JoSTrans
The Journal of Specialised Translation

www.jostrans.org · ISSN: 1740-367X

Trurnit-Verbic, R. & Ledouble, H. (2012). Amalgames de la terminologie allemande : Problématique pour l'accès aux données terminologiques multilingues et la traduction. *The Journal of Specialised Translation, 18*, 93-112. https://doi.org/10.26034/cm.jostrans.2012.440

This article is publish under a *Creative Commons Attribution 4.0 International* (CC BY): https://creativecommons.org/licenses/by/4.0

© Renate Trurnit-Verbic, Hélène Ledouble, 2012

# Amalgames de la terminologie allemande : Problématique pour l'accès aux données terminologiques multilingues et la traduction

**Renate Trurnit-Verbic, Université de Nice-Sophia Antipolis et Hélène Ledouble, Université du Sud Toulon-Var**

**RESUME**

Cet article montre les différentes problématiques que posent les amalgames de la terminologie médicale allemande. La faculté, en allemand, de créer des nouveaux termes en amalgamant des termes qui peuvent eux-mêmes être des composés, pose des questions en traduction médicale et en matière d'accès aux données terminologiques. En effet, la structure opposée de la syntaxe française et de la syntaxe allemande aboutit à un classement différent. Le corpus français est trié sur la base du terme complexe, alors que le corpus allemand est trié sur le déterminant. La consultation de dictionnaires bilingues ou de bases de données électroniques s'avère donc souvent infructueuse. Nous proposons l'introduction d'un séparateur entre les différents constituants d'un amalgame et sa reconnaissance en tant que limite de mot. Des tests avec des linguiciels ont montré les avantages qu'il y a à faire figurer les différents constituants de l'amalgame en tant que termes-vedette d'un dictionnaire. L'accès à la consultation dans un dictionnaire est alors facilité par un classement plus logique des termes amalgamés, dans le cadre d'une approche ontologique du domaine médical étudié.

**ABSTRACT**

This paper deals with German compound words and translation units within the framework of medical terminology. The German language creates compound words often based on (already) complex terms. From a bilingual point of view, the opposed syntactic structures between French and German phraseological units raise issues when considering the composition of bilingual or multilingual databases: French complex terms are listed under the main term of the compound, i.e. the head of the noun phrase, whereas German compounds are listed under the modifier, specifying the head of the noun phrase. Access to information in bilingual dictionaries or electronic databases is therefore often unsuccessful. We suggest the addition of a partition sign between the constituents of the compounds as a word separator. A test with the ZLOC software showed the advantage of presenting the different terms constituting a German compound word as main entries in the dictionary. Looking for information in a dictionary is therefore facilitated by a more logical classification of "sealed" German translation units, within a general ontological approach of medical terminology.

**MOTS-CLÉS**

Amalgames allemands, mots-composés, terminologie médicale, unités de traduction, corpus bilingue, alignement.

**KEYWORDS**

German compound words, medical terminology, translation units, bilingual corpus, aligned corpus.

## Introduction

Nous proposons une étude sur les lexèmes composés de la terminologie médicale allemande dans le cadre des travaux de recherche liés à la phraséologie médicale et à sa traduction[1].

Nous portons une réflexion sur les différents constituants des termes complexes en allemand et en français et aux difficultés liées à leur mise en regard au sein de bases de données bi– ou multilingues.

Un certain nombre de problématiques ont été soulevées lors de la constitution de ces bases de données : elles concernent d'une part l'accès aux données lors de la consultation des bases et d'autre part, la traduction des termes amalgamés.

En effet, le fonctionnement 'amalgamé' de l'allemand crée une difficulté pour l'accès aux données. La structure syntaxique opposée des deux langues aboutit à une différence de classement lors d'un tri alphabétique : le corpus français est trié selon la base des unités terminologiques et le corpus allemand selon le déterminant, positionné à gauche du déterminé. L'élément significatif principal en allemand n'est donc pas pris en compte en tant que base de recherche. Le terme allemand *Integralmembranproteasen* et son équivalent français ('protéases membranaires intégrales') illustre bien le placement opposé des constituants.

Une comparaison entre le tri effectué en français et en allemand montre ainsi un déséquilibre quant aux possibilités de recherche sur la base d'un amalgame.

Une solution simple à ce problème serait de trier sur les différents constituants d'un terme amalgamé et de présenter les bases en tant que termes-vedettes. Des tests avec deux linguiciels ont donné des résultats encourageants[2]. Quant à l'interrogation de systèmes de traduction accessibles sur Internet, un test montre qu'il n'est pas toujours aisé de trouver la traduction des termes amalgamés allemands.

L'objectif de cette étude est de réfléchir à la notion de 'terme-vedette' et à un meilleur accès au sens des amalgames en langue allemande, en vue de la constitution d'une base de données bilingue ou multilingue.

## 1. Présentation du corpus et méthodologie

Le corpus médical à la base de cette étude concerne le domaine des maladies respiratoires et de l'asthme en particulier.

Nous avons initialement constitué un corpus de textes parallèles à partir de documents mis à disposition par les publications Van Zuiden Communications B.V.

Il s'agit de textes spécialisés en allemand et en français, traductions d'articles publiés le plus souvent en anglais. Ces publications présentent les derniers progrès relatifs au traitement de l'asthme. Nous avons également utilisé des bi-textes, publications d'un même texte aligné en deux langues, anglais-allemand ou anglais-français, mis en ligne par les sites médicaux *Medpress* ou *SMW* (*Swiss Medical Weekly*). Les documents en langue française comportent 4 191 formes et les documents en langue allemande 5 311 formes[3].

Ces publications d'un même article en français et en allemand ont permis la constitution d'un second corpus. Il s'agit d'un corpus d'unités phraséologiques (UP) présentant les termes médicaux dans leur co-texte, principalement en français, en allemand, mais également en anglais.

Des textes monolingues sur la même thématique ont permis de confirmer l'utilisation de ces unités, d'élargir les paradigmes d'emploi d'un terme et d'enrichir les expressions. Le fait de montrer les termes dans leur combinatoire lexicale et les verbes dans leurs associations préférentielles constitue une démarche préparatoire à la traduction de ces unités (Isabelle et Warwick-Armstrong 1993). Notre corpus d'UP constitué à partir de ces textes dans le domaine de l'asthme comprend actuellement un total de 5791 enregistrements partiellement alignés avec une des autres langues.

En vue de la constitution d'unités de traduction à partir de ces UP, nous avons jugé important de définir les concepts du domaine étudié (Gerzymisch-Arbogast 1994; Sabah 2000; Trurnit-Verbic 2006).

Pour faire la description du domaine d'étude nous avons procédé par étapes. Les mots-clés des textes ont permis d'identifier les thématiques et de les hiérarchiser selon leur fréquence, constituant ainsi une première description notionnelle du domaine.

L'analyse lexicométrique a ensuite révélé les termes les plus fréquents. Le regroupement sémantique de ces termes a confirmé les thématiques mises en évidence initialement et a permis d'affiner la description du domaine par de nouveaux concepts :

| Concepts | Fréquence des mots-clés | Fréquence des termes |
|---|---|---|
| [maladie] | (3 468) | asthme (288), BPCO (56), asthmatiques (50) |
| [agent pathogène] | (36) | acarien (25), allergène (26), mycoplasme (6), |
| [symptôme] | (30) | expectoration (60), inflammation (56) |
| [médicament] | (20) | Singulair[i] (213), cortisone (17), corticoïde (11) |
| [test] | (19) | VEMS/FEV1 (30) |
| [malade] | - | patient (126) |
| [localisation/faculté atteints] | - | bronchique (87), respiratoire (98) |
| [étude] | - | étude (37) |
| [soins] | - | traitement (98) |
| [cellule] | - | éosinophiles (63) |

Singulair, traitement préventif de l'asthme induit par l'effort (Vidal 2000 : 1184)

**Tableau 1 : Concepts du domaine étudié**

Ces concepts spécifiques au domaine de l'asthme participent à la construction d'une structure ontologique du domaine, fondamentale pour la traduction (Zweigenbaum *et al.* 1997; Nazarenko *et al.* 2001). L'ontologie est une représentation du domaine d'étude à partir des connaissances acquises lors de l'analyse linguistique du corpus. Chaque concept de l'ontologie regroupe des termes avec leur paradigme d'emploi. L'alignement des UP, en vue de l'obtention des unités de traduction (UT) (Janicijevic 1997)[4] dans les deux langues s'appuie ainsi sur les concepts du domaine en relation avec l'ontologie sous-jacente.

A l'intérieur de notre base de données d'unités phraséologiques, nous avons pu aligner 850 unités de traduction dans les 3 langues. Les problèmes d'alignement et de découpage des textes sont à l'origine de la réflexion sur l'unité de traduction.

## 2. Terminologie allemande

En allemand, trois niveaux terminologiques cohabitent en médecine. Il y a d'abord le degré le plus spécialisé, une terminologie savante gréco-latine, dont les termes sont appelés *termini technici*. Ces termes sont les plus adaptés, selon le dictionnaire des termes de la médicine Duden (1998 : 26), à la communication écrite entre spécialistes. Cette terminologie savante est doublée d'une terminologie, toujours d'origine gréco-latine, mais dont l'orthographe et les déclinaisons ont été adaptées à l'allemand (*Trivialbezeichnungen* 'désignations ordinaires'). Une terminologie d'origine allemande, destinée à la communication avec des non-spécialistes (*volkstümliche Bezeichnungen* 'désignations populaires') est considérée comme l'échelon le plus bas des termes spécialisés. Cette

terminologie d'origine allemande qui est utilisée dans les publications de vulgarisation de l'information sur la santé, fait l'objet de notre étude.
Voici un exemple de ces différences de terminologie à partir de 4 termes :

| TERMINUS | TRIVIALBEZEICHNUNG | VOLKSTÜMLICHE BEZEICHNUNG |
|---|---|---|
| *noxa* | *Noxe* | *Schadstoff* |
| *Cella pl.* | *Zelle* | --- |
| *Pneumonia* | *Pneumonie* | *Lungenentzündung* |
| *Trachea* | --- | *Luftröhre* |

**Tableau 2 : Différents niveaux de spécialisation en allemand**

## 2.1. Importance du terme amalgamé ou composé

La langue allemande fait appel à la création d'amalgames. On parle en grammaire allemande d'amalgame syntaxique (Schanen et Confais 1989: 361). Il s'agit de la « nominalisation d'un groupe verbal, d'une nominalisation d'une lexie verbale désignant le sujet de l'action, ou désignant l'action même ou de la nominalisation d'autres syntagmes (GN et G ADJ principalement) » : *Gewebeveränderungsprozess* ('processus de remodelage tissulaire'), *Therapietreue* (lit. 'être fidèle à un traitement') 'observance thérapeutique').

Certains termes de la langue médicale allemande sont à la frontière entre un amalgame syntaxique et un composé, tels que les exemples suivants : *Entzündungszellen* ('cellules inflammatoires'), *Sputumoberschicht*, ('surnageant de l'expectoration'). L'expression 'terme composé' ne souligne pas la spécificité de l'allemand qui 'soude' les composés et engendre ainsi la difficulté d'alignement et de traduction. Pour simplifier et souligner cette problématique, nous avons regroupé tous nos termes, amalgames ou composés, sous la même dénomination : amalgames.

Cette dynamique lexicale existe depuis le Moyen-Age dans l'histoire de la langue allemande (Braun 1979/1993 : 168), mais elle s'est fortement accentuée dans la langue moderne. Ceci se vérifie aussi dans la langue médicale.

Lors de l'analyse lexicométrique de notre corpus parallèle de textes sur l'asthme[5], nous avons constaté un nombre plus important de formes et d'hapax[6] dans le corpus allemand par rapport au corpus français.

Sur 5 311 formes du corpus allemand, 3 057 sont des hapax, soit 57,5 % des formes du corpus. Le nombre important d'hapax en allemand est dû à la possibilité d'amalgamer deux termes pour n'en faire qu'un seul. Ces

candidats au statut de terme, bien qu'ils ne soient utilisés qu'une seule fois dans le corpus, participent pourtant à la thématique de l'asthme.

Voici quelques hapax issus de notre corpus :

| Premier élément du terme | Base du terme et hapax dans le Corpus | Traduction du sens |
|---|---|---|
| *Asthma\|~* | *~\|anfall* | crise d'asthme |
| | *~\|exazerbationen* | exacerbations d'asthme |
| | *~\|fällen* | cas d'asthme |
| | *~\|komponente* | composantes de type asthmatiques |
| | *~\|kontrolle* | contrôle de l'asthme |
| | *~\|morbiditât* | mortalité due à l'asthme |
| *Atem\|~* | *~\|äquivalenz* | équivalence respiratoire |
| | *~\|beschwerden* | difficultés respiratoires |
| | *~\|holen* | prendre sa respiration |
| | *~\|not* | détresse respiratoire |
| | *~\|minutenvolumen* | volume respiratoire par minute |
| *Bronchial\|~* | *~\|biopsie* | biopsie bronchique |
| | *~\|kontraktion* | contraction bronchique |
| | *~\|mukosa* | muqueuse bronchique |
| | *~\|muskulatur* | muscle (lisse) bronchique |
| | *~\|schleimhaut* | muqueuse des bronches |

**Tableau 3 : Hapax formés avec « *Asthma* », « *Atem*» et « *Bronchial* » comme premier élément d'un terme composé (Le tilde ~ marque l'emplacement de l'élément lexical manquant)**

Il ne s'agit ici que de quelques exemples mais chaque terme de base de notre corpus allemand peut produire des hapax par la faculté de pouvoir s'amalgamer avec d'autres lexèmes. Il est donc difficile, sinon impossible, de trouver l'ensemble de ces amalgames dans les dictionnaires.

## 2.2. Construction de l'amalgame

L'amalgame allemand suit la règle de construction des lexèmes nominaux composés, qui se construisent par juxtaposition de deux lexèmes simples. Mais dès qu'un lexème composé n'est plus ressenti par l'utilisateur comme un composé, il peut s'agglutiner avec d'autres lexèmes et entrer dans de

nouvelles compositions lexicales (Braun 1979/1993 : 168) (voir exemples ci-après).

Une autre origine de la formation des lexèmes nominaux est la préposition à la base d'un complément de nom au génitif avec la marque conventionnelle du 's' du cas génitif singulier masculin et neutre. Ce 's' du génitif, quasiment un 'marqueur de joncture,' se rencontre même lorsque le déterminant est un féminin dont le paradigme flexionnel ne comprend pas de 's,' comme les dérivés en – *heit*, et – *keit* (Pérennec 1999) :

- *die Sicherheit* (N, f*), der Sicherheit* (G sg), → *Sicherheit**s**vorkehrungen* ('conditions de sécurité').
- *die Krankheit* (N, f), *der Krankheit* (G sg), → *Krankheit**s**prozess* ('processus de la maladie').

Les nombreux exemples ci-dessous nous permettent d'illustrer et d'approfondir cette construction syntaxique où un composé de 2 termes est susceptible de passer à 3 ou 4 constituants.

En langue médicale, on use largement de ce puissant moyen de création terminologique. A partir de *Atem* ('respiration,' 'souffle') et *Weg* ('voie,' 'chemin,' 'conduit'), deux lexèmes de la langue générale, nous obtenons le terme *Atemwege* ('voies respiratoires'). Le premier lexème détermine le second en précisant sa finalité : '*Weg für den Atem*' ('voie pour la respiration') constituerait une paraphrase à cet amalgame.

*Atemwege* s'oppose ainsi à d'autres *Wege* ('voies'):

- *Leber- Gallenwege* ('voies biliaires'),
- *Harnwege* ('voies urinaires'),
- *Luftweg* ('voie aérienne'), etc.

Le terme *Atemwege* participe à son tour à la création de nouveaux termes, en devenant déterminant d'un processus pathologique dans le domaine de l'asthme :

- *Atemwegsobstruktion* ('obstruction des voies respiratoires').

Ce nouveau composé entrera en opposition avec les divers processus ou états pathologiques dont les voies respiratoires peuvent être atteintes :

- *Atemwegsentzündung* ('inflammation des voies respiratoires'),
- *Atemwegsveränderung* ('remodelage des voies respiratoires'),
- *Atemwegssensibilität* ('sensibilité des voies respiratoires'),
- *Atemwegserkrankungen* ('maladies des voies respiratoires'), etc.

L'état ou le processus pathologique fournissent la base à la formation d'autres termes composés :

- *Obstruktion* (‘obstruction’) : *Bronchialobstruktion* (‘obstruction bronchique’), etc.
- *Entzündung* (‘inflammation’) : *Schleimhautentzündung* (‘inflammation muqueuse’), *Lungenentzündung* (‘inflammation pulmonaire’), etc.

Le processus pathologique est déterminé par ses localisations à l’intérieur du domaine des maladies respiratoires : ‘poumons,’ ‘bronches,’ ‘muqueuses.’

En continuant d’explorer les possibilités combinatoires de *Atemwege* nous trouvons ce terme composé en tant que déterminant à *Schleimhaut* (‘muqueuse’), *Wand* (‘paroi’) et *Zellen* (‘cellules’), ce qui permet la construction d’une classification à partir des constituants organiques des voies respiratoires :

- *Atemwegswand* (‘paroi des voies respiratoires’),
- *Atemwegsschleimhaut* (‘muqueuse des voies respiratoires’),
- *Atemwegszellen* (‘cellules des voies respiratoires’), etc.

Ce système de création lexicale peut continuer en liant *Atemwegswand* (‘paroi des voies respiratoires’) au processus pathologique qui lui est propre :

- *Atemwegswandverdickung* (‘épaississement de la paroi des voies respiratoires’) :

Ce terme composé de quatre lexèmes simples est pourtant ressenti comme un composé de deux unités logiques, ce qui pose de vrais problèmes de reconstruction syntaxique lors de la traduction:

- (1) [[lieu [fonction physique + lieu] + [précision sur le lieu]] + (2.) [processus pathologique]]

## 3. Problèmes liés à la structure de l’amalgame

Les amalgames représentent une concentration d’informations, par l’ajout successif de déterminants et posent de ce fait la question de l’alignement lors de la constitution de bases de données bi- ou multilingues et donc de leur traduction.

Une seconde problématique est l’accès aux données lors de la consultation de dictionnaires bilingues ou de bases de données avec tri alphabétique. En effet, la base de ces termes se situe à la fin du composé et n’est pas accessible par le tri alphabétique, qui commence sur la première lettre du terme composé, donc sur les compléments. La consultation de bases terminologiques multilingues est également peu satisfaisante. Nous donnons ci-après quelques exemples pour chaque problématique évoquée.

## 3.1. Problèmes d'alignement

Lors de l'alignement avec le français, le terme amalgamé nous oblige parfois à élargir l'unité sélectionnée pour intégrer la totalité des informations véhiculées. Par exemple l'amalgame syntaxique :

- *Hausstaubsanierung* (*das Haus vom Staub sanieren*/'assainir la maison de la poussière').

L'expression française équivalente est : 'lutte contre les poussières ménagères.' Lors de la collecte des UP nous recherchons le prédicat avec ses arguments possibles propres à un domaine spécialisé. Pour notre exemple ce serait : lutte contre qqch.

La variable < qqch > pourra prendre la valeur [objet] pouvant s'appliquer à < objet : cancer, maladie, mort, poussière de maison > ou dans d'autres domaines : lutte contre < objet : terrorisme/pollution/etc. >. Cette solution appliquée à l'allemand donnerait < objet : |*sanierung* >. Mais dans le cas de l'allemand, *Sanierung* est bien plus sélectif que 'lutte.' Il s'agit d'assainissement de bâtiments. On ne peut l'employer aussi facilement que 'lutte.' Quand le français évoque une action énergique contre un adversaire, le terme allemand renvoie à une action d'hygiène publique. Nous n'avons pu séparer l'amalgame de son objet inhérent. L'amalgame complet a donc été inséré en tant qu'unité de traduction (UT) dans notre base de données :

- UT : *Hausstaubsanierung* : ('lutte contre les poussières ménagères')

Dans la phrase allemande : *« Eine Untersuchung an 118 gesunden Erwachsenen in jüngerer Zeit zeigt, daβ der Sputumeosinophilenanteil normalerweise nicht über 2,2 % liegt*. », en français : « Une étude récente regroupant 118 adultes en bonne santé a montré que les taux normaux d'éosinophiles dans l'expectoration n'excèdent pas 2,2 %. ». Nous voulons collecter *Sputumeosinophilenanteil* ('taux d'éosinophiles dans l'expectoration').

L'information « normale » est associée au verbe en allemand, sous la forme de l'adverbe *normalerweise* ('le taux d'éosinophiles dans l'expectoration n'excède normalement pas 2,2 %.'). Comment aligner alors la base française avec son qualifiant et ses compléments et le terme composé allemand sans qualifiant ? Il est difficile d'inclure cette information dans le terme qui est déjà formé par trois constituants :

- *Sputum*|*eosinophilen*|*anteil* →**Sputum*|*eosinophilen*|*normal*|*anteil.*

Le traducteur ou rédacteur n'a pas choisi non plus de mettre l'adjectif « normal » en épithète **normaler Sputum*|*eosinophilen*|*anteil*, la

qualification étant loin de la base qualifiée. Nous avons décidé de faire deux entrées dans notre base, l'une pour l'unité nominale, l'autre pour l'unité verbale :

- UT : 'taux d'éosinophiles dans l'expectoration' : *Sputum*|*eosinophilen*|*anteil*
- UT : qqch < taux normaux d'éosinophiles dans l'expectoration > 'ne pas excéder qqch < % >' : *etwas < Sputumeosinophilenanteil > (normalerweise) über etwas < % > liegen.*

Les problèmes d'alignement proviennent aussi d'une différence de traduction d'un même article qui est la plupart du temps rédigé en anglais.

## 3.2. Problèmes de traduction

Notre corpus a également été constitué à partir de bi-textes, c'est-à-dire de publications d'un même texte aligné en deux langues, et nous avons pu observer une différence dans les traductions ou la réécriture des articles. Un terme composé en allemand ne retrouve pas systématiquement son équivalent en français.

Ainsi nous avons : *Makrophagenspülung* :*'lavage des macrophages'. En parlant de lavage dans le domaine de l'asthme on fait référence à une intervention qui consiste à recueillir des macrophages par lavage broncho-alvéolaire à des fins d'examen. Il s'agit donc du lavage des bronches et non des macrophages. Nous retenons deux traductions :

- UT : *Makrophagenspülung* : 'macrophages recueillis par lavage broncho-alvéolaire'
- UT : *Makrophagenspülung* : 'lavage broncho-alvéolaire'

Le terme 'macrophage(s)' est absent de la partie française de la seconde UT ci-dessus. La compréhension de l'amalgame et de sa traduction ne peut se faire que par une connaissance du domaine.

Les différences de traduction viennent quelquefois d'une sous- ou d'une sur-traduction. Le sème non traduit est compris implicitement pour la sous-traduction, la sur-traduction provient d'une nécessité de précision supplémentaire. Pour décider s'il s'agit d'une sur- ou sous-traduction il faut partir de la langue source. Dans le cas des bi-textes on ne peut parler ni de langue source, ni de langue cible. Il s'agit de deux langues mises en parallèle.

La différence vient alors d'une différence stylistique ou d'une conception différente de la réalité, comme dans les exemples ci-après :

- *Freizeitsport* ('sport du temps libre') : activités sportives[7]
- *Normalfall* ('cas normal') : sujets normaux

En français il n'est pas nécessaire de préciser que le sport se fait pendant le temps libre, et que le malade est un sujet et non seulement un « cas ». Les exemples suivants montrent que le traducteur / rédacteur français cherche à alléger la traduction du terme composé allemand. Il préfère des termes simples, compréhensibles grâce au contexte :

- *Therapiemassnahmen* ('prise de mesures thérapeutiques') : 'traitement'
- *Behandlungsmethode* ('méthode de traitement') : 'traitement'
- *Wirkungsweise* ('manière de réagir') : 'mécanismes'

La connaissance du domaine et le co-texte permettent également de traduire le terme suivant qui est plus explicite en français qu'en allemand:

- *Matrixmoleküle* ('molécules de la matrice') : 'macromolécules de la matrice extracellulaire.'

Le co-texte de *Matrixmoleküle* (Vignola 2000) traitant des anomalies de la matrice extracellulaire dans l'asthme, l'amalgame se décode alors automatiquement par le spécialiste du domaine comme macromolécules de la matrice extracellulaire.

## 3.3. Problèmes d'accès aux données

La consultation d'un dictionnaire bilingue ou un tri alphabétique dans une base de données présentent régulièrement des résultats insatisfaisants. Nous montrons ci-après les difficultés d'accès aux données terminologiques dans le domaine médical.

### 3.3.1. Dictionnaires bilingues

En nous mettant dans la situation d'un apprenant français qui chercherait la traduction du terme allemand *Atemwege*, nous voudrions montrer la difficulté d'accéder à la traduction des termes composés en utilisant les dictionnaires bi- ou multilingues à notre disposition. Il s'agit d'une difficulté qui peut ralentir considérablement le temps de rédaction ou la traduction d'un article.

En consultant le dictionnaire Harrap's sous l'entrée *Atem* (respiration, souffle), première partie du terme composé, on pourrait penser trouver tous ses composés. Le dictionnaire y donne les composés *Atembeklemmung* ('suffocation' f), *Atemholen* ('prendre sa respiration' f), *Atemlähmung*. ('paralysie respiratoire' f, 'asphyxie' f), mais non la lexie *Atemwege*. Curieusement, on trouve un peu plus loin sous le terme vedette *Atemluft* ('air' m. ['respirable']) le terme composé *Atemwegserkrankung* ('affection des voies respiratoires'). Avec quelques connaissances grammaticales on peut déduire que *Atemwegs~* est le

génitif (singulier) de *Atemwege* et on a trouvé, non sans difficulté, la traduction de sa lexie.

Une recherche sur 'voies respiratoires' dans un dictionnaire électronique médical allemand – français (Unseld *et al.* 1999-2000), renvoie huit occurrences avec voie au singulier et 2 occurrences avec voies au pluriel :

- 'voies biliaires' = *Gallenwege* f.pl.
- 'voies hépatiques et biliaires' = *Leber- und Gallenwege* m. pl.

mais pas le terme 'voies respiratoires.'

En approfondissant notre recherche, nous trouvons dans le texte intégral 11 occurrences avec 'voies' en face de leurs traductions allemandes, mais toujours pas de 'voies respiratoires,' terme pourtant courant en médecine. En vérifiant dans la partie allemand – anglais, et en utilisant l'allemand *Wege* ('voies') comme passerelle entre le français et l'anglais, nous ne trouvons qu'un seul composé avec *Wege* : *Wegegeld* = 'mileage grant' ('indemnité de déplacement' f.), un composé éloigné de nos préoccupations médicales.

Devant l'absence du terme dans nos dictionnaires, il serait possible d'envisager l'existence d'un composé sur le modèle de *biliary ducts*. Le terme *airducts* existe, mais n'est pas la traduction de 'voies respiratoires.' Le dictionnaire des termes médicaux et biologiques (Lepine 1974) donne comme équivalent français 'canal aérien' (poissons, etc.). Nous passons alors du domaine médical vers la zoologie, ce qui ne permet pas une traduction efficace du terme recherché.

Cet exemple, qui peut sembler un peu caricatural, est pourtant typique des difficultés de recherche dans les dictionnaires. Maurice Rouleau (2001 : 49) précisait déjà en 2001 que « la consultation des dictionnaires médicaux par le traducteur est une opération pénible et chronophage. »

### 3.3.2. Bases de données

Le tri alphabétique sur l'allemand dans un logiciel de base de données n'est pas plus aisé. En effet, comme spécifié auparavant, la base des amalgames se situe à la fin du lexème et n'est pas accessible par le tri alphabétique qui classe les unités phraséologiques en fonction de la lettre du premier des termes de cette unité.

En triant dans notre base de données multilingue sur *Atemweg* ('voies respiratoires'), nous trouvons 2 entrées sous *Atem* ('souffle'), 7 entrées sous *Atemweg* ('voies respiratoires') et Ø entrées sous ~*weg* ('voie'), base du terme. Pour chercher éventuellement les combinaisons lexicales dans lesquelles *Weg* ('voie') peut figurer, il convient de faire appel à des

connaissances du domaine médical ou de trier sur les termes correspondants du français pour trouver le paradigme d'emploi de *Weg*. Effectivement, en triant sur le terme français 'voie,' nous trouvons 13 occurrences.

Cette différence montre bien la non-équivalence de la recherche et de l'accès à la traduction à partir des tris effectués. L'interrogation d'une base de données est donc plus difficile à partir de l'allemand.

### 3.3.3. Accès aux données avec le Linguiciel ZLOC

Le traitement du corpus d'unités phraséologiques (UP) a été effectué à l'aide d'un formalisme compatible avec une exploitation des données par les linguiciels ZText ou ZLOC, destinés à l'analyse de corpus et au traitement de données phraséologiques (constitution de bases de données) (Zinglé 1998, 1999).

Une recherche avec le linguiciel ZLOC, qui trie les UP en fonction de chaque lexème d'une unité, n'a pas donné de meilleurs résultats au départ.

À partir de la liste des lemmes, c'est-à-dire des mots-vedette simples, descripteurs du corpus, nous avons choisi : « voie », « respiratoire » et « asthmatique ». Sous « voie » nous avons trouvé 56 entrées, sous « respiratoire » 70 entrées, sous « asthmatique » 13 entrées. Nous y repérons donc très rapidement les diverses combinaisons lexicales dans lesquelles peuvent entrer les trois termes du syntagme en question avec leurs équivalents en allemand. Mais, en triant sur l'allemand, aucune occurrence n'est séléctionnée sous le terme vedette *Wege* ('voies'), *Wege* (~wege) étant la base du terme composé et n'étant pas reconnu comme lemme ou terme vedette par le linguiciel.

Voici les 4 premiers résultats de la recherche sur le terme 'voie' (sur 56 entrées) :

- 'agents irritants des voies respiratoires' *: Reizstoffe der Atemwege*
- 'augmentation de la sensibilité des voies respiratoires' : *erhöhte Atemwegssensibilität*
- 'calibre de base des voies respiratoires' : *Luftröhrenweite*
- 'calibre des voies respiratoires' : *Atemwegsweite*

En regardant du côté français sous 'respiration' (1), 'respiratoire' (70) et 'respiré' (1) nous trouvons 72 occurrences en tout. Aucune occurrence n'apparaît du côté allemand sous *Atem* ('respiration,' 'souffle'), le terme n'entrant dans ce domaine que dans des termes composés.

Nous trouvons à la suite des termes composés avec *Atem* une occurrence sous l'entrée *Atmen* ('respiration'), nominalisation du verbe *atmen*

(‘respirer’). Un déséquilibre apparaît en faveur du français avec 72 occurrences pour seulement 2 occurrences en allemand.

En y ajoutant les 26 occurrences dans lesquelles on trouve le terme *Atem* agglutiné à une autre base, nous obtenons toujours 72 occurrences en français pour seulement 28 en allemand : voici 26 occurrences du terme *Atem* agglutiné à une autre base (les fréquences sont précisées entre parenthèses) :

- *Atemluft* : ‘air respiré’ (1)
- *Atemweg* : ‘voies respiratoires’ (10)
- *Atemweg-Hyperreaktivität* : ‘hyper réactivité bronchique’ (1)
- *Atemwegsentzündung* : ‘inflammation des voies respiratoires’ (1)
- *Atemwegslumen* : ‘cavité des voies respiratoires’ (1)
- *glatte Atemwegsmuskulatur* ; ‘muscle lisse des voies respiratoires’ (2)
- *Atemwegsobstruktion* : ‘obstruction des voies respiratoires’ (4)
- *Atemwegssensibilität* : ‘sensibilité des voies respiratoires’ (1)
- *Atemwegsstrukturen* : ‘structures des voies respiratoires’ (1)
- *Atemwegsveränderung* : ‘remodelage des voies respiratoires’ (1)
- *Atemwegsweite* : ‘calibre des voies respiratoires’ (1)
- *Atemwegszellen* : ‘cellules des voies respiratoires’ (2)

D’autres unités de traduction avec respiratoire existent, mais le terme allemand diffère d’ « *Atem* » comme « *Luft* » (‘flux d’air,’ ‘flux aérien’) par exemple ou n’est pas accessible par un tri alphabétique.

## 4. Proposition

Pour pouvoir faire des recherches sur la base d’un terme composé, il faut que lors d’un tri électronique, la base d’un amalgame soit reconnue par le moteur de recherche d’un linguiciel en tant que mot isolé. L’idée d’un séparateur entre les différents constituants d’un terme amalgamé est apparue en consultant le dictionnaire médical (Pschyrembel 1998), qui utilise un séparateur pour faciliter la lecture et la compréhension de la terminologie médicale : *An*|*algesie, Ana*|*lyse, Anal*|*ekzem,* etc.

En adaptant le même principe au corpus allemand d’unités phraséologiques, il devenait alors facile de faire reconnaître le séparateur comme limite de lemme par le logiciel ZLOC.

### 4.1. Application au corpus

Nous avons introduit des séparateurs entre les différents constituants des amalgames du corpus. Pour un terme comme *Atemwegsobstruktion,* nous pouvons choisir deux ou trois séparateurs selon la finesse d’analyse désirée : *Atemwegs*|*obstruktion* ou *Atem*|*wegs*|*obstruktion*. En introduisant deux séparateurs, on sépare la base de son déterminant, en introduisant trois séparateurs, chaque constituant du terme apparaîtra lors du tri dans son paradigme d’utilisation.

## 4.2. Application au linguiciel ZLOC

Le choix de séparateurs existe dans le programme de reconnaissance de mots du linguiciel ZLOC. Il a été facile de faire reconnaître le séparateur en tant que limite de mot au même titre que le blanc entre deux groupes de lettres. Nous avons alors pu refaire un tri sur notre corpus.

Termes composés :

- *Atemwegs*|*wand*|*dicke (*'épaisseur de la paroi des voies respiratoires'*)*
- *Bronchial*|*biopsie*|*zählungen (*'taux obtenus dans les biopsies bronchiques'*)*

Lemmes correspondants :

- *Atemweg(s), Wand, Dicke*
- *bronchial, Biopsie, Zählung (en)*

Nous trouvons réunies les unités phraséologiques de notre corpus sous ces lemmes, nouveaux termes-vedette d'un dictionnaire facile à éditer. Les bases des termes composés, comme les différents constituants, pourront être observés dans leurs paradigmes d'utilisation.

### 4.2.1. Résultats

Un nouveau tri avec ZLOC après l'introduction du séparateur, permet d'améliorer le tri à partir de *Weg* en obtenant 26 entrées au lieu d'aucune précédemment :

> *Weg*
> *« Fibrose » der Atem*|*wege* : « fibrose » 'des voies respiratoires'
> *Atem*|*weg mit verdickter Wand* : 'voies respiratoires épaissies'
> *Atem*|*wege der Kontroll*|*gruppe* > : 'voies respiratoires des sujets du groupe témoin'
> *Atem*|*wege bei Asthmatikern* : 'voies respiratoires asthmatiques'
> *Atem*|*wege jeder Größe* : 'voies respiratoires de toutes les tailles'
> *Atem*|*weg-Hyper*|*reaktivität* : 'hyper|réactivité bronchique'
> *chronische, entzündliche Störung der Atem*|*wege* : 'trouble inflammatoire chronique des voies respiratoires'
> etc.

De même, les actions pathologiques du domaine sont mises en évidence par leur fréquence dans le corpus :

> *Entzündung (*'inflammation'*)* avec 52 entrées,
> *Veränderung (*'remodelage'*)* avec 23 entrées.
> *Obstruktion (*'obstruction'*)* avec 11 entrées,

Une hiérarchie d'importance des processus liés à l'asthme se dessine avec l'inflammation comme processus pathologique le plus important, le remodelage ou l'altération en second et l'obstruction en troisième position.

Ces observations confirment l'importance du tri sur la base des termes amalgamés, qui facilite l'analyse linguistique ainsi que l'alignement de corpus. Il permet également une recherche plus efficace dans des dictionnaires pour la compréhension ou la traduction des termes médicaux.

## 4.3. Tests sur des systèmes de traduction

L'introduction d'un séparateur pourrait également être une solution pour les systèmes de traduction en ligne que nous avons testés, Systran et Reverso. Nous constatons qu'un certain nombre de composés sont correctement traduits par l'un ou l'autre système, dès lors qu'il s'agit de termes relativement courants :

*Lungenentzündung*/'pneumonie,'
*Schleimhautentzündung/* 'inflammation de muqueuse,'
*Atemwegsstrukturen*/'structures de voies respiratoires,'
*Atemwegswand /'*paroi des voies respiratoires,'
*Atemwegsschleimhaut /'muqueuses des voies respiratoires, etc.*

Par contre, la traduction d'autres termes plus spécifiques est plus problématique : le terme demandé n'est pas traduit ou la traduction est inadaptée au domaine médical. Quelques-uns de ces résultats sont présentés dans le tableau suivant :

| Terme allemand | Traduction Systran | Traduction Reverso | Notre Corpus |
|---|---|---|---|
| *Bronchialobstruktion* | *Bronchialobstruktion* | obstruction bronchique | obstruction bronchique |
| *Sputumeosinophilenanteil* | *Sputumeosinophilenanteil* | *Sputumeosinophilenanteil* | taux d'éosinophiles dans l'expectoration |
| *Atembeklemmung* | oppression d'haleine | oppression d'haleine | suffocation |
| *Bronchialbiopsiezählungen* | comptes de biopsie bronchitiques | comptages bronchiques de biopsie | taux d'éosinophiles dans l'expectoration |
| *glatte Atemwegsmuskulatur* | musculature de voies respiratoires égale | musculature de chemin d'haleine lisse | muscle lisse des voies respiratoires |
| *Luftröhrenweite* | largeur de trachée | étendue de conduit respiratoire | calibre (de base) des voies respiratoires |
| *Makrophagenspülung* | macrorinçage | *Makrophagenspülung* | macrophages recueillis par lavage broncho-alvéolaire |
| *Atemwegsentzündung* | inflammation de voies respiratoires | inflammation de chemin d'haleine | inflammation de voies respiratoires |
| *Atemwegsveränderung* | modification de voies respiratoires | changement de chemin d'haleine | remodelage des voies respiratoires |

**Tableau 4 : Test de traduction avec Systran et Reverso**

Pour traduire ~|*zählungen* les moteurs de traduction renvoient « comptes » ou « comptages », alors que dans ce contexte, « taux » serait l'équivalent terminologique de ce terme. De la même manière, *Atemwegs*|~ est traduit par « chemin d'haleine », une traduction littérale issue de la langue générale. C'est la collocation « voies respiratoires » qui est utilisée dans le domaine médical. Ces erreurs de traduction sont généralement dues à l'absence d'indexation de domaine (cf. Pognan 1999). Une version professionnelle de ces mêmes logiciels serait sans doute susceptible d'améliorer les propositions de traduction.

En tout état de cause, ces exemples confirment les difficultés que posent les amalgames allemands pour la traduction et l'accès aux données dans le domaine médical.

## Conclusion

L'introduction d'un séparateur entre les divers constituants d'un amalgame a prouvé son intérêt, celui d'identifier chaque constituant d'un terme composé comme une lexie simple. La reconnaissance de chaque constituant comme une entrée lexicale dans un dictionnaire ou une base de données permet alors de regrouper sous chaque nouveau terme-vedette son paradigme d'emploi. L'accès à l'information recherchée dans un dictionnaire bilingue allemand-français retrouve ainsi une base commune. L'alignement des unités sur la base de ces termes permet alors d'établir leurs emplois synonymiques et d'identifier les unités de traduction. L'ontologie sous-jacente à notre travail délimite le cadre du domaine d'étude dans lequel s'inscrivent les unités phraséologiques de chaque langue. Les concepts identifiés à partir des mots-clés et des termes les plus fréquents sont représentatifs du domaine étudié. Ils fonctionnent tel un "langage pivot" entre deux langues. Nous pouvons alors poursuivre l'analyse et étendre l'appariement des unités phraséologiques à d'autres langues afin de faciliter l'accès aux données terminologiques multilingues relatives à l'asthme et augmenter ainsi les ressources de traduction.

## Remerciements

Nous souhaitons dédier cet article au regretté Professeur Henri Zinglé, notre Directeur de thèse et Directeur du laboratoire d'Ingénierie Linguistique et de Linguistique appliquée de l'université de Nice-Sophia Antipolis (LILLA) qui nous a quittés en 2007.

Henri Zinglé a toujours su nous transmettre sa passion pour les langues, sa rigueur dans la Recherche, son goût pour l'innovation et son ouverture d'esprit. Nous voulons ici lui rendre hommage tout en essayant de perpétuer les valeurs qu'il défendait et l'énergie qui le caractérisait.

## Références bibliographiques


- **Braun, Peter** (1979/1993). *Tendenzen in der deutschen Gegenwartssprache, Sprachvarietäten*. 3e Ausgabe. Stuttgart: Kohlhammer.

- **Duden** (1998). *Das Wörterbuch medizinischer Fachausdrücke, Bedeutung, Aussprache, Rechtschreibung, Silbentrennung, Abkürzung*. 6e Ausgabe. Mannheim: Dudenverlag.

- **Gerzymisch-Arbogast, Heidrun** (1994). *Übersetzungswissenschaftliches Propädeutikum.* Coll. UTB (Uni-Taschenbücher) für Wissenschaft. Tübingen/Basel: A. Franke Verlag.

- **Harrap's** (1996). *Universal Dictionnaire Allemand-Français/Français-allemand*. Edinburgh/Stuttgart: Harrap Publishing Group Ltd. / Ernst Klett Verlag für Wissen und Bildung GmbH.

- **Isabelle, Pierre et Warwick-Armstrong Susan** (1993). "Les corpus bilingues : une nouvelle ressource pour le traducteur." Pierrette Bouillon et André Clas (eds) (1993). *La Traductique, études et recherches de traduction par ordinateur.* Montréal: Presses de l'Université de Montréal, 288-306.

- **Janicijevic, Tatjana** (1997). "L'approche informatisée du dépistage des unités de traduction." Paper presented at *Colloque interdisciplinaire sur L'informatique dans les études françaises LIL97* (Canada, Queen's 15.03.1997).

- **Lepine, Pierre** (1974). *Dictionnaire Français-Anglais/Anglais Français des Termes médicaux & biologiques*. Paris: Ed. Flammarion.

- **Nazarenko, Adeline *et al*.** (2001). "Corpus bases extension of a terminological semantic lexicon." Didier Bourigault, Christian Jacquemin and Marie-Claude L'Homme (eds) (2001). *Recent Advances in Computational Terminology. Coll. Natural Language Processing*. Amsterdam/Philadelphia: John Benjamins, 327-351.

- **Pérennec, Marcel** (1998). "(Dé)nomination, phrasème et terme." Paper presented at C.R.T.T. : Centre de recherche en terminologie et traduction. EA 4162. (Université Lumière Lyon 2, le 15 mars 1999, 1-9).

- **Pognan, Patrice** (1999). " 'Autoéclairage' des mots en tchèque. Application à l'indexation automatique." *Travaux du LILLA* (Laboratoire d'ingénierie linguistique et de linguistique appliquée de la faculté de lettres, arts et sciences humaines de Nice-Sophia Antipolis, France): *JILA'99* (Journées internationales de linguistiques appliquées, Nice, France 24-25 juin 1999), 220-224.

- **Pschyrembel** (1998). *Pschyrembel Klinisches Wörterbuch*. 258. Auflage. Berlin/New York: Walter de Gruyter.

- **Rouleau, Maurice** (2001). "La facture des principaux dictionnaires médicaux français : le point de vue d'un traducteur." *Meta : journal des traducteurs* XLVI(1), 35-55.

- **Sabah, Gérard** (2000). "Sens et traitement automatiques des langues." Jean-Marie Pierrel (ed.) *Ingénierie des langues. Coll. Informatique et Systèmes d'information*. Paris: Hermes Science Europe, 77-108.

- **Schanen, François et Jean Paul Confais** (1989). *Grammaire de l'allemand, formes et fonctions.* Paris: Nathan Université.

- **Trurnit-Verbic, Renate** (2006). "L'incidence du contexte sur la traduction médicale." Daniel Blampain, Philippe Thoiron, Marc Van Campenhoudt (eds) (2006). *Mots, termes et contextes. Actes des septièmes Journées scientifiques du réseau des chercheurs Lexicologie, Terminologie, Traduction.* Bruxelles, 8 au 10 septembre 2005 : Coll. Actualité scientifique et Agence universitaire de la Francophonie. Paris : Archives Contemporaines, 483-495.

- **Unseld *et al.*** (1999-2000). *Medical Dictionary - Dictionnaire Médical — Medizinisches Wörterbuch – Englich. Français. Deutsch*. CD-ROM. Stuttgart: Medpharm Scientific Publishers.

- **Vignola, Antonio Maurizio** (2000). "Le remodelage des voies respiratoires dans l'asthme." *Practical Issues in Asthma Management* N° 19, 1-9.

- **Zinglé, Henri** (1998). "ZText : un outil pour l'analyse de corpus." *Travaux du Lilla.* Publications de la Faculté de lettres, arts et sciences humaines de l'université de Nice-Sophia Antipolis, N° 3, 69-78.

- **Zinglé, Henri** (1999). "Terminologie et ingénierie linguistique." *Travaux du Lilla : JILA' 99, Journées internationales de linguistique appliquée*, Nice 24-25 juin 1999, Publications de la Faculté de lettres, arts et sciences humaines de l'université de Nice-Sophia Antipolis, 281-283.

- **Zweigenbaum, Pierre *et al*.** (1997). "Coopération apprentissage en corpus et connaissances du domaine pour la construction d'ontologies." *1ères Journées Scientifiques et Techniques FRANCIL*, Avignon, April 1997 : Réseau Francophone de l'Ingénierie de la Langue de l'AUPELF-UREF, 501-508.


## Websites


- "**Medpress**." http://www.medpress.ch (consulted 27.09.2005)

- "**Swiss Medical Weekly**" (SMW). http://www.smw.ch (consulted 15.01.2011)

- "**Systran**." http://www.systranet.fr/translate (consulted 29.09.2011)

- "**Reverso.**" http://www.reverso.net/text_translation.aspx?lang=FR (consulted 18.01.2012)


## Biographies des auteurs

Renate Trurnit-Verbic :
1995 – 2004 : Etudes à l'Université de Nice – Sophia Antipolis en Sciences du Langage, mention Traitement Automatique des Langues et Lettres et Civilisations étrangères (allemand), dont 6 ans au Laboratoire d'Ingénierie Linguistique et de Linguistique Appliquée (LILLA) de l'université de Nice Sophia-Antipolis
2004 Thèse en Sciences du Langage

2005 et 2009 Participation aux Journées scientifiques du Réseau Lexicologie, terminologie, traduction. Contact : jr.verbic@orange.fr.

Hélène Ledouble est Maître de Conférences à l'Université du Sud Toulon-Var. Après 6 années de Recherche au Laboratoire d'Ingénierie Linguistique et de Linguistique Appliquée (LILLA) de l'université de Nice Sophia-Antipolis, elle rejoint le laboratoire Babel à l'Université du Sud Toulon-Var. Ses recherches portent sur la phraséologie et la traduction à partir de corpus bilingues ou multilingues. Contact : ledouble@univ-tln.fr.

**Notes**

[1] « Analyse de documents médicaux en français et en allemand : Incidence sur la traduction »Thèse présentée en 2004 sous la direction du Professeur Henri Zinglé, Directeur du Laboratoire d'Ingénierie Linguistique et de Linguistique Appliquée (L.I.L.L.A.), Université de Nice-Sophia-Antipolis,maintenant CIRCPLES/EA3159.

[2] linguiciel ZLOC (Zinglé 1998), Lexico 3, conçu en 2003 par l'équipe universitaire SYLED-cl2AT, de l'université de la Sorbonne Nouvelle-Paris 3.

[3] En parlant de forme nous faisons référence à un même segment textuel qui apparaît comme entrée du dictionnaire créé lors de la segmentation du texte en base de données (les verbes : formes conjuguées, à l'infinitif ou participes, pour les substantifs : formes au singulier, au pluriel ou les formes déclinées).

[4] Janicijevic (1997) définit l'unité de traduction comme « une structure compacte, inséparablement bilingue, composée de deux signifiants et d'un seul signifié : UT = Sé [Sa {F} + Sa {A}]». Sa formule concernait le français {F} associé à l'anglais {A}, mais elle est valable pour toutes les paires de langues.

[5] Logiciel utilisé : Lexico 3.

[6] Le terme hapax renvoie à un segment textuel unique.

[7] Entre parenthèses nous donnons la traduction littérale, après les deux points la formulation trouvée dans le bi-texte.