

De perceptione :
du protocole Ucherek 1982 à un grand corpus
étiqueté bilingue 2022

De perceptione :
from the Ucherek protocol to the large bilingual
multi-label corpus 2022

Fabrice Marsac

Université de Strasbourg, France, UR 1339 LiLPa – Linguistique, Langues & Parole

f.marsac@unistra.fr

ORCID : 0000-0002-4120-5526

Witold Ucherek

Université de Wrocław, Pologne

witold.ucherek@uwr.edu.pl

ORCID : 0000-0002-7954-7206

Abstract

This paper is made in continuation of the article of Fabrice Marsac, Witold Ucherek and Magdalena Dańko “De l’infinitive de perception dans la pratique traductologique” (*Studia Romanica Posnaniensia* 46/1, 2019); both papers fall within the framework of the bilateral Polonium program “*On the translation of French perception structures into Polish*” (n° PPN/BIL/2018/1/00181), implemented and financed by the Polish National Agency for Academic Exchange (NAWA), the French Ministry for Europe and Foreign Affairs (MEAE) and the Ministry of Higher Education, Research and Innovation (MESRI). In the present case, we intend to report on the content and progress of the part “multi-dimensional labelling” of this long-term French-Polish project. Therefore, we describe the formal encoding of structures, categories and functions of the items in the large multi-label corpus (French-Polish) we have been collecting since 2019.

Keywords: French-Polish corpus, multi-dimensional labelling, poly-informational tag, translation, infinitive subordinate clause, verbs of perception

1. INTRODUCTION

Le présent article prolonge celui de Fabrice Marsac, Witold Ucherek et Magdalena Dańko intitulé « De l’infinitive de perception dans la pratique traductologique » et publié en 2019 dans le numéro 46 (1) de la revue *Studia Romanica Posnaniensia*, les deux contributions s’inscrivant dans le cadre du Programme bilatéral Polonium n° PPN/BIL/2018/1/00181 : « *On the translation of French perception structures into Polish* », mis en œuvre et financé par l’Agence nationale pour l’échange académique (*Narodowa Agencja Wymiany Akademickiej* – NAWA) en Pologne et par les Ministères de l’Europe et des Affaires Étrangères (MEAE) et de l’Enseignement Supérieur, de la Recherche et de l’Innovation (MESRI) en France.

En l’espèce, c’est du contenu et de l’avancée du pan « étiquetage multidimensionnel » de ce projet franco-polonais au long cours dont nous nous apprêtons à rendre compte. Aussi décrirons-nous l’encodage formel des structures, des catégories et des fonctions des items du grand corpus étiqueté bilingue (français-polonais) que nous récoltons depuis 2019, et dont voici, à titre d’exemple, un premier tag, pris au hasard :

[FR] J’étais désolée de **voir** s’interrompre si vite une carrière prometteuse.

$CP^A \supset ICP : [SN1^-(\emptyset) \cup V1^-voir(I) \cup ICP^-(\text{☑})]_{CA}$

[PL] Byłam niepokieszona, **że** moja obiecująca kariera tak szybko się zakończyła.
że P

L’objectif ultime que nous poursuivons par là est de préparer linguistiquement le terrain, afin qu’un programme informatique puisse bientôt venir corrélérer les étiquettes des deux langues et mettre au jour, ce faisant, les pratiques effectives des traducteurs dans le passage du français au polonais s’agissant des constructions régies par un verbe de perception.

La présente contribution s’articulera en deux temps : après avoir rappelé les tenants, les aboutissants et les fondements méthodologiques de la recherche en cours (sect. 2), nous procéderons à la typologie commentée des différents types de tags employés pour l’étiquetage de notre corpus (sect. 3).

2. SYNTHÈSE DE LA RECHERCHE

Cette section est un auto-résumé de Marsac, Ucherek et Dańko (2019), contribution avancée à laquelle nous nous contenterons de renvoyer ici les lecteurs qui voudraient avoir devant les yeux le panorama scientifique du projet Polonium précité.

2.1. OBJET D'ÉTUDES, CADRE D'INVESTIGATION ET PROBLÉMATIQUE

Nous prenons pour objet d'étude l'Infinitive de Compte rendu de Perception (ICP), suivant la terminologie de Marsac (2006) :

- (1) J'entends les oiseaux chanter.
- (2) SN₁ V1 [SN₂ V2]_{ICP}

Le cadre dans lequel nous nous intéressons à l'ICP est celui de la traduction, en l'occurrence du français vers le polonais.

Cette structure n'existant pas en polonais, la langue de Mickiewicz compte au moins huit traductions différentes attestées dans notre corpus, dont la plus fréquente est *jak P* :

- (3) J'entends les oiseaux chanter/chanter les oiseaux.
- (4) *Słyszę ptaki śpiewać/śpiewać ptaki.
- (5) Słyszę, jak ptaki śpiewają/śpiewają ptaki.

D'où ces deux questions fondamentales (parmi d'autres) auxquelles nous aimerions trouver réponse : (a) comment le traducteur opère-t-il son choix, factuellement, parmi les différentes structures que lui offre sa langue naturelle, au hasard ou de manière contrainte (linguistiquement) ? ; (b) la structure polonaise *jak P*, la plus sollicitée dans nos relevés, possède-t-elle les diverses propriétés ou spécificités de l'ICP ?

2.2. HYPOTHÈSE DE RECHERCHE, CADRE THÉORIQUE ET OBJECTIFS

L'hypothèse que nous faisons à ce sujet est que la langue polonaise propose trop d'options pour que le choix du traducteur soit aléatoire ; toutes ces structures (au moins huit) ne sauraient être (strictement) équivalentes sémantiquement, en effet, sans compter que suivant notre corpus d'étude, en termes de fréquence, le taux de sortie de *jak P* (environ 50 %) est bien trop supérieur à celui des autres constructions (*gdy P*, *kiedy P*, *że P*, etc.), qui oscille entre 5 à 9 % maximum.

Nous avons choisi comme cadre théorique, en l'adaptant à l'ICP, le protocole traductologique établi par le Professeur Eugeniusz Ucherek (Université de Wrocław, Pologne) en 1982. Fixée, à l'origine, pour la description contrastive des prépositions du français et du polonais, cette procédure a déjà été reprise, par la suite, pour être appliquée avec succès à l'analyse confrontative d'autres constructions françaises. Inspirée,

notamment, des réflexions de Bausch (1970, 1971), de Perrot (1974) et de Bogusławski (1976), la méthode Ucherek se fonde sur la comparaison entre les textes originaux et leur(s) traduction(s). Il va sans dire qu'une telle analyse est à la fois orientée et unidirectionnelle, ce qui, lorsque la langue de départ (celle du texte) coïncide avec la langue maternelle de l'auteur, nous permet, ainsi qu'au lecteur, de suivre la direction des interférences. Quant à la traduction de textes originaux, deux grands cas d'école sont à distinguer : soit le travail est effectué par un chercheur (le but en est alors exclusivement métalinguistique), soit il est mené par un autre (type de) traducteur dans un autre but (c'est-à-dire non métalinguistique) et les produits ainsi obtenus fonctionnent alors comme de véritables textes dans l'univers culturel de la langue d'arrivée... et c'est précisément cette deuxième possibilité que propose de retenir Ucherek (1982) afin que soit garantie l'objectivité de l'analyse.

Chemin faisant (celui du projet Polonium), nous poursuivons deux objectifs principaux : établir la typologie exhaustive des constructions polonaises susceptibles de pouvoir transposer l'ICP, d'une part, et montrer, d'autre part, au moyen du protocole précité, quels types de contraintes (linguistiques) exactement pèsent sur les épaules du traducteur lors du passage au polonais. Et c'est précisément à la poursuite de ce dernier objectif que le présent article entend apporter sa contribution.

2.3. CORPUS D'ÉTUDES ET PROTOCOLE EXPÉRIMENTAL

À ce stade, notre corpus est à la fois en cours de prélèvement et d'analyse : plus précisément, 3 000 items ont d'ores et déjà été amassés (sur les 5 000 que nous escomptons pour l'étude complète) et 2 000 analysés. Dans la pratique, nous choisissons les œuvres à prélever de la sorte : les exemples ne doivent pas être construits mais attestés, les auteurs doivent être francophones natifs ou assimilés et ceux des transpositions polonophones natifs ou assimilés et divers types/genres de texte et registres de langue doivent être représentés. S'agissant, plus particulièrement, du mode de prélèvement, nous nous sommes fixé comme ligne de conduite d'extraire les items en contexte large, c'est-à-dire en retenant suffisamment de matière cotextuelle afin de prévenir toute difficulté de compréhension ou d'interprétation relativement à l'ICP.

S'agissant, précisément, de l'application du protocole Ucherek, chaque item est soumis à une série de trois questions, toujours les mêmes et toujours posées dans le même ordre.

1^{re} question

Le traducteur avait-il le choix entre différentes structures ?

Cette question n'est pas à polarité sémantique mais syntaxique. En d'autres termes, il s'agit de savoir si, au moment du passage vers le polonais, le site d'accueil en tant

que tel, c'est-à-dire le cotexte immédiat de la traduction en cours, offrait différentes options structurellement parlant. à ce stade, si la réponse est négative, c'est-à-dire s'il est jugé que le traducteur n'avait pas d'autre choix que celui constaté, l'item est alors étiqueté comme exemple de contrainte syntaxique (selon la terminologie Ucherek) puis classé. Si, au contraire, la réponse est positive, nous passons alors à la question n° 2.

2^e question

Les différentes structures à disposition sont-elles sémantiquement équivalentes ?

Pour le dire autrement, y a-t-il une différence de sens, si ténue soit-elle, entre les diverses structures possibles ? Si oui, c'est-à-dire s'il est estimé que le traducteur a opéré un choix sémantiquement motivé entre différentes structures envisageables, l'item est alors étiqueté comme exemple de contrainte sémantique (suivant la terminologie Ucherek) puis classé. Sinon, c'est-à-dire s'il est admis que le traducteur n'a pas opéré de choix sémantiquement motivé entre diverses structures possibles, l'item est alors étiqueté comme exemple de contrainte stylistique (d'après la terminologie Ucherek) puis analysé plus avant dans le cadre de la question n° 3.

3^e question

De quel type de contrainte stylistique s'agit-il ?

Précisons que la version originale du protocole Ucherek ne prévoit pas cette question avancée ; il s'agit d'un ajout, d'une adaptation de notre fait. Il nous paraît linguistiquement profitable, en effet, d'indiquer à quel(s) type(s) de facteur(s) précis la contrainte stylistique engagée est plus précisément due (au registre de langue, à la phonétique, à la prosodique, ... ?).

3. ÉTIQUETAGE DU CORPUS

On l'aura compris, l'ICP et ses transpositions possibles en polonais, tel est l'objet d'étude de notre Polonium.

Par ailleurs, la voie que nous avons décidé de suivre est celle, actuellement en vogue dans les sciences du langage, des grands corpus étiquetés. Ce type de recueils de données n'étant guère représenté s'agissant des structures et des langues qui nous intéressent, en effet, c'est cette friche-là que nous avons donc entrepris, en répondant à l'appel du Programme bilatéral Polonium de 2018, de rendre fertile. Et c'est ainsi que nous devrions bientôt livrer – normalement fin 2022 – à la communauté scientifique franco-polonophone un corpus attesté de 5 000 items marqués d'une estampille grammaticale de type alphanumérique.

S'agissant de l'éventail, du spectre du corpus, nous avons notamment prélevé, à ce jour, nos occurrences chez Amélie Nothomb, Georges Simenon, Marcel Proust, Philippe Delerm, Anna Gavalda, Gérard de Villiers, Jean Giono, Frédéric Dard, François Mauriac, Cizia Zykë, Daniel Pennac, Marc Levy, Michel Tournier, Françoise Sagan, Bernard Werber, Fred Vargas, Pierre Rey et bien d'autres¹.

Rappelons-le, l'objectif ultime de notre étiquetage bilingue est de préparer le terrain pour un prochain traitement robotisé de notre corpus, mise en corrélation automatisée des tags dont nous attendrons notamment qu'elle confirme ou infirme par la statistique les pratiques (des traducteurs lors du passage du français au polonais s'agissant des constructions régies par un verbe de perception) que nous avons déjà appréhendées en soumettant une grande partie de notre corpus au protocole Ucherek de manière naturelle (traitement par l'Homme).

En décrivant le mécanisme et les principes sous-tendant l'étiquetage de ses items, la présente section tâchera de montrer quels peuvent être la valeur, l'intérêt scientifique et les perspectives du grand corpus bilingue *DE PERCEPTIONE*.

En pratique, les estampilles alphanumériques des items de notre corpus s'établissent au fur et à mesure d'une suite immuable de questions (toujours les mêmes, toujours dans le même ordre), cinq en tout et que voici.²

1^{re} question

La CP est-elle autonome ?

La première question de notre protocole d'étiquetage porte sur l'incrustation de la construction de la perception dans son tissu syntaxique cotextuel.

En des termes plus grammaticaux, il s'agit de savoir si la CP est une proposition indépendante ou une structure dépendante quelle qu'elle soit, l'autonomie syntaxique caractérisant, il va sans dire, le premier cas de figure.

À ce stade, si la réponse est positive quant à l'autonomie, deux premières griffures (toujours les mêmes) intègrent l'étiquette en cours d'élaboration, à savoir celle représentant l'indépendance de la construction de la perception, notée : **CP^{+A}** (pour 'autonome'), et celle en représentant la nature et/ou fonction, notée : **[...] _{PI}** (pour 'proposition indépendante') :

- (6) (a) [FR] Plusieurs heures après, la délégation s'en alla. La voix tonitruante de l'énorme monsieur Omochi cria : – Saito-san ! Je *vis* monsieur Saito se lever d'un bond, devenir livide et courir dans l'antre du vice-président.

¹ Tous les exemples cités dans cette étude viennent du sous-corpus Amélie Nothomb.

² Dans les questions comme dans les tags à venir, le terme *CP* signifie 'construction de la perception'.

$CP^{+A} \supset ICP : [SN1^+ \cup V1^+ \textit{voir} \cup ICP^-(3V2)]_{PI}$

(b) [PL] Po kilku godzinach delegacja wyszła. Grzmiący głos olbrzymiego pana Omochi zakrzyknął: – Saito-san! Zobaczyłam, **jak** pan Saito zrywa się z miejsca, robi się trupioblady i biegnie do jaskini wiceprezesa.

jak P

Si la réponse est négative, en revanche, c'est-à-dire si la CP ne constitue pas une proposition indépendante mais une structure dépendante quelconque, deux autres griffures investissent alors le tag, à savoir celle représentant la dépendance de la construction de la perception, notée : CP^{-A} (pour 'non autonome'), et celle en représentant la nature et/ou fonction, variable quant à elle, par exemple notée : $[...]_{CA}$ (pour 'complément de l'adjectif') (7a), $[...]_{AS}$ (pour 'attribut du sujet') (8a), $[...]_{PP}$ (pour 'proposition principale') (9a : 1^{er} item), $[...]_{CN}$ (pour 'complément du nom') (9a : 2nd item) ou $[...]_{PIC}$ (pour 'proposition indépendante coordonnée') (10a) :

(7) (a) [FR] J'étais désolée de **voir** s'interrompre si vite une carrière prometteuse.

$CP^{-A} \supset ICP : [SN1^-(\emptyset) \cup V1^- \textit{voir}(I) \cup ICP^-(\square)]_{CA}$

(b) [PL] Byłam niepokieszona, **że** moja obiecująca kariera tak szybko się zakończyła.

że P

(8) (a) [FR] Le plus insupportable, c'était de **voir** mon bienfaiteur humilié par ma faute.

$CP^{-A} \supset PPé : [SN1^-(\emptyset) \cup V1^- \textit{voir}(I) \cup PPé^+]_{AS}$

(b) [PL] Największą przykrość sprawiało mi patrzeć, **jak** mój dobroczyńca jest z mojej winy upokarzany.

jak P

(9) (a) [FR] Je fus convoquée dans le bureau de monsieur Omochi : je m'y rendis sans la moindre appréhension, ignorant ce qu'il me voulait. Quand je pénétrai dans l'antre du vice-président, je **vis** monsieur Tenshi assis sur une chaise. [...]. Mais vint un moment où l'obèse lui dit : – Vous n'avez jamais eu d'autre but que de saboter la compagnie ! [...]. Je me jetai sous le flot grondant des cris du vice-président : – [...]. Je suis l'unique responsable. J'eus juste le temps de **voir** le regard effaré de mon compagnon d'infortune se tourner vers moi. Dans ses yeux, je lus : « Taisez-vous, par pitié ! » – hélas, trop tard.

$CP^{-A} \supset PPé : [SN1^+ \cup V1^+ \textit{voir} \cup PPé^+]_{PP}$

$CP^{-A} \supset ICP : [SN1^-(\emptyset) \cup V1^- \textit{voir}(I) \cup ICP^+]_{CN}$

(b) [PL] Zostałam wezwana do gabinetu pana Omochi: udałam się tam bez cienia lęku, nie wiedząc, czego ode mnie chce. Kiedy weszłam do jaski-

ni wiceprezesa, zobaczyłam *siedzącego* na krześle pana Tenshi. [...]. Aż nadszedł moment, kiedy grubas zagrzmiał: – Pana jedynym celem zawsze było sabotowanie firmy! [...]. Rzuciłam się w grzmiący strumień krzyków wiceprezesa: – [...]. Tylko ja jestem za to odpowiedzialna. Zdążyłam jedynie zobaczyć, *jak* mój towarzysz niedoli posyła mi wystraszone spojrzenie. W jego oczach wyczytałam: „Niech pani milczy, na litość boską!” – niestety, było już za późno.

Participe présent

jak P

- (10) (a) [FR] Je *vis* mon bienfaiteur fermer les yeux et je compris que je venais de prononcer l'irréparable.

$CP^{-A} \supset ICP : [SN1^+ \cup V1^+voir \cup ICP^+]_{PIC}$

- (b) [PL] Zobaczyłam, *jak* mój dobroczyńca zamyka oczy i pojęłam, że to, co powiedziałam, jest nie do naprawienia.

jak P

2^e question

Quel type de sous-CP la CP emporte-t-elle ?

La deuxième question de notre protocole de taguage est relative aux fondements de la construction de la perception.

En pratique, il s'agit de déterminer quel type de structure régie par un verbe de perception (désormais : sous-CP) emporte exactement la CP.

À ce stade, une troisième griffure (bipartite) investit le tag en cours, représentant à la fois l'agencement des ensembles (CP et sous-CP) engagés, noté : \supset (pour 'inclus/contient'), et la nature ou catégorie syntaxique de la sous-CP, variable quant à elle, notamment notée : **ICP** (pour 'infinitive de compte rendu de perception') (11a), **SP** (pour 'syntagme prépositionnel') (12a : 1^{er} item), **SN** (pour 'syntagme nominal') (12a : 2nd item), **RCP** (pour 'relative de compte rendu de perception') (13a), **PPé** (pour 'participe passé') (14a) ou **que P** (pour 'subordonnée conjonctive pure') (15a) :

- (11) (a) [FR] Quelques minutes plus tard, je l'*entendis* s'exclamer : – Je n'y crois pas ! Je n'y crois pas !

$CP^{+A} \supset ICP : [SN1^+ \cup V1^+entendre \cup ICP^+]_{PI}$

- (b) [PL] (b) Po kilku minutach usłyszałam, *jak* krzyczy: – To nie do wiary! To nie do wiary!

jak P

- (12) (a) [FR] Du haut de ma tour de Babel, je *regarde* vers le parc d'Ueno et je *vois* des arbres enneigés : des cerisiers en fleur – oui, ce doit être Pâques.

$CP^{-A} \supset SP \rightarrow [SN1^+ \cup V1^+regarder \cup SP^+]_{PIC}$

$CP^{-A} \supset SN \rightarrow [SN1^+ \cup V1^+voir \cup SN2^-(R)]_{PIC}$

- (b) [PL] Z wyżyn mojej wieży Babel patrzę **na** park Ueno i widzę ośnieżone *drzewa*: kwitnące wiśnie – tak, musi być Wielkanoc.

Syntagme prépositionnel

Syntagme nominal

- (13)(a) [FR] La porte du bureau de l'énorme Omochi s'ouvrit et j'**entendis** la voix de l'infâme qui me hurlait : – Qu'est-ce que vous fichez là ? On ne vous paie pas pour traîner dans les couloirs !

$CP^A \supset RCP \rightarrow [SN1^+ \cup V1^+entendre \cup RCP^+]_{PIC}$

- (b) [PL] Drzwi gabinetu monstrualnego Omochi otworzyły się i rozległ się wrzask niegodziwca: – Co pani tu jeszcze robi? Nie płacę pani za wystawianie w korytarzu!

Traduction par contournement/évitement

- (14)(a) [FR] Il y a quelque temps, je vous en ai voulu, c'est exact, et cependant, quand je vous **ai vue** si bassement humiliée, il n'y a plus eu de place en moi que pour la compassion primitive.

$CP^A \supset PPé : [SN1^+ \cup V1^+voir \cup PPé^+]_{CCT}$

- (b) [PL] Jakiś czas temu miałam do pani żal, to prawda, ale kiedy zobaczyłam panią tak niegodziwie **upokorzoną**, ogarnęło mnie głębokie współczucie.

Participe passé

- (15)(a) [FR] Il m'attrapa comme King Kong s'empare de la blondinette et m'entraîna à l'extérieur. J'étais un jouet entre ses bras. Ma peur atteignit son comble quand je **vis** qu'il m'emportait aux toilettes des messieurs. Me revinrent à l'esprit les menaces de Fubuki : « Vous ne savez pas ce qui pourrait vous arriver ».

$CP^A \supset que P : [SN1^+ \cup V1^+voir \cup que P^+]_{CCT}$

- (b) [PL] Złapał mnie jak King Kong blondyneczkę i wyciągnął na korytarz. Byłam zabawką w jego rękach. Mój strach sięgał szczytu, kiedy zorientowałam się, **że** wlecze mnie do męskiej toalety. Przypomniały mi się groźby Fubuki: „Nie wie pani, co może panią spotkać”.

że P

3^e question

Quelle est l'architecture de la CP ?

La troisième question de notre procédure de bombage vise la structuration de la construction de la perception.

Plus précisément, il s'agit de réécrire la CP suivant les codes syntagmatique et paradigmaticque qui la sous-tendent, transcription donnant à voir les différentes catégories syntaxiques programmées et leur distribution fondamentale.

À ce stade, un quatrième graffiti (complexe et variable), respectivement marqué du double point (indiquant que la réécriture suit immédiatement), des crochets (indi-

quant le début et la fin de celle-ci) et du symbole \cup (pour ‘union/réunion’), fait son entrée dans l’estampille en cours, où apparaissent alors, dans l’ordre linéaire de base, les étiquettes catégorielles des paradigmes successifs.

Aussi lisons-nous – au stade de la 3^e question de notre protocole d’étiquetage – les tags de (16a) comme suit :

- (16)(a) [FR] Par malheur, il faisait très chaud. Le Hollandais avait tombé la veste et sa chemise arborait aux aisselles de vastes auréoles de sueur. Je *vis* Fubuki changer de figure. Elle s’efforça de parler normalement, comme si elle ne *s’était aperçue* de rien.

$CP^{+A} \supset ICP : [SN1^+ \cup V1^+voir \cup ICP^+]_{PI}$

$CP^{-A} \supset SP : [SN1^+ \cup -V1^+s’apercevoir \cup SP^+]_{CCCondition}$

- (b) [PL] Na nieszczęście było bardzo gorąco. Holender zdjął marynarkę i pod pachami jego koszuli dały się dostrzec rozległe plamy potu. Zobaczyłam, *jak* Fubuki zmienia się na twarzy. Starła się mówić normalnie, *jakby* niczego nie zauważyła.

jak P

jakby P

1^{er} item :

- la CP est autonome (qui constitue grammaticalement une proposition indépendante) et repose sur une structure régie par un verbe de perception de type ICP ;
- l’architecture de la CP répond à ce schéma : syntagme nominal sujet + verbe de perception *voir* + infinitive de compte rendu de perception ;

2nd item :

- la CP n’est pas autonome (qui occupe grammaticalement la fonction de complément circonstanciel de condition) et repose sur une structure régie par un verbe de perception de type SP ;
- l’architecture de la CP répond à ce schéma : syntagme nominal sujet + verbe de perception *s’apercevoir* à la forme négative + syntagme prépositionnel.

Reste à savoir ce que symbolisent les exposants (en l’occurrence des plus) accolés aux étiquettes catégorielles SN1, V1, ICP et SP, ce qui constitue précisément l’objet de l’ultime question de notre procédure de taguage.

4^e question

Les catégories constitutives de la CP sont-elles prototypiques ?

La quatrième et dernière question de notre protocole d’étiquetage interroge la prototypicité de la construction de la perception.

Plus précisément, il s'agit de voir comment les catégories en question sontinstanciées en discours : de manière prototypique ou plus ou moins éloignée de leur prototype ? Sachant que nous considérons comme prototypique une instantiation où la catégorie est remplie comme il faut et où il faut, et inversement ; dans la phrase *Une pomme, Pierre en a déjà mangé une il y a cinq minutes*, par exemple, SN2 (que nous avons souligné), à la fois disloqué (ou thématisé) et quantifié, s'éloigne sensiblement de son prototype, que l'on trouverait dans *Pierre a mangé une pomme*.

S'agissant de la prototypicité, la règle du jeu est simple : une instantiation prototypique de catégorie sera affublée du signe + en exposant, comme cela est respectivement le cas des catégories SN1, V1, ICP et SP de (16a), et une instantiation non prototypique du signe –, comme cela est respectivement le cas des catégories SN1, V1 et ICP dans l'exemple suivant :

- (17)(a) [FR] Fus-je la seule à *entendre* s'élever une frêle voix, une voix de fillette de huit ans, qui gémit par deux fois : – Okoruna. Okoruna.

$CP^A \supset ICP : [SN1^-(\emptyset) \cup V1^-entendre(I) \cup ICP^-(\mathbb{N})]_{CA}$

- (b) [PL] A może to tylko ja usłyszałam słaby *głosik*, głosik ośmioletniej dziewczynki, która dwukrotnie zajęczała : – Okoruna. Okoruna.

Syntaxme nominal

De plus, le tag indique entre parenthèses la raison de la non-prototypicité de (l'instanciation de) la catégorie : en l'occurrence, SN1 n'est pas prototypique car n'est pas instancié en discours (ce que symbolise l'élément vide : \emptyset), V1 ne l'est pas car n'est pas conjugué mais à l'infinitif (d'où la lettre *i* en capitales : **I**) et ICP ne l'est pas car l'infinitif y précède SN2 (d'où le signe de l'inversion ou permutation : \mathbb{N}). Bien entendu, les facteurs susceptibles de sanctionner la non-prototypicité de (l'instanciation de) la catégorie sont nombreux et divers ; en voici deux autres exemples pour en clore provisoirement la liste :

- V1 est un infinitif modalisé (notation : **IM**)

- (18)(a) [FR] Par exemple, même quand tu seras isolée aux toilettes pour l'humble besoin de soulager ta vessie, tu auras l'obligation de veiller à ce que personne ne puisse *entendre* la chansonnette de ton ruisseau : tu devras donc tirer la chasse sans trêve.

$CP^A \supset SN : [SN1^-(\emptyset) \cup -V1^-entendre(IM) \cup SN2^+]_{COI}$

- (b) [PL] Na przykład kiedy znajdziesz się na osobności w toalecie z przyziemnego powodu ulżenia pęcherzowi, musisz czuwać nad tym, *by* nikt nie usłyszał ciurkania twojej stróżki: będziesz zatem zmuszona bez przerwy spuszczać wodę.

by P

– l'ICP contient deux verbes à l'infinitif (notation : **2V2**)

(19)(a) [FR] Certes, personne ne m'*avait vue* parcourir les bureaux toute nue, en marchant sur les mains, ni rouler un patin à un honnête ordinateur. Mais on m'*avait* quand même retrouvée endormie sous le contenu de la poubelle. Dans d'autres pays, on m'eût peut-être mise à la porte pour ce genre de comportement.

$CP^{+A} \supset ICP : [SN1^+ \cup -V1^+ \textit{voir} \cup ICP^-(2V2)]_{P1}$

(b) [PL] Wprawdzie nikt nie widział, *jak* na golasa, chodząc na rękach, przemieszczałam się po biurze, ani jak obcałowywałam porządny komputer, mimo to jednak znaleziono mnie śpiącą pod zawartością kosza na śmieci. W innym kraju pewnie wyrzucono by mnie z pracy za takie zachowanie.

jak P

4. CONCLUSION

S'agissant du caractère innovant de notre recherche, il s'exprimera sans doute le plus nettement dans les points suivants : (a) les acteurs du projet appartiennent à cinq corps professionnels distincts mais complémentaires en l'occurrence, à savoir la recherche, l'enseignement, la traduction-interprétation, l'informatique et l'édition (☞ synergie des compétences) ; (b) s'il s'agit, plus particulièrement, des domaines relatifs aux sciences du langage, les compétences des mêmes acteurs en couvrent au moins une dizaine, de la syntaxe à la traduction-traductologie en passant par la sémantique, la pragmatique, la cognition, le traitement automatique des langues, la sociolinguistique, la grammaire, l'orthographe, le français pratique ou la didactique (☞ pluridisciplinarité ; spectre de compétences) ; (c) pour ce qui est de l'analyse du corpus d'étude en elle-même, celle-ci est progressive, méthodique et systématique, d'une part, et s'appuie sur un protocole de traduction préexistant et ayant déjà fait ses preuves en la matière, d'autre part (☞ économie, cohérence, optimisation de la recherche) ; (d) ce projet nous semble également innover quant aux (deux familles de) langues qu'il met en contraste : car, si le français a régulièrement été comparé par le passé à l'espagnol ou l'italien (langues romanes) ou à l'anglais ou l'allemand (langues germaniques), il l'a beaucoup moins été, en effet, avec le polonais, le slovaque ou le serbo-croate, par exemple (☞ langues slaves) ; (e) l'aspect le plus innovant de ce projet étant certainement que celui-ci ne se veut pas une réflexion théorique sur telle ou telle méthode de traduction (traductologie) mais consiste, à l'inverse, dans une démarche méthodique d'observation-déduction-description scrupuleuse et systématique de la pratique effective de traducteurs modernes et contemporains dans leur passage du français (langue non maternelle) à une langue cible maternelle donnée (☞ pratiques de traduction ; linguistique de la traduction).

Si les acteurs du projet s'accordent tous à y voir une recherche d'importance dans le domaine des sciences du langage, c'est notamment parce que : (a) le projet constituera un témoignage anthropologique notable qui offrira un ensemble de productions attestées (exemples d'auteur) et important (5.000 paires d'items, sans compter celles des corpus témoins à suivre) d'un fait de langue particulier (les constructions relatives à la perception) saisi à un moment donné de son évolution (le français moderne et contemporain, le polonais moderne et contemporain) ; (b) le projet constituera un témoignage syntaxique important, également, puisqu'il donnera à voir et à appréhender les constructions de la perception dans toute l'étendue de leurs architectures et de leurs fonctions, d'une part, et de leur distribution, d'autre part ; (c) étant donné qu'elle renfermera les différentes configurations de sens inhérentes aux structures de la perception (point de vue sensoriel *vs* épistémique, aspects lexical et verbal, portée de la négation, etc.), l'entreprise constituera de plus un témoignage sémantique d'envergure ; (d) sans compter qu'elle incarnera de fait une véritable grammaire pratique de l'accord – notamment du participe passé – dans le cadre desdites constructions (témoignage orthographique) ; (e) l'étude donnera lieu aussi à un étiquetage très avancé de l'ensemble des structures de la perception, lequel, sous la forme d'un encodage tridimensionnel, affichera de manière synthétique la structure, la catégorie et la fonction simultanément engagées (☞ tags poly-informationnels) ; (f) sans compter qu'elle engendrera une base de données électronique considérable sur lesdites constructions, dont l'interface, en libre accès, permettra aux spécialistes (mais aussi aux amateurs, nous y veillerons) de lancer des requêtes nombreuses et diverses (☞ linguistique de corpus), d'établir les taux de sortie de constructions ou de termes donnés (☞ linguistique statistique), de contre-vérifier les résultats issus de l'application du protocole Ucherek (☞ linguistique informatique) et certainement de nombreuses autres possibilités encore ; (g) le français pratique, le FLE et la didactique gagneront également beaucoup à voir ce projet aboutir, notamment lors de la parution des triptyques didactiques prévus. Ainsi la dernière étape convoquera-telle une équipe de spécialistes de linguistique appliquée, de français pratique, de grammaire et de didactique afin qu'ils scrutent les enseignements linguistiques issus du projet, voient lesquels pourraient être didactisés de manière utile et en initient précisément la didactisation. Les travaux ainsi obtenus donneront lieu à la publication de triptyques didactiques consacrés à l'expression de la perception dans le cadre de l'enseignement du français.

Une fois le corpus tagué de manière tridimensionnelle, l'avant dernier mouvement du projet donnera lieu à la création d'une base données et de son interface d'utilisation. En pratique, le grand corpus tagué (français-polonais) sera intégralement numérisé pour être déposé en ligne (sur des serveurs publics nationaux afin d'en assurer la pérennité et le libre accès à la communauté d'utilisateurs). L'équipe d'informaticiens en charge de cette étape du projet entrera ensuite en collaboration étroite avec des linguistes spécialistes des structures de la perception en vue de la création d'un logiciel et une interface de requêtes linguistiques et statistiques à destination des utilisateurs de la nouvelle base de données.

BIBLIOGRAPHIE

- Bausch, K.-R. (1970). Qualité en traduction et linguistique dite « différentielle ». *Babel*, XVI, 13-20.
- Bausch, K.-R. (1971). Linguistique comparative, linguistique appliquée et traduction. *Meta : Journal des traducteurs*, 16 (1-2), 45-55.
- Bogusławski, A. (1976). Problem *tertium comparationis* w porównaniu lingwistycznym. *Kwartalnik Neofilologiczny*, XXIII (3), 295-300.
- Marsac, F. (2006). *Les constructions infinitives régies par un verbe de perception* (Thèse de doctorat NR, version originale). Université Lille 3 : Atelier National de Reproduction des Thèses – ANRT.
- Marsac, F., Ucherek, W. & Dańko, M. (2019). De l'infinitive de perception dans la pratique traductologique. *Studia Romanica Posnaniensia*, 46/1, 115-136.
- Perrot, J. (1974). Problèmes méthodologiques en description contrastive. *Études finno-ougriennes*, XI, 219-227.
- Ucherek, E. (1982). Pour une description des prépositions susceptible d'applications en didactique des langues étrangères et en lexicographie bilingue. *Beitraege zur Romanischen Philologie*, XXI, 169-175.