

Michał Hrabia

ITÉRATIVITÉ

DANS LA TRADUCTION AUTOMATIQUE

ANALYSE ORIENTÉE OBJETS
DES VERBES DE MOUVEMENT FRANÇAIS
TRADUITS EN POLONAIS PAR *JECHAĆ* / *JEŹDZIĆ*
(SUR LA BASE DES FORMES DU PRÉSENT)



WYDAWNICTWO
UNIwersytetu śląskiego
KATOWICE 2016

Itérativité
dans la traduction automatique

Analyse orientée objets
des verbes de mouvement français
traduits en polonais par *jechać / jeździć*
(sur la base des formes du présent)



NR 3487

Michał Hrabia

**Itérativité
dans la traduction automatique**

**Analyse orientée objets
des verbes de mouvement français
traduits en polonais par *jechać* / *jeździć*
(sur la base des formes du présent)**

Redaktor serii: Językoznawstwo Neofilologiczne
Maria Wysocka

Recenzent
Grażyna Vetulani

Table des matières

Introduction	11
------------------------	----

PARTIE I

Aspect : une question toujours actuelle

1. Questions préliminaires	17
2. Système aspectuel en polonais	19
3. À l'aube de l'aspectologie cognitive : la classification de Zeno Vendler	22
3.1. Schèmes temporels	22
3.2. Catégories verbales	23
4. Sémantique de l'aspect verbal : l'approche de Francesco Antinucci et Lucy- na Gebert	26
4.1. Structures sémantiques des verbes	26
4.2. Classes de verbes et aspect	31
4.3. Verbes de mouvement	34
5. Modèle topologique de l'aspect de Jean-Pierre Desclés	37
5.1. De la relation prédicative à l'énoncé	37
5.2. Intervalles topologiques de validation	38
5.3. État, événement et processus	39
5.3.1. État	40
5.3.2. Événement	40
5.3.3. Processus	42
5.4. « Aspect lexical »	45
5.5. Aspect et modalités d'action	46
5.5.1. Modalités d'action en français	46
5.5.2. Schéma énonciatif	48
5.5.3. Itérativité	49
6. Aspect dans la grammaire à base sémantique de Stanisław Karolak	51
6.1. Composante catégorielle	51

6.1.1. Concept et expression	51
6.1.2. Concepts simples et complexes	52
6.2. Composante combinatoire	53
6.2.1. Combinatoire des concepts	53
6.2.1.1. Structure prédicat-argument	54
6.2.1.2. Formes logiques et propositions	55
6.2.1.3. Contraintes sur les positions d'argument	56
6.2.2. Combinatoire des expressions	57
6.3. Temps et modalité	58
6.3.1. Caractéristique temporelle de la proposition	58
6.3.2. Modalité	59
6.4. Structure thème-rhème	60
6.5. Questions aspectuelles	61
6.5.1. Contre les conceptions classiques	61
6.5.2. « Porteurs » de l'aspect	62
6.5.3. Aspects simples et configurations d'aspects	63
6.5.3.1. Valeur aspectuelle des infinitifs français	63
6.5.3.2. Valeur aspectuelle des <i>verba finita</i> en français	65
6.5.3.3. Valeur aspectuelle des verbes polonais	68
7. Itérativité : essai de synthèse et quelques remarques supplémentaires	71

PARTIE II

Vers l'avenir de la linguistique : lexicographie computationnelle

1. Questions préliminaires	75
2. Théorie Sens-Texte d'Igor Mel'čuk	77
2.1. Trois postulats de la théorie	77
2.2. Représentations linguistiques	78
2.2.1. Niveaux de représentation	78
2.2.2. Formalismes de représentation	79
2.3. Trois particularités du modèle	81
2.4. Structure fonctionnelle du modèle	81
2.4.1. Composante déclarative	82
2.4.1.1. Grammaire	82
2.4.1.2. Dictionnaire	83
2.4.2. Composante procédurale	85
2.5. Fonctions lexicales	85
2.5.1. Concept de fonction lexicale	86
2.5.2. Fonctions lexicales standard	86
2.5.3. Fonctions lexicales paradigmatiques et syntagmatiques	87
2.5.4. Fonctions lexicales dans la traduction automatique	89
3. Classes d'objets et dictionnaire électronique de Gaston Gross	91
3.1. Unité minimale d'analyse	91
3.2. Comment décrire les prédicats ?	92
3.3. Classes d'objets	94

3.3.1. Propriétés des classes d'objets	94
3.3.2. Applications des classes d'objets	96
3.4. Dictionnaire électronique	97
3.4.1. Description des substantifs arguments	97
3.4.2. Description des prédicats verbaux	98
3.4.3. Description des prédicats nominaux	99
3.4.4. Exemples de descriptions	101
4. Approche orientée objets de Wiesław Banyś	103
4.1. Extensibilité et réutilisabilité	103
4.2. Architecture modulaire	104
4.2.1. Décomposabilité modulaire	105
4.2.2. Composabilité modulaire	105
4.2.3. Compréhensibilité modulaire	106
4.2.4. Continuité modulaire	107
4.3. Description de la langue : fondements théoriques	107
4.3.1. Prédicats ou arguments ?	107
4.3.2. Attributs et opérations	108
4.3.3. Classes d'objets et leur hiérarchie	109
4.3.4. Opérations (fonctions)	111
4.3.5. Cadres (<i>frames</i>) et scénarios (<i>scripts</i>).	112
4.4. Schéma de la description	114
4.5. Fonctionnement du système	116
4.6. Approche orientée objets et d'autres modèles linguistiques	117
4.6.1. Classes d'objets de Gaston Gross	117
4.6.2. <i>Qualia structure</i> de James Pustejovsky	117
4.6.3. WordNet	118

PARTIE III

Analyse orientée objets des verbes de mouvement

1. Questions préliminaires	123
1.1. Verbes de mouvement français potentiellement itératifs	123
1.2. Étapes de l'analyse orientée objets	123
2. Schémas syntaxico-sémantiques des emplois des verbes de mouvement	125
2.1. [ANM <hum>] en position de sujet	127
2.1.1. <i>Aller</i>	127
Section A ₁	127
Section B ₁	129
Section C ₁	133
Section D ₁	136
Section E ₁	139
Section F ₁	140
Section G ₁	142
Section H ₁	145
Section I ₁	145
Section J ₁	147

Section K_1	149
Section L_1	150
Section M_1	152
Section N_1	154
Section O_1	155
Section P_1	157
Section Q_1	158
Section R_1	159
Section S_1	160
2.1.2. <i>Rouler</i>	162
Section A_2	162
Section B_2	163
Section C_2	164
Section D_2	166
Section E_2	167
Section F_2	168
Section G_2	170
Section H_2	171
Section I_2	172
Section J_2	173
Section K_2	176
Section L_2	176
Section M_2	147
Section N_2	177
Section O_2	178
2.1.3. <i>Circuler</i>	179
Section A_3	179
Section B_3	180
Section C_3	181
2.1.4. <i>Conduire</i>	181
Section A_4	182
Section B_4	183
2.1.5. <i>Prendre</i>	184
Section A_5	184
Section B_5	185
Section C_5	186
Section D_5	186
Section E_5	187
Section F_5	187
Section G_5	188
2.2. [CONC <moyen de transport>] en position de sujet	188
2.2.1. <i>Aller</i>	188
Section α_1	189
Section β_1	190
Section γ_1	190
Section δ_1	191
Section ε_1	192
Section ζ_1	192
Section η_1	193

2.2.2. <i>Rouler</i>	194
Section α_2	194
Section β_2	195
Section γ_2	196
Section δ_2	198
Section ε_2	200
Section ζ_2	200
Section η_2	201
Section θ_2	202
Section ι_2	203
Section κ_2	204
2.2.3. <i>Circular</i>	205
Section α_3	205
Section β_3	207
Section γ_3	209
Section δ_3	210
Section ε_3	211
2.3. Inventaire des schémas syntaxico-sémantiques	212
2.4. Classes d'objets employées dans la désambiguïsation	228
En guise de récapitulation	237
Références citées	239
Streszczenie	245
Summary	246

Introduction

La langue de l'Europe, c'est la traduction.
Umberto Eco

L'Union Européenne compte actuellement vingt-quatre langues officielles. Un véritable creuset linguistique, dirait-on. Or, on ne peut pas oublier que c'est effectivement dans cette diversité que réside notre identité culturelle en tant qu'Européens. Il est donc fort souhaitable que la richesse linguistique, porteuse du grand héritage culturel de l'Europe, soit précieusement préservée dans le cadre de l'intégration. Mais est-ce que les citoyens de la Communauté peuvent bâtir l'Europe unie sans se comprendre mutuellement? Sûrement pas. Ainsi, c'est la traduction qui devient le seul moyen possible de communication, la seule « langue » commune de tous les Européens.

Cette « langue » de traduction remplit dans la société une double fonction : d'un côté, elle est moyen d'expression artistique (traduction littéraire), et de l'autre — moyen de communication quotidienne ou professionnelle (traduction non littéraire). Ces deux types de traduction se distinguent l'un de l'autre d'une manière prononcée. Il est hors de doute que la traduction littéraire est un art et en tant que tel, elle exige que l'esprit humain pénètre profondément le texte source afin de s'imprégner de toutes ses subtilités et ses nuances de façon à pouvoir le réécrire (et non pas traduire) dans la langue cible. En revanche, la traduction non littéraire se contente grosso modo de garantir l'équivalence sémantique entre le texte dans la langue source et le texte dans la langue d'arrivée. Il est évident que cette équivalence peut être obtenue, tout comme dans le cas de la traduction littéraire, uniquement grâce aux efforts de la pensée humaine. Toutefois, il est possible (ou plutôt recommandé) de faciliter le travail du traducteur des textes non littéraires en mettant à sa disposition des outils informatiques tels que les logiciels de traduction assistée par ordinateur (p.ex. *Trados*, *MemoQ*, *Wordfast*, *Déjà Vu*), voire même des traducteurs automatiques (p.ex. *Google Traduction*, *Systran*).

De nos jours, l'emploi de l'ordinateur dans le domaine de traduction, ou plus généralement dans la linguistique, est une pratique ordinaire. En effet, déjà en 1962, on a fondé *The Association for Machine Translation and Computational Linguistics* dont le but majeur était d'œuvrer pour le développement des méthodologies et des recherches linguistiques mettant à profit la puissance de calcul de l'ordinateur. Cette date est considérée comme une date symbolique qui témoigne de la naissance d'une nouvelle branche de la linguistique appliquée — linguistique informatique (computationnelle) — un «hybride» se situant à la frontière entre l'informatique et les sciences du langage.

Au XXI^e siècle, la linguistique informatique connaît son plus grand essor. Les recherches se focalisent sur des questions telles que : la linguistique de corpus, la lexicographie computationnelle, le traitement automatique des langues, la traduction automatique ou la synthèse vocale (cf. D. SŁAPEK, A. CHRUPAŁA, 2010). L'éventail des modes et des «lieux» d'application de cette nouvelle branche scientifique est tellement large qu'il va de soi que la linguistique informatique constitue le présent et, sans aucun doute, l'avenir des sciences linguistiques.

Par le présent travail nous voudrions contribuer à la formation de cet avenir. Nous nous occuperons de la traduction automatique du français vers le polonais en nous concentrant sur la question de l'itérativité des verbes de mouvement, qui n'a jusqu'à présent pas été résolue de manière définitive et satisfaisante. En effet, il est indéniable (et facile à prouver) que le choix entre la forme sémelfactive et itérative (c'est-à-dire le choix entre *jechać* et *jeździć*, *isć* et *chodzić*, etc.), tout naturel pour un natif polonais, constitue toujours une véritable difficulté pour les traducteurs automatiques contemporains.

L'objet de notre livre est donc d'étudier le phénomène de l'itérativité en vue du perfectionnement de la traduction automatique. Étant donné que nous considérons l'itérativité comme un phénomène aspectuel, nous commencerons notre travail par la présentation de quelques conceptions de l'aspect verbal. Nous focaliserons notre attention essentiellement sur les théories onomasiologiques, et ce en raison du fait que celles-ci nous semblent mieux adaptées aux études comparatives.

Dans la deuxième partie, nous aborderons la question de la traduction automatique. Nous y détaillerons trois méthodologies de la lexicographie computationnelle centrées sur la sémantique. L'une d'elles — l'approche orientée objets de Wiesław Banyś — servira de méthode d'analyse de l'itérativité dans la suite de ce travail.

La troisième partie, foncièrement pratique, se consacrera à l'étude du problème de l'itérativité, appliquée à un cas particulier : celui des verbes de mouvement traduits en polonais par *jechać* ou *jeździć*, et envisagée dans le cadre de l'approche orientée objets. L'analyse présentée sera la première description opérationnelle de l'itérativité créée exclusivement à des fins de traduction automatique de textes.

Par la méthode d'analyse choisie, notre travail s'inscrit tout naturellement dans le courant des recherches linguistiques menées depuis des années dans l'Institut des Langues Romanes et de Traduction à l'Université de Silésie. Il s'en distingue pourtant par son orientation nettement aspectologique. En effet, en règle générale, les travaux de désambiguïsation « traditionnelle » effectués dans le cadre de l'approche orientée objets ne prennent pas en considération les différences de nature aspectuelle (cf. p.ex. K. GABRYSIAK, 2009). Toutefois, il serait bien hâtif de constater que la désambiguïsation orientée objets est un outil insuffisant (ou bien inapproprié) pour lever l'ambiguïté du type *jechać / jeździć*, puisque, à y regarder de plus près, il apparaît qu'il suffit seulement de modifier légèrement la méthode en question pour qu'elle devienne parfaitement adaptée aux besoins du traitement de l'aspect. Plus précisément, il suffit d'admettre *a priori* que certains verbes français (tels que *aller*, *rouler* ou *circuler*) sont susceptibles d'être traduits en polonais par *jechać* ou *jeździć* et de restreindre la désambiguïsation à ces deux emplois uniquement. Il sera ainsi possible de porter plus d'attention aux conditions linguistiques responsables du choix de la forme sémelfactive ou itérative en polonais. Une fois ces conditions trouvées, elles pourront être par la suite convenablement schématisées à l'aide des dispositifs théoriques proposés par l'approche orientée objets, de façon qu'elles soient informatiquement implémentables, c'est-à-dire compréhensibles pour la machine.

Conscient de l'ampleur du travail à exécuter, nous espérons pouvoir atteindre, ne serait-ce que partiellement, l'objectif fixé et contribuer ainsi à un réel enrichissement des capacités des traducteurs automatiques.

Michał Hrabia

Iteratywność w przekładzie automatycznym
Zorientowana obiektowo analiza francuskich czasowników ruchu
tłumaczonych na język polski przez *jechać* / *jeździć*
(na materiale form czasu teraźniejszego)

Streszczenie

Przedmiotem niniejszej rozprawy jest zorientowany obiektowo opis złożonego problemu iteratywności czasowników ruchu dla celów przekładu automatycznego z języka francuskiego na język polski.

Praca składa się z trzech zasadniczych części, z których dwie początkowe mają charakter opisowo-teoretyczny. Z racji tego, że iteratywność jest postrzegana przez autora jako uniwersalne zjawisko aspektologiczne, pierwsza część w całości poświęcona jest wybranym onomazjologicznym teoriom aspektu czasownikowego, a mianowicie: klasyfikacji czasowników według Zeno Vendlera, semantycznej teorii aspektu Francesca Antinucciego i Lucyny Gebert, modelowi topologicznemu Jean-Pierre'a Desclés'go oraz teorii aspektu w ujęciu Stanisława Karolaka (przedstawionej na tle modelu gramatyki o podstawach semantycznych). W części drugiej opisane są ukierunkowane semantycznie teorie leksykografii komputacyjnej, mogące znaleźć zastosowanie w opisie aspektu, czyli: model Sens-Tekst Igora Mielczuka, teoria klas obiektowych Gastona Grossa i ujęcie zorientowane obiektowo Wiesława Banysia.

Część trzecia rozprawy ma charakter praktyczny. Przedstawione są w niej wyniki zastosowania metody zorientowanej obiektowo do opisu dezambiguacyjnego francuskich czasowników ruchu tłumaczonych na język polski przez *jechać* lub *jeździć* (*aller, rouler, circuler, conduire, prendre*). Opierając się na wnikliwej analizie konkretnych przykładów użyć zaczerpniętych z obszernego korpusu, autor konstruuje 38 schematów syntaktyczno-semantycznych warunkujących tłumaczenie badanych czasowników za pomocą formy semelfaktywnej *jechać* i 40 schematów dla wariantu iteratywnego *jeździć*. Pracę zamyka wyliczenie przykładowych elementów tworzących wykorzystane w dezambiguacji klasy obiektowe.

Michał Hrabia

Iterativity in the Machine Translation
Object-oriented Analysis of French Verbs of Movement
Translated into Polish as *Jechać / Jeździć*
(on the Basis of Present Tense Forms)

Summary

The subject of the following study is the object-oriented description of the complex problem of movement verbs' iterativity for the purpose of the automated French-Polish translation.

The study consists of three fundamental parts, where the two first are theoretical and descriptive. Since the author perceives iterativity as a universal aspectological phenomenon, the first part is entirely devoted to the chosen onomasiological theories of verb aspect: Zeno Vendler's verb classification, Francesco Antinucci's and Lucyna Gebert's semantic theory of aspect, the topological model by Jean-Pierre Desclés and the theory of aspect as seen by Stanisław Karolak (presented on the backdrop of the semantic-based grammar model). The second part is focused on semantically-directed theories of computational lexicography, which can be applied in the aspect description: the Meaning-Text model by Igor Mel'čuk, Gaston Gross's object classes' theory and Wiesław Banyś's object-oriented approach.

The third part concentrates on practice. It is a presentation of results of the object-oriented approach method application in the disambiguation of the French verbs of movement translated into Polish as *jechać* or *jeździć* (*aller, rouler, circuler, conduire, prendre*). The author, founding his study on a detailed analysis of concrete examples of possible uses derived from an extensive corpus, creates thirty-eight syntactic-semantic schemes conditioning the translation of the analysed verbs by the semelfactive form (*jechać*) and forty schemes for the iterative form (*jeździć*). The study also includes a listing of exemplary elements of the object classes used in the disambiguation.

Redakcja
ALEKSANDRA MAŃKA-CHMURA

Projekt okładki
AGATA AUGUSTYNIK

Redakcja techniczna
BARBARA ARENHÖVEL

Korekta
WIESŁAWA PISKOR

Łamanie
ALICJA ZAŁĘCKA

Copyright © 2016 by
Wydawnictwo Uniwersytetu Śląskiego
Wszelkie prawa zastrzeżone

ISSN 0208-6336
ISBN 978-83-226-3000-6
(wersja drukowana)
ISBN 978-83-226-3001-3
(wersja elektroniczna)

Wydawca
Wydawnictwo Uniwersytetu Śląskiego
ul. Bankowa 12B, 40-007 Katowice
www.wydawnictwo.us.edu.pl
e-mail: wydawus@us.edu.pl

Wydanie I. Ark. druk. 15,5. Ark. wyd. 20,0. Papier
offset. kl. III, 90 g Cena 20 zł (+ VAT)

Druk i oprawa:
Druk i oprawa: „TOTEM.COM.PL Sp. z o.o.” Sp.K.
ul. Jacewska 89, 88-100 Inowrocław



Więcej o książce



CENA 20 ZŁ
(+VAT)

ISSN 0208-6336
ISBN 978-83-226-3001-3