actuelle il est important de disposer de corpus représentatifs des langues analysées, et aussi d'utiliser des systèmes de recherche efficaces, mais par ailleurs, elle nous fait davantage réfléchir sur les préférences des usagers de la langue dans le choix des éléments pour exprimer une idée. Parmi toutes les formes renvoyant aux propriétés et aux traits de caractère que nous avons analysées, un bon nombre se conforment rarement au modèle proposé dans nos travaux exercés jusqu'à présent. Pourtant, le fait mentionné ne remet pas en cause la pertinence du modèle car des emplois qui lui correspondraient sont tout à fait possibles et acceptables, p. ex. dans le discours scientifique, recherché ou spécialisé. D'un autre côté, les résultats de nos analyses peuvent tout simplement résulter du fait que, dans l'usage courant, quand les usagers de la langue parlent des propriétés et des traits de caractère, ils préfèrent utiliser les structures à prédicat adjectival (plus simples, du type : *być dobrym*, en fr. *être bon*) à celles comportant un prédicat nominal (p. ex. *okazywać dobroć*, en fr. *être d'une bonté*). La réflexion sur les rapports entre la prédication nominale et adjectivale va accompagner nos prochaines analyses.

¹⁴ Nous retenons tout de même de telles structures, en envisageant d'introduire une notation sur leur valeur émotionnelle.

BIBLIOGRAPHIE

- Carnap, R. (1934). *Logische Syntax der Sprache*. In series *Schriften zur wissenschaftlichen Weltanschauung*, Band 8. Berlin Heidelberg : Springer-Verlag. DOI : 10.1007/978-3-662-25375-5.
- Danlos, L. (1980). *Représentation d'informations linguistiques : constructions N être Prép X. Thèse de 3 cycle*. Paris : L.A.D.L. Université Paris VII. <https://www.sudoc.fr/04109459X> [accès : 15.03.2022].
- Giry-Schneider, J. (1978). *Les nominalisations en français : l'opérateur « faire » dans le lexique*. Genève-Paris : Librairie DROZ.
- Gross, G. (1987). *Les constructions converses en français*. Genève-Paris : Librairie DROZ.
- Gross, G. & Vivès, R. (eds.). (1986). Syntaxe des noms. *Langue Française*, 69, 5-27. DOI : 10.3406/lfr.1986.6360.
- Gross, M. (1975). *Méthodes en syntaxe : régime des constructions complétives*. Paris : Hermann.
- Gross, M. (1981). Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63, 7-52. DOI : 10.3406/lge.1981.1875.
- Grzegorzczak, A. (1969). *Zarys logiki matematycznej* (wyd. 2). Warszawa : PWN.
- Guillet, A. (1991). Dictionnaires électroniques et lexique-grammaire. *Studia Romanica Posnaniensia*, XVI, 117-128.
- Habert, B., Nazarenko, A. & Salem, A. (1997). *Les linguistiques de corpus*. Paris : A. Colin.
- Kakoyianni-Doa, F. (ed.). (2014). *Penser le Lexique-Grammaire. Perspectives actuelles*. Paris : Honoré Champion.
- Przeziórkowski, A. (2004). *Korpus IPIPAN. Wersja wstępna*. Warszawa. http://nlp.ipipan.waw.pl/~adam/Papers/2004-corpus/book_pl.pdf [accès : 15.03.2022].
- Przeziórkowski, A., Bańko, M., Górski, R. & Lewandowska-Tomaszczyk, B. (2012). *Narodowy Korpus Języka Polskiego*. Warszawa : PWN. http://nkjp.pl/settings/papers/NKJP_ksiazka.pdf [accès : 15.03.2022].
- Silberstein, M. (1993). *Dictionnaires électroniques et analyse automatique de textes. Le système IN-TEX*. Paris, Milan, Barcelone, Bonn : Masson.
- Szymczak, M. (ed.). (1978). *Słownik języka polskiego*. Warszawa : PWN.
- Vetulani, G. (2000). *Rzeczowniki predykatywne języka polskiego. W kierunku syntaktycznego słownika rzeczowników predykatywnych na tle porównawczym*. Poznań : Wydawnictwo Naukowe UAM.
- Vetulani, G. (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, G. (2018). Les démarches et problèmes à résoudre dans les études comparatives – la prédication nominale en français et en polonais. *Neophilologica*, 30, 340-351.
- Vetulani, Z. & Vetulani, G. (2020). Polish Lexicon-Grammar Development Methodology as an Example for Application to other Languages. In *Proceedings of the LREC 2020 WILDRE5 – 5th Workshop on Indian Language Data : Resources and Evaluation* (pp. 51-59). Paris : European Language Resources Association (ELRA). EAN : 9791095546672. <https://aclanthology.org/2020.wildre-1.10/> [accès : 15.03.2022].
- Vetulani, G., Obrębski, T. & Vetulani, Z. (2007). Towards a Lexicon-Grammar of Polish : Extraction of Verbo-Nominal Collocations from Corpora. In *Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference, May 7-9, 2007, Key West, Florida, USA* (pp. 267-268). AAAI Press. <https://www.aaai.org/Papers/FLAIRS/2007/Flairs07-055.pdf> [accès : 15.03.2022].
- Vetulani, G., Vetulani, Z. & Obrębski, T. (2006). Syntactic Lexicon of Polish Predicative Nouns. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation, 24-26.05.2006, Genoa, Italy* (pp. 1734-1737). Paris : European Language Resources Association (ELRA).

- Vetulani, G., Vetulani, Z. & Obreński, T. (2008). Verb-Noun Collocations Syntax Dictionary – Corpus-Based Approach. In *Proceedings of 6th International Conference on Language Resource and Evaluation, Marrakech, Morocco* (pp. 1561-1564). Paris : European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/pdf/398_paper.pdf [accès : 15.03.2022].
- Vivès, R. (1983). *Avoir, prendre, perdre : constructions à verbe support et extension aspectuelle, Thèse de 3a cycle*. Paris : L.A.D.L. Université Paris VII.

