Document generated on 10/12/2024 2:48 a.m.

Meta
Journal des traducteurs
Translators' Journal

META

Aide au transfert lexical dans une perspective de TAO : expérimentation sur un lexique non-terminologique

Arlette Attali, Guy Bourquin, Marie-Christine Bourquin-Launey, Annette Euvrard and Christiane Vigroux

Volume 37, Number 4, décembre 1992

Études et recherches en traductique / Studies and Researches in Machine Translation

URI: https://id.erudit.org/iderudit/001917ar
DOI: https://doi.org/10.7202/001917ar

See table of contents

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)
1492-1421 (digital)

Explore this journal

Cite this article

Attali, A., Bourquin, G., Bourquin-Launey, M.-C., Euvrard, A. & Vigroux, C. (1992). Aide au transfert lexical dans une perspective de TAO : expérimentation sur un lexique non-terminologique. *Meta*, *37*(4), 770–790. https://doi.org/10.7202/001917ar

Article abstract

Is it possible to devise an automatic procedure that could become part of an MAT programme and at the same time be used independently by a human translator? The present paper is an attempt to deal with a specific problem in lexical transfer: translating English and Spanish verbs into French. Through a systematic investigation of verbal contexts (valency-frames, word meanings) semantic classes are empirically defined and common semantic labels assigned to lexical units associated with each target-equivalent of the source verb. A tentative list of such classes is being drawn up. It is assumed that the semantic class of a subject and/or object (possibly adjunct) lexical unit can help find out the adequate translation of the source verb. Semantico-cognitive descriptors thus become part of the transfer-dictionary decision making. Being more especially based on lexical verb/valency associations, such a dictionary might also be of some help to a human translator.

Tous droits réservés © Les Presses de l'Université de Montréal, 1992

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

AIDE AU TRANSFERT LEXICAL DANS UNE PERSPECTIVE DE TAO: EXPÉRIMENTATION SUR UN LEXIQUE NON TERMINOLOGIQUE

ARLETTE ATTALI, GUY BOURQUIN,
MARIE-CHRISTINE BOURQUIN-LAUNEY, ANNETTE EUVRARD, CHRISTIANE VIGROUX
CELTA, URA 1035 du CNRS et Université de Nancy II, Nancy, France

Résumé

Peut-on construire un outil automatisé qui puisse à la fois s'insérer comme module dans un programme de TAO et indépendamment être utile à un traducteur humain? Le présent article tente de répondre à certains problèmes rencontrés en transfert lexical, plus particulièrement lors de la traduction en français de verbes anglais et espagnols. L'exploration systématique de l'environnement immédiat (schèmes d'actance, contenus lexicaux) permet de dégager empiriquement des classes notionnelles, chacune regroupant, sous un même descripteur sémantico-cognitif, des lexèmes de même fonction syntaxique reconnus compatibles avec telle traduction spécifique du verbe source. Une liste provisoire de ces classes (encore évolutives) a été élaborée. On fait l'hypothèse que la classe notionnelle du lexème sujet et / ou objet (voire circonstant) peut aider à sélectionner l'équivalent français réclamé par le verbe source. Les descripteurs sémantico-cognitifs sont intégrés aux procédures qui, dans un dictionnaire automatique de transfert, conditionnent les choix de traduction. Le traducteur humain, de son côté, peut trouver là un nouveau type de dictionnaire plus spécifiquement axé sur les associations sémantiques lexème verbal / lexèmes actants.

Abstract

Is it possible to devise an automatic procedure that could become part of an MAT programme and at the same time be used independently by a human translator? The present paper is an attempt to deal with a specific problem in lexical transfer: translating English and Spanish verbs into French. Through a systematic investigation of verbal contexts (valency-frames, word meanings) semantic classes are empirically defined and common semantic labels assigned to lexical units associated with each target-equivalent of the source verb. A tentative list of such classes is being drawn up. It is assumed that the semantic class of a subject and / or object (possibly adjunct) lexical unit can help find out the adequate translation of the source verb. Semantico-cognitive descriptors thus become part of the transfer-dictionary decision making. Being more especially based on lexical verb / valency associations, such a dictionary might also be of some help to a human translator.

1. CADRE GÉNÉRAL

De la traduction humaine autonome (THA) à la traduction automatique autonome (TAA), on postule généralement le continuum:

I	II	III	IV
Traduction humaine autonome	Traduction humaine assistée par l'automate	Traduction automatique assistée par l'homme	Traduction automatique autonome
(THA)	(THAA)	(TAAH)	(TAA)

Idéalement, chaque point sur le continuum induit sa problématique et ses méthodes, qui conditionnent à leur tour procédures et outils. Mais les frontières sur un continuum ne sont jamais étanches : des raisonnements conçus dans le cadre de l'étape II (THAA) peuvent fort bien trouver à s'appliquer jusqu'au terme de l'étape IV (TAA). Le présent article tentera de montrer que **la démarche inverse (depuis IV vers III et II) est elle aussi possible.** En d'autres termes, on fait l'hypothèse qu'un algorithme de transfert lexical initialement conçu comme module informatisé à insérer dans un éventuel programme de traduction automatique entièrement autonome pourrait, moyennant quelques aménagements, être réutilisé sur le mode virtuel (c'est-à-dire hors implémentation machine) par un traducteur humain qui en exécuterait les instructions pour ses propres besoins.

2. PROBLÉMATIQUE

On décrira une expérience encore en pleine évolution conduite au CELTA. L'objectif initial était de construire un **module de transfert lexical** devant trouver sa place dans une chaîne de traduction anglais-français entièrement automatisée. On dispose, à ce stade du programme, des sorties de l'analyseur morphosyntaxique : unités lexicales et syntagmes, fonctions syntaxiques (sujet et objet profonds, circonstants). Quand elle porte sur le lexique dit de spécialité (terminologie), la tâche du transfert lexical est relativement aisée. Il en va autrement du vocabulaire **non terminologique**, habituellement polysémique. C'est sur celui-ci qu'a porté l'expérience.

2.1. Cependant le vrai problème, en traduction automatique, n'est pas celui de la polysémie du terme dans sa langue d'origine. L'opération de transfert lexical, en effet, consiste non pas à interpréter du signifié, mais à remplacer du **signifiant** (source) par du **signifiant** (cible). Le choix entre les multiples équivalents cibles candidats à la traduction d'un terme polysémique source ne se décide pas nécessairement à partir de telle ou telle valeur sémantique qu'on aura pu isoler. Il est bien plus souvent dicté par des usages dépendant du contexte d'emploi. Les termes monosémiques eux-mêmes requièrent des traductions variables selon l'environnement immédiat. La tâche de l'automate traducteur est avant tout d'orienter sur un **signifiant** cible acceptable pour un **contexte** donné.

En outre, les signifiants cibles qui figurent dans un dictionnaire de transfert automatique ont un statut radicalement différent de ce que peut offrir un dictionnaire destiné à un lecteur humain. Là où le dictionnaire humain se contente de proposer, le dictionnaire automatique **impose** ses choix. Dans un dictionnaire bilingue ordinaire, les signifiants de langue cible ont un rôle ambigu : ils sont au moins aussi souvent simples pointeurs de sens que mots à réinsérer **tels quels.** Le traducteur doit souvent, à partir des suggestions du dictionnaire, trouver par lui-même le mot adapté au contexte. À l'opposé, un dictionnaire automatique présente une grande rigidité : les signifiants cibles suggérés n'y sont pas à prendre comme les points de départ interchangeables ou complémentaires d'une négociation sens-forme mais comme les **points d'arrivée** d'un calcul. Et ce calcul, dans le cas présent, consiste non pas, avons-nous dit, à analyser le signifié du mot à traduire mais à tirer parti de **régularités contextuelles** dûment répertoriées et explicitées.

Pour ces raisons on avait décidé, dans un premier temps — afin de simplifier les opérations de l'automate traducteur —, de ne pas soumettre le terme à traduire à une analyse sémantique interne, mais plutôt de s'attacher immédiatement à l'exploration de son **contexte** (syntaxique, lexical, **et sémantique**) et d'en extraire toutes informations utiles au choix de l'équivalent cible. Le problème se posait dans les termes suivants :

> Étant donné, d'une part, un signifiant A appréhendé dans un **environnement** source spécifique E et, d'autre part, des signifiants cibles B_1, B_2...B_n proposés par un dictionnaire automatique comme équivalents possibles (mais non interchangeables localement) de A,

quelle information minimale l'environnement E doit-il porter pour permettre à un programme automatique de choisir l'équivalent B_i le plus apte à restituer en phrase cible la valeur de sens de A dans E ?

Le problème est celui du **contexte pertinent minimal** dans les choix de traduction. Et c'est cette pertinence qu'il faut tenter de dégager.

On a été ainsi amené à construire un dictionnaire de transfert qui soit autre chose qu'une simple liste de correspondances et dont chaque entrée lexicale reçoit la configuration suivante :

SIGNIFIANT SOURCE	==> {	Procédure d'exploitation du contexte	=	TESTS HIÉRARCHISÉS (si...,alors, sinon, etc)	} ==>	SIGNIFIANT CIBLE

2.2. La relation au contexte se présente différemment selon la classe grammaticale de l'unité à traduire. Le **lexème verbal** a semblé, pour commencer l'expérience, particulièrement intéressant. Avec le verbe, en effet, le réseau des contextes potentiellement utilisables s'élargit à l'ensemble des relations que met en jeu la morphosyntaxe actancielle (sujet profond, objets direct et indirect profonds, circonstants plus ou moins attendus). Le contexte pertinent est à la fois morphosyntaxique et sémantique : plus précisément, il est constitué par la sémantique des **unités lexicales** qui occupent les fonctions actancielles dont le verbe est le pivot. De la conjonction entre tel rôle actanciel et le sémantisme de telle unité lexicale dotée de ce rôle, on inférera le mot de langue cible le plus apte à restituer le sens contextuel du verbe source.

Reste à construire la procédure pour y parvenir. Tout repose, on le voit, sur une codification sémantique préalable des mots **autres** que le verbe. Mais il faut éviter de faire porter au dictionnaire une charge excessive et donc déterminer tout d'abord quel est le minimum sémantique à prendre en compte, c'est-à-dire ne retenir que les informations utiles au calcul du lexème verbal cible. Or un même mot[1] non seulement peut renvoyer à plus d'une notion (ou représentation culturelle), mais chaque notion est elle-même un agrégat de composantes (ou propriétés) notionnelles. C'est donc le plus souvent des **traits** constitutifs plutôt que l'ensemble d'un sémantisme qu'il suffira de mettre en évidence et c'est sur la présence de ces traits dans le cadre contextuel prédéfini que porteront les tests effectués par l'automate traducteur. On donnera aux traits constitutifs le nom de «descripteurs sémantiques».

La définition du contenu des descripteurs ne doit rien à un système sémantique préétabli (tel que Nedobejkine 1990, Steiner 1988). La méthode, purement empirique, a été dictée par la seule finalité de l'expérience. Autour de chaque verbe de langue **source** concerné, on a réuni un corpus plus ou moins important de ses attestations en contexte et on a associé intuitivement à chaque attestation un **signifiant cible** acceptable par le contexte. On se demande alors quel(s) mot(s) de l'environnement actanciel a (ont) pu motiver le choix du signifiant cible. Au terme de l'inventaire on dispose, pour chaque équivalent verbal **cible**, d'une liste de mots dont la présence dans les divers contextes où il est apparu a été jugée discriminante (c'est-à-dire à mettre en corrélation avec lui). On recherche alors ce que les mots d'une même liste ont **sémantiquement** en commun. Le trait notionnel commun, une fois identifié, reçoit le statut de «descripteur sémantique». La liste des lexèmes regroupés sous un même descripteur constitue donc **de facto** une classe notionnelle (ou cognitive). On voit comment la traduction humaine a pu servir dans un premier temps d'outil d'investigation et comment, inversement, dans un second temps, les classes notionnelles obtenues par le biais de la traduction humaine, vont à leur tour servir, dans le programme automatisé, de critères de sélection des signifiants cibles.

Rappelons que le descripteur ne rend pas compte de la totalité du sémantisme d'une notion, mais n'en représente qu'une des composantes dont la délimitation est toujours tributaire des critères de minimalité et de pertinence définis ci-dessus[2].

3. DESCRIPTION DES EXPÉRIENCES CONDUITES SUR L'ANGLAIS

La recherche sur l'anglais s'est effectuée jusqu'ici en deux temps. On fera donc état ci-après de deux expériences successives, la seconde ayant pour objectif principal de tester la validité et les possibilités de généralisation de la première.

3.1. EXPÉRIENCE INITIALE

La première expérience (Attali *et al.* 1989) s'est concentrée sur cinq verbes anglais, (*draw, hold, lie, turn, work*), appréhendés en contexte (limité à la phrase) dans des textes réels (c'est-à-dire non fabriqués) extraits d'un corpus de concordances à l'intérieur duquel ils représentent 5 465 attestations[3].

Les verbes ont été choisis pour les raisons suivantes :

a) ils appartiennent à des champs sémantiques différents (qu'on n'a d'ailleurs — conformément à l'esprit de l'expérience — pas cherché à caractériser) ;
b) l'éventail des effets de sens de chacun est suffisamment ouvert (concret / abstrait, propre / figuré) ;
c) chacun donne lieu à un nombre assez élevé de traductions françaises différentes ;
d) les réseaux actanciels de chacun sont suffisamment diversifiés ;
e) chacun a une fréquence d'emploi assez élevée dans le corpus pour que l'étude soit significative.

Les procédures d'exploration des contextes sont de longueur et de complexité variables. Les tests portent selon les cas sur :

- une unité lexicale interrogée en clair :

  (1) MOVE + FORWARD —> 'avancer'

- une composante de la structure syntaxique :

  (2) HOLD + adj. —> 'rester'

- un descripteur sémantique (cas le plus fréquent) :

  (3) WORK + {sujet = MACHINE } —> 'fonctionner'

- une combinaison des cas précédents.

  (4) HOLD (voix : passif) + {sujet = PHÉNOMÈNE SOCIAL TEMPORAIRE} —> 'avoir lieu'

Les descripteurs construits dans le cadre de cette première expérience sont en nombre relativement élevé (44 pour les cinq verbes étudiés). La liste en est donnée en **Annexe I** (descripteurs précédés du chiffre 1). Ils présentent les caractéristiques suivantes :

1. Ils ne sont pas de même nature :

Auprès de traits d'une grande généralité tels que *'humain'*, *'animal'*, *'abstrait'*, *'objet'*, etc., figurent des notions très spécifiques : *'finance'*, *'affectivité'*, *'porteur de virtualités'*, *'moyen de transport'*...

2. Ils ne sont pas hiérarchisés :

Un embryon de hiérarchisation n'a été réalisé qu'à l'intérieur de quelques catégories :

(5) a. HUMAIN —> COLLECTIF HUMAIN STRUCTURE, etc.
——> GÉNÉRATEUR DE LUMIÈRE-CHALEUR
b. OBJET —> OBJET FABRIQUÉ —> MACHINE —
——> MOYEN DE TRANSPORT

3. Ils n'éclairent qu'un aspect du contenu notionnel des unités lexicales auxquelles ils sont affectés. Aussi certains lexèmes figurent-ils dans plusieurs listes sémantico-notionnelles différentes et relèvent donc de plus d'un descripteur :

(6) a. *book* : CONTENANT
OBJET
PORTEUR D'INFORMATION

b. *position* : ENTITÉ SPATIALE
LIEU
OPINION
SOCIO-PROFESSIONNEL ABSTRAIT

Inversement, sous un même descripteur peuvent être regroupés des mots perçus intuitivement comme très différents :

(7) a. dollar / rate / pension —> FINANCE
b. comparison / parallel —> MISE EN RELATION
c. phone / idea —> PORTEUR D'INFORMATION
letter / statement
d. destruction / programme —> PROCESSUS

4. Ils sont de rentabilité variable :

Un descripteur est rentable lorsqu'il est fréquemment utilisé : soit parce qu'il coiffe une liste importante d'unités lexicales, soit parce qu'il est interrogé dans le contexte de plusieurs verbes différents, soit parce qu'il répond à ces deux conditions à la fois.

Le descripteur '*humain*' satisfait à ce double critère : il est utilisé pour les 5 verbes, couvre un nombre élevé de lexèmes, eux-mêmes répartis entre les trois fonctions GN1, GN2, GN3[4]. Les descripteurs '*partie du corps*', '*porteur d'information*' ou '*machine*' sont eux aussi sollicités pour plusieurs verbes. En revanche, le descripteur '*phénomène social temporaire*' ne l'est que pour *hold*, mais couvre une classe de lexèmes nombreuse tout en étant sémantiquement spécifique et bien délimité. À l'inverse, '*entité spatiale + géographique + géologique*', sollicité par les verbes *draw, hold, lie*, a une extension sémantique plus difficile à cerner. D'autres descripteurs comme '*vêtement / partie de vêtement*' ou '*bruit*' ou '*phénomène naturel*' sont pour l'instant encore faiblement rentables :

(8) a. TURN + TO + GN2
si GN2 = BRUIT ==> 'se transformer en'
The signal TURNED TO *a* **buzz**
Le signal **se transforma en** bourdonnement

b. TURN + UP + GN2
si GN2 = VÊTEMENT / PARTIE DE VÊTEMENT ==> 'remonter'
He was TURNING UP his coat **collar** against the rain.
Il **remontait** le col de son manteau pour se protéger de la pluie.

On fait ici le pari que ces descripteurs accroîtront leur rendement au fur et à mesure qu'augmentera le nombre de verbes étudiés.

3.2. DEUXIÈME EXPÉRIENCE

L'expérience dont il vient d'être fait état a permis de dégager 44 descripteurs de poids inégal. L'objectif à long terme est néanmoins d'en limiter la prolifération. La deuxième expérience s'appuie donc délibérément sur les acquis de la précédente afin d'en tester la résistance et la réutilisation.

Cinq nouveaux verbes (*drive, go, move, run, walk*) ont été choisis, prélevés dans les mêmes conditions contextuelles que les premiers. Mais, à la différence de ceux-ci, ils relèvent tous les cinq — par leur sens d'origine — d'un même domaine notionnel : le déplacement spatial.

L'expérience, comparée à la précédente, conduit aux constatations suivantes :

3.2.1. EXTENSION DU SCHÉMA ACTANCIEL

On est amené à interroger plus souvent des descripteurs situés dans des groupes prépositionnels ou adverbiaux circonstants :

(9) a. — The wasting of natural resources has been GOING ON **for centuries** (circ. = DURÉE)
— Le gaspillage des ressources naturelles **continue** depuis des siècles
b. — Unemployment continues to RUN **at 6 percent** of the labour force (circ. = MESURE)
— Le chômage continu d'**atteindre** 6 pour cent de la main-d'œuvre

3.2.2. RÉEMPLOI ET / OU CRÉATION DE DESCRIPTEURS

Les rapports numériques entre descripteurs conservés, descripteurs non réutilisés et descripteurs nouvellement créés s'établissent comme suit :

Nombre de descripteurs utilisés dans l'expérience 1	44
Nombre de descripteurs de l'expérience 1 réutilisés dans l'expérience 2	28
Nombre de descripteurs non réutilisés	16
Nombre de descripteurs créés pour la deuxième expérience	13

On constate que les deux tiers environ des descripteurs de l'expérience 1 sont réemployés. D'autre part le nombre des descripteurs nouvellement créés équilibre celui des descripteurs non réutilisés. On trouvera en **Annexe I** la liste totale des descripteurs utilisés dans les deux expériences et, pour chacun, la mention de son appartenance à l'une et / ou à l'autre.

La rentabilité des descripteurs communs aux deux expériences est variable. En prenant comme critère le nombre de fois qu'un descripteur est sollicité par l'automate traducteur, voici, à titre d'exemple, quelques indications chiffrées :

	1re exp.	2e exp.
• Plus fréquemment sollicités dans exp. 2 que dans exp. 1 :		
'*voie de passage*'	3	6
'*humain*'	40	67
'*porteur d'information*'	5	10
• Interrogés à égalité dans les deux expériences :		
'*partie du corps*'	9	8
'*phénomène social temporaire*'	4	4
'*finance*'	2	2

• Moins sollicités dans exp. 2 que dans exp. 1 :		
'générateur lumière-chaleur'	5	2
'mesure'	7	3

Des descripteurs comme *'humain'*, *'animal'*, *'lieu'*... confirment leur importance. D'autres (*'voie de passage'*, *'moyen de transport'*, *'partie du corps'*, *'machine'*, *'porteur d'information'*, *'socio-professionnel abstrait'*, etc.), dont la rentabilité semblait incertaine lors de la première expérience, se voient beaucoup plus fréquemment sollicités (certains, peut-être, du fait de leur affinité avec tel ou tel effet de sens du verbe concerné).

(10) VOIE DE PASSAGE
a. — ...an evergrowing **network of railroads** was MOVING [grain] to the ports
— ...un réseau ferroviaire en expansion **transportait** [des céréales] jusqu'aux ports
b. — I could WALK OUT the **door**
— Je pouvais **sortir par** la porte

(11) MOYEN DE TRANSPORT
a. — The 1958 **carriers** that MOVE material from the yard to the job site
— Les transporteurs de 1958 qui **transportent** les matériaux depuis le dépôt jusqu'au chantier
b. — He RAN the **jeep** DOWN TO headquarters
— Il **conduisit** la jeep jusqu'au QG

(12) PARTIE DU CORPS
a. — White's **arm** WENT UP
— Le bras de White **se leva**
b. — It is easy to RUN your **finger** ACROSS a map
— Il est facile de **passer** le doigt **sur** une carte
c. — The man... trying to talk without MOVING his **lips**
— L'homme... essayant de parler sans **bouger** les lèvres

(13) MACHINE
a. — ...producing steam that DRIVES the **turbine** that DRIVES the **generator**
— ...produisant de la vapeur qui **fait marcher** la turbine qui **fait marcher** le générateur
b. — a **compressor** will RUN continuously during hot weather
— Un compresseur **marchera** sans arrêt par temps chaud
c. — a small **motor** that GOES ON as soon as it is plugged in
— Un petit moteur qui **se met en marche** dès qu'il est branché

(14) PORTEUR D'INFORMATION (très rentable pour : GO, MOVE et RUN)
a. — The first day of deliberation was spent in GOING OVER the **evidence**
— Le premier jour des débats se passa à **examiner** les dépositions
b. — A **story** GOES that...
— Une histoire **dit** que...
c. — Millikin MOVED an **amendment**
— Millikin **proposa** un amendement
d. — He RAN DOWN his **check list**
— Il **passa en revue** sa liste de contrôle

(15) SOCIO-PROFESSIONNEL ABSTRAIT
a. — His decision to RUN FOR **office** should have been revoked
— Sa décision de **briguer** la fonction aurait dû être annulée
b. — We need science generalists, not to GO INTO **politics**, but...
— Nous avons besoin de scientifiques non spécialistes, non pour **se mettre à** la politique, mais...

3.2.3. Les classes notionnelles nouvelles (au nombre de 13) ont toutes un sémantisme très spécifique (v. **Annexe I**, descripteurs précédés du chiffre 2 seul) :

(16) ÉLÉMENT LONGITUDINAL
— Street car **tracks** RUN DOWN the center of Pennsylvania
— Les rails de tramway **descendent** le centre de Pennsylvania

(17) ÉLÉMENT LIQUIDE
— This **water** RUNS DOWN the sloping surface
— Cette eau **coule le long** de la pente

3.2.4. Le nombre des descripteurs consultés par le programme varie beaucoup d'un verbe à l'autre. À l'opposé de *walk* (qui en utilise 5), les verbes *go* et *run*, à cause de leur forte polysémie, en font intervenir au contraire un très grand nombre. On en dénombre 26 dans le programme de *run* (dont *'humain'*, *'animal'*, *'porteur d'information'*, *'élément liquide'*, *'couleur'*, *'machine'*, *'comportement'*...), et 29 dans celui de *go* (dont *'humain'*, *'porteur d'information'*, *'machine'*, *'onomatopée'*...).

(18) a. Russell's **horse** RAN towards the building
(ANIMAL)
Le cheval de Russell **courut** vers le bâtiment
b. «...», RUNS an old **adage**
(PORTEUR D'INFORMATION)
«...», , **dit** un vieil adage
c. The club RUNS regular skiing **trips**
(PHÉNO. SOCIAL TEMPORAIRE)
Le club **organise** des voyages de ski réguliers
d. It was thought advisable to RUN the **rotors**
(MACHINE)
On a pensé souhaitable de **faire marcher** les rotors
e. So the **legend** GOES, ...
(PORTEUR D'INF.)
Ainsi **dit** la légende
f. ... expect the squirrel to GO OFF **nuts**
(ALIMENT)
... s'attendre à ce que l'écureuil **n'ait plus envie de** noisettes
g. GO THROUGH his **pocket**, will you?
(CONTENANT)
Fouille dans sa poche, veux-tu ?
h. He could not GO THROUGH a **field**
(ENTITÉ SPATIALE)
Il ne pouvait pas **traverser** un champ

Les exemples (18)a et h appartiennent au sémantisme du mouvement dans l'espace ; les exemples (18)d et f conservent l'idée de mouvement mais sans déplacement ; les autres exemples (18)b, c, et e ont perdu toute référence à la notion de mouvement. Le plus intéressant est la transformation de *go et run* en verbes de «dire» lorsque le sujet est caractérisé comme *'porteur d'information'* (voir exemple (18)b et e).

On remarque l'importance, pour la traduction, des postpositions ou prépositions associées à ces verbes (en particulier *go* et *run*). Des tests sur ces mots suffisent assez souvent à sélectionner l'équivalent cible du verbe :

(19) a. ... my father RAN him OFF six years ago
... mon père l'a **chassé** il y a six ans
b. ... people waving as you WENT PAST
... des gens saluant de la main comme on **passait**

c. The coachman WALKED OFF briskly in search of...
Le cocher **partit** sans tarder à la recherche de...
d. Bobby, DRIVING BACK from an afternoon's golf
Bobby, **revenant** d'un après-midi de golf
e. ... Parisians who have GONE AWAY for the week-end
... les Parisiens qui **sont partis** pour le week-end

3.3. CHOIX DES TRADUCTIONS

Malgré la présence des tests morpho-syntaxiques et sémantiques, les choix de traductions résultent souvent de compromis dictés par les limites mêmes de l'expérience, à savoir :

1. prise en compte uniquement des attestations recensées dans un corpus, certes vaste et représentatif mais évidemment non exhaustif ;
2. traduction de lexème à lexème sans remise en cause des structures syntaxiques ;
3. exploration du seul contexte actanciel aux dépens d'autres composantes de l'énoncé ;
4. contraintes sur la hiérarchisation des tests liées à l'automatisation.

Dans ce cadre restrictif, le problème est de proposer des équivalents cibles qui reproduisent le plus possible, **contexte par contexte**, les effets de sens du verbe source.

Plusieurs cas se présentent :

1. Il existe un équivalent cible exact :

(20) a. LIE + Adj. = ABSENCE DE DYNAMIQUE => 'rester'
Fields ... were LYING **fallow**
Les champs... **restaient** en friche
b. RUN DOWN + GN2.
Si GN1 = ÉLÉMENT LIQUIDE => 'couler le long de'
... The **tears** were RUNNING DOWN her cheeks
... Les larmes **coulaient le long de** ses joues
c. WORK + GN1 = SUBSTANCE => 'agir'
... to know why and how the **aspirin** WORKS.
... savoir pourquoi et comment **agit** l'aspirine

2. On a le choix entre un terme français **exact** mais ne couvrant pas tous les cas attestés, et un équivalent **moins précis** mais partout recevable. On privilégie alors ce dernier :

(21) a. The **meeting** of the council WAS HELD at Coldfield
La réunion du conseil **eut lieu** à Coldfield
b. The **century dinner** WAS HELD at Eton
Le dîner du centenaire **eut lieu** à Eton

Hold au passif + GN2 => *'phénomène social temporaire'* => 'avoir lieu'. L'équivalent 'se tenir' serait plus précis dans (21)a, mais ne convient pas à (21)b.

3. On choisit le terme acceptable pour la majorité des cas en sachant que quelques exemples auront une traduction **approximative** ou même **incorrecte**.

(22) a. ... stinking pus that DREW buzzing **flies**
... du pus nauséabond qui **attirait** des mouches bourdonnantes
b. He DREW **birds** in his copy-book
Il ***attirait** les oiseaux dans son cahier

La règle majoritaire : *draw* + *'animal'* = > 'attirer' donne une traduction correcte dans le premier cas mais incorrecte dans le second.

(23) a. To RUN THROUGH the **books** in chronological sequence is...
Parcourir les livres dans l'ordre chronologique est...
b. Professional stars ... RAN THROUGH their **words** 5 or 6 times
Les vedettes professionnelles **parcouraient** leur texte 5 ou 6 fois

Run through + GN2 => *'porteur d'information'* => 'parcourir'. La traduction est correcte dans le premier exemple, approximative mais compréhensible dans le second où 'revoir' ou 'répéter' serait meilleur.

4. Sur la masse des cas examinés il en reste toujours un certain nombre pour lesquels on n'a pas pu dégager de critères contextuels. On suggère alors une traduction «par défaut». Deux possibilités s'offrent :

a) On cherche s'il existe un «cover word» suffisamment neutre pour couvrir les **exemples qui n'ont pu être traités.** Ainsi, dans le cas de *work*, le verbe français 'marcher' se revèle être un équivalent par défaut satisfaisant pour les exemples restant après qu'ont été épuisés les tests contextuels conduisant sur 'travailler', 'agir' (exemple (20)c), 'fonctionner' (exemple (3)), etc.

(24) a. His plan WORKED perfectly
Son plan **marcha** parfaitement

De même les séquences *walk up* + GN et *walk up to* se traduisent respectivement par 'monter', 'aller jusqu'à'. Partout ailleurs la traduction par défaut 'arriver' est considérée comme acceptable : même s'il existe, cas par cas, des traductions plus précises, elles ne sont pas (pour l'instant) calculables :

(24) b. Mickey WALKED UP and volunteered an anecdote.
Mickey **arriva** et y alla d'une anecdote

b) On propose deux (ou trois) traductions, laissant au lecteur / réviseur le soin de trancher à l'aide de sa compréhension du contexte:

(25) a. Steve DREW a glass of water
Steve **tira / dessina** un verre d'eau
b. Shall we flip a coin to see which of us GOES first
Est-ce qu'on joue à pile ou face pour voir lequel de nous **y va / part** le premier

Une vérification manuelle sur les corpus a permis une évaluation des traductions obtenues. Sur l'ensemble des deux expériences, les résultats donnent satisfaction dans plus des trois quarts des cas, la fiabilité des traductions pour la première expérience atteignant en réalité 85 %. C'est avec des verbes comme *go* (et dans une moindre mesure *move*) que se rencontrent le plus de traductions approximatives ou erronées. La raison en est que ces verbes ont un éventail sémantique très ouvert, et qu'on ne peut souvent cerner leur valeur d'emploi (et donc choisir la traduction adaptée) sans faire appel à un contexte élargi, voire à la situation extra-linguistique, chose possible pour un traducteur humain, mais irréalisable dans le cadre de cette expérience. Ainsi, dans le cas de *move*, il est difficile de distinguer entre mouvement sur place ('bouger', 'se tourner'...).

(26) He did not MOVE except to roll a little
Il ne **bougea / se tourna** pas si ce n'est pour rouler un peu

ou déplacement dans l'espace ('se déplacer', 'partir').

(27) You'd better MOVE a little or you'll get ... squashed flat
Vous feriez mieux de **vous déplacer / partir** un peu ou vous allez vous faire écraser.

Une séquence GN1 *go to* GN2 pose le même problème. Si un test GN1 = *'humain'*, GN2 = *'humain cat. socio-professionnelle'* permet de traduire correctement :

(28) ... she WENT TO a judge to see what could be done
... elle **alla voir** un juge pour voir ce qui pouvait être fait

par 'aller voir', en revanche, dans les deux énoncés suivants, faute de précision sur la situation extérieure, il est impossible de trancher entre 'aller voir' et 'aller vers' :

(29) a. At dawn Argiento WENT TO Jacopo Galli
À l'aube Argiento **alla voir** Jacopo Galli
b. After a minute he WENT TO Roberts, looked at one of his eyes and felt for a pulse
Une minute après il **alla vers** Roberts, examina l'un de ses yeux et chercha son pouls

La traduction par défaut *go to* GN2 = 'aller (jusqu')à', acceptable bien qu'approximative dans le deuxième énoncé, est incorrecte dans le premier.

De même, dans la phrase :

(30) If somebody were to drop a match here, this place would GO UP like a haystack
Si quelqu'un laissait tomber une allumette ici, cet endroit **s'enflammerait** comme une meule de foin

là où l'anglais se contente de noter un simple mouvement vers le haut (*go up*), le français demande une traduction très spécifique ('partir en flammes' 's'enflammer'...) suggérée par la présence de *match* et *haystack*, mais non détectable à partir des tests prévus. La double traduction par défaut 'monter / aller' se révèle incorrecte.

Les postpositions, de leur côté, font problème. Elles peuvent parfois traduire une intention de l'énonciateur. Ainsi, dans les exemples suivants :

(31) a. You go DOWN to this Cobbold boy
Tu **vas** vers le jeune Cobbold
b. Speaking in a low voice of loathing, she went UP to the girls
Parlant d'une voix basse et haineuse, elle **alla** vers les jeunes filles
c. The colonel would not go DOWN to see his men
Le colonel ne voulut pas **aller** voir ses hommes
d. I went UP to see him only a few days ago
Je **suis allé** le voir il y a quelques jours seulement

une modalisation est apportée par l'énonciateur sur le mouvement, mais aucun critère contextuel ne permet de traduire *go up* ou *go down* par autre chose qu'un simple 'aller'.

L'impossibilité, dans certains cas, de distinguer entre les différentes valeurs (spatiales, aspectuelles, énonciatives) acquises en contexte a donc conduit à proposer deux (parfois trois) traductions.

4. DESCRIPTION DE L'EXPÉRIENCE CONDUITE SUR L'ESPAGNOL

Une expérience similaire a été tentée sur le transfert espagnol-français de quelques verbes. Disposant de corpus beaucoup moins étendus[5], elle a été conduite dans un esprit quelque peu différent. Il s'agissait non pas tant de viser l'exhaustivité (trouver tous les équivalents français d'un verbe espagnol) que d'examiner les convergences avec l'étude déjà réalisée sur l'anglais. Autrement dit, dans quelle mesure les descripteurs déjà créés pour l'anglais étaient-ils généralisables ? Quels autres étaient à créer ? À travers la comparaison entre deux expériences, on s'oriente déjà vers la recherche de classes sémantiques universelles.

4.1. LES VERBES RETENUS

Ils sont au nombre de douze (*componer, tratar, llamar, manejar, poner, cargar, levantar, mover, llevar, sacar, mantener, adquirir*). Ils appartiennent à des classes sémantiques variées : modification de l'objet direct (*componer*), non modification d'un objet, le plus souvent humain (*tratar, llamar*), mouvement (*poner, cargar, levantar, mover, llevar, sacar*), possession (*mantener, adquirir*) et manipulation d'instrument (*manejar*).

Ces verbes diffèrent les uns des autres quant à leur degré de polysémie. La moitié d'entre eux requièrent une dizaine d'équivalents français. *Poner, cargar, levantar* et *llevar* en admettent une vingtaine. À l'exception de *manejar* et *adquirir*, tous ces verbes ont des constructions syntaxiques diversifiées.

4.2. LA MÉTHODE ADOPTÉE

Comme pour l'anglais, sont associées syntaxe et sémantique. Sont retenus comme éléments à explorer tous les éléments du schéma actanciel :

(32) a. *levantar algo* : lever quelque chose
b. *cargarle a uno con algo* : accuser quelqu'un de quelque chose

On a procédé à un premier classement syntaxique :

niveau 1 : + GN2
niveau 2 : + GN2 + préposition + GN3
niveau 3 : + A + GN2 (HUMAIN)
niveau 4 : + A + GN2 (HUMAIN) + préposition + GN3
etc.

Les prépositions sont elles aussi examinées : *a, de, con, en, por,* etc. Les renseignements d'ordre sémantique qu'elles apportent peuvent parfois suffire pour trouver la traduction française d'un verbe espagnol :

(33) a. *mantener a uno* = *entretenir quelqu'un*
b. *tratar en algo* = *faire le commerce de quelque chose*
c. *sacarle algo a uno* = *obtenir* quelque chose de quelqu'un

Mais, dans la majorité des cas, on recourt, comme pour l'anglais, aux indications sémantiques données par les descripteurs affectés aux unités lexicales du contexte actanciel. La seule différence avec l'anglais réside dans le fait qu'on peut être amené à utiliser non pas un mais deux descripteurs affectés à la même unité lexicale (v. exemple (43)b, (47)a,b,c).

Exemples de niveau 1 :

(34) a. LEVANTAR *la mano* (PARTIE DU CORPS) = **lever** la main
b. LEVANTAR *un liebre* (ANIMAL) = **lever** un lièvre (à la chasse)
c. LEVANTAR una casa (CONSTRUCTION) = **construire** une maison
d. LEVANTAR la piedra (OBJET) = **soulever** la pierre
e. LEVANTAR odio (AFFECTIVITÉ) = **susciter** de la haine
f. LEVANTAR ruido (BRUIT) = **provoquer** du bruit

Exemples de niveau 2 :

(35) a. SACAR la espada (OBJET) de la vaina (CONTENANT) = **tirer** l'épée du fourreau
b. SACAR la funda (CONTENANT) del paraguas (OBJET) = **retirer** la gaine du parapluie

Les écueils liés à la mise au point de cette méthode sont du même ordre en espagnol qu'en anglais. La détermination des lexies pose les mêmes problèmes. Le degré de lexicalisation est en raison inverse de la création de descripteurs adaptés. Les descripteurs s'avèrent notamment peu efficaces lorsqu'on traite d'expressions lexicalisées :

(36) a. SACAR la lengua (PARTIE DU CORPS) = **sortir** la langue
b. SACAR la barbilla (PARTIE DU CORPS) = **sortir** le menton
c. SACAR el busto (PARTIE DU CORPS) = **sortir** le torse

au lieu de 'tirer' la langue, 'redresser' le menton, 'bomber' le torse.

4.3. LES DESCRIPTEURS RÉUTILISÉS

Les **deux tiers** environ des descripteurs créés pour l'anglais (les deux expériences confondues) sont réutilisés pour l'espagnol. Ici aussi, leur rentabilité, calculée en fonction du nombre de verbes concernés par chacun d'eux, est variable (voir **Annexe II**).

1. Vient en tête le descripteur *'humain individu'* (présent pour tous les verbes). Ensuite on a, dans l'ordre, *'abstrait'* (9 verbes), *'objet'* (7 verbes), *'partie du corps'* (6 verbes) :

(37) a. CARGAR con un paquete (OBJET) = **porter** un paquet
b. CARGAR con una responsabilidad (ABSTRAIT) = **se charger** d'une responsabilité
c. Este caldo te COMPONDRA el estomago (PARTIE DU CORPS) = ce bouillon te **remettra** l'estomac
d. La bala le LLEVO la cabeza (PARTIE DU CORPS) = la balle lui **arracha la tête**

2. Puis, curieusement, cinq des douze verbes étudiés, tous de mouvement, utilisent un descripteur lié à une vision subjective, *'connotation péjorative'*, et un autre, très spécifique, *'finance'* :

(38) a. Nos PUSIERON muchos deberes (CONN PÉJ.) = On nous a **imposé** beaucoup de devoirs
b. Los tributos (FINANCE) CARGAN sobre la riqueza = Les impôts **pèsent** sur la richesse

3. La suite du classement réserve moins de surprises : *'affectivité'*, *'animal'*, *'catégorie socio-professionnelle'*, *'collectif humain structuré'*, *'processus'*, *'machine'* (4 verbes).

L'expérience de l'espagnol confirme le bien-fondé de la subdivision des descripteurs liés à l'humain (*'humain individu'*, *'catégorie socio-professionnelle'*, *'collectif humain structuré'*) :

(39) a. SACAR + GN2 (HUMAIN) = **sortir** ou inviter
b. SACAR + GN2 (CAT. SOCIO-PROF.) = **nommer**
c. LLEVAR + GN2 (HUMAIN) = **conduire**
d. LLEVAR + GN2 (COLL. HUM. STRUCT.) = **diriger**

Les descripteurs *'affectivité'*, *'processus'* et *'machine'* sont essentiellement liés à des verbes de mouvement :

(40) Este bordado (ACTIVITÉ) LLEVA mucho trabajo (PROCESSUS)
= cette broderie **demande** beaucoup de travail

Le descripteur *'affectivité'* permet de dégager des significations figurées, telles que 'provoquer', 'susciter', 'inspirer'.

Le sémantisme très spécifique de *'machine'* oriente la traduction du mouvement vers le fonctionnement :

(41) MOVER la turbina (MACHINE) = **faire marcher** la turbine

4. La rentabilité de descripteurs aussi spéciaux que *'phénomène social temporaire'* et *'porteur d'information'* (3 verbes) se trouve confirmée :

(42) a. Le SACARON en las Fallas (PHÉNOMÈNE SOCIAL TEMPORAIRE)
= on l'a **montré** aux fêtes de Saint-Joseph
b. Le PUSE un telegrama (PORTEUR D'INFORMATION) = je lui ai **envoyé** un télégramme

5. Les autres descripteurs de l'anglais réutilisés pour l'espagnol le sont dans une moindre mesure (voir **Annexe II**). Notons que les descripteurs *'voie de passage'*, *'moyen de transport'* et *'lieu'* sont relativement peu employés en dépit de la forte proportion de verbes de mouvement étudiés.

Si ce classement de rentabilité n'a pour l'instant qu'une valeur indicative, en revanche la liste elle-même des descripteurs réutilisés, dans la mesure où elle est confirmée par une double expérience, revêt le plus grand intérêt.

4.4. DESCRIPTEURS CRÉÉS POUR L'ESPAGNOL

Aux 36 descripteurs empruntés à l'expérience anglaise s'ajoute une vingtaine d'autres créés pour l'espagnol. Ils sont, comme les précédents, très hétérogènes tant du point de vue de leur contenu sémantique que du degré de généralité.

1. Quelques-uns sollicitent des classes sémantiques très voisines des classes déjà prévues pour l'anglais.

À côté de :

'absence de dynamique'	l'espagnol ajoute	*'dynamique'*
'élément liquide'		*'élément solide'*
'avantage'		*'bien matériel'*
'sol + produit du sol'		*'plante'*

(43) a. Nosotros LLEVABAMOS un ritmo (DYNAMIQUE) acelerado = Nous **maintenions** un rythme accéléré
b. COMPONER una ensalada (ALIMENT + SOLIDE) = **préparer** une salade
c. Ya en Abril empiezan a MOVER los cerezos (PLANTE)
= En avril déjà, les cerisiers commencent à **bourgeonner**

2. Deux descripteurs sont issus d'une reclassification de descripteurs très généraux. Le descripteur *'concret'*, que l'anglais n'interrogeait pas en tant que tel, s'est révélé utile :

(44) a. La cubierta (CONCRET) CARGA sobre los pilares = la toiture **repose** sur les piliers

s'oppose à :

b. Los impuestos (FINANCE) CARGAN sobre la riqueza = les impôts **pèsent** sur la richesse

La classe *'animé'* a été instaurée pour les cas de neutralisation de l'opposition *'humain'* # *'animal'* :

(45) Les (ANIMÉ) LLEVARE por el camino mas corto = je les **conduirai** par le chemin le plus court

3. Les autres descripteurs créés pour l'espagnol peuvent se classer en trois groupes d'importance décroissante : ceux qui s'appliquent indifféremment à toutes les classes sémantiques de verbe énumérées ci-dessus, ceux qui sont plus particulièrement liés à une de ces classes et enfin, ceux qui, jusqu'ici, n'ont été rencontrés qu'avec un seul verbe.

Au groupe 1 appartient le descripteur *'totalité'* :

(46) a. Jose SUELE cargar con la maleta (OBJET) del amo
= José a l'habitude de **porter** la valise du patron

s'oppose à :

b. CARGO con todos (TOTALITÉ) mis libros (OBJET) = Il m'**a pris** tous mes livres

Le groupe 2 inclut des descripteurs figurant dans le contexte de verbes de déplacement dans l'espace : *'contact'*, *'verticalité'*, *'construction'* et *'installation provisoire'*.

(47) a. LEVANTAR el mantel (OBJET + CONTACT) = **enlever** la nappe
b. LEVANTAR la cosecha (SOL + CONTACT) = **faire** la récolte
c. LEVANTAR los postes (OBJET + VERTICALITÉ) = **relever** les poteaux
d. LEVANTAR una casa (CONSTRUCTION) = **construire** une maison
e. LEVANTAR una tienda (INSTALLATION PROVISOIRE) = **démonter** une tente

Le descripteur *'espace de temps'* transforme le verbe de mouvement *llevar* en verbe de durée :

(48) LLEVABAMOS tres dias (ESPACE DE TEMPS) en Paris = Nous **étions** à Paris depuis trois jours

Les descripteurs *'défaut'*, *'qualité'*, *'maladie'* s'associent plus particulièrement aux verbes d'objet affecté humain :

(49) a. Le TRATO la bronquitis (MALADIE) en poco tiempo = il lui a **soigné** sa bronchite en peu de temps
b. Le LLAMARON tonto (DÉFAUT) = on l'**a traité** d'idiot
c. Te LLAMARON guapa (QUALITÉ) = on **t'a dit** que tu étais belle

Le descripteur *'apparence'* se retrouve avec des verbes de possession :

(50) Su amigo ADQUIRIO un aspecto (APPARENCE) sombrio = son ami **prit** un air sombre

Enfin, le verbe instrumental, *manejar*, a entraîné la création de *'instrument'*, distinct de *'objet'* ou *'machine'* :

(51) a. MANEJAR la espada (INSTRUMENT) = **manier** l'épée
b. MANEJAR pucheros (OBJET) = **manipuler** des pots
c. MANEJAR un barco (MACHINE) = **diriger** un bateau

Dans le groupe 3, on citera seulement le cas de *llevar* employé pour tout ce qu'on **porte** sur soi (vêtements, armes, lunettes, mais aussi puces, poux, et autres parasites !). D'où le descripteur *'parasite'* :

(52) a. LLEVAR un traje (VÊTEMENT) = **porter** un costume
b. LLEVAR pulgas (PARASITE) = **avoir** des puces

Les descripteurs créés pour l'espagnol sont d'importance inégale (v. **Annexe III**). Une bonne partie des notions essentielles a déjà été générée à propos de l'anglais.

5. RÉFLEXIONS SUR LES EXPÉRIENCES PRÉSENTÉES

Les classes sémantico-cognitives dégagées n'ont pour l'instant qu'un statut très provisoire. Leur survie dépend de leurs possibilités de réemploi. Ce sont des molécules qui devront, au fur et à mesure de l'extension de l'expérience, éclater en atomes et qu'il faudra ensuite recomposer (redéfinir) autrement et en nombre fini à partir de la combinatoire des atomes. Faudra-t-il, par exemple, éclater le descripteur *'collectif humain structuré'* en *'collectif'* + *'humain'* + *'structure'*. Dans ce cas, le trait *'structure'* pourrait entrer comme composante dans le descripteur *'machine'*; il en va de même du trait *'humain'*, qu'on retrouverait aussi dans *'objet fabriqué'*... Parallèlement, le descripteur affecté au vocable *'système'* inclurait le trait *'structure'* (combinable avec d'autres traits). Les descripteurs *'source d'énergie'*, *'porteur d'information'*, *'porteur de virtualités'*, *'générateur de lumière / chaleur'* ont-ils en commun (à partir des notions de source, porteur, générateur) un trait comme *'implication'*? *'Repérage spatial'*, pour sa part, n'est sans doute qu'un cas particulier du descripteur (plus général) *'mise en relation'*.

On répondra, en outre, à deux objections. La première consisterait à mettre en doute la valeur de généralité des descripteurs. L'expérience espagnole, au contraire, tend à montrer qu'il existerait, d'une part, un **noyau stable** de descripteurs assez généraux pour dépasser le stade du transfert vers une seule langue-cible: les deux tiers des descripteurs de l'anglais sont réutilisés pour l'espagnol (v. **Annexe II**); d'autre part, une **zone** plus **floue** de descripteurs utilisés soit pour l'une soit pour l'autre langue, mais beaucoup moins rentables (comparer **Annexes I** et **III**).

La seconde objection porterait sur le trop grand nombre de descripteurs à créer. Ne risque-t-on pas d'avoir à générer presque autant de classes que de cas à résoudre? L'expérience, ici aussi, prouve que non. On constate très vite, après le traitement de quelques verbes seulement, une certaine stabilisation et la réutilisation des mêmes descripteurs principaux (voir détails en **Annexes I** et **II**).

S'il n'est pas fait appel, dans cette étude, au sémantisme du verbe à traduire, on n'en constate pas moins des phénomènes de compensation entre contenu sémantique du verbe et contenu du descripteur affecté à tel ou tel actant. On peut penser que les caractéristiques sémantiques respectives du verbe et du descripteur actanciel sont en raison inverse l'une de l'autre: plus le sémantisme du lexème verbal tendra vers le général, plus la spécificité des descripteurs nécessaires aux prises de décision augmentera (v. plus haut, exemples 52a, b). Il est néanmoins un seuil de généralité au-delà duquel le contexte actanciel à lui seul ne pourra suffire (v. ci-dessus, 3.3.4., ce qui est dit sur *go* et *move*).

Dans le choix des critères de création de descripteurs, on a surtout visé en effet jusqu'ici l'environnement lexical. Les critères syntaxiques ou lexico-sémantiques ne sont évidemment pas les seuls. S'agissant de traduire du verbal, la prise en compte de données grammaticales telles que l'aspect, le modal, le temps (*time* et *tense*), la personne, la voix, ainsi que le nombre et le degré de détermination et de définitude des lexèmes actants peut se révéler indispensable. Par exemple, dans les phrases suivantes:

(53) a. The captain ... was fighting to HOLD his seat...
Le capitaine... se battait pour ***tenir** son siège...
b. A critic striving to HOLD his franchise...
Un critique s'efforça de ***détenir** son contrat

la prise en compte de la valeur aspecto-modale (effort...) apportée par les verbes *fight, strive*, aurait permis d'opter pour les équivalents français 'garder' ((53)a) ou 'conserver' ((53)b), tandis que la seule exploration des traits sémantiques des groupes nominaux à fonction actancielle oriente sur les très peu satisfaisants 'tenir' (phrase (53)a) et 'détenir' (phrase (53)b).

Jusqu'où, cependant, pourra-t-on reculer les limites de l'automatisable? Et corrélativement: dans quels cas faut-il (ou peut-on) **élargir** le contexte pertinent? Jusqu'où faut-il étendre le corpus de référence? Ces deux problèmes sont étroitement liés. Plus on dispose pour un verbe donné (et plus précisément pour un couple signifiant-source/signifiant-cible) de contextes d'emploi en nombre très élevé et plus on est en mesure de voir apparaître, et donc de capter, des régularités non soupçonnées. Aurait-on tiré les mêmes conclusions d'un dépouillement des cent millions d'occurrences du corpus COBUILD de Birmingham (au lieu des trois millions et demi d'occurrences des corpus décrits ci-dessus?). L'abondance de données sur un problème pointu s'apparente au grossissement du microscope. C'est en parcourant l'infra-perceptuel (révélé par l'observation microscopique) qu'on peut avancer dans la recherche de l'automatisable et donc dans l'élaboration de règles contextuelles pour la traduction. Ces règles peuvent demander l'exploration (sémantique ou syntaxique) d'environnements beaucoup plus vastes que ceux qui sont retenus ici.

Si on a choisi à la fois de se concentrer sur le lexique **non terminologique** et de ne pas se limiter à un domaine de spécialité particulier, c'est afin de mieux exploiter la problématique et notamment l'hypothèse sous-jacente: il existe quelque part dans tout texte des régularités sémantico-cognitives que les opérations de traduction peuvent contribuer à déceler. Il n'est pas nécessaire cependant d'avoir obtenu des résultats complets à ce niveau de généralité pour passer à des applications ponctuelles. On peut déjà penser à coupler le genre d'approche non terminologique préconisé ici avec les dictionnaires de micro-domaines spécifiques. En combinant les certitudes du vocabulaire de spécialité et la rigueur de l'exploration contextuelle imposée par le dictionnaire général, on diminuerait considérablement les risques d'erreur et on améliorerait la qualité des traductions.

6. DE L'HUMAIN À L'HUMAIN VIA L'AUTOMATE

Le dictionnaire général de transfert présenté ici se distingue du dictionnaire bilingue humain de type classique sur plusieurs points essentiels. Le plus original est sans doute le raffinement et la systématisation des critères de contextualisation et la plus grande attention portée à leur fréquence: le dictionnaire de transfert hiérarchise là où le dictionnaire humain uniformise. Un autre trait différenciateur est la précision et la diversité des descripteurs, voire le caractère quelque peu insolite de certains. Enfin l'étude contextuelle non seulement fait apparaître des configurations orientant sur des choix de traductions non répertoriées dans le dictionnaire humain, mais encore, à la différence de celui-ci, les traductions proposées, tout en étant plus diversifiées, excluent par définition toute synonymie: aucun des équivalents cibles proposés n'est substituable à un autre. Tous ces traits résultent de la finalité même du projet: on a cherché à construire quelque chose qui soit un véritable **dictionnaire contextuel interactif**.

La notion d'interactivité suggère un dialogue (questions-réponses) entre partenaires, ici entre le dictionnaire et le programme général. Un dictionnaire interactif se prend en quelque sorte lui-même en main et dirige la procédure de consultation, ne livrant sa décision qu'une fois en possession de l'information contextuelle qui la conditionne.

C'est à ce stade qu'on peut se demander s'il y aurait avantage à introduire l'utilisateur humain comme partenaire dans ce processus. On sait que le paramètre contextuel préconisé par la présente approche tient son originalité de ce qu'il s'appuie sur des données langagières reflétant les tendances d'une discursivité authentique. Les descripteurs sémantiques ne sont pas définis **a priori** mais sont extraits de l'univers cognitif et culturel vivant, et la méthode s'oriente vers une symbiose entre l'humain et l'automate: on cherche à la fois à «automatiser» l'humain et à «humaniser» l'automate.

On s'est aperçu, en effet, lors des premiers tests en grandeur réelle, que l'outil, conçu pour un automate et encore imparfait de ce point de vue, pouvait, s'il était manipulé par un être humain, rendre bien au-delà de ce qui avait été prévu, en particulier parce que l'être humain jouit d'un accès au cognitif qu'un automate ne possède pas pour l'instant. Par delà la «lettre» des règles, l'être humain, à l'inverse de l'automate, est à même d'en dégager l'«esprit». Ainsi, pour un automate informatisé, l'étiquette d'un descripteur sémantique n'est guère autre chose qu'un code renvoyant à la liste des mots auxquels il a été affecté : le libellé du descripteur pourrait tout aussi bien être un chiffre. Au contraire, l'automate humanisé (c'est-à-dire le traducteur humain recourant aux règles d'un éventuel programme automatisé) peut traiter le descripteur comme un indicateur d'une valeur notionnelle qu'**il appréhende alors en tant que telle** : il part ensuite de cette valeur et la compare intuitivement à la valeur locale du mot. Les règles de l'automate, si elles sont guidées par la sensibilité culturelle de l'homme, se révéleront parfois plus efficaces que lorsqu'elles s'appuient sur les seules informations données par le dictionnaire automatique.

Par delà l'automate, l'être humain traducteur rejoint alors la visée du concepteur (lui aussi humain) du programme. Un traducteur (surtout chevronné) qui accepterait de jouer le jeu de l'automate prendrait peu à peu conscience de toute une activité d'exploration sémantico-contextuelle à laquelle il se livre inconsciemment quand il traduit. Cette prise de conscience le pousserait alors à s'expliciter pour lui-même cette activité et participer à son tour à l'affinement ultérieur de l'automate. C'est la raison pour laquelle on a choisi de s'étendre un peu longuement (dans les sections 3. et 4.) sur des expériences en cours : leur genèse éclaire la manière dont peut naître la symbiose entre l'homme et l'automate.

Les pronostics ci-dessus sont évidemment à nuancer. Les utilisateurs humains forment une population très hétérogène. Les adaptations différeront selon que la traduction se fait depuis la langue étrangère vers la langue maternelle ou vice versa, selon le degré de compétence dans la langue étrangère et, dans le cas des textes de spécialité, selon le degré de familiarité avec le domaine. Par exemple, s'il traduit vers sa langue maternelle mais à partir d'une langue étrangère qu'il connaît très mal, le traducteur humain ne sera guère mieux armé que l'automate pour déterminer intuitivement l'affinité de tel mot du contexte avec tel ou tel descripteur : il devra, comme l'automate, opérer au niveau du code du lexème tel qu'il est donné dans le dictionnaire. En revanche, grâce à sa connaissance de la langue cible, il retrouve, en fin de parcours, sa supériorité et pourra affiner les traductions proposées.

Il ne devrait donc pas être impossible de concevoir et de construire un outil d'aide à la décision, qui puisse à la fois s'insérer comme module dans un programme de TAA et, indépendamment, répondre, moyennant divers ajustements, aux multiples besoins de traducteurs humains. Perspective d'autant plus souhaitable que — quel que soit le point du continuum (évoqué en 1.) où prennent naissance procédures et outils — l'utilisateur du produit demeure (même au terme de l'étape IV) le traducteur humain.

Plus que d'une gamme d'outils de TAO conçus indépendamment, c'est d'un **méta-outil** (générateur à son tour d'outils plus spécifiques) que le traducteur humain devrait pouvoir disposer. Il y aurait donc intérêt, dès avant l'élaboration de procédures étroitement finalisées ou informatisées, à déterminer et à unifier le plus possible **en amont** les principes de résolution automatique d'un problème donné. Automatiser n'est pas informatiser. Les problèmes de l'automatisation **antécèdent** ceux de l'informatisation. Un automate est essentiellement une suite hiérarchisée et strictement codifiée d'opérations de calcul conduisant à diverses prises de décision. Il peut ensuite s'actualiser selon le cas soit comme un guide de conduite qu'un être humain, agissant seul (c'est-à-dire en dehors

de toute contrainte informatique), s'impose et dont il s'astreint à exécuter les règles, soit au contraire comme un programme informatisé entièrement autonome, soit encore comme une synergie homme-machine. Le véritable gradient est moins celui — horizontal — des modes d'actualisation de l'automate que celui — vertical — des stades préliminaires à cette actualisation.

ANNEXE I

LISTE DES DESCRIPTEURS UTILISÉS DANS L'EXPÉRIENCE SUR L'ANGLAIS (AVEC EXEMPLES DE MOTS CONCERNÉS)

(Présentés en ordre alphabétique, sans hiérarchisation notionnelle)

Le chiffre à gauche d'un descripteur indique que le descripteur est utilisé soit dans la première expérience (<1>), soit dans la deuxième (<2>), soit dans les deux à la fois (<1> <2>).

<1>		ABSENCE DE DYNAMIQUE (*becalmed, stagnant, untouched...*)
<1>	<2>	ABSTRAIT (*consequence, interest, leadership...*)
<1>		AFFECTIVITÉ (*devotion, hatred, fury...*)
	<2>	ALIMENT (*drink, salami...*)
<1>	<2>	ANIMAL (*horse, pet, cattle...*)
<1>		APPRÉCIATIF (*precious, meaningless, worthy...*)
<1>		ATTITUDE + DÉPLACEMENT (*smirk, follow, watch...*)
<1>	<2>	AVANTAGE (*monopoly, title, power...*)
	<2>	BATEAU (*liner, ship...*)
<1>		BRUIT (*buzz*)
	<2>	COMPORTEMENT (*insane, loose, free...*)
<1>	<2>	CONNOTATION PÉJORATIVE (*poverty, danger, obstacle...*)
<1>	<2>	CONTENANT (*bag, socket, church...*)
<1>	<2>	COULEUR (*colour, green...*)
<1>	<2>	DIRECTION (*north, westwards, right...*)
	<2>	DURÉE (*endlessly, forever...*)
	<2>	ÉLÉMENT LIQUIDE (*bath, paraffin, milk...*)
	<2>	ÉLÉMENT LONGITUDINAL (*column, road, line...*)
	<2>	ÉLÉMENT DE SÉPARATION (*barrier, membrane...*)
<1>	<2>	ENTITÉ SPATIALE + GÉOGRAPHIQUE + GÉOLOGIQUE (*territory, glacier, cretaceous arc, island...*)
<1>	<2>	FINANCE (*mortgage, cost, gold, share...*)
<1>	<2>	GÉNÉRATEUR DE LUMIÈRE + CHALEUR (*lamp, burner, fluorescent tube...*)
<1>	**<2>**	**HUMAIN** recouvrant :
<1>	<2>	HUMAIN : CATÉGORIE SOCIO-PROFESSIONNELLE + PARENTÉ + NATIONALITÉ (*Californian, technologist, wife, driver, legislator...*)
<1>	<2>	HUMAIN : COLLECTIF HUMAIN STRUCTURE (*committee, tribe, government...*)
	<2>	HUMAIN : INDIVIDU (*girl, partner, immigrant...*)
	<2>	INITIATIF (*beginning...*)
<1>	<2>	LIEU (*countryside, locality, neighbourhood...*)
<1>	<2>	MACHINE + ÉLÉMENT DE MACHINE (*air conditioner, turbine, rotor, wheel, piston...*)
<1>		MANIÈRE (*breathlessly, differently...*)
<1>	<2>	MESURE : UNITÉ + NOTION MESURABLE (*temperature, speed, level, inch, degree...*)
<1>		MISE EN RELATION (*comparison, analogy, balance...*)
<1>	<2>	MOYEN DE TRANSPORT (*ambulance, phaeton, bus, bicycle...*)
<1>	<2>	OBJET (*stone, razor, branch, chunk...*)
<1>		OBJET = OBJET FABRIQUÉ (*stocking, axe, handcuffs...*)
	<2>	OCCUPATION + ACTIVITÉ (*business, task, job...*)
	<2>	ONOMATOPÉE (*aaawww...*)
<1>		OPINION (*assumption, belief, view...*)
<1>	<2>	PARTIE DU CORPS (*heel, lip, muscle,...*)
<1>		PARTIE DU CORPS : TÊTE ET MEMBRES (*hand...*)
<1>		PHÉNOMÈNE NATUREL (*wave lift...*)
<1>	<2>	PHÉNOMÈNE SOCIAL TEMPORAIRE (*conference, auction, exhibition, party, rehearsal...*)

<1>	<2>	PORTEUR D'INFORMATION (*data, book, wisdom, riddle, sciences, memory...*)
<1>	<2>	PORTEUR DE VIRTUALITÉS (*hope, menace, possibility, promise, seed...*)
<1>	<2>	PROCESSUS (*erosion, growth, development, movement...*)
<1>		REPÉRAGE SPATIAL (*adjacent, sandwiched, uppermost...*)
<1>		REPRÉSENTATION NON SCHÉMATIQUE (*cartoon, portrait, illustration...*)
<1>		REPRÉSENTATION SCHÉMATIQUE (*graph, line, curve, outline...*)
<1>	<2>	SOCIOPROFESSIONNEL ABSTRAIT (*politics, trade, dentistry...*)
<1>		SOL + PRODUIT DU SOL (*coal, potato, clod, mine...*)
	<2>	SOURCE D'ÉNERGIE (*battery, atomic power, gas...*)
<1>	<2>	SUBSTANCE + ENTITÉ CHIMIQUE + BIOLOGIQUE (*acid, drug, enzyme, molecule, pheromone...*)
<1>	<2>	TEMPS (*moment, morning, life, september...*)
<1>		TERMINATIF (*end, halt, standstill...*)
<1>	<2>	VÊTEMENT + PARTIE DE VÊTEMENT (*sarong, cuff, slipper...*)
<1>	<2>	VOIE DE PASSAGE (*entrance, stairs, path...*)
	<2>	VOYAGE (*stroll, touring holiday, trip...*)

ANNEXE II

RENTABILITÉ DES DESCRIPTEURS DE L'ANGLAIS RÉUTILISÉS EN ESPAGNOL

(évaluée d'après le nombre de verbes concernés)

12 verbes : HUMAIN
09 verbes : ABSTRAIT
07 verbes : OBJET
06 verbes : PARTIE DU CORPS
05 verbes : CONNOTATION PÉJORATIVE, FINANCE
04 verbes : AFFECTIVITÉ, ANIMAL, CAT. SOCIO-PROFESSIONNELLE, COLLECTIF HUMAIN STRUCTURE, PROCESSUS, MACHINE
03 verbes : PHÉNOMÈNE SOCIAL TEMPORAIRE, PORTEUR D'INFORMATION
02 verbes : ALIMENT, BRUIT, CONTENANT, DIRE, ENTITÉ SPATIALE, MANIÈRE, MESURE, OBJET FABRIQUÉ, ACTIVITÉ, OPINION, SOL + PRODUIT DU SOL, MOYEN DE TRANSPORT, SUBSTANCE, VÊTEMENT, VOIE DE PASSAGE
01 verbe : AVANTAGE, DIRECTION, ÉLÉMENT LIQUIDE, LIEU, PHÉNOMÈNE NATUREL, SOCIO-PROFESSIONNEL ABSTRAIT, TEMPS

ANNEXE III

RENTABILITÉ DES DESCRIPTEURS CRÉÉS POUR L'ESPAGNOL

(évaluée d'après le nombre de verbes concernés)

06 verbes : CONCRET, ANIMÉ
04 verbes : BIEN MATÉRIEL
03 verbes : DÉFAUT, DYNAMIQUE, ESPACE DE TEMPS
02 verbes : APPARENCE, CONSTRUCTION, MALADIE, PLANTE, QUALITÉ, TOTALITÉ
01 verbe : CONTACT, ÉLÉMENT SOLIDE, INSTALLATION PROVISOIRE, INSTRUMENT, PARASITE, VERTICALITÉ, VOLUME

Notes

1. Le concept de mot ou plus exactement de lexie, dans la problématique de cette recherche, se définit empiriquement : devient lexie (c'est-à-dire est mise telle quelle au dictionnaire) toute séquence qui résiste à l'effort d'automatisation de la traduction terme à terme de chacun de ses constituants.
2. On trouvera, dans les annexes, la liste des classes sémantico-cognitives dégagées lors des différentes expériences (anglais et espagnol).
3. Le corpus se décompose comme suit :
   1. *Brown Corpus* : 1 000 000 occurrences ; 2. *LOB Corpus* : 1000 000 occurrences ; 3. *Geographical Magazine* : 340 000 occurrences ; 4. *Scientific American* : 600 000 occurrences ; 5. *Impact* : 512 000 occurrences.
4. Le sigle GN1 — sujet profond (c'est-à-dire du verbe actif), GN2 — objet direct profond ; GN3 — objet indirect.
5. Ce corpus consiste en une centaine d'articles d'*El Correo de la UNESCO*, représentant environ 500 000 mots. Il inclut aussi, pour les verbes étudiés, tous les exemples figurant dans le *Diccionario de uso del espanol* de Maria Moliner.

BIBLIOGRAPHIE

ATTALI, A. *et al.* (1989) : *Transfert lexical et contexte sémantique en traduction automatique. Les traductions françaises des verbes anglais DRAW, HOLD, LIE, TURN, WORK*, Nancy, Cahiers du CRAL, 46, 158 p.

ATTALI, A. et A. EUVRARD (1992) : «L'environnement sémantique, aide au transfert lexical en T.A. Problèmes posés par le calcul des traductions françaises de quelques verbes anglais», *Actes du Colloque international «L'environnement traductionnel. La station de travail du traducteur de l'an 2001»*, Mons, 1991.

BOURQUIN, G. (1990) : «Du transfert à la traduction : Quelle(s) sémantique(s) ?», *T.A. Information*, 31-1, 1990, pp. 71-91.

BOURQUIN, G. (1990) : «Sur quelques aspects du transfert lexical en traduction automatique», *Mélanges Cling*, Paris, CELDA (Université de Paris XIII), pp. 35-48.

BOURQUIN, G. et J. CHAUCHÉ (1990) : «Élaboration de dictionnaires dans un programme de traduction assistée par ordinateur», *Cahiers de l'APLIUT*, déc. 1990, pp. 49-60.

NEDOBEJKINE, N. (1990) : «Représentation des informations lexicales dans les dictionnaires électroniques», *T.A. Information*, 31-1, pp. 5-17.

STEINER, E., SCHMIDT, P. et C. ZELINSKY-WIBBELT (éds.) (1988) : *From Syntax to Semantics. Insights from Machine Translation*, London,Pinter.

VIGROUX, C. (1985) : *Bilan d'une expérience de traduction automatique trilingue de l'espagnol vers le français et l'anglais*, Cahiers du CRAL, 42, 139 p.

VIGROUX, C. (1990) : «Calcul du référent du sujet non exprimé en espagnol», *T.A. Information*, 31-1, pp. 71-91.