

Problèmes de traduction automatique dans les sous-langages des bulletins d'avalanches

Pierrette Bouillon and Katharina Boesefeldt

Volume 37, Number 4, décembre 1992

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

URI: <https://id.erudit.org/iderudit/002108ar>

DOI: <https://doi.org/10.7202/002108ar>

[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

Bouillon, P. & Boesefeldt, K. (1992). Problèmes de traduction automatique dans les sous-langages des bulletins d'avalanches. *Meta*, 37(4), 635–646.
<https://doi.org/10.7202/002108ar>

Article abstract

In this article we describe the current status of the Swiss Avalanche Bulletins machine translation project. After a short presentation of the project, we highlight the translation problems which we have encountered during project development and the way to solve them with ELU, a Unification-based Linguistic Development Environment. We will then show that a close collaboration while writing the grammars for the different languages generally allows us to obtain the same semantic representation, which models the domain of avalanches in Switzerland to obtain good quality translation that does not induce modifications in the bulletins' style.

PROBLÈMES DE TRADUCTION AUTOMATIQUE DANS LE SOUS-LANGAGE DES BULLETINS D'AVALANCHES

PIERRETTE BOUILLON et KATHARINA BOESEFELDT
ISSCO, Genève, Suisse

Résumé

Dans cet article, nous décrivons l'état actuel du projet de traduction automatique des bulletins d'avalanches de la Suisse. Après un bref aperçu du projet, nous mettons en évidence les problèmes de traduction, auxquels nous avons été confronté lors du développement du projet et la manière de les résoudre avec ELU, environnement linguistique d'unification. Nous montrerons qu'une collaboration étroite lors de l'écriture des différentes grammaires nous permet généralement d'obtenir une même représentation sémantique dans les différentes langues, laquelle modélise le domaine des avalanches en Suisse. Mais nous maintiendrons aussi qu'un système de transfert est indispensable pour obtenir une traduction de qualité, qui ne provoque pas de modification du style des bulletins.

Abstract

In this article we describe the current status of the Swiss Avalanche Bulletins machine translation project. After a short presentation of the project, we highlight the translation problems which we have encountered during project development and the way to solve them with ELU, a Unification-based Linguistic Development Environment. We will then show that a close collaboration while writing the grammars for the different languages generally allows us to obtain the same semantic representation, which models the domain of avalanches in Switzerland to obtain good quality translation that does not induce modifications in the bulletins' style.

1. INTRODUCTION

Cet article décrit le projet de traduction automatique des bulletins d'avalanches de la Suisse, qui utilise ELU, l'environnement linguistique d'unification développé à l'ISSCO pour des expérimentations linguistiques. Nous montrons que la méthode du transfert, couplée à une représentation sémantique cohérente du monde des avalanches, nous permet une traduction de qualité, conforme aux habitudes des météorologistes. Nous commençons par donner un bref aperçu général du projet et par mettre en évidence les différents problèmes posés lors de la traduction automatique des bulletins d'avalanches, au niveau de l'analyse des textes et de leur traduction. Nous nous attardons ensuite sur la mise en œuvre du projet, c'est-à-dire le logiciel ELU, sur la représentation sémantique du monde des avalanches et sur les règles de transfert nécessaires pour résoudre les problèmes de divergences et d'équivalences traductionnelles.

2. LE PROJET DE TRADUCTION AUTOMATIQUE DES BULLETINS D'AVALANCHES

Ce projet vise à la traduction automatique des bulletins d'avalanches, émis par «l'Institut Fédéral pour l'Étude de la Neige et des Avalanches» (IFENA)¹, pour informer la population sur les conditions de neige et les dangers d'avalanches dans les montagnes

suisses. Ces bulletins paraissent une à plusieurs fois par semaine, pendant l'hiver, en fonction des conditions atmosphériques. Rédigés initialement en allemand à Davos, par des guides de montagne, ils doivent ensuite être rapidement traduits dans les deux autres langues officielles : le français et l'italien. Depuis 1990, le projet se concentre sur l'allemand et le français. La possibilité de la création d'un institut semblable en région francophone, lequel rédigerait les bulletins en français, nous a décidés à adopter une approche bidirectionnelle et à traiter les deux couples de langues allemand-français et français-allemand. Pour l'instant, cependant, nous mettons surtout l'accent sur la qualité de la traduction française, la traduction allemande servant seulement de test au système.

Ces bulletins d'avalanches constituent une application idéale pour la traduction automatique (Bouillon *et al.* 1991a; Bouillon *et al.* 1991b; Boesefeldt *et al.* 1991). Tout d'abord, ils présentent une structure rigide. Ils sont généralement subdivisés en trois alinéas qui informent respectivement la population de la situation générale, des degrés de danger et des différentes mesures à prendre dans les régions suisses. Ensuite, ils traitent du domaine limité de la neige et des avalanches en Suisse et utilisent de ce fait une sémantique et une syntaxe particulières. Enfin, les bulletins constituent un corpus trilingue, qui définit un sous-langage (Salm 1982). Nous pouvons donc limiter le traitement automatique à ce sous-langage, éviter ainsi de modifier le style des bulletins et obtenir une traduction de qualité, sans postédition et conforme aux exigences des météorologistes et à leur habitude langagière. Nous pouvons aussi contraindre le système à produire une seule traduction par phrase et éviter ainsi que les utilisateurs allemands de Davos n'aient à effectuer un choix parmi plusieurs solutions.

3. LE CORPUS DES BULLETINS D'AVALANCHES

Le corpus des bulletins d'avalanches est intéressant pour deux raisons : tout d'abord, les bulletins présentent une syntaxe très complexe, qui nous contraint à aborder des problèmes peu traités en linguistique informatique. Contrairement aux notices météorologiques traditionnelles, ils ne sont pas écrits en langage télégraphique et contiennent des descriptions assez longues, très précises et rigoureuses des conditions atmosphériques, du type de neige, des dangers éventuels, du climat et des lieux géographiques. Une étude de corpus montre, par exemple, qu'il existe plus de trois cents concepts différents qui contiennent le mot *neige*. Souvent, ces expressions constituent des hapax, qui correspondent à des situations exceptionnelles d'enneigement. Le français, par exemple, distingue des groupes nominaux, avec des structures syntaxiques aussi diverses que :

[art n adj] : *la neige fraîche*

[art adj n [[adv adj] et adj]] : *la vieille neige peu consolidée et humide*

[art adj n adj pp] : *l'ancienne couverture humide de neige*

[art n et art n pp] : *le tassement et la consolidation de la couverture de neige*

[n et n [adj pp]] : *la consolidation et le tassement favorables de la couverture de neige*

[art adj n adj [pp ou pp]] : *les minces couches défavorables de givre ou de croûtes de fusion.*

Ces derniers constituants posent d'importants problèmes au niveau de la grammaire française, comme celui de la place des adjectifs en français (le français dit *une grande quantité de neige*, mais postpose le même adjectif en présence de certains adverbes : *une quantité relativement grande de neige*), de l'attachement des modificateurs et de la cohérence des syntagmes nominaux.

En second lieu, les bulletins sont intéressants au niveau de la traduction. Ils présentent en effet un grand nombre de divergences traductionnelles et posent différents problèmes d'équivalence entre les langues (Barnett *et al.* 1991 : 25).

Par **divergence**, nous entendons ici deux contenus identiques qui sont exprimés dans les langues cible et source par des structures syntaxiques différentes (Barnett *et al.* 1991 et Dorr 1990). Les exemples sont nombreux. Tout d'abord, l'allemand fait un usage fréquent de participes présents qui sont peu satisfaisants en français et sont plutôt remplacés par d'autres tournures, comme des propositions relatives. Par exemple, la phrase :

1. *die anhaltenden Niederschläge und die Setzung der Schneedecke haben zu einer Abnahme der Lawinengefahr geführt*

contient le participe *anhaltenden* qui est traduit dans les bulletins français par la relative *qui continuent* :

-> *les précipitations qui continuent et le tassement de la couverture de neige ont causé une diminution du danger d'avalanches.*

Ensuite, les adverbes sont aussi la cause de nombreuses autres divergences. Très fréquents en allemand, ils sont moins utilisés en français, qui les remplace fréquemment par des adjectifs. L'adverbe de phrase *vereinzelt*, par exemple, est traduit en français par l'adjectif épithète *isolé* :

2. *Grundlawinen sind vereinzelt noch möglich*
-> *les avalanches isolées de fond sont toujours possibles.*

De même, certains adjectifs allemands se traduisent par des noms ou des groupes nominaux en français. L'adjectif *übrig*, par exemple, se traduit dans la plupart des cas par le nom *le reste de*. L'adjectif *östlich*, quant à lui, sera de préférence rendu en français par le groupe nominal *la partie est de* et il en va de même pour tous les adjectifs de ce type. Par exemple :

3. *in den übrigen Alpen ist die Gefahr noch gering und lokal*
-> *dans le reste de la région des Alpes, le danger est encore faible et local.*

Nous dirons au contraire qu'il y a un **problème d'équivalence** si les informations transmises par les deux textes français et allemand sont différentes, les unes étant moins précises, plus vagues que les autres (Kameyama *et al.* 1991). Les problèmes d'équivalence lexicale et syntaxique sont eux aussi nombreux dans les bulletins d'avalanches. Il arrive tout d'abord que l'allemand et le français distinguent deux notions différentes, mais que l'une des langues les exprime par deux termes différents et que l'autre utilise un terme unique polysémique. Dans les bulletins, par exemple, le français utilise indifféremment le terme «variable» pour désigner «ce qui est susceptible de se modifier, de changer souvent au cours d'une durée» (par exemple : *vents variables, temps variable, etc.*) ou «quelque chose qui se réalise diversement» (par exemple : *couches de neige variables*), deux notions qui se traduisent en allemand par deux termes différents : «*unterschiedlich*» et «*wechselhaft*».

Il arrive aussi qu'un terme d'une langue n'ait pas d'équivalent exact dans une autre et nécessite une périphrase ou doive être ignoré dans la traduction. Certains groupes adverbe-adjectif se traduisent par de simples adjectifs en français, comme en témoigne l'exemple suivant, où *örtlich beschränkt* se traduit par *limité* :

4. *am östlichen Alpennordhang und am Alpensüdhang ist die Lawinengefahr gering und örtlich beschränkt*
-> *dans la partie est du versant nord des Alpes et sur le versant sud des Alpes, le danger d'avalanches est faible et limité.*

De même, les syntagmes nominaux que nous avons vus plus haut et qui servaient à décrire les différents types de neige correspondent souvent en allemand à des mots composés, c'est-à-dire à des unités graphiques :

5. *neige fraîche* -> *Neuschnee*
couverture de neige -> *Schneedecke*
danger de plaques de neige -> *Schneebrettgefahr*.

Ces derniers ne donnent aucune indication sur la préposition ou le type d'article du complément en français (défini, comme dans *constitution de la couverture de neige* ou nul, comme dans *danger d'avalanches*). Ce problème d'équivalence est loin d'être aisé et nous oblige à trouver une technique qui établisse une relation systématique et générale entre les unités graphiques de l'allemand et les structures nominales complexes utilisées par le français.

Enfin, une langue peut exprimer une notion par un ensemble de synonymes ou de quasi synonymes, pour lequel il n'existe pas de correspondance mot à mot dans l'autre langue. Dans le cadre de ce projet, où nous ne voulons obtenir qu'une seule traduction par phrase, il faudra nécessairement opérer un choix entre les différents synonymes. S'il s'agit de quasi synonymes qui ne peuvent pas s'utiliser indifféremment, la traduction devra tenir compte des contextes d'utilisation. La préposition de lieu allemande *in*, par exemple, qui exprime la situation, se traduit en français par *en*, *dans* ou *à*, en fonction du nom de lieu qui suit la préposition. De même, l'allemand utilise indifféremment le terme *gross* pour désigner l'importance d'un danger ou des accumulations de neige, alors que le français préfère nettement deux termes distincts *grand* et *gros*, ce qui nous oblige à établir les correspondances suivantes :

6. *eine grosse Gefahr* -> *un grand danger*
7. *die grossen Schneeverfrachtungen* -> *les grosses accumulations de neige*

4. MISE EN ŒUVRE DU PROJET

La mise en œuvre du projet se fait avec ELU, environnement linguistique d'unification, développé à l'ISSCO pour des expérimentations linguistiques. Inspiré de PATR II, il utilise comme domaine d'informations des structures de traits, c'est-à-dire des ensembles d'attributs et de valeurs, combinées entre elles par l'opération d'unification (Shieber 1986). ELU comprend trois modules principaux : un analyseur qui permet d'obtenir pour une phrase les représentations en structures d'attribut permises par la grammaire, un générateur, capable d'interpréter les règles de la grammaire pour trouver les chaînes de caractères qui peuvent être associées à une structure de traits et enfin un module de transfert réversible, qui établit une relation entre deux représentations en structures de traits pour permettre le passage d'une langue à l'autre. Comme ces trois modules utilisent la technique d'unification, chacun d'eux est réversible. La même grammaire peut donc être utilisée pour l'analyse et la génération, et des règles de transfert identiques peuvent être utilisées dans les deux directions. La traduction avec ELU consiste à établir avec des règles de transfert des relations entre les structures de traits de la langue source et cible. Ces règles sont de deux types, atomiques et structurelles, pour transformer respectivement les atomes et les structures de traits complexes. Par exemple, les deux structures de traits suivantes qui représentent les syntagmes nominaux allemand *die Gefahr* et français *le danger*² et qui indiquent le prédicat du groupe nominal (PRED)³ et le type de l'article (DETYPE) :

8. *pred = gefahr*
detype = definite

pred = danger
 detype = definite

sont mises en relation par les quatre règles suivantes :

```
:TA: gefahr danger
:TA: definite definite

:T:pred
:L1:< *pred > = X
:L2:< *pred > = Y
:X: X = Y

:T:detype
:L1:< *detype > = X
:L2:< *detype > = Y
:X: X = Y
```

Ces dernières établissent un rapport entre les paires d'atomes Gefahr — danger et definite — definite et les deux chemins PRED et DETYPE de deux langues différentes L1 et L2. Les deux règles **pred** et **detype** sont indépendantes des valeurs des chemins, et leur application dépendra de celle de toutes les règles nécessaires pour le transfert des valeurs liées aux variables X et Y. Le transfert se fait donc de manière réursive, en commençant par les racines des structures de traits et en terminant par leurs feuilles. La règle est appliquée si la représentation de la langue source est subsumée par la structure de traits décrite dans la règle de transfert et si la représentation de la langue cible s'unifie avec la structure de traits de la même règle de transfert (Estival *et al.* 1990b et Russell *et al.* 1991).

Il est clair que la syntaxe des règles de transfert permet a priori d'effectuer n'importe quelle transformation. Il serait donc possible de construire indépendamment les représentations allemandes et françaises et de multiplier ensuite les règles de transfert pour permettre le passage d'une représentation à une autre. Cette solution semble toutefois peu efficace et souhaitable. Tout d'abord, la multiplication des règles de transfert, module moins bien connu et expérimenté en linguistique informatique, ralentit considérablement le système. D'autre part, elle ne rend pas compte de la sémantique générale du monde des avalanches. Les bulletins contiennent en effet les mêmes informations et véhiculent un sens identique. Pour ces deux raisons, il nous a donc semblé préférable de définir une représentation sémantique commune, d'écrire en parallèle les grammaires des langues source et cible, pour obtenir, chaque fois que c'est possible, une même représentation sémantique et simplifier le plus possible les règles de transfert. Ces dernières ne sont complexifiées que dans le cas où il nous a semblé impossible ou peu souhaitable d'obtenir la même représentation dans les deux langues. Dans la suite de cet article, nous commencerons par présenter la représentation sémantique du monde des avalanches et les différents problèmes de traduction qu'elle nous permet de résoudre. Nous mettrons ensuite en évidence les autres problèmes que nous n'avons pas pu traiter de cette manière et nous montrerons comment des règles de transfert complexes nous permettent d'y apporter une solution élégante.

4.1. UNE REPRÉSENTATION SÉMANTIQUE COHÉRENTE DU MONDE DES AVALANCHES

Le but de la représentation sémantique est de créer un lien entre des expressions linguistiques de langues différentes qui traitent d'un même contenu. Elle permet donc de faire abstraction des particularités syntaxiques de chaque langue et de ne représenter que les informations sémantiques nécessaires à la traduction. Comme le domaine des

avalanches est relativement bien connu et rigide, nous pouvons déterminer à l'avance quelles informations pertinentes pour la traduction doivent figurer dans le domaine d'informations. Citons, à titre d'exemple, le prédicat logique de la phrase, l'argument, le lieu, le climat, l'altitude, le temps, le mode, le caractère positif ou négatif de la phrase. Chacune de ces informations est encodée dans un chemin d'attributs particulier, qui permet d'avoir un accès aisé à la valeur.

Par exemple, la phrase *dans les Alpes subsiste un grand danger généralisé d'avalanches* recevra la structure sémantique suivante (simplifiée ici) :

```

9.  args = [<81>args : args = []
      detype = non
      pred = avalanches
      detype = indefinite
      mod : alt : pred = []
      expo : pred = []
      nom : [<1>pred = grand
            <2>pred = general ]
      pred = danger]
      mod : alt : pred = []
      loc : pred = Alpes
      detype = definite
      top = yes
      temps : pred = []
      climat : pred = []
      morph : temps = present
      voix = actif
      positif = yes
      pred = subsister

```

Cette structure signifie que la phrase est positive (POSITIF), au présent actif (MORPH) et indique que le prédicat logique (PRED) est le verbe intransitif *subsister*. Celui-ci a pour argument (ARGS) le sujet indéterminé *un grand danger général d'avalanches*, encodé dans une liste⁴. Un seul modifieur est réalisé, le complément de lieu *dans les Alpes*, de type défini (DETYPE). Le trait **<*top> = yes** signale qu'il est topicalisé dans la phrase. Les listes vides ([]) indiquent que la phrase ne contient pas d'autres modifieurs.

Dans ce cas-ci, l'équivalent allemand *in den Alpen besteht eine grosse allgemeine Lawinengefahr* recevra la même structure sémantique. Seules les valeurs des traits, les atomes français, sont remplacées par d'autres valeurs : les atomes allemands.

```

10. args = [<81>args : args = []
      detype = non
      pred = lawinen
      detype = indefinite
      mod : alt : pred = []
      expo : pred = []
      nom : [<1>pred = gross
            <2>pred = allgemein ]
      pred = gefahr]
      mod : alt : pred = []
      loc : pred = alpen
      detype = définitive
      top = yes
      temps : pred = []
      climat : pred = []

```

morph : temps = present
 voix = actif
 positif = yes
 pred = bestehen

Comme les représentations sémantiques des deux phrases sont identiques, le transfert est aisé et peut se faire au moyen de règles très générales. Pour établir une correspondance entre les représentations des phrases *in den Alpen besteht eine grosse allgemeine Lawinengefahr* et *dans les Alpes subsiste un grand danger généralisé d'avalanches*, il nous suffit simplement de créer autant de règles atomiques qu'il y a d'atomes (puisqu'il y a de chemins (de la même manière que dans l'exemple 8).

Obtenir une même représentation dans les deux langues n'est pas toujours aisé et compliqué souvent les grammaires et les lexiques. Dans l'exemple cité plus haut, les lexiques et les grammaires contrôlent, par exemple, les prépositions de lieu, qui ne font pas partie de la représentation sémantique. La plupart des systèmes qui essaient d'établir une correspondance entre les prépositions se heurtent en effet à des problèmes importants et sont obligés de définir des algorithmes très complexes, qui, souvent, ralentissent et complexifient considérablement le système (Japkowics 1991). Dans notre système, donc, les noms de lieu sont codés dans les lexiques avec la préposition qu'ils requièrent (par exemple : *Engadine* -> *en*, *Alpes* -> *dans* et *crêtes* -> *sur*) et la règle de la grammaire qui construit le groupe prépositionnel de lieu vérifie que la préposition de la phrase courante correspond bien à la préposition exigée par le nom de lieu.

De même, les mots composés sont représentés de la même manière dans les deux langues. Dans le lexique allemand, nous avons assigné aux noms composés des représentations complexes qui rendent compte de leur composition interne. La représentation du nom composé *Lawinengefahr*, par exemple, indique ses deux éléments constitutifs : la tête (*gefahr*) et l'argument (*lawinen*), ainsi que le type d'article du complément (ici nul), tous nécessaires pour la traduction française et italienne :

11. args : args = []
 pred = lawinen
 detype = non
 pred : gefahr

En français, par contre, la grammaire se charge du contrôle de la préposition requise par la tête du mot composé. Les noms sont donc codés dans le lexique avec une sous-catégorisation complexe, qui stipule le type sémantique du complément pour lequel ils sous-catégorisent⁵ la préposition requise par le complément et le type d'article du complément, défini ou nul. Cette technique nous permet un traitement général, unifié et efficace du problème des noms composés⁶.

Enfin, le mode et le genre, qui varient en fonction des langues ont été exclus de la représentation sémantique.

4.2. DES DIVERGENCES TRADUCTIONNELLES ET AUTRES PROBLÈMES D'ÉQUIVALENCE EN TRADUCTION

Dans certains cas cependant, il est peu souhaitable et même parfois impossible d'obtenir la même représentation dans les deux langues, sans encourir des problèmes de surgénération, et il est dès lors indispensable de complexifier et de multiplier les règles de transfert.

Reprenons l'exemple 1 du participe présent *anhaltend* de la section trois. Dans ce cas-ci, il est impossible d'obtenir la même représentation dans les deux langues, sans

s'exposer à des problèmes de surgénération, puisque les deux types de constructions, participiales et relatives, coexistent dans les deux langues. Avec une même représentation, le générateur, qui produit toutes les phrases qui s'unifient avec la représentation de départ, générerait en effet deux solutions, l'une avec une relative et l'autre avec un participe, ce que nous voulons à tout prix éviter dans le cadre de ce projet. Nous avons donc décidé de représenter différemment ces deux phrases dans les deux langues. Dans la représentation allemande, le participe *anhaltenden* est considéré comme un modifieur du nom *Niederschläge* (MOD:PRED), alors qu'en français, une structure réentrante (exemple 13.) (Shieber 1986: 13) représente les relatives et indique que le sujet de la relative (la variable #12) contenu dans REL:ARGS est le même que le prédicat principal (#12) (PRED).

12. pred = Niederschläge
 mod : pred = anhalten
 rel = []
13. pred = #2 précipitation
 mod = []
 rel : args : [pred = #12]
 pred = continuer

Comme ces deux structures sont assez différentes, la règle de transfert qui établit la correspondance entre ces deux structures est complexe : elle stipule que le prédicat du modifieur en allemand Z1 correspond au prédicat de la relative Z2. Cette dernière a pour argument une liste [R], dont le prédicat X2 est semblable au prédicat de la phrase nominale et correspond au nom X1 auquel se rapporte le participe allemand Z1.

14. :T: part_rel
 :L1: <*pred> = X1
 <*rel> = []
 <*mod pred> = Z1
 :L2: <*pred> = X2
 <*mod> = []
 <*rel> = W
 <Wargs> = [R]
 <Rpred> = X2
 <Wpred> = Z2
 <*reversible> = non
 :X: X1 = X2
 Z1 = Z2

Le trait <*reversible> = non, qui ne subsumera aucune représentation sémantique en français, sert simplement à bloquer la réversibilité de la règle et à empêcher que les relatives françaises ne se traduisent par des participiales en allemand. De cette manière, les participes verbaux allemands se traduiront par des relatives en français, tandis que les relatives françaises seront rendues par des relatives en allemand, ce qui nous semble la meilleure solution. Par exemple, la relative française dont nous venons de parler sera traduite en allemand de la manière suivante :

15. *die Niederschläge, die anhalten, und die Setzung der Schneedecke führten zu einer Abnahme der Lawinengefahr.*

Ce problème de surgénération se pose aussi lors du traitement d'autres cas de divergences. Prenons cette fois l'exemple de l'adverbe *vereinzelt*, qui se traduit par l'adjectif *isolé* en français (cf. section 3., exemple 2.). Comme la forme dictionnaire des adverbes en allemand est semblable à celle des adjectifs, il est impossible d'obtenir la même

représentation pour l'adjectif et l'adverbe allemands, sans provoquer la surgénération de plusieurs solutions, l'une avec l'adverbe et l'autre avec l'adjectif :

16. *Grundlawinen sind vereinzelt noch möglich*
17. *vereinzelte Grundlawinen sind noch möglich.*

La phrase *Grundlawinen sind vereinzelt noch möglich* et sa traduction recevront donc deux représentations distinctes :

18. args:[<1>args: args = []
 pred = grund
 detype = non
 mod = []
 detype = indefinite
 pred = lawinen]
 mod: adverbial: pred = vereinzelt
 phrase: pred = noch
 pred: möglich
19. args:[<1>args: args = []
 pred = fond
 detype = non
 mod: pred = isole
 detype = indefinite
 pred = avalanches]
 mod: adverbial: pred = []
 phrase: pred = encore
 pred: possible

La mise en correspondance de ces deux structures sera cette fois encore assurée par une règle de transfert. Cette dernière établit le lien entre l'adverbe modifieur de phrase contenu dans le chemin MOD:ADVERBIAL:PRED (*vereinzelt*) et le modifieur MOD:PRED du premier argument (A2) de la liste d'arguments de ARG^S⁷. La queue de la liste (B1 et B2) reste inchangée dans les deux langues.

20. :T: isoles_vereinzelt
 :L1: <*args> = [A1|B1]
 <*mod adverbial pred> = vereinzelt
 <A1 args> = C1
 <A1 mod> = []
 <A1 detype> = D1
 <A1 pred> = J1
 :L2: <*mod adverbial pred> = []
 <*args> = [A2|B2]
 <A2 mod pred> = isole
 <A2 args> = C2
 <A2 detype> = D2
 <A2 pred> = J2
 <*reversibilite> = non
 :X: B1 = B2
 C1 = C2
 D1 = D2
 J1 = J2

Comme dans l'exemple précédent, le trait <*reversibilite> = non bloque la réversibilité de la règle. Si nous traduisons cette même phrase du français vers l'allemand, le système appliquera cette fois les règles générales et produira une représentation

identique au français où *vereinzelt* figure comme adjectif épithète, ce qui permettra la génération de la phrase :

21. *vereinzelte Grundlawinen sind noch möglich.*

Enfin, pour ne citer qu'un dernier exemple, nous avons vu que certains adjectifs en allemand se traduisent de préférence par des groupes nominaux complexes en français :

22. *die übrigen Alpen* -> le reste des Alpes

23. *die östlichen Alpen* -> la partie est des Alpes

Il est clair qu'il serait également possible, dans certains cas, d'utiliser des adjectifs en français, mais ceci n'est pas conforme aux habitudes des bulletins d'avalanches et doit de ce fait être évité.

Nous avons donc préféré définir une règle de transfert qui transformera la structure allemande suivante :

24. args = []
 pred = alpen
 detype = definite
 mod : pred = ostlich

en une structure qui permettra la génération de la traduction proposée ci-dessus :

25. args : args = []
 detype = definite
 pred = alpes
 pred = partie
 detype = definite
 mod : pred = est

Cette règle de transfert présente la syntaxe suivante et établit une correspondance entre le prédicat principal allemand X1 (*alpen*, dans notre exemple) et l'argument du nom *partie* (X2, c'est-à-dire *alpes*) :

26. :T: ostlichen_partie_est_de
 :L1: <*args> = Y1
 <*pred> = X1
 <*mod pred> = ostlich
 :L2: <*args args> = Y2
 <*args pred> = X2
 <*args detype> = definite
 <*pred> = partie
 <*mod pred> = est
 :X: Y1 = Y2
 X1 = X2

Notons toutefois qu'un syntagme nominal comme *les Alpes est* ne sera pas exclu pour autant par la grammaire française. L'analyseur produira une structure semblable à celle que nous avons vue pour l'allemand, et ce syntagme sera lui aussi traduit en allemand par *die östlichen Alpen*.

Les problèmes d'équivalence se posent différemment. Il s'agit cette fois de définir des règles de transfert qui sont sensibles au contexte. Prenons comme exemple le traitement du mot *gross* (section 3., exemples 6. et 7.). Il est nécessaire de rédiger deux règles de transfert : la première, atomique, stipule que *gross* se traduit de manière générale par *gros*.

27. :TA: gross gros

La seconde (exemple 29.) traite du cas plus spécifique de *gross* qui modifie le nom *Gefahr*. Elle stipule que si l'adjectif *gross* modifie le nom *Gefahr*, c'est-à-dire dans le cas suivant :

28. mod : pred = gross
pred : gefahr

ce dernier se traduit en français par *grand*.

29. :T: gross_grand
:L1: <*mod pred> = gross
 <*pred> = gefahr
:L2: <*mod pred> = grand
 <*pred> = danger
:X: -

Dans le système actuel, en effet, une règle spécifique bloque l'application d'une règle générale.

Les règles de transfert spécifiques nous permettent donc d'obtenir une traduction de qualité, tout en évitant les problèmes de surgénération.

5. CONCLUSION

Dans cet article, nous avons mis en évidence les problèmes de traduction que nous avons rencontrés lors de l'implémentation du système de traduction automatique des bulletins d'avalanches et nous avons montré comment nous avons pu les résoudre avec le logiciel ELU. Nous avons démontré que, bien que nous utilisions la méthode du transfert, une représentation sémantique cohérente des bulletins d'avalanches est indispensable pour résoudre un grand nombre de problèmes de divergences et d'équivalences, sans complexifier le système. Grâce à cette représentation, la définition des règles de transfert spécifiques se limite aux seuls cas où l'obtention d'une même représentation provoque la surgénération ou lorsque le choix d'une traduction dépend du contexte.

Notes

1. Nous remercions l'IFENA qui subventionne partiellement ce projet de traduction automatique des bulletins d'avalanches de la Suisse.
2. Les chemins qui permettent d'accéder aux valeurs sont indiqués ici par des majuscules.
3. Les noms sont codés sans majuscules ni accents dans le prédicat.
4. La syntaxe des listes en ELU est presque identique à celle du langage de programmation Prolog. Une liste vide s'indique par [], tandis qu'une liste non vide consiste en une tête (le premier élément) et une queue (tout le reste). Une liste avec une tête Y et une queue X se note [Y|X]. Cf. Estival (1990a : 23).
5. Les noms sont codés dans le lexique avec un type sémantique (*neige*, **danger**, *lieu*, etc.).
6. Comme l'adverbe *vereinzelt* peut modifier n'importe quel type de verbe (transitif, bitransitif, etc.), il est possible que la liste d'arguments contienne plusieurs éléments.
7. Pour plus de détails, voir Bouillon *et al.* (1992).

BIBLIOGRAPHIE

- BARNETT, J., MANI, I., RICH, E., AONE, CH., KNIGHT, K. et J.C. MARTINEZ (1991) : «Capturing Language-Specific Semantic Distinctions in Interlingua Based MT», *Proceedings of Machine Translation Summit III*, Washington, pp. 25-32.
- BOESEFELDT, K. et P. BOUILLON (1991) : «Le rôle de la représentation sémantique dans un système de traduction multilingue», *Working Paper 58*, ISSCO, Genève.
- BOUILLON, P. et K. BOESEFELDT (1991a) : «Applying an Experimental MT System to a Realistic Problem», *Proceedings of Machine Translation Summit III*, Washington, pp. 45-49.

- BOUILLON, P. et K. BOESEFELDT (1992): «La traduction automatique des bulletins d'avalanches de la Suisse», Actes du Colloque *L'environnement traductionnel. La station de travail du traducteur de l'an 2001*, Les Presses de l'Université du Québec et l'AUPELF, pp. 69-78.
- BOUILLON, P., BOESEFELDT K. et G. RUSSELL (1992): «Compound Nouns in a Unification-Based MT System», *Proceedings of the 3rd Conference on Applied Natural Language Processing*, Trento, pp. 209-215.
- DORR, B. (1990): «Solving Thematic Divergences in Machine Translation», *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, Pittsburgh, pp. 127-134.
- ESTIVAL, D. (1990a): «Elu User Manual», *Technical Report*, ISSCO, Genève.
- ESTIVAL, D., BALLIM, A., RUSSELL, G. et S. WARWICK (1990b): «A Syntax and Semantics for Feature-Structure Transfer», *The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Language*, Austin, pp. 131-143.
- JAPKOWICZ, N. et M. WIEBE (1991): «A System for Translating Locative Prepositions from English to French», *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, pp. 153-160.
- KAMEYAMA, M., OCHITANI, R. et S. PETERS (1991): «Resolving Translation Mismatches With Information Flow», *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, pp. 193-200.
- RUSSELL, G., BALLIM, A., ESTIVAL, D. et S. WARWICK (1991): «A Language for the Statement of Binary Relation over Feature Structures», *Proceedings of Fifth European Association for Computational Linguistics*, Berlin, pp. 287-292.
- SALM, B. (1982): *Lawinenkunde für den Praktiker*, Bern, Verlag des SAC.
- SHIEBER, S.M. (1986): *An Introduction to Unification Based Grammar*, Stanford, CSLI.