

Article

« La traduction multilingue des noms propres dans PROLEX »

Claire Agafonov, Thierry Grass, Denis Maurel, Nathalie Rossi-Gensane and Agata Savary
Meta : journal des traducteurs / Meta: Translators' Journal, vol. 51, n° 4, 2006, p. 622-636.

To cite the digital version of this article, use the following address:

<http://id.erudit.org/iderudit/014330ar>

Notice: citation formatting rules may vary according to different fields of knowledge.

This document is subject to copyright. All services operated by Érudit available for your use are also subject to the terms and conditions set forth in this document <http://www.erudit.org/documentation/eruditUserPolicy.pdf>

Érudit is a non-profit multi-institutional publishing consortium comprising the Université de Montréal, the Université Laval and the Université du Québec à Montréal. Its mission is to produce and disseminate scholarly journals. Érudit offers digital publishing services for scientific journals since 1998.

To contact the Érudit team : erudit@umontreal.ca

La traduction multilingue des noms propres dans PROLEX

CLAIRE AGAFONOV, THIERRY GRASS, DENIS MAUREL,
NATHALIE ROSSI-GENSANE, AGATA SAVARY

Université de Tours, Tours, France
claire.agafonov@univ-tours.fr

RÉSUMÉ

Après avoir défini ce que nous entendons par nom propre, nous évoquons quelques-uns des problèmes généraux posés par cette catégorie en traduction – considérée sous un angle multilingue. Nous proposons une typologie des noms propres sur deux niveaux puis passons en revue les différents procédés que recouvre la translation : l'emprunt – c'est-à-dire la non-traduction –, le calque, la transposition, l'adaptation, la modulation, l'incrémentialisation, mais aussi, lorsque les deux langues envisagées s'écrivent avec des alphabets différents, la translittération et la transcription. Enfin, nous nous penchons sur certains aspects morphologiques des noms propres en rapport avec la traduction automatisée.

ABSTRACT

After defining what we mean by proper nouns, we will mention some general problems raised by this category in relation to – multilingual – translation. We will establish a two-level typology of proper nouns and then examine various devices, going from loan translation – which amounts to no translation – to free translation, through literal translation – not forgetting transliteration and transcription between two languages using different alphabets. We will finally study some aspects of the morphology of proper nouns in relation to machine translation.

MOTS-CLÉS/KEYWORDS

multilingue, nom propre, traduction automatique, transcription, translittération

1. Introduction

Il est peu courant d'envisager la traduction sous un angle multilingue et non bilingue, chaque langue ayant ses propres particularités. C'est pourtant le parti que nous avons pris ici en évoquant quelques-uns des problèmes généraux posés par les noms propres en traduction. Nous illustrerons notre propos par des exemples issus de différentes langues européennes (allemand, anglais, français, polonais et russe).

1.1. Les noms propres

Comment définir les noms propres ? Les linguistes semblent s'accorder sur une définition classique, que Jonasson appelle celle des noms propres « purs » et qui est bien adaptée aux noms de personnes, de pays ou de villes. Citons, par exemple, celle de Grevisse dans *Le Bon Usage* (Grevisse et Goosse 1986 : 751) : « Le nom propre n'a pas de signification véritable, de définition ; il se rattache à ce qu'il désigne par un lien qui n'est pas sémantique, mais par une convention qui lui est particulière. »

Cependant, dans la réalité des textes, des expressions complexes apparaissent, souvent par composition d'un nom propre avec une expansion, comme *Muséum d'histoire naturelle* ou *Organisation des Nations Unies*. Celles-ci semblent être des descriptions définies figées (ou en cours de figement) et forment une classe ouverte à la création lexicale. Jonasson parle alors de noms propres « descriptifs ».

Nous adopterons donc la définition suivante du nom propre (Jonasson 1994: 21): « Toute expression associée dans la mémoire à long terme à un particulier en vertu d'un lien dénommatif conventionnel stable. » C'est cette définition étendue qui est traditionnellement utilisée dans le cadre du traitement automatique des langues et ce, depuis les conférences *MUC (Message Understanding Conferences)*, qui appellent les noms propres des « entités nommées » (Chinchor 1997).

1.2. Le projet PROLEX

Le projet PROLEX est un projet de recherche initié en 1994 (Maurel *et al.* 1996) et coordonné à partir du Laboratoire d'Informatique de l'Université de Tours (LI). Il a pour objet le traitement automatique des noms propres et a commencé par la création d'un dictionnaire relationnel de toponymes contenant plus de 70 000 entrées en langue française. Depuis 2000, il s'est étendu à tous les noms propres et à leurs traductions (Grass *et al.* 2002).

Ce travail est accompagné de la construction d'une typologie des noms propres qui sera brièvement présentée à la section 2.

1.3. La translation des noms propres

La translation des noms propres constitue le passage d'une langue à une autre sans qu'il y ait nécessairement changement de forme graphique. La translation recouvre quatre cas de figure selon que le nom propre est traduit, transcrit, translittéré ou reste inchangé. Des combinaisons de ces différents procédés sont aussi possibles. De plus, la forme obtenue par translation interagit avec les caractéristiques morphologiques de la langue cible. Tous ces phénomènes favorisent l'existence de variantes de différents types qu'il convient d'énumérer dans le but d'une reconnaissance automatique de textes.

Nous traiterons tout d'abord de la traduction (section 3), puis de la transcription et de la translittération (section 4), pour aborder enfin les aspects morphologiques (section 5).

2. Typologie des noms propres

2.1. Une hiérarchie à deux niveaux

La classification que nous avons établie possède deux niveaux hiérarchiques; le premier, celui des hypertypes, correspond aux traits sémantiques primitifs (humain, lieu, concret et événement); le second, celui des types, comprend des champs lexicaux relativement homogènes, en relation d'hyponymie avec les hypertypes. Nous avons aussi défini des liens entre les noms propres, ce qui nous a conduits à une base de données relationnelle et non à une simple liste de mots.

À partir de la classification de Bauer (1998: 53-59), nous avons adopté une répartition des noms propres en quatre hypertypes, les anthroponymes (humains), les toponymes (lieux), les ergonymes (objets et travaux) et les pragmonymes (événements). L'intérêt de ce premier niveau est d'attribuer à un nom propre une information minimale, sans connaissances encyclopédiques, ce qui permettra d'utiliser des procédures automatiques non supervisées.

Le second niveau de typologie comprend vingt-neuf types (Tableau 1); nous avons comparé la classification de Bauer avec celle de Koß (1999: 422-444), pour les noms propres économiques, et avec celle de Zabeeh (1968: 53). Nous avons souhaité réduire au minimum le nombre de types, afin de respecter une certaine homogénéité, sans tomber dans l'encyclopédisme et, surtout, de permettre la définition de procédures automatiques, qu'il faudra certainement, dans certains cas ambigus, superviser. Chaque nom propre n'est associé qu'à un type. Deux homonymes sont dupliqués (par exemple, *Loire* apparaît deux fois, comme hydronyme et comme région).

TABLEAU 1

Types et hypertypes

| | |
|--|--------------------------------|
| Célébrité Personnage mythique ou fictif Gentilé ou ethnonyme Dynastie Prénom | Anthroponyme |
| Association ou parti Ensemble artistique ou club sportif Organisation internationale | Anthroponyme/Ergonyme |
| Entreprise Établissement public ou privé | Anthroponyme/Toponyme/Ergonyme |
| Pays Région Groupe de pays Hydronyme Géonyme Objet céleste Lieu mythique ou fictif | Toponyme |
| Ville Quartier, voie Bâtiment | Toponyme/Ergonyme |
| Appellation commerciale Œuvre Objet mythique ou fictif Vaisseau | Ergonyme |
| Événement historique ou politique Phénomène météorologique Catastrophe | Pragmonyme |
| Manifestation artistique ou sportive Fête | Pragmonyme/Ergonyme |

2.2. Désambiguïsation par le type conceptuel

Dans la perspective du traitement automatique des langues, des informations d'ordre sémantique sont nécessaires pour lever les ambiguïtés dues à la présence d'homographes. Prenons l'exemple du mot allemand *Paris* qui désigne soit une ville, soit un personnage mythique (Tableau 2).

TABLEAU 2

Translations du mot allemand *Paris*

| Type | Français | Polonais |
|---|--------------|--------------|
| Ville | <i>Paris</i> | <i>Paryż</i> |
| Divinité, personnage mythique ou fictif | <i>Pâris</i> | <i>Parys</i> |

Le type conceptuel donne habituellement une information suffisante. La définition encyclopédique est inutile pour la translation. Il existe bien sûr des exceptions, comme celle du mot anglais *London* (Tableau 3).

TABLEAU 3

Translations du mot anglais *London*

| Type | Français | Allemand |
|--------------------|----------------|---------------|
| Ville (Angleterre) | <i>Londres</i> | <i>London</i> |
| Ville (Canada) | <i>London</i> | <i>London</i> |

Parfois, c'est la morphologie de la langue source qui permet le rattachement à un type conceptuel et, donc, une translation appropriée, comme pour le mot *France* en français (Tableau 4).

TABLEAU 4

Translations du mot français *France*

| Français | Type | Allemand | Polonais |
|------------------|----------|-------------------|---------------------------|
| <i>La France</i> | pays | <i>Frankreich</i> | <i>Francja</i> |
| <i>Le France</i> | vaisseau | <i>das France</i> | <i>(statek) Le France</i> |
| <i>France</i> | prénom | <i>France</i> | <i>France</i> |

3. Traduire ou ne pas traduire, telle est la question

La translation des noms propres d'une langue vers une autre peut consister en un emprunt, une traduction littérale (calque) ou une série d'autres procédés que nous passerons en revue (Vinay et Darbelnet 1958; Demanueli 1991; Ballard 1995).

3.1. L'emprunt

Souvent, les noms propres ne varient pas d'une langue à l'autre et sont conservés tels quels. Dans ce cas, on peut parler d'emprunt, bien que le nom propre provienne parfois d'une langue tierce qui n'est, dans une perspective traductionnelle, ni la langue

source, ni la langue cible, comme lorsqu'on importe un patronyme tel que *Bush*, un toponyme tel que *Washington*, un nom de firme tel que *Enron* dans une traduction de l'allemand vers le français. *A contrario*, on traduit à partir du moment où l'on opère une quelconque modification sur le texte en langue source.

3.2. *Le calque*

Le calque consiste en une traduction littérale, totale ou partielle (Tableau 5).

TABLEAU 5

Exemples de calques

| | Allemand | Français | Polonais | Anglais |
|----------------|-------------------------------|-------------------------------|--------------------------------|----------------------------|
| Calque total | <i>Kap der Guten Hoffnung</i> | <i>Cap de Bonne-Espérance</i> | <i>Przystań Dobrej Nadziei</i> | <i>Cape of Good Hope</i> |
| Calque partiel | <i>Universität Tours</i> | <i>Université de Tours</i> | <i>Uniwersytet w Tours</i> | <i>University of Tours</i> |

La translation par calque dépend évidemment du couple de langues choisi. Par exemple, dans le cadre de la traduction anglais-allemand, *Strasse von Dover* est un calque de *Straits of Dover*, dont il reproduit la structure lexico-syntaxique; en revanche, dans le cadre de la traduction français-allemand ou de la traduction français-anglais, ni *Strasse von Dover* ni *Straits of Dover* ne sont respectivement des calques de *Pas de Calais*.

3.3. *La transposition*

La transposition consiste à utiliser une tournure sémantiquement identique, mais syntaxiquement différente du lexème en langue cible (Tableau 6).

TABLEAU 6

Exemples de transpositions

| Allemand | Français | Polonais | Anglais |
|-------------------|------------------------------|-------------------------|--------------------------|
| <i>Vorpommern</i> | <i>Poméranie occidentale</i> | <i>Pomorze Przednie</i> | <i>Western Pomerania</i> |

Dans ce toponyme composé allemand, l'élément *Vor* de nature prépositionnelle est restitué par un adjectif dans les trois autres langues.

3.4. *L'adaptation*

L'adaptation est une traduction libre. Elle est utilisée principalement pour la traduction des noms d'œuvres au sens large (Tableau 7).

TABLEAU 7

Exemples d'adaptations

| Anglais | Français | Polonais |
|--|------------------------|---|
| <i>Baywatch</i> (litt. <i>Surveillance de la baie</i>) | <i>Alerte à Malibu</i> | <i>Słoneczny patrol</i> (litt. <i>Patrouille du soleil</i>) |

Dans cet exemple, le nom en anglais de la série télévisée américaine n'est pas évocateur en français et en polonais, parce qu'incompréhensible par la plupart des locuteurs. Il a donc été adapté. En allemand, en revanche, il est conservé tel quel et constitue un emprunt.

3.5. La modulation

La modulation consiste en un changement de point de vue. Ce procédé se rencontre très rarement, et principalement pour les noms de batailles où les victoires des uns sont les défaites des autres et vice versa (Tableau 8).

TABLEAU 8

Exemples de modulations

| Allemand | Français | Polonais | Anglais |
|--|--------------------------|---|--|
| <i>Sieg bei Königgrätz</i> (litt. victoire de Königgrätz) | <i>défaite de Sadowa</i> | <i>bitwa pod Sadową</i> (litt. bataille de Sadowa) | <i>Battle of Sadowa</i> <i>Battle of Königgrätz</i> |

Dans cette bataille de la guerre austro-prussienne en 1866, un Allemand voit une victoire alors qu'un Autrichien voit une défaite. Le français, bien qu'historiquement non directement concerné, parle aussi de *la défaite de Sadowa* et non de *la victoire de Königgrätz*. En revanche, l'anglais et le polonais restent neutres et parlent simplement de *bataille*.

3.6. L'incrémentialisation

Enfin, l'incrémentialisation est une traduction explicative destinée à éclairer le lecteur ignorant de la réalité culturelle de la région de la langue cible. Elle est fréquemment utilisée dans la presse et suit la formulation en langue source. Elle permet ainsi d'éviter une note de bas de page.

La Studienstiftung des Deutschen Volkes (organisation qui accorde des bourses aux meilleurs étudiants allemands)...

Boursière de la « Studienstiftung des deutschen Volkes », l'organisation allemande la plus renommée pour encourager les jeunes talents...

The Studienstiftung des deutschen Volkes (German National Merit Foundation) is an institution supported...

Welcome to the Studienstiftung des deutschen Volkes (German National Academic Foundation).

Studienstiftung des Deutschen Volkes, niemiecka organizacja wspierająca finansowo najlepszych niemieckich studentów... (litt. Organisation allemande aidant financièrement les meilleurs étudiants allemands).

Dans ces exemples, la traduction littérale *Fondation étudiante du peuple allemand* n'évoque rien. Elle revêt même une connotation nationaliste presque négative. L'incrémentialisation permet non seulement de neutraliser la connotation, mais aussi de restituer le sens caché derrière le nom de l'institution.

4. Translittérations et transcriptions

Dans le cas où les deux langues envisagées s'écrivent avec des alphabets différents, le « degré zéro » n'existe pas. En effet, tout passage d'une langue à l'autre implique alors un changement de code graphique. Aux cas de figure précédemment mentionnés (section 3), s'ajoute donc, pour la grande majorité des noms propres, le problème de la transcription. Nous l'évoquerons ici à travers l'exemple de la langue russe.

Si l'on exclut la transcription phonétique scientifique, qui utilise l'alphabet phonétique international et dont la rigueur s'accompagne d'une complexité la rendant inutilisable en dehors d'une étude linguistique, il existe deux grands types de translations entre le russe et les langues à alphabet latin : les translittérations et les transcriptions empiriques. Notons qu'il ne s'agit ici que de la translittération du cyrillique au latin. L'opération inverse n'a aucun sens, car aucune équivalence ne pourrait rendre compte de l'extrême diversité des correspondants phonétiques d'un même graphème ou des réalisations graphiques d'un même phonème dans les différentes langues utilisant l'alphabet latin ou même à l'intérieur d'une seule langue : on ne voit pas comment la même lettre russe pourrait correspondre à la fois au C de *Chantal*, *Cécile* ou *Claire* et pourquoi la suite EAU de *Marceau* serait notée par trois lettres cyrilliques.

4.1. Translittérations

La translittération est, en théorie, le système le plus simple et le plus fiable pour passer de l'alphabet cyrillique à l'alphabet latin. La norme internationale ISO 9, établie en 1995, fait correspondre à chaque graphème cyrillique un graphème latin, de façon arbitraire, en se basant sur la proximité des sons représentés. L'alphabet russe comportant plus de signes que l'alphabet latin (trente-trois vs vingt-six), sont également utilisés des signes diacritiques, inspirés de la façon dont les langues slaves à alphabet latin (en particulier le tchèque) notent des sons équivalents (voir la Tableau 9 pour quelques exemples).

TABLEAU 9

Exemples de translittérations

| Lettre cyrillique | Prononciation (API) | Translittération |
|-------------------|---------------------|------------------|
| а | [a] | a |
| г | [g] | g |
| ж | [ʒ] | ž |
| ц | [ts] | c |
| ш | [ʃ] | š |

Les avantages de ce système sont importants. La relation établie entre les deux alphabets est bijective et l'opération de passage de l'un à l'autre peut être effectuée dans les deux sens sans perte d'information. Elle peut être automatisée sans difficulté. La représentation d'un nom propre russe sera donc (en théorie) unique dans toutes les langues à alphabet latin. À cause de ces avantages, la translittération est utilisée dans les bibliothèques et les ouvrages scientifiques.

Mais ce système présente des inconvénients majeurs :

1. Pour l'écriture : les signes diacritiques sont généralement absents des claviers informatiques. L'écriture en est donc complexe, ce qui augmente le temps passé à la saisie et le risque d'erreur. De plus, l'affichage à l'écran de ces signes dépend beaucoup du logiciel utilisé et la restitution en est hasardeuse.
2. Pour la lecture : l'alphabet latin est lu des façons les plus diverses dans les différents pays qui l'utilisent. Par exemple, le même signe J sera interprété comme [Z] par les francophones, [j] par les germanophones, [x] par les hispanophones, etc. Ainsi, pour un francophone, il sera difficile de reconnaître *Tchaïkovski* dans la graphie *Čajkovskij*.

Pour ces raisons, certains assouplissements interviennent dans les normes de translittération : une lettre russe correspondra à deux lettres latines sans signes diacritiques, au lieu d'une seule lettre latine accompagnée d'un signe diacritique inhabituel (*š, ž, ŭ* seront respectivement remplacés par *sh, zh, ju*). C'est le cas de la norme russe GOST 1983, qui supprime l'accent circonflexe, et de la norme américaine ALA-LC, qui supprime presque tous les signes diacritiques. Ces variantes ont l'avantage de faciliter la saisie au clavier et de rapprocher l'écriture de la transcription anglo-saxonne, mais la relation n'est plus bijective signe à signe (le codage n'est plus réversible). En outre, l'existence même de variantes remet en cause l'intérêt principal de la translittération : faciliter la recherche documentaire.

4.1. Transcriptions empiriques

Chaque langue note les noms propres d'une autre langue selon un certain usage qui existe parfois depuis des siècles. On parlera alors de transcription empirique. Cet usage a pu varier dans l'histoire, il est toujours en mutation, mais certaines constantes peuvent être dégagées, en particulier grâce à l'existence d'encyclopédies qui codifient les noms propres étrangers.

Une langue note les sons d'une autre langue à travers le prisme de sa propre phonétique et au moyen des correspondances phonétique-graphie qui lui sont propres. Il existe donc, pour chaque couple de langues à alphabet différent, deux systèmes de règles entièrement distincts, dans un sens et dans l'autre (par exemple, cyrillique-latin et latin-cyrillique), qui codifient avec plus ou moins de succès la langue-source avec les moyens de la langue-cible. Le Tableau 10 présente trois exemples de transcriptions à partir du russe et le Tableau 11 montre les différents équivalents russes de la suite graphique CH selon la langue source.

TABLEAU 10

Exemples de transcriptions cyrillique-latin

| | | | |
|---------------------|-----|------|----------|
| Lettre russe | ш | ц | ю |
| Prononciation (API) | [ʃ] | [ts] | [ju] |
| Anglais | sh | tz | yu |
| Allemand | sch | z | ju |
| Français | ch | ts | you, iou |

TABLEAU 11

Exemples de transcriptions latin-cyrillique

| Lettres latines: CH | Lettre russe | Prononciation (API) |
|---------------------|--------------|---------------------|
| Anglais | ч | [tʃ] |
| Allemand | х | [x] |
| Français | ш | [ʃ] |
| Italien | к | [k] |

Si l'on allonge la liste des langues à alphabet latin, le tableau devient impressionnant. Le Tableau 12 présente les différents avatars des noms de *Tchekhov* et *Pouchkine* dans diverses langues européennes. On voit tout l'intérêt d'un dictionnaire multilingue de noms propres, qui permettrait de se retrouver dans ce foisonnement...

TABLEAU 12

***Tchekhov* et *Pouchkine* dans diverses langues européennes**

| | | |
|-------------------------------|--------------------|-----------|
| Russe | Чехов | Пушкин |
| <i>Translittération ISO 9</i> | Čehov | Puškin |
| Allemand | Tschechow | Puschkin |
| Anglais | Chekhov | Pushkin |
| Catalan | Txèkhov | Puixkin |
| Espagnol | Chejov, Chéjov | Pushkin |
| Finnois | Tsehov, Tshehov | Pushkin |
| Français | Tchekhov, Tchékhev | Pouchkine |
| Italien | Cechov | Puskin |
| Néerlandais | Tsjechov | Poesjkin |
| Polonais | Czechow | Puszkín |
| Portugais | Tchekhov | Púchkin |
| Suédois | Tjechov | Pusjkin |
| Tchèque | Čechov | Puškin |
| Turc | Çehov | Puşkin |

Cependant, sous l'influence de la mondialisation et de la domination du monde anglo-saxon, la transcription russe-anglais est de plus en plus utilisée dans les langues européennes pour les patronymes des célébrités (acteurs, sportifs, hommes politiques). Ceci est illustré par le Tableau 13: les chiffres entre parenthèses représentent le nombre d'occurrences de quelques patronymes sur Yahoo! France (pages uniquement en français). Il n'est pas exclu que cette influence vienne concurrencer les usages traditionnels dans les années à venir.

TABLEAU 13

Différentes transcriptions sur Yahoo! France

| | |
|------------------------|------------------------|
| Usage français | Usage anglais |
| Marat Safine (7) | Marat Safin (4570) |
| Oleg Menchikov (239) | Oleg Menshikov (239) |
| Boris Berezovski (683) | Boris Berezovsky (500) |

Si la transcription est indispensable pour la translation entre deux langues utilisant des alphabets différents, elle peut aussi avoir lieu, à un moindre degré, entre deux langues utilisant le même alphabet. Certains noms de célébrités sont transcrits en polonais afin que soit restituée au locuteur non initié la prononciation la plus proche possible de celle de la langue d'origine, par exemple *Szopen* (fr. *Chopin*), *Szekspir* (ang. *Shakespeare*), *Russo* (fr. *Rousseau*), *Waszyngton* (ang. *Washington*).

5. La morphologie des noms propres lors de translations

La mise en œuvre de la traduction automatisée de noms propres nécessite une description formelle détaillée de leurs caractéristiques morphologiques aussi bien dans la langue source que dans la langue cible. Cette description doit concerner le comportement des unités monolexicales, comme celui des unités polylexicales.

5.1. Flexion de noms propres monolexicaux

Les noms propres apparaissent dans un dictionnaire sous leur forme canonique (le « lemme »), habituellement au singulier (sauf par exemple *les États-Unis*) et, pour les langues concernées par la déclinaison, au nominatif. Cependant, la traduction s'applique à des textes où les mots sont fléchis. C'est pourquoi il est nécessaire d'inclure dans la base des informations suffisantes pour la création de deux types d'outils :

1. Analyseur morphologique des mots en langue source, pour attacher une forme fléchie d'un mot à traduire (trouvée dans un texte) à sa forme canonique (figurant dans la base de données). Ceci est nécessaire pour la recherche d'équivalents dans la langue cible.
2. Générateur morphologique de la langue cible qui permet d'obtenir les formes fléchies à partir de la forme canonique de la translation.

Il existe différentes approches de la morphologie computationnelle des unités monolexicales (par exemple Koskeniemi 1997, Silberztein 1993, Clémenceau 1997). Notre choix porte sur la méthode employée déjà pour le français (Courtois 1990) et d'autres langues européennes, qui consiste à attacher à chaque unité monolexicale un « code flexionnel » qui décrit explicitement quelles modifications doit subir la terminaison de la forme canonique pour que toutes les formes fléchies soient créées. Les codes flexionnels permettent notamment d'engendrer des dictionnaires de formes fléchies, employés par des analyseurs morphologiques.

Du point de vue d'un projet multilingue tel que PROLEX, les avantages majeurs de l'approche par codes flexionnels sont sa fiabilité (chaque entrée fait objet d'un classement « contrôlé »), sa facilité d'extension (de nouvelles entrées ainsi que de nouveaux codes peuvent être ajoutés facilement) et son indépendance de la langue : le formalisme morphologique choisi s'applique à de nombreuses langues européennes et asiatiques (ce qui a été démontré notamment par les différents projets du réseau RELEX : <<http://infolingu.univ-mlv.fr>>). Ainsi, les analyseurs et les générateurs morphologiques basés sur ce formalisme peuvent être eux aussi indépendants de la langue.

La difficulté majeure de l'approche par codes flexionnels réside dans le grand nombre de noms propres existants. Pour chaque nom entré dans la base, toutes les formes fléchies doivent être analysées, avant le choix du code flexionnel correct. Pour

les langues slaves, le nombre de formes fléchies d'un nom propre peut atteindre quelques dizaines. Par exemple (Tableau 14), en polonais, les noms de famille en *-ski* et *-cki* possèdent vingt-huit formes (deux nombres, sept cas, deux genres).

TABLEAU 14

Formes fléchies du nom de famille polonais *Malczewski*

| | Masculin humain singulier | Féminin singulier | Masculin humain pluriel | Féminin pluriel |
|--------------|---------------------------|---------------------|-------------------------|---------------------|
| Nominatif | <i>Malczewski</i> | <i>Malczewska</i> | <i>Malczewscy</i> | <i>Malczewskie</i> |
| Génitif | <i>Malczewskiego</i> | <i>Malczewskiej</i> | <i>Malczewskich</i> | <i>Malczewskich</i> |
| Datif | <i>Malczewskiemu</i> | <i>Malczewskiej</i> | <i>Malczewskim</i> | <i>Malczewskim</i> |
| Accusatif | <i>Malczewskiego</i> | <i>Malczewską</i> | <i>Malczewskich</i> | <i>Malczewskie</i> |
| Instrumental | <i>Malczewskim</i> | <i>Malczewską</i> | <i>Malczewskimi</i> | <i>Malczewskimi</i> |
| Locatif | <i>Malczewskim</i> | <i>Malczewskiej</i> | <i>Malczewskich</i> | <i>Malczewskich</i> |
| Vocatif | <i>Malczewski</i> | <i>Malczewska</i> | <i>Malczewscy</i> | <i>Malczewskie</i> |

Si une ou plusieurs translations d'un nom propre figurent également dans la base (tel est le but final de notre projet), elles doivent toutes être accompagnées de codes flexionnels dans chacune des langues cibles concernées. Or, la flexion des mots étrangers est un problème difficile pour un locuteur moyen. Il n'aura pas de mal à classer certains noms étrangers qui s'adaptent « fortuitement » au système morphologique de la langue cible, comme par exemple le nom de la capitale de l'Allemagne qui reste inchangé en polonais, et qui se décline selon le même paradigme que le nom d'une ville polonaise *Konstancin* ayant une terminaison identique (*Berlin* comme *Konstancin* au nominatif, *Berliną* comme *Konstanciną* au génitif, *Berlinowi* comme *Konstancinowi* au datif, etc.). Néanmoins, souvent, l'emploi des noms étrangers nécessite la création de paradigmes flexionnels atypiques pour la langue cible. Par exemple, le *e* muet n'existant pas en polonais, l'apostrophe marque alors l'apparition de ce phénomène dans certaines formes fléchies de mots étrangers. *Sartre* se déclinera donc en *Sartre'a* (génitif), *Sartre'owi* (datif), mais on écrira au locatif *Sartrze* (**Sartre'ze* est jugé impossible sûrement à cause du fait que le *r* et le *z* doivent apparaître ensemble afin de correspondre au son [ʃ] nécessaire ici).

La difficulté de la flexion des noms étrangers tient également à d'autres facteurs qu'il faut prendre en considération lors du choix du paradigme flexionnel, comme, par exemple en polonais (Grzenia 1998) :

- La prononciation d'origine (souvent approximative ou inconnue); par exemple, *Charles* devient au génitif *Charles'a* s'il s'agit d'un prénom français (car le *e* muet exige l'apostrophe) et *Charlesa* s'il s'agit d'un prénom anglais (le *s* se prononce).
- La langue d'origine; par exemple, pour les noms de famille terminés par *-y*, le génitif se construit sans apostrophe s'il s'agit d'un nom hongrois (par exemple *Mindszenty*, *Mindszentyego*), tandis que l'apostrophe est nécessaire s'il s'agit d'une autre langue (*Kennedy*, *Kennedy'ego*).
- Le degré de polonisation; par exemple, les noms hongrois terminés par *-y* gardent ce *y* final dans les formes fléchies (*Mindszenty*, *Mindszentyego*, etc.), ce qui est atypique pour les paradigmes polonais (par exemple *Wincenty*, *Wincentego*, etc.); néanmoins, dans le cas de haut degré de polonisation, ces noms peuvent s'adapter à un paradigme polonais, comme pour *Batory*, *Batorego*, etc. (roi de Pologne).

- Le type conceptuel ; par exemple, le pluriel du nom propre *Flawiusz* est *Flawiuszowie* s'il s'agit d'un prénom, et *Flawiusze* s'il s'agit d'une dynastie.

5.2. Mono- et polylexicalité des translations

Les noms propres sont des unités monolexicales (par exemple *Bordeaux*) ou polylexicales (par exemple *Lac de Constance*). Sur le plan graphique, la définition précise de la frontière entre ces deux types se heurte à des problèmes semblables à ceux concernant la distinction entre les unités simples (ou « mots simples ») et les unités complexes (ou « mots composés ») dans certaines applications informatiques du traitement automatique de langues (comme INTEX [Silberztein 1993] ou UNITEX [Paumier 2003]). Ces applications distinguent graphiquement les unités simples des unités complexes grâce à l'existence ou non d'au moins un « séparateur » tel que l'espace, le tiret, l'apostrophe, etc. Selon cette approche, le mot *aujourd'hui* serait une unité complexe composée de deux unités simples *aujourd* et *hui* séparées par une apostrophe. Or, dans le contexte des noms propres, nombreux sont les cas où différentes formes fléchies du même mot sont, d'après ce critère formel, tantôt des unités simples, tantôt des unités complexes, comme l'exemple déjà mentionné du nom *Sartre* en polonais : *Sartre'a* serait une unité complexe, *Sartrze* une unité simple. Le même problème concerne certains noms allemands terminés par *-s* : *Engels* au nominatif serait une unité simple, tandis que son génitif *Engels'* serait complexe grâce à la présence d'une apostrophe. On peut en outre constater que souvent, lors de la translation des noms propres, un mot a pour équivalent plusieurs mots dans une autre langue (par exemple *Charlemagne* en français = *Karl der Große* en allemand) ou, inversement, plusieurs mots ont pour équivalent un seul mot (*Mer d'Aral* en français = *Aralsee* en allemand).

Rappelons également que la traduction automatisée des noms propres suppose d'abord leur reconnaissance correcte dans le texte source. Pour les noms polylexicaux, ceci pose notamment le problème majeur des limites gauches et droites, ce que décrit Nathalie Friburger dans ce numéro.

5.3. Flexion de noms propres polylexicaux

La morphologie des noms propres polylexicaux, comme celle des mots composés communs (cf. Savary (2000)), nécessite une description à deux niveaux :

- Comment fléchir les composants d'une unité polylexicale, pris séparément ?
- Comment fléchir l'unité polylexicale entière ?

Prenons deux exemples du polonais. Pour décliner le nom du roi *Wilhelm Zdobywca* (*William the Conqueror*), il faut décliner les deux composants. Un code flexionnel pour chacun est donc indispensable. De plus, les deux composants s'accordent : le génitif du tout s'obtient avec le génitif de *Wilhelm* suivi du génitif de *Zdobywca*, ce qui donne *Wilhelma Zdobywcy* ; pour obtenir le datif du tout, on aligne les datifs des deux composants : *Wilhelmowi Zdobywcy*, etc. On dit que les deux composants de *Wilhelm Zdobywca* sont ses « constituants caractéristiques » car leurs genre, nombre et cas sont les mêmes que ceux du nom propre entier. Dans un autre nom propre, *Ryszard Lwie Serce* (*Richard the Lionheart*), seul le premier composant se fléchit : *Ryszarda Lwie Serce*, *Ryszardowi Lwie Serce*, etc. Le prénom *Ryszard* est donc le seul constituant caractéristique de ce nom polylexical.

La description du comportement morphologique des noms propres polylexicaux peut être réalisée à l'aide d'un formalisme semblable à celui proposé pour les mots composés par (Silberztein 1990) et étendu par (Savary 2000 : 48-74). Néanmoins, elle se heurte à certains problèmes décrits par ce dernier auteur, dus notamment à des irrégularités de constituants caractéristiques, ainsi qu'à l'existence de variantes flexionnelles. Par exemple, dans un patronyme polylexical polonais comme *Rydz Śmigły*, les deux composants se fléchissent (*Rydz Śmigłego*, *Rydzowi Śmigłemu*, etc.). Un autre patronyme, *Ostoja-Malinowski*, se fléchit selon deux modèles différents : soit le dernier composant seul se fléchit, soit les deux, ce qui donne au génitif ou *Ostoja-Malinowskiego* ou *Ostoji-Malinowskiego* (mais jamais **Ostoji-Malinowski*). Les deux noms ont donc une structure semblable, mais leur comportement flexionnel est différent et difficile à prévoir sans une description formelle explicite au cas par cas.

5.4. Variantes flexionnelles de noms propres monolexicaux et polylexicaux

Les codes flexionnels des noms monolexicaux doivent permettre l'engendrement de toutes leurs variantes flexionnelles. Celles-ci peuvent avoir différentes origines. Voici quelques exemples en polonais :

- Une forme fléchie obtenue selon un paradigme polonais peut coexister avec une forme invariable :
Jorge au génitif donne *Jorgego* ou *Jorge*.
- Une forme fléchie peut garder ou non son caractère diacritique provenant de la langue d'origine :
Fauré au génitif donne *Faurégo* ou *Faurego*.
- Une forme déclinée peut se construire avec ou sans apostrophe :
Jacques à l'instrumental donne *Jakiem* ou *Jacque'em*.
- Une forme correcte peut coexister avec une forme incorrecte mais fréquemment employée :
Mariusz au pluriel donne *Mariuszowie* ou ?*Mariusze* ;
Bruno au génitif donne *Brunona* ou ?*Bruna*.

Les variantes flexionnelles des noms polylexicaux incluent les exemples déjà cités dans la section précédente 5.3 :

Ostoja-Malinowski au génitif donne soit *Ostoji-Malinowskiego* soit *Ostoja-Malinowskiego* (jamais **Ostoji-Malinowski*) ;
Ludwika-Maria Gonzaga au datif donne *Ludwice-Marij Gonzaga* ou *Ludwice-Marij Gonzadze* (jamais **Ludwice-Maria Gonzaga*).

Notons enfin que les noms propres, aussi bien mono- que polylexicaux, peuvent posséder d'autres variantes – variantes du degré de translation et variantes orthographiques.

5.5. Variantes du degré de translation

Elles concernent notamment les noms de célébrités et leur formation peut s'avérer arbitraire. Voici quelques exemples provenant du polonais. Le nom anglais *George Washington* peut soit rester inchangé, soit subir à la fois la traduction du prénom et la transcription du patronyme :

George Washington ou *Jerzy Waszyngton* (jamais **Jerzy Washington* ni **George Waszyngton*).

Dans *Sigmund Freud*, seul le prénom est traduit :

Zygmunt Freud (jamais **Frojd*).

Dans *William Shakespeare*, le patronyme peut être transcrit ou non, mais le prénom n'est jamais traduit :

William Shakespeare ou *William Szekspir* (jamais **Wilhelm Shakespeare* ni **Wilhelm Szekspir*).

5.6. Variantes orthographiques

Si l'on examine plus particulièrement la transcription du russe en français, on constate que certaines règles d'usage sont difficiles à utiliser de par leur complexité. Par exemple, le nom de l'ancien chef d'État soviétique, en translittération *Hru ev* (six graphèmes, comme en russe), devient, selon les règles de transcription russe-français, *Khrouchtchev* (douze graphèmes) : il n'est pas rare que certains soient oubliés. Les variantes sont alors considérées comme fautives ; cependant, étant donné leur fréquence (la graphie *Kroutchev* représente 613 occurrences dans le moteur de recherche Yahoo! France, contre 3160 pour la graphie correcte, soit près de 20 %), il serait bon qu'elles soient reconnues.

6. Conclusion et perspectives

Ont été évoqués dans cet article quelques-uns des problèmes généraux posés par les noms propres en traduction dans une perspective multilingue. En nous appuyant sur divers types de noms propres, nous avons examiné les différents procédés de translation : l'emprunt, c'est-à-dire la non-traduction, le calque, la transposition, l'adaptation, la modulation, l'incrémentalisation, mais aussi la translittération et la transcription. Nous avons également insisté sur la nécessité de la prise en compte des caractéristiques morphologiques des noms propres, en raison notamment de la difficulté à fléchir certains noms étrangers.

De nombreuses pistes apparaissent à l'issue de ce travail. Une fois la base de données relationnelle multilingue constituée, le projet Prolex à un stade plus avancé devrait en particulier :

- Fournir une aide à la traduction.
- En traduction automatique, assurer le remplacement des noms propres par leurs équivalents correctement fléchis en langue cible.
- Favoriser la recherche d'informations et l'indexation de textes.
- Permettre l'alignement de textes multilingues à partir des noms propres qu'ils contiennent.

RÉFÉRENCES

- BALLARD, M. (1995) : *La traduction, de l'anglais au français*, Paris, Nathan.
- BAUER, G. (1998) : *Namenkunde des Deutschen*, Berne, Germanistische Lehrbuchsammlung Band 21 (1^{re} édition : 1985).
- CHINCHOR, N. (1997) : *Muc-7 Named Entity Task Definition*, consultable sur le site <http://www.itl.nist.gov/iaui/894.02/related_projects/muc/proceedings/muc_7_toc.html#appendices>.

- CLEMENCEAU, D. (1997): « Finite-State Morphology: Inflections and Derivations in a Single Framework Using Dictionaries and Rules », in ROCHE, E. and Y. SCHABES (eds): *Finite-State Language Processing*, Cambridge, MIT Press.
- COURTOIS, B. (1990): « Un système de dictionnaires électroniques pour les mots simples du français », *Langue Française* 87, p. 11-22.
- DEMANUELLI, C. et J. (1991): *Lire et traduire, anglais-français*, Paris, Masson.
- GRASS, T., MAUREL, D. et O. PITON (2002): « Description of a Multilingual Database of Proper Names », *Actes de PorTal 2002, Lecture Notes in Computer Science* 2389, p. 137-140.
- GREVISSE, M. et A. GOOSSE (1986): *Le Bon Usage*, Gembloux, Duculot.
- GRZENIA, J. (1998): *Słownik nazw własnych. Ortografia, wymowa, słotwórstwo i odmiana (Dictionnaire de noms propres. Orthographe, prononciation, formation de mots et flexion)*, Warszawa, Wydawnictwo naukowe PWN.
- JONASSON, K. (1994): *Le nom propre. Constructions et interprétations*, Paris, Duculot.
- KOSKENNIEMI, K. (1997): « Representations and Finite-State Components in Natural Language », in ROCHE, E. and Y. SCHABES (eds): *Finite-State Language Processing*, Cambridge, MIT Press, p. 383-402.
- KOSS, G. (1999): « Was ist ‚Ökonomie‘? », *Beiträge zur Namensforschung* 34-4, p. 373-444.
- MAUREL, D., BELLEIL, C., EGGERT, E. et O. PITON (1996): « Le projet PROLEX », *Actes du séminaire Représentations et Outils pour les Bases Lexicales, Morphologie Robuste de l'action Lexique du GDR-PRC CHM*, Grenoble, p. 164-175.
- PAUMIER, S. (2003): *De la reconnaissance de formes linguistiques à l'analyse syntaxique*, Thèse de doctorat en Informatique, Université de Marne-la-Vallée.
- SAVARY, A. (2000): *Recensement et description des mots composés – méthodes et applications*, Thèse de doctorat en Informatique Fondamentale, Université de Marne-la-Vallée.
- SILBERZTEIN, M. (1990): « Le dictionnaire électronique des mots composés », *Langue française* 87, p. 71-83.
- SILBERZTEIN, M. (1993): *Dictionnaires électroniques et analyse automatique de textes: le système INTEX*, Paris, Masson.
- VINAY, J.-P. et J. DARBELNET (1977): *Stylistique comparée de l'anglais et du français*, Paris, Didier (1^{re} édition: 1958).
- ZABEEH, F. (1968): *What's in a Name? An Inquiry into the Semantics and Pragmatics of Proper Names*, La Haye, Martinus Nijhoff.