

Les dictionnaires techniques français-néerlandais de *Metal*

Beaucoup de logiciels d'informatique linguistique charrient un volume non négligeable de vocables spécialisés. Un système de traduction automatique, notamment, ne peut fonctionner si ses utilisateurs n'y encodent pas leur propre vocabulaire technique. Qui plus est, la microstructure des dictionnaires de certains logiciels de traduction assistée par ordinateur détermine jusqu'à un certain point la qualité syntaxique des phrases traduites. Le présent article met en exergue le volet terminologique du traducteur assisté *Metal* et celui de son utilisation. Pour la paire de langues français-néerlandais de ce logiciel, une partie du vocabulaire français technique de base a été introduite dans le système au sein de la Communauté française. De plus, Siemens Nixdorf s'est constituée en courtier de traductions automatiques, donc en centre d'alimentation de termes pour ses futures entreprises-clientes.

Termes-clés:
terminotique, traduction
automatique, microstructure,
macrostructure.

1 Terminologie pour *Metal*

M*etal* est le système de traduction assistée de Siemens (depuis 1990, Siemens Nixdorf). Il est commercialisé pour les stations de travail Sun dotées du système d'exploitation Unix (Solaris), éventuellement connectées à des clients «PC». Rappelons que les systèmes de traduction automatique à transfert comme *Metal* requièrent quatre entrées de dictionnaire pour chaque terme, une entrée unilingue dans les deux langues constituant la paire et deux entrées bilingues. Ces deux entrées, dites de transfert, précisent l'équivalent en langue cible du terme de la langue source et vice-versa.

1.1 Langues

La littérature spécialisée a décrit le fonctionnement et les composantes de ce système de traduction automatique (cf. bibliographie).

La combinaison de langues français-néerlandais et néerlandais-français de *Metal* a été développée par Siemens Nixdorf, en collaboration avec les universités de Leuven, de

Liège et de Mons. Avec l'anglais-français, le français-anglais, l'allemand-français et le français-allemand, cette paire de langues est commercialisée depuis deux ans en Belgique.

1.2 Vocabulaire scientifique et technique

Metal est livré avec un vocabulaire français-néerlandais de 70 000 mots et termes, dont un lexique bancaire d'environ 8 000 unités terminologiques et un lexique d'informatique de 15 000 unités.

Siemens Nixdorf termine l'installation de *Metal* au ministère fédéral belge de la Justice. Le personnel de la section française de son Service de traduction s'apprête à encoder un volume important de vocabulaire. Ce service dispose, outre les dictionnaires juridiques, d'un fichier terminologique manuel très étendu, alimenté pendant de nombreuses années. Le ministère fédéral de la Fonction publique a, quant à lui, encodé un grand nombre de termes et de phraséologismes administratifs. Deprez, Adriaens, Depoortere, de Braekeleer (1994) constatent que 40% du vocabulaire administratif sont constitués de collocations, dont seules 20% sont traduisibles plus ou moins littéralement. Siemens Nixdorf utilise notamment *Metal* pour traduire les

Communauté française de
Belgique

procès-verbaux des réunions de son conseil d'entreprise.

Siemens Nixdorf se propose de faire traduire elle-même par *Metal* les documents d'entreprises et d'autres types d'organismes dont les besoins en traduction automatique seraient ponctuels.

2 Articles des dictionnaires

On sait que la polysémie et l'équivalence d'une langue à l'autre dépendent avant tout, pour un vocable, du secteur du savoir ou de l'activité dont relèvent les textes à traduire.

2.1 Les domaines, clients et produits

Metal traduira la phrase «*la bascule est un circuit logique*» par «*de wip is een logisch circuit*». L'utilisateur peut préalablement lui signaler le domaine de spécialité du texte, en pointant son libellé *DP-Data processing* dans la liste des domaines, sous-domaines et sous-sous-domaines du système. *Metal* traduira alors cette phrase par «*de flip-flop is een logische schakeling*».

Dans *Metal*, les domaines sont à trois niveaux. Sous le vocabulaire général se situent, dans le système tel qu'il est livré aux utilisateurs, le *CTV-Common Technical Vocabulary* ou tronc commun du vocabulaire technique, le *CSV-Common Social Vocabulary*, c'est-à-dire le tronc commun du vocabulaire social. Ces domaines sont eux-mêmes subdivisés

en sous-domaines, par exemple *ECON-Economics pour le CSV*. Des sous-sous-domaines sont subordonnés à ces sous-domaines; ainsi, *ECON-BUS-Business Management*, *ECON-FIN-Finance and Banking*, *ECON-POL-Political Economics* ont été placés sous *ECON-Economics*. Lorsqu'il consulte les dictionnaires traductifs, *Metal* choisit d'abord les acceptions les plus spécialisées des termes polysémiques. Plusieurs domaines ou sous-domaines peuvent donc guider le choix d'une équivalence, comme *publication assistée* et *traitement de texte* pour le terme *justification*. Les utilisateurs du système peuvent ajouter et supprimer à leur guise des domaines, sous-domaines et sous-sous-domaines. La société Siemens Nixdorf conseille de ne pas attribuer d'étiquettes idiosyncrasiques de domaines aux fichiers des vocabulaires destinés aux échanges inter-clients. Des vocabulaires circulent en effet au sein du groupe international d'utilisateurs de *Metal*.

Commandé et optimisé par une société utilisant un jargon de firme, *Metal* permet d'opérer une distinction supplémentaire dans les aires d'application des termes en précisant le client et le produit dont relèvent les textes à traduire. Cette spécification est notamment utile en informatique, qui charrie souvent des terminologies parallèles entières. Le système prend toujours cette distinction en considération avant tous les autres critères, lors du choix d'un équivalent dans l'autre langue.

Pour certains domaines, ces marques d'usage ne sont pas pertinentes. Elles peuvent cependant être détournées de leur fonction première pour restreindre l'usage des

termes à des types et sous-types de textes. Ou bien descendre à un niveau inférieur au sous-sous-domaine, comme pour *code*, qui ne doit être traduit par *instructions* qu'en programmation.

2.2 Préférences

On connaît en médecine les vingt-quatre dénominations alternatives à *syndrome de Down*. Pour faire choisir à *Metal* un de ces synonymes stricts, l'utilisateur peut indiquer des préférences, au sein d'un domaine, sous-domaine ou sous-sous-domaine. Un chiffre dans l'entrée de transfert de ces synonymes, c'est-à-dire l'entrée mentionnant l'équivalent du terme dans l'autre langue, influence directement l'ordre dans lequel le système examine les équivalents possibles. Plus cet entier est élevé, plus rapidement l'équivalence correspondante sera examinée.

2.3 Catégories syntaxiques et traits sémantiques

La traduction automatique met en lumière le caractère contextuel de la terminologie. La traduction d'un terme peut, avec *Metal*, être conditionnée par la présence ou l'absence d'une fonction syntaxique dans l'environnement. «*On a grossi l'affaire*», dont le verbe est suivi d'un complément direct, doit se traduire en néerlandais par de «*zaak werd opgeblazen*». Par contre, «*elle a grossi*» se traduit par «*ze is verdikt*» parce que cette phrase est sans objet direct.

La traduction peut aussi être conditionnée par la catégorie

syntactique du mot ou syntagme revêtu d'une fonction syntaxique précise. Ainsi, le verbe *se résoudre* doit se traduire différemment si son complément prépositionnel est un syntagme nominal ou un syntagme comportant un verbe conjugué. « *Il se résout à prendre des mesures* » devient « *hij besluit om maatregelen te treffen* » tandis que « *les nuages se résolvent en eau* » se traduit par « *de wolken zetten zich om in water* ».

La traduction d'un terme ou d'une collocation peut dépendre du trait sémantique d'un lexème ou autre groupe de mots doté d'une fonction syntaxique précise. L'utilisateur de *Metal* conditionnera la traduction de *couiner* par *krijzen* à la présence du trait 'humain' pour le sujet de la phrase et par *piepen* à l'absence de ce trait chez le sujet. Il subordonnera la traduction de *juger* par *beoordelen* ou par *veroordeelen* à la présence de ce trait pour l'objet direct. Dès lors, voici les traits auxquels la valeur « plus » (présence) ou « moins » (absence) est attribuable, pour les substantifs du dictionnaire unilingue français ou néerlandais: 'animé', 'humain', 'intervention humaine' (« *la banque a décidé d'abaisser le taux d'escompte* »), 'temporel' (*cycle d'horloge*), 'concret' (*frein à main*), 'abstrait' (*subsidiarité*), 'quantitatif' (*joule*), 'locatif' (*champ de manœuvres*). Les adjectifs peuvent quant à eux se voir étiquetés des valeurs « plus » ou « moins » pour les traits 'manière' (*continu/alternatif*), 'modalité' (*irrecevable*), 'couleur/forme' (*auburn*), 'locatif' (*externe*), 'temporel' (*précédent*). L'utilisateur peut qualifier l'adverbe d'une valeur positive ou négative pour les attributs 'conjonctif' (*d'autre part*), 'locatif', 'manière', 'temporel',

'modalité' (*assurément*), 'degré' (*peu*). Les traits du verbe peuvent être 'volonté', 'état d'âme', 'doute', 'déclaration/opinion'.

Enfin, la traduction d'un terme peut être conditionnée par le radical même du lexème ou constituant principal d'un syntagme auquel est attribuée une fonction syntaxique précise, surtout pour les bases (têtes) des collocations. L'utilisateur de *Metal* précise par exemple que *commettre* se traduit par *plegen* si son objet direct est le lexème *délit*, *assassinat*, *meurtre* ou *suicide*.

Ne pourrait-on également imaginer une typologie similaire des équivalences, dans les dictionnaires de spécialités destinés aux humains? Elle présente l'avantage d'être mnémonique.

2.4 Orthographe

Pour les vocables de spécialités, l'orthographe est également un critère de choix de l'équivalent.

En raison de l'échec relatif d'une réforme, la langue néerlandaise connaît deux orthographe, respectivement qualifiées de *progressief* (*filozofie*) et de *standaard* (*filosofie*). La différence réside notamment dans le traitement des confixes, dont le *c* initial et le *ph* sont par exemple remplacés respectivement par *k* et *f* dans l'orthographe progressive. *Metal* prévoit en outre certaines variantes peu usuelles pour les lexèmes (*philosofie*). Ce sont souvent des graphies archaïques.

Lorsqu'il y a lieu, l'utilisateur de *Metal* encode plusieurs variantes, comme *kollationering* et *collationering*.

2.5 Formes abrégées

Metal fait préciser à l'utilisateur le statut acronymique de certains termes. Il définit l'acronyme comme le mot formé des premières lettres des mots constituant l'expression abrégée: *laser*. Les acronymes n'ont pas de variante dans les textes: on ne trouve jamais l'expression *Light Amplification by Stimulated Emission of Radiation*. Pour *Metal*, les abréviations se distinguent de ces acronymes par l'occurrence possible de leur forme complète dans les textes, comme *l* et *litre*.

2.6 Régionalismes

Les régionalismes peuvent servir de critère pour le choix d'équivalents. L'utilisateur précisera donc qu'*ordenador* est en usage en espagnol d'Espagne et *computadora*, en espagnol colombien. Les régionalismes du français de Belgique et de France ne peuvent être libellés comme tels, peut-être parce qu'ils sont moins nombreux.

2.7 Informations morphosyntaxiques

Les dictionnaires de spécialités destinés aux traducteurs et autres utilisateurs humains devraient comporter eux aussi plus d'informations morphologiques sur leurs vedettes. Des substantifs comme *fenouil* ne peuvent en effet être mis au pluriel. Certains substantifs au singulier se traduisent par un pluriel; c'est le cas de *marchandise*, traduit par *goederen*. Le genre d'autres noms

n'est pas prédictible de leur forme. Ou bien il est double, soit masculin soit féminin, comme celui d'astérisque.

2.8 Termes syntagmatiques

Les termes plurilexématiques comme *motion de confiance* et les collocations comme *éteindre une pension*, *rabattre un défaut*, *rétracter un arrêt* et *révoquer une faillite* peuvent être encodés d'un bloc. Ces unités devront alors être étiquetées comme termes plurilexématiques. Le programme demande quelle est leur tête, qui est par exemple *entrer* pour *entrer en vigueur*.

Comme on l'a vu plus haut, une alternative à la consignation en bloc du syntagme conditionne la traduction d'un lexème à la présence d'un autre lexème dans son entourage syntaxique. L'entrée de transfert pour *uitkering* mentionnera que ce terme devra être traduit par *allocation* s'il est suivi de *bij zwangerschap*, par *dividende* s'il l'est de *van dividend*, par *capital-décès* s'il l'est de *bij overlijden*, etc.

3 Assistance à l'alimentation des dictionnaires

3.1 Intercoder

Siemens Nixdorf a développé un sous-programme d'aide à l'alimentation des dictionnaires. Intercoder évite l'encodage en Lisp des entrées de dictionnaires. Ce

programme pourrait inspirer la conception d'un module d'alimentation pour les gestionnaires de glossaires. En effet, ce sont surtout des fonctions de consultation qu'offrent à profusion les *Multiterm*, *Mtx* et autres *Term-PC*. Les commandes d'alimentation des fichiers de termes y sont relativement rudimentaires. Dans les logiciels de traduction automatique, ces commandes ont par contre été prévues parce que les caractéristiques des verbes, des adjectifs, des adverbes, notamment, revêtent une importance stratégique pour la qualité des traductions.

Lors de l'introduction d'une entrée de dictionnaire, qu'elle soit unilingue ou de transfert, la fenêtre intitulée *Specify* apparaît la première. Elle enregistre les données saisies. Si l'entrée-candidate n'existait pas encore dans le dictionnaire unilingue ou de transfert, la fenêtre *Edit* s'affiche. Elle comporte une série de valeurs par défaut, modifiables, pour les données morphologiques, comme le lemme du mot, la conjugaison pour les verbes, le pluriel proposé pour les substantifs, etc.

Si le terme dactylographié, seul ou avec un équivalent, avait déjà fait l'objet d'une entrée, c'est la fenêtre *History* qui s'affiche automatiquement. Elle permet à l'utilisateur de lire, modifier, compléter ou supprimer cette entrée. Au fur et à mesure que se déroule le travail du terminographe, *History* affiche en outre le récapitulatif de toutes les entrées ajoutées, modifiées, copiées, restituées dans leur forme originelle ou supprimées au cours de la session

en cours. *History* assure donc la cohérence de l'alimentation.

3.2 MPM

Metal Pattern Matcher est un sous-programme de recherche et remplacement global sophistiqué de chaînes de caractères. Ces chaînes peuvent être dans le texte à traduire ou dans le texte traduit par *Metal*. *MPM* permet d'empêcher la traduction des constituants de chaînes, comme certains noms propres, certains phraséologismes. *MPM* peut faire traduire à *Metal* des syntagmes ou des phrases entières telles que «*gelieve zo spoedig mogelijk te betalen*» en s'écartant de l'original, donc par «*ayez l'obligance d'effectuer le paiement dans les plus brefs délais*». Dans des fichiers *MPM* consultés automatiquement par *Metal* avant ou après la traduction de tous les textes ou de certains types de textes, l'utilisateur peut stocker des instructions remplaçant des expressions. Ces instructions peuvent être conditionnelles et inclure des masques ou des jokers, ce qui en accroît la portée. Par exemple, une seule commande permet de remplacer plusieurs variantes dénomminatives de la langue source par un équivalent unique de la langue d'arrivée. *REPLACE-C* [(*()*) «*gestionnaire de bases*» [(*()*) «*de données*» [(*()*)? «*terminologiques*» [(*()*)] ⇒ «*terminological database management system*» fera remplacer dans un texte à traduire à la fois *gestionnaire de bases terminologiques* et *gestionnaire de bases de données terminologiques* par *terminological database management system*.

3.3 Fichier des termes manquants

Lorsqu'un texte lui est soumis pour traduction, *Metal* consulte d'abord ses propres dictionnaires. Il liste ensuite dans un fichier les termes qui lui sont inconnus. L'utilisateur peut alors ajouter les entrées de dictionnaires manquantes avant de demander la traduction du texte. Cette préconsultation automatique est également programmée dans les logiciels maladroïtement dénommés « mémoires de traductions ». Ces derniers consultent les traductions antérieures archivées sur support électronique, pour y recopier les phrases équivalentes ou quasi équivalentes aux phrases à traduire. Avec ces logiciels de traduction assistée, comme *Translator's workbench* ou *Translation manager*, et de traduction automatique, comme *Metal*, un terminographe peut donc préparer le travail des traducteurs pour des textes précis. Il lance cette consultation terminologique automatique puis complète ponctuellement les dictionnaires unilingues et/ou de transfert.

3.4 Listes de contrôle

Comme pour les banques de terminologie, des fichiers de travail ou « patches » peuvent être conçus puis versés dans les dictionnaires de *Metal* après les vérifications diverses qui s'imposent.

Divers listages sont également prévus. L'utilisateur peut demander les entrées encodées par une personne, avant ou après une date précise. Il peut faire afficher celles

qui portent sur un radical précis, celles qui ont été enregistrées sur tel site utilisateur de *Metal*, celles qui relèvent d'un domaine, du jargon propre à un produit, d'une catégorie morphosémantique flanquée d'un trait sémantique précis, etc. *Find-target-entries can « inflation »* entraînera par exemple l'affichage des entrées unilingues de la langue précisée qui portent sur le lemme *inflation*. *Find-source-entries (and (cat nst) (tyn abs))* fera lister les substantifs (nst) dont le type sémantique (tyn) est abstrait (abs).

4 Collections de termes

On l'a dit, l'introduction massive de collections terminologiques est la règle en traduction automatique.

4.1 Stratégies d'introduction

Le ministère de l'Intérieur du gouvernement fédéral belge et Siemens Nixdorf ont déployé diverses stratégies d'encodage des syntagmes terminologisés, disent Deprez, Adriaens, Depoortere et de Braekeleer (1994). La solution la plus économique pour traduire les combinaisons de participes présents ou passés, comme *lettre recommandée* par *registered letter*, s'avéra par exemple l'introduction d'une condition dans le dictionnaire de transfert. Elle vérifie la présence d'un modificateur adjectival. La traduction du syntagme se fait ensuite par suppression du participe *recommandé* et la création-ajout concomitante du participe *registered*.

À l'issue d'une phase importante d'encodage, le ministère de l'Intérieur a soumis à *Metal* des corpus de textes pour tester la qualité et la cohérence des entrées terminologiques encodées par collections entières. Le premier corpus comportait 20 000 unités de traduction artificiellement construites pour chaque direction de traduction. Le second comptait environ 500 pages de textes à traduire au sein du ministère pour chaque direction de traduction.

4.2 Importations et exportations automatiques

Les fusions de collections de *Metal* avec celles d'autres provenances sont assistées. Le traitement des doubles est paramétrable. Toutes les entrées de *Metal* peuvent être soit gardées soit remplacées par les entrées-candidates. L'utilisateur peut également choisir pour chaque double l'entrée à préférer.

5 Débouché pour les terminographes d'entreprises

Lawson (1993) fait observer qu'en 1993, 3 300 000 mots ont été traduits automatiquement par 38 utilisateurs, soit la plupart des utilisateurs de 17 principaux systèmes de traduction automatique. Ces phrases traduites relevaient de la technique. Les types de documents soumis aux logiciels de traduction automatique étaient surtout des articles, manuels, questionnaires, listes, formulaires, index, annuaires, catalogues. En Communauté

française, Siemens Nixdorf est jusqu'à présent la seule utilisatrice du secteur privé, pour la terminologie de *Metal*. Mais par le courtage en traductions qu'elle a entrepris, elle pourrait bien se muer en centre commercial de terminologie.

D'une manière générale, les terminographes d'entreprise travaillent exclusivement pour les consommateurs humains. Demain, les usages de tout dictionnaire de mots ou de termes seront pourtant multiples et les cibles de ces ouvrages seront mixtes, humaine et automatique. Comme le font les lexicographes, les professionnels de la terminographie ne devraient-ils pas s'insérer en plus grand nombre dans la chaîne de production, de traitement et de circulation électroniques des documents?

Caroline de Schaetzen,
Centre de Terminologie de Bruxelles,
Institut Libre Marie Haps,
Bruxelles
et Geert Adriaens,
Siemens Nixdorf,
Centre de linguistique informatique
de Liège
et Katholieke Universiteit Leuven,
Centrum voor Computer-ondersteunde
taalkunde.

Bibliographie

Cormier (Monique), *Estival* (Dominique), éd., 1992: *Méta, numéro spécial, Études et recherches en traductique=Studies and researches in machine translation*, Montréal, Presses de l'université de Montréal.

Deprez (F.), Adriaens (G.), Depoortere (B.), de Braekeleer (G.), 1994: «Experiences with *Metal* at the Belgian ministry of the interior», dans *Méta, numéro spécial, La traduction et l'interprétation dans la Belgique multilingue*, Montréal, Presses de l'université de Montréal.

Hutchins (W. John), Somers (Harold), 1992: *An introduction to machine translation*, London, Academic Press.

Lawson (Véronique), 1993: «Forty ways to skin a cat: users report on machine translation», dans *Machine translation today. Translating and the computer* 15, London, ASLIB.

Nirenburg (Sergei), 1987: *Machine translation. Theoretical and methodological issues*, Cambridge, Cambridge university Press.

Paillet (Alain), 1992: «The application of a machine translation system in a R&D environment», dans *Translating and the computer* 14. *Quality standards in the implementation of technology in translation*, London, ASLIB.

Picken (Catherine), éd., 1990: *Machine translation in practice. Translating and the computer II*, London, ASLIB.

Slocum (John), 1988: *Machine translation systems*, Cambridge, Cambridge university Press.