



**UNIVERSITÉ
DE GENÈVE**

Archive ouverte UNIGE

<https://archive-ouverte.unige.ch>

Master

2008

Open Access

This version of the publication is provided by the author(s) and made available in accordance with the copyright holder(s).

Les Interlangues en TA : l'exemple de MedSLT

Gerlach, Johanna

How to cite

GERLACH, Johanna. Les Interlangues en TA : l'exemple de MedSLT. Master, 2008.

This publication URL: <https://archive-ouverte.unige.ch/unige:5563>

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.

Johanna Gerlach

Les Interlangues en TA : l'exemple de MedSLT

Mémoire présenté à l'Ecole de Traduction et d'Interprétation pour
l'obtention du Master en traduction – mention traductique.

Directeur : Prof. Pierrette Bouillon
Juré : Mme Marianne Santaholma

Université de Genève
Septembre 2008

Introduction.....	1
1 Traduction automatique.....	2
1.1 Introduction	2
1.2 Différentes approches en TA	2
1.2.1 Systèmes bilingues et systèmes multilingues	3
1.2.2 Architecture	3
1.2.2.1 Systèmes directs	4
1.2.2.2 Systèmes indirects.....	5
1.2.2.3 Systèmes fondés sur des informations non linguistiques.....	10
1.2.3 La traduction automatique en domaine limité	12
1.2.4 Le langage contrôlé.....	13
1.2.5 Conclusion	14
1.3 La traduction automatique de la parole.....	14
1.3.1 Reconnaissance de la parole	14
1.3.1.1 Approche statistique.....	15
1.3.1.2 Approche linguistique	16
1.3.2 Traduction.....	17
1.3.3 Synthèse de la parole	17
1.3.4 Conclusion	18
1.4 Conclusion.....	18
2 L'interlangue	19
2.1 Introduction	19
2.2 Domaine et contexte d'application	20
2.3 Langues du système	21
2.4 Difficultés.....	21
2.4.1 Représentation des objets lexicaux.....	21
2.4.1.1 Non-isomorphie	21
2.4.1.2 Trous lexicaux.....	22
2.4.1.3 Non-compositionalité.....	22
2.4.2 Représentation des relations temporelles et des aspects.....	23
2.4.3 Représentation des structures et fonctions.....	24
2.4.4 Représentation du contenu subjectif.....	26
2.5 Exemples d'approches interlangue	26
2.5.1 Rosetta – grammaire de Montague.....	26
2.5.2 Interlangues restrictives	27
2.5.2.1 NESPOLE - domain actions	28
2.5.3 DIONYSUS – interlangue à base de connaissances.....	29
2.5.4 UNL – interlangue maximaliste.....	31
2.6 Conclusion.....	33
3 Application de la TA de la parole au domaine médical	34
3.1 Introduction	34
3.2 Exemples de systèmes.....	35
3.2.1 Converser.....	35
3.2.2 S-MINDS.....	35
3.2.3 IBM MASTOR	36

3.3	MedSLT	37
3.3.1	Introduction.....	37
3.3.2	Architecture du système	37
3.3.3	Outils	39
3.3.3.1	Reconnaissance et synthèse de la parole: la plateforme Nuance	39
3.3.3.2	Développement et compilation des ressources linguistiques: Regulus 2.....	39
3.3.4	Etapes de la traduction dans MedSLT	43
3.3.4.1	Reconnaissance et analyse	43
3.3.4.2	Traduction	47
3.3.4.3	Description de l'interlangue.....	48
3.3.4.4	Règles de mise en correspondance.....	52
3.3.4.5	Retrotraduction	52
3.3.4.6	Génération.....	52
3.3.4.7	Synthèse de la parole	53
3.4	Conclusion.....	53
4	L'interlangue utilisée par MedSLT.....	54
4.1	Introduction	54
4.2	Caractéristiques de l'interlangue	54
4.2.1	Une interlangue lisible.....	54
4.2.1.1	Développement monolingue	54
4.2.1.2	Évaluation monolingue	55
4.2.2	Une interlangue vérifiable	56
4.2.2.1	L'interlangue comme source de connaissances.....	56
4.2.3	Conception de l'interlangue MedSLT	56
4.2.3.1	Une grammaire Regulus	56
4.2.3.2	Forme de surface de l'interlangue.....	57
4.2.3.3	L'interlangue comme source de connaissances.....	58
4.2.3.4	Interlingua checker.....	58
4.2.4	Conclusion	59
4.3	Evaluation de l'interlangue MedSLT	60
4.3.1	Expériences.....	60
4.3.1.1	Déroulement des expériences	62
4.3.2	Première expérience.....	62
4.3.2.1	Description de l'expérience	63
4.3.2.2	Résultats.....	64
4.3.2.3	Récapitulation des difficultés.....	75
4.3.2.4	Comparaison de l'interlangue structurée avec sa forme de surface.....	78
4.3.2.5	Conclusion	79
4.3.3	Deuxième expérience.....	79
4.3.3.1	Erreurs introduites.....	81
4.3.3.2	Résultats obtenus	83
4.3.3.3	Conclusion	94
4.3.4	Troisième expérience.....	94
4.3.4.1	Description de l'expérience	94
4.3.4.2	Résultats obtenus	95
4.3.4.3	Récapitulation des erreurs.....	106
4.3.4.4	Conclusion	109

5	Conclusion.....	110
6	Bibliographie	112

Annexe 1 : Expériences

Annexe 2 : Résultats 1^{ère} expérience

Introduction

La stratégie par interlangue présente plusieurs avantages pour les systèmes de traduction automatique. Elle facilite notamment le développement de systèmes multilingues destinés à fournir des traductions entre de nombreuses paires de langues. La traduction ne se faisant pas individuellement pour chacune des paires de langues, mais via une représentation sémantique abstraite centrale, le nombre de modules à développer pour l'ajout d'une nouvelle langue est réduit. Le développement d'un système par interlangue est cependant relativement complexe, puisqu'une grande partie du développement repose sur la manipulation d'une représentation abstraite, l'interlangue.

Dans ce travail, nous nous proposons d'étudier de plus près l'interlangue, en particulier sa lisibilité et compréhensibilité. Ces caractéristiques ont en effet une grande importance aussi bien pour le développement des ressources utilisées pour la traduction que pour l'évaluation des traductions elles-mêmes.

Pour ce faire, après une introduction générale à la traduction automatique (chapitre 1), nous nous intéresserons de plus près à la stratégie de traduction par interlangue et présenterons différents systèmes appliquant cette stratégie (chapitre 2). Nous allons ensuite, après une brève introduction à la traduction automatique de la parole pour le domaine médical, présenter le système MedSTL, développé à l'Université de Genève pour la traduction de dialogues de diagnostic médical (chapitre 3). Enfin, nous entrerons plus en détail sur l'interlangue utilisée dans MedSLT (chapitre 4).

Après cette partie théorique, nous présenterons plusieurs expériences destinées à évaluer la compréhensibilité de l'interlangue MedSLT pour des humains (chapitre 5). Ces expériences nous permettront d'une part de déceler les difficultés rencontrées par des humains non-spécialistes confrontés à l'interlangue utilisée dans MedSLT, d'autre part à voir quelles sont les possibilités d'évaluation d'une interlangue.

1 Traduction automatique

1.1 Introduction

La traduction prend de plus en plus d'importance dans nos sociétés multilingues. En effet, le nombre de documents qui doivent être disponibles en plusieurs langues ne cesse de croître. Au delà des besoins de communication écrite, il y a également de plus en plus de situations où des personnes parlant des langues différentes doivent communiquer par oral.

De ce grand besoin de traductions, et de l'impossibilité de le combler de façon économiquement intéressante par la traduction humaine, est né l'intérêt pour la traduction automatique (TA). De plus, les continuels progrès des performances des systèmes informatiques qui permettent de stocker et de traiter des quantités considérables de données a permis le développement de systèmes de plus en plus complexes.

La recherche en TA a vu passer de nombreuses approches et stratégies, plus ou moins complexes, employant entre autres des méthodes linguistiques, des calculs statistiques ou des ressources bilingues. Toutes les approches ont leurs avantages et inconvénients. Chaque contexte d'utilisation impose des contraintes qui influenceront le choix de la stratégie. Nous allons brièvement décrire les différentes stratégies usuelles dans la première partie de ce chapitre (1.2).

Un des objectifs de plus ou moins tous les systèmes de traduction automatique est de produire des traductions de haute qualité, qui ne nécessitent pas de postédition humaine. Cet objectif est cependant encore loin d'être atteint, du moins pour les systèmes destinés à traiter des domaines illimités. Pour améliorer la qualité des traductions, une stratégie souvent employée est d'imposer des restrictions aux textes source, que ce soit au niveau du domaine traité ou du langage utilisé. Nous parlerons de ces possibilités à la fin de la section 1.2.

Une variante particulière de la TA est la traduction automatique de la parole, dont nous parlerons à la fin de ce chapitre (1.3).

1.2 Différentes approches en TA

Dans cette section, nous allons présenter brièvement les différentes approches utilisées en TA. Ces approches se différencient par leur architecture, les ressources utilisées, mais aussi par le

résultat recherché.

Dans la suite nous allons d'abord parler des différences au niveau des langues que les systèmes sont destinés à traiter (1.2.1). Puis nous allons nous intéresser aux architectures des systèmes (1.2.2). Finalement nous parlerons des possibilités de restriction, au niveau du domaine ou du langage traité par les systèmes (1.2.3 et 1.2.4).

1.2.1 Systèmes bilingues et systèmes multilingues

On distingue les systèmes bilingues, destinés à traduire entre une seule paire de langues, des systèmes multilingues, conçus pour traduire entre plus de deux langues.¹

Les systèmes bilingues peuvent être réversibles ou non-réversibles. Dans un système bilingue réversible, le processus employé pour l'analyse d'une langue A peut être directement inversé sans modifications pour générer du texte cible dans la langue A. Les mêmes connaissances, par exemple sous forme d'une grammaire, sont utilisées pour l'analyse et la génération. Seuls les algorithmes utilisés diffèrent. La conception d'un tel système nécessitant des grammaires non-procédurales, la plupart des systèmes bilingues commerciaux comportent cependant deux systèmes unidirectionnels, parfois similaires.

Tout système implémentant plus que deux langues est considéré comme étant un système multilingue. Il y a toutefois des variations. Un système peut être prévu pour traiter 5 langues, et traduire dans les deux sens entre toutes les combinaisons de langues possibles. Mais on peut aussi imaginer un système ne traitant qu'une partie des combinaisons ou des directions.

Pour qu'un système soit réellement multilingue, et non un assemblage de systèmes bilingues unidirectionnels, il devrait avoir pour une langue donnée des composantes d'analyse et de génération fixes, indépendamment des autres langues impliquées. L'uniformité pourrait être étendue à l'approche linguistique appliquée aux différentes langues.

1.2.2 Architecture

Au niveau architectural, différentes approches sont possibles. La première étape est la compréhension, donc l'analyse du texte source. Nous allons voir que les différentes approches se différencient principalement au niveau de la profondeur de l'analyse. Les systèmes plus

1 Hutchins&Somers (1992)

simples, les systèmes directs, n'effectuent qu'une analyse minimaliste, tandis que les systèmes indirects effectuent une analyse plus approfondie, comme nous allons le voir par la suite.

1.2.2.1 Systèmes directs

Les systèmes directs, aussi appelés systèmes de première génération² ou minimalistes, sont généralement conçus pour une paire de langues donnée, ainsi que pour une seule direction de traduction.

Dans ces systèmes, la traduction se fait en deux étapes. La première consiste à effectuer une analyse minimaliste du texte source ; la deuxième à transposer le texte source en langue cible à l'aide d'un dictionnaire bilingue contenant les informations sur les composants lexicaux de la langue source avec leurs équivalents en langue cible. La traduction se fait pour ainsi dire mot à mot.

Pour analyser le texte source, ce type de système procède à une segmentation, ensuite à une analyse morphologique, et finalement à une désambiguïsation. Celle-ci repose sur des règles linguistiques définissant quel types de mots peuvent être suivis/précédés par quels autres types de mots. La plupart de ces systèmes ne font pas d'analyse de la structure syntaxique ni des relations sémantiques. La désambiguïsation lexicale ne repose donc que sur l'analyse morphologique. Certains systèmes effectuent toutefois une analyse de surface, permettant d'identifier les fonctions syntaxiques.

Après avoir traduit les mots par leur équivalent en langue cible, les systèmes directs appliquent des règles de réarrangement, qui servent à adapter à la langue cible l'ordre des mots dans la phrase résultante.

Cette approche de la traduction est très robuste puisqu'elle permet de traduire n'importe quel énoncé, même incorrect selon les règles de la langue source, pour autant que les mots du texte source figurent dans le dictionnaire du système. La qualité des traductions produites n'est par contre pas toujours satisfaisante, en particulier lorsqu'il y a d'importants décalages et divergences entre la langue source et la langue cible, qui ne peuvent pas être traités avec ce type d'approche.

Beaucoup de systèmes commerciaux fonctionnent selon ce principe.

2 Hutchins&Somers (1992)

1.2.2.2 Systèmes indirects

Les résultats peu satisfaisants obtenus avec les systèmes de première génération ont mené au développement de modèles linguistiques plus sophistiqués en partant de l'hypothèse qu'on ne peut obtenir une traduction de bonne qualité qu'en disposant de connaissances linguistiques de la langue source et de la langue cible, ainsi que des différences entre les deux langues. Plutôt que de traduire mot à mot, les systèmes indirects vont effectuer une analyse approfondie du texte source, pour en extraire une représentation qui servira de base pour la traduction. On distingue deux approches parmi les systèmes indirects: les systèmes de transfert et ceux par interlangue.

Les systèmes indirects utilisent des modèles linguistiques. Il s'agit de grammaires qui décrivent les constructions de phrases possibles dans une langue. Ces grammaires sont utilisées par des parseurs pour analyser les phrases source, et par un module de génération pour produire les phrases cibles. Dans les systèmes de transfert, il faut de plus une grammaire contrastive pour transposer la représentation source en représentation cible.

Le passage de la représentation source à la représentation cible sera d'autant plus simple que l'analyse linguistique sera approfondie, c'est-à-dire plus les représentations seront abstraites. Si on arrive à un niveau d'analyse suffisamment profond, on peut se passer entièrement de grammaire comparative. On a alors un système par interlangue. La distinction entre les systèmes de transfert et par interlangue est donc plus une question de degré d'analyse, que de différence de stratégie.

Les systèmes motivés linguistiquement ont pour désavantage d'être moins robustes que les systèmes directs, puisqu'ils ne peuvent analyser que les phrases qui correspondent aux constructions couvertes par leurs ressources linguistiques, donc qui sont décrites dans les grammaires. Par contre, ces systèmes utilisant aussi une grammaire pour la génération de la langue cible, ils produiront moins fréquemment des phrases cibles incorrectes au niveau syntaxique.

Nous allons maintenant revenir plus en détail sur ces systèmes

1.2.2.2.1 Systèmes de transfert

Ces systèmes procèdent en trois étapes: analyse, transfert et génération.

Ils vont d'abord effectuer une analyse approfondie du texte source, pour arriver à une

représentation syntaxique, qui est spécifique à la langue source.³ Cette étape nécessite une grammaire de la langue source indiquant comment les éléments peuvent être combinés pour former des phrases syntaxiquement correctes, et un lexique contenant les mots de la langue. La grammaire et le lexique forment une description linguistique de la langue source. La représentation source résultant de l'analyse contient donc, sous une forme hiérarchisée, toutes les composantes lexicales de la phrase source ainsi que les liens structurels entre ces éléments.

Vient ensuite l'étape de transfert, où le système applique des règles de transfert pour transformer la représentation source en une représentation équivalente pour la langue cible. Ces règles de transfert contiennent les informations sur les équivalences lexicales et structurelles pour une paire de langues.

Finalement, le système va générer une phrase cible à partir de la représentation cible, en utilisant une description linguistique de la langue cible. Le processus de génération est l'inverse de l'analyse.

Pour passer de la représentation source à la représentation cible, ces systèmes doivent effectuer un transfert lexical ainsi qu'un transfert structurel. Le transfert lexical consiste à remplacer chaque objet lexical source par un objet lexical cible. Ce remplacement est fort simple lorsqu'il existe une correspondance lexicale une à une, sauf lorsque le remplacement entraîne des différences au niveau syntaxique. Toutefois, hormis dans un contexte technique, où l'on aura une correspondance exacte au niveau des termes, cette correspondance lexicale n'est généralement pas donnée. On parle alors de décalages lexicaux. Une langue peut avoir un seul mot pour décrire plusieurs concepts, tandis qu'une autre langue aura un mot par concept. Le verbe 'know' en anglais, par exemple, peut être traduit en français par 'savoir' ou par 'connaître', selon le contexte. Prenons l'exemple suivant :

I know the answer.	I know what the answer is.
Je connais la réponse.	Je sais quelle est la réponse.

Dans la première phrase, 'know' doit être traduit par 'connaître', dans la deuxième par 'savoir'. Dans ce cas précis, donc la traduction de l'anglais vers le français, le choix peut être basé sur l'objet du verbe, qui dans la première phrase est un syntagme nominal, dans la deuxième une clause avec un infinitif. Un autre cas plus délicat serait le mot 'river', qui en français sera traduit soit par 'rivière' soit par 'fleuve', selon la taille du cours d'eau en

3 Hutchins&Somers (1992)

question. Dans ce cas, il n'y aucune information disponible au niveau de la phrase qui permettrait à un système informatique de faire le bon choix.⁴

Une autre difficulté rencontrée lors du transfert lexical sont les ambiguïtés de traduction. Celles-ci apparaissent lorsqu'un mot de langue source peut avoir plusieurs sens. Pour obtenir un résultat correct, le transfert requiert dans ce cas une inspection du contexte, instinctive pour un traducteur humain. Un système de traduction automatique pourra très facilement choisir la bonne solution parmi plusieurs si l'information de contexte se rapporte à une composante lexicale de la phrase, par exemple si un verbe ne peut prendre qu'un certain type de sujet. Pour que le système puisse choisir la bonne traduction pour la phrase suivante, il suffira de lui indiquer que le verbe 'manger' ne peut pas prendre un objet humain ('lawyer') :

Je mange un avocat.

I eat a lawyer. / I eat an avocado.

Cela devient plus difficile lorsque la phrase ne comporte pas d'informations utiles à la désambiguïsation. Si l'on prend par exemple la phrase suivante :

Les jumelles de l'opticien grossissent.

Elle pourrait être interprétée de plusieurs façons : les jumelles peuvent soit être des sœurs jumelles, filles de l'opticien, soit des instruments optiques. Ce n'est que grâce à la phrase qui la suit que l'on peut déduire qu'il s'agit des filles de l'opticien.

Les jumelles de l'opticien grossissent. Elles n'arrêtent pas de manger.

On ne peut traduire la première phrase correctement que si l'on prend en compte la phrase qui la suit, il faut donc avoir un moyen de traiter les informations extraphrasiques.

Un autre problème exigeant le recours à des informations extraphrasiques est la pronominalisation.⁵ Pour pouvoir choisir le bon pronom (genre, nombre), il faut savoir à quel nom il se réfère. Prenons les phrases suivantes :

The prohibition on torture is well established. It is also unambiguous and absolute.

L'interdiction de torturer est bien établie. Elle est sans équivoque et absolue.

Dans la deuxième phrase, 'it' doit être traduit par 'elle', puisque ce pronom renvoie à 'interdiction'. Pour pouvoir traduire ce pronom correctement, il faut donc mettre cette phrase en relation avec la phrase précédente.

4 Hutchins&Somers (1992)

5 L'Homme (2001)

En plus du transfert lexical, les systèmes de transfert effectuent un transfert structurel. Celui-ci devient nécessaire lorsque la structure de la langue source n'est pas correcte pour la langue cible. Les règles de transfert servent à changer les fonctions syntaxiques des différents éléments, par exemple à transformer un sujet en objet, mais aussi à nominaliser des constructions, entre autres. Théoriquement, plus l'analyse est approfondie, moins ce problème devrait être fréquent, puisque l'analyse vise à gommer les distinctions entre les langues.

Une possibilité est d'utiliser une représentation sémantique. Une approche utilisée fréquemment est celle des cas profonds (ou rôles sémantiques), basée sur la Grammaire des cas développée par le linguiste C.J. Fillmore⁶. Selon la grammaire des cas, une phrase est constituée d'un verbe ainsi que d'un ou plusieurs cas profonds. Les cas profonds marquent les relations sémantiques telles que Agent, Patient, Instrument etc. Ce type de représentation permet d'éliminer une partie des problèmes rencontrés lors du transfert structurel. En effet, le résultat d'une analyse selon la grammaire des cas sera usuellement représenté sous forme d'une structure de traits, et non sous forme d'un arbre reflétant la structure de la phrase. Il n'y a toutefois pas de consensus au niveau des cas profonds à utiliser, ils sont donc souvent définis pour chaque système de TA⁷. Ce genre de représentation est proche d'une interlangue, mais seulement en ce qui est des attributs structurels. Le transfert lexical doit encore être effectué de façon bilingue.

Les systèmes de transfert ont donc des modules bilingues pour chaque paire de langues, servant à transposer les représentations qui peuvent être plus ou moins approfondies. Il n'y a toutefois pas de représentation indépendante de la langue. L'ajout d'une nouvelle langue à un tel système implique la création de modules d'analyse et de génération pour la nouvelle langue, ainsi que d'un certain nombre de modules de transfert, en fonction du nombre de langues supportées par le système.

1.2.2.2 Systèmes par interlangue

Si l'analyse du texte source est poussée encore plus loin, jusqu'à l'extraction du sens, on parle de systèmes par interlangue. Dans ces systèmes, la traduction se fait par l'intermédiaire d'une représentation sémantique abstraite et idéalement indépendante des langues naturelles, appelée interlangue.

6 Hutchins&Somers (1992)

7 *ibid.*

Plutôt que d'utiliser des modules de transfert bilingues pour passer de la représentation de langue source à celle de langue cible, ces systèmes utilisent une représentation intermédiaire, la représentation interlingue. Cette représentation doit comporter tout le contenu sémantique nécessaire à la génération du texte cible, tout en étant aussi neutre que possible par rapport au texte source.

Tout comme les systèmes de transfert, les systèmes par interlangue analysent le texte source pour obtenir une représentation dépendante de la langue source. Ils appliquent ensuite des règles de mise en correspondance pour transformer celle-ci en représentation interlingue. Cette représentation interlingue est ensuite transformée avec d'autres règles de mise en correspondance pour obtenir une représentation dépendante de la langue cible qui servira à la génération de la langue cible.

Dans ce type de système, le processus de génération est généralement totalement indépendant du processus d'analyse, puisqu'il n'est plus possible de se reporter au texte source lors de la génération du texte cible, mais ce n'est pas forcément le cas. Dans MedSLT par exemple, les grammaires sont réversibles et sont utilisées aussi bien pour l'analyse que pour la génération.

L'approche par interlangue présente plusieurs avantages au niveau du développement des ressources linguistiques pour un système multilingue. D'une part, il n'est pas nécessaire de créer des modules de transfert pour chaque paire de langues ; d'autre part, les développeurs n'ont plus besoin de connaître toutes les langues traitées par le système.

Si l'on voulait créer un système de TA multilingue destiné à traiter toutes les combinaisons possibles entre n langues, avec une stratégie par transfert, il serait nécessaire de créer $n*(n-1)$ modules de transfert en plus de n modules d'analyse et n modules de génération. L'ajout d'une langue supplémentaire nécessiterait la création de n nouveaux modules de transfert⁸. Un système par interlangue n'ayant pas de modules de transfert, il ne nécessite que deux grammaires par langue, une pour l'analyse et une pour la génération. Un système avec n langues aurait donc $2n$ modules au total. L'ajout d'une langue supplémentaire à un tel système ne demande que l'ajout de deux nouveaux modules: une grammaire d'analyse et une grammaire de génération. Ce calcul repose toutefois sur une stratégie par interlangue idéale, où l'analyse de la langue source aboutirait directement à la représentation interlingue, qui pourrait directement être convertie en langue cible, ce qui n'est pas le cas. En réalité, en plus

8 Hutchins&Somers (1992)

des modules d'analyse et de génération, il faut développer pour chaque langue les règles de mise en correspondance pour passer de la représentation dépendante de la langue à la représentation interlingue et inversement. Ces règles ne devant toutefois être développés qu'une fois par langue, les systèmes par interlangue gardent un avantage certain par rapport aux systèmes par transfert lorsqu'il s'agit de développer des systèmes multilingues.

A condition d'avoir une représentation interlingue suffisamment explicite pour être utilisable par des humains comme une langue à part entière, le développement des grammaires et règles de mise en correspondance est simplifié. En effet les personnes développant les modules de transfert langue naturelle – interlangue n'auraient plus besoin de maîtriser une paire de langues, mais seulement une langue ainsi que l'interlangue. Cela présente aussi des avantages au niveau de l'évaluation de la qualité des traductions. Nous reviendrons sur cet aspect de l'interlangue dans le chapitre 4.

L'approche par interlangue comporte cependant plusieurs difficultés. D'une part la définition de l'interlangue, même pour des langues proches, telles que les langues romanes. En effet, il est difficile de créer une représentation qui soit totalement indépendante des langues naturelles, donc composée d'universaux sémantiques communs à toutes les langues. D'autre part, la complexité des modules d'analyse et de transfert nécessaires pour arriver à une représentation très éloignée des attributs caractéristiques des textes source et cible. En comparaison, les modules d'analyse et de génération d'un système de transfert sont plus simples, puisque les représentations intermédiaires sont des abstractions dépendantes de la langue.

Nous décrivons les interlangues plus en détail dans le chapitre 2.

1.2.2.3 Systèmes fondés sur des informations non linguistiques

Une approche totalement différente des systèmes directs et indirects est celle basée sur les corpus, parfois aussi appelée analogique. Elle se base sur le fait que l'on dispose de corpus bilingues, donc d'équivalents traductionnels, que l'on peut exploiter pour la traduction moyennant des méthodes statistiques⁹. Cette approche devient d'autant plus intéressante que le nombre de documents parallèles disponibles en format électronique ne cesse de croître, et que les progrès technologiques permettent de traiter des quantités considérables de données.

9 Somers (1992)

On distingue généralement trois formes d'utilisation des corpus bilingues¹⁰ : l'utilisation dans la traduction humaine assistée par ordinateur (TAO), l'utilisation dans des systèmes hybrides et l'utilisation dans des systèmes purement statistiques.

Une difficulté commune à toutes les approches reposant sur les corpus est celle de l'alignement des textes bilingues. En effet, dans les corpus bilingues, traduits par des traducteurs humains, la traduction n'est généralement pas une traduction mot à mot ou même phrase à phrase. Une unité lexicale source peut être traduite par une construction syntaxique en langue cible. La tâche d'extraction des équivalences lexicales bilingues est donc complexe.

1.2.2.3.1 La TA basée sur la statistique

Un exemple de cette approche sont les travaux de Brown *et al.* (1990) qui utilisent un corpus constitué des procès-verbaux bilingues du Parlement canadien. Cette approche repose sur l'idée que toute phrase d'un corpus en langue A a une probabilité plus ou moins grande d'être une traduction d'une phrase appartenant au corpus en langue B. Pour la traduction, il s'agit donc de trouver la phrase dans la langue cible ayant une probabilité maximale. La probabilité repose sur la probabilité de correspondance des mots individuels des phrases ainsi que sur l'ordre dans lequel les mots se trouvent.

Un avantage de cette approche est le fait qu'il n'est pas nécessaire de créer de modèles linguistiques, de grammaires, de lexiques, etc., puisqu'on exploite des traductions existantes.

1.2.2.3.2 Les systèmes hybrides

Il existe aussi des systèmes hybrides, combinant une approche linguistique avec des méthodes utilisant des corpus, afin d'obtenir une meilleure qualité de traduction. Ces systèmes utilisent généralement les stratégies linguistiques usuelles pour analyser les textes source et obtenir les informations syntaxiques et sémantiques, puis appliquent des méthodes basées sur les corpus pour trouver les meilleures traductions. Ils exploitent ainsi les connaissances disponibles dans les corpus bilingues pour optimiser le processus de traduction.

Un exemple d'approche hybride est celle de Sato et Nagao (1990), appelée 'TA basée sur la mémoire'. Elle consiste à analyser l'entrée, en produire toutes les traductions possibles par transfert lexical, puis comparer celles-ci avec des exemples stockés dans une base de données pour trouver la traduction la plus adéquate.

10 Somers (1992)

Une autre possibilité est de n'utiliser les exemples que pour traiter certains problèmes particuliers. C'est ce que font Sumita et Iida (1991), qui emploient les exemples pour résoudre les cas spéciaux, comme les idiomes, ou lorsque la traduction ne peut pas se faire de manière compositionnelle.

Le projet MedSLT utilise également des corpus, pour spécialiser les grammaires utilisées pour la reconnaissance de la parole ainsi que pour choisir les meilleures phrases lors de la génération. Nous décrivons ces processus plus en détail dans la section 3.3.3.

1.2.3 La traduction automatique en domaine limité

Un système exhaustif destiné à traduire la langue générale nécessiterait énormément d'informations syntaxiques, lexicales et sémantiques et serait probablement incapable de traduire correctement tous les cas où il y a ambiguïté lexicale ou structurelle, si fréquents dans la langue naturelle. De plus, il paraît difficile de concevoir un système capable de traiter la quasi-infinité d'énoncés possibles en langue naturelle. Pour cette raison, le développement de systèmes de traduction s'est tourné vers les possibilités permettant de restreindre la complexité des textes source, afin d'améliorer la qualité et la fiabilité de la traduction automatique.¹¹

Une possibilité est de restreindre le domaine d'application du système à un domaine particulier, donc créer un système destiné à ne traiter que des textes écrits dans le sous-langage propre à ce domaine. C'est le cas de nombreux systèmes, par exemple TAUM-METEO, développé à la fin des années 70 à l'Université de Montréal au Canada, destiné à la traduction des rapports météorologiques.

La restriction du contexte d'application du système à un domaine précis des connaissances humaines présente l'avantage que l'input soumis au système sera restreint aux textes relevant du domaine en question. Plutôt que de traiter la langue naturelle, on traite un sous-langage propre à ce domaine. Il s'ensuit une diminution du lexique, des constructions syntaxiques et des ambiguïtés que le système devra gérer. Le temps de développement d'un tel système diminue également, puisqu'il n'est plus nécessaire d'implémenter un nombre presque infini de constructions et d'entrées lexicales. La restriction du domaine d'application d'un système de traduction automatique permet de mieux cibler les ressources linguistiques utilisées.

Un autre avantage des domaines limités est la 'disparition' de bon nombre d'ambiguïtés au

¹¹ Kittredge (1987)

niveau lexical et structurel. D'une part, un mot polysémique de la langue générale n'aura souvent qu'un seul sens dans un domaine particulier. D'autre part, plus le domaine sera technique, plus il y aura de termes spécifiques au domaine, exempts de toute ambiguïté, et donc aisément traductibles.

1.2.4 Le langage contrôlé

Une autre possibilité pour améliorer la qualité de la traduction automatique est le contrôle du langage source, qui permet de réduire les ambiguïtés et la complexité des phrases.¹² Un langage contrôlé simplifie les constructions syntaxiques et limite le lexique afin que le sens des phrases soit très facilement compréhensible. Ces langages ont été développés principalement pour des domaines techniques. Ils sont nés du besoin des grandes compagnies qui opèrent à un niveau multinational, dans des domaines tels que l'aéronautique, et dont les communications doivent être claires et sans aucune ambiguïté, afin de diminuer les erreurs de manipulation et les risques d'accident.

Les langages contrôlés ont principalement été développés pour faciliter la compréhension par des lecteurs humains. Toutefois, les documents rédigés en langage contrôlé se prêtent également mieux à la traduction, qu'elle soit automatique ou humaine.¹³

Prenons par exemple le Simplified English, développé par l'AECMA¹⁴ pour la rédaction de manuels d'entretien aéronautique. Ce langage contrôlé est destiné à améliorer la compréhension humaine des instructions de maintenance, en particulier pour les personnes dont l'anglais n'est pas la langue maternelle. Il définit un vocabulaire de base et des règles syntaxiques décrivant les constructions autorisées, afin d'éliminer toute ambiguïté. Ce faisant, il permet de diminuer les erreurs d'interprétation et de rendre l'utilisation et l'entretien des installations plus sûrs. Pour la plupart des mots, une seule catégorie grammaticale est admise. De plus, chaque mot ne peut avoir qu'un seul sens. Des règles syntaxiques définissent entre autres la longueur des phrases, l'usage des verbes, la disposition du texte et la ponctuation. Cela rend non seulement les textes moins ambigus pour les lecteurs humains, mais présente aussi des avantages pour la traduction, puisque l'étape de désambiguïsation lexicale n'est pratiquement plus nécessaire. D'autre part, les phrases suivant des constructions limitées et précisément définies, l'analyse syntaxique sera beaucoup moins complexe.

12 Arnolds (1994)

13 Kittredge (2003)

14 European Association of Aerospace Manufacturers (1986)

Un désavantage des langages contrôlés est l'importance du travail de préparation des textes. Les textes à traduire doivent être pré-processés pour correspondre aux règles du langage contrôlé, ce qui demande un travail considérable de la part des rédacteurs. Cet effort est toutefois compensé par le gain de temps lors de la traduction. Les langages contrôlés sont donc particulièrement intéressants lorsqu'il faut produire un important volume de textes dans de nombreuses langues.

1.2.5 Conclusion

Nous avons vu que les différentes approches en TA ont toutes leurs avantages et inconvénients, les rendant plus ou moins intéressantes selon le contexte d'application. Les systèmes de TA les plus simples, les systèmes directs se distinguent par leur robustesse mais peinent à produire des traductions de bonne qualité. Les systèmes indirects produisent généralement de meilleures traductions mais sont moins robustes. Les systèmes par interlangue se prêtent particulièrement aux systèmes multilingues. Finalement, une dernière catégorie utilise des documents multilingues et des méthodes statistiques plutôt que des ressources linguistiques. Souvent les systèmes de TA combinent plusieurs approches, pour profiter des avantages de chaque stratégie.

Dans la section suivante, nous allons parler d'une application particulière de la TA, la traduction automatique de la parole.

1.3 La traduction automatique de la parole

Pour la traduction automatique de la parole, un système doit tout d'abord pouvoir reconnaître la parole, c'est-à-dire analyser le son entrant afin d'obtenir la phrase source correcte. Il doit ensuite traduire cette phrase, tout comme le font les autres systèmes de TA. Finalement, il synthétise la parole à partir de la phrase cible. Nous allons discuter ces étapes plus en détail dans les chapitres suivants.

1.3.1 Reconnaissance de la parole

La reconnaissance de la parole consiste à transformer le signal acoustique enregistré par un microphone en une chaîne de caractères. Elle est sujette à plusieurs difficultés. Un mot peut avoir une multitude de prononciations différentes en fonction du contexte, de l'accent du

locuteur, du type de phrase, etc. De plus, tous les mots ne sont pas accentués de la même manière: certains mots pourront être distingués aisément tandis que d'autres seront peu accentués, les articles par exemple, qui en plus d'être très courts sont souvent contractés avec les noms qui les suivent. Ces mots sont donc très difficiles à reconnaître. A l'inverse, il y a également des sons (par exemple « euh ») qui peuvent être reconnus, mais qui ne devraient pas être traités comme s'il s'agissait de mots.

Une autre difficulté de la traduction automatique de la parole est liée au fait que, contrairement à la langue écrite, la langue parlée comporte souvent des phrases elliptiques ou mal formées, voire des mots manquants. Un système de traduction automatique de la parole doit donc pouvoir traiter ces cas¹⁵. Nous décrirons plus en détail cette problématique en rapport avec MedSLT dans le chapitre 3.

A ces difficultés liées à la langue elle-même, s'ajoutent les facteurs purement techniques, tels que le bruit de fond ou la qualité d'enregistrement par exemple.

Les systèmes de reconnaissance vocale utilisent deux sources de connaissances : un modèle acoustique, qui spécifie les correspondances entre les phonèmes du langage et les ondes acoustiques, et un modèle du langage, qui spécifie quelles séquences de mots sont permises. En effet, la combinatoire étant très grande et une séquence de sons pouvant correspondre à différentes combinaisons de mots, il faut trouver un moyen de décrire les séquences de mots valides. Deux approches sont possibles: l'approche statistique basée sur des corpus et l'approche linguistique basée sur des grammaires.¹⁶

1.3.1.1 Approche statistique

Beaucoup de systèmes de traduction automatique de la parole emploient un modèle statistique du langage (SLM, statistical language model) pour la reconnaissance de la parole¹⁷. Cette approche repose sur l'hypothèse que chaque élément dans une séquence de sons peut être interprété de différentes façons, mais qu'une seule combinaison de ces interprétations est probable d'un point de vue statistique. On assigne des probabilités aux combinaisons de mots d'après leur fréquence d'apparition dans des corpus. Cette approche fonctionne avec des paires

15 Kitano (1993)

16 Torckola (1994)

17 Rayner *et al.* (2005)

ou des trigrammes de mots. Si pour un son, plusieurs mots sont possibles, on choisira le mot qui a la plus grande probabilité de suivre le mot précédant. Un point faible de cette approche est le manque de contraintes globales sur les phrases. Il est donc tout à fait possible d'obtenir des phrases agrammaticales.

De plus, comme toute approche reposant sur des corpus, la qualité des données à disposition est déterminante pour le résultat. Suivant le domaine, il peut être difficile d'obtenir des corpus suffisamment vastes pour être exploitables. Dans ces cas, l'approche linguistique est souvent la plus pratique.

1.3.1.2 Approche linguistique

Le modèle du langage peut être défini par une grammaire, qui encode des contraintes syntaxiques ainsi que des contraintes sémantiques ou liées au domaine. On parle alors de GLM (grammar-based language model). La grammaire décrit tous les énoncés admis. Le reconnaisseur utilise la grammaire pour trouver une séquence de mots correcte qui corresponde à la séquence de sons enregistrée. Les contraintes syntaxiques définies par la grammaire assurent que les phrases sont grammaticalement correctes. Les contraintes sémantiques sont ajoutées à la grammaire sous forme de restrictions de sélection, sans nécessiter de modification de la structure de la grammaire. Les restrictions de sélection permettent entre autres de définir quels types d'objets un verbe peut avoir, quels adjectifs peuvent qualifier quel nom, etc. On réduit ainsi le nombre d'énoncés incorrects d'un point de vue sémantique. Les contraintes sémantiques ont toutefois des limites, puisque les grammaires ne permettent pas d'inclure des contraintes portant sur le contexte extraphrasique. L'approche linguistique présente l'avantage par rapport à l'approche statistique que toutes les phrases reconnues seront grammaticalement correctes et cohérentes.

Par contre, comme pour toute approche reposant sur des grammaires, il n'est pas possible de créer un système exhaustif capable de traiter tous les énoncés de la langue naturelle. Les énoncés non-couverts par la grammaire ne pourront pas être reconnus. Cette approche est donc adaptée lorsque le discours à reconnaître est relativement stéréotypé. L'approche linguistique peut être très efficace lorsqu'il est possible d'entraîner les utilisateurs à l'emploi du système de reconnaissance¹⁸.

Certains systèmes utilisent les deux approches. MedSLT par exemple associe une

18 Rayner *et al.* (2004)

reconnaissance principale basée sur des grammaires à un module d'aide utilisant des méthodes statistiques pour traiter les énoncés non-couverts par les grammaires. Nous reviendrons plus en détail sur la reconnaissance de la parole dans MedSLT dans le chapitre 3.

1.3.2 Traduction

Les systèmes de traduction automatique de la parole emploient généralement les mêmes stratégies que les systèmes de traduction automatique de textes écrits. Les exigences au niveau du résultat de la traduction ne sont cependant pas les mêmes. En effet, dans le contexte d'un dialogue on préférera généralement des traductions plus idiomatiques que littérales.

L'approche par interlangue, utilisée par exemple dans le projet NESPOLE, se prête bien à la traduction de la parole, puisqu'elle produit plutôt des paraphrases. Elle permet de plus de retourner à l'énonciateur une paraphrase de son énoncé, lui permettant ainsi de valider l'énoncé reconnu avant qu'il ne soit traduit.¹⁹ Cet aspect est également exploité dans MedSLT, comme nous le verrons dans le chapitre 3.

1.3.3 Synthèse de la parole

La synthèse de la parole permet de transformer la phrase cible en un signal acoustique. L'objectif est de générer un signal aussi naturel et intelligible que possible. Il faut donc reproduire la coarticulation entre les sons et une prosodie naturelle.²⁰

La phrase doit généralement d'abord être transposée en une séquence de phonèmes de la langue cible. Différentes techniques peuvent ensuite être utilisées pour générer le signal acoustique. La plus simple consiste à concaténer des morceaux de parole naturelle enregistrés. On recompose l'énoncé à partir de mots, de syllabes, de diphones, d'expressions prononcées par un humain. Ces sons peuvent ensuite être déformés pour gérer la prosodie. Une autre possibilité est la synthèse vocale par formants. La voix est créée artificiellement, basée sur une modélisation du spectre sonore de la parole humaine. Les phonèmes sont générés à partir des plages de fréquence qui leurs sont propres, appelées formants.²¹

19 Lavie *et al.* (2001)

20 Dutoit *et al.* (2002)

21 Styger, Keller (1994)

1.3.4 Conclusion

La traduction automatique de la parole présente des difficultés supplémentaires puisque déjà l'acquisition des énoncés source est sujette à erreurs. Pour éviter les erreurs, on utilise un modèle du langage, qui peut être linguistique ou statistique. La traduction en tant que telle repose sur les stratégies usuelles, avec la difficulté supplémentaire du traitement des énoncés mal reconnus, ainsi que des phrases incomplètes ou agrammaticales typiques du langage parlé. Finalement, la parole peut être générée en utilisant différentes méthodes.

1.4 Conclusion

Les approches en TA sont au moins aussi nombreuses que les domaines d'application de la TA. A chaque domaine correspondent des exigences, que soit au niveau de la précision, du nombre de langues à traiter, du langage à couvrir, etc. Il n'y a pas de stratégie universellement applicable, chacune a ses avantages et inconvénients.

Après cette introduction générale à la TA, nous allons dans le chapitre suivant parler plus en détail de la stratégie par interlangue.

2 L'interlangue

2.1 Introduction

Comme nous l'avons vu dans le chapitre précédant, certains systèmes de traduction automatique utilisent une stratégie par interlangue. Celle-ci consiste à transposer le texte source vers une représentation intermédiaire, la représentation interlingue, à partir de laquelle sont produits les textes cibles de toutes les langues du système. Il n'y a donc pas de transfert direct entre chaque paire de langues, mais un transfert depuis chaque langue vers l'interlangue et inversement. Dans ce chapitre, nous allons nous intéresser de plus près à cette représentation intermédiaire.

On a vu que l'interlangue doit contenir tous les éléments sémantiques apportés par la phrase source et nécessaires à la production de la phrase cible. Idéalement, elle devrait représenter le sens transmis par les mots ainsi que par les structures dans lesquelles ils se trouvent, tout en étant indépendante des langues naturelles. On voudrait donc représenter aussi bien les aspects sémantiques que pragmatiques. Cela implique plusieurs difficultés, et les avis sur la définition d'une interlangue sont très partagés.

La définition d'une interlangue sera forcément dépendante du domaine et du contexte d'utilisation du système de traduction dans lequel elle sera utilisée. La traduction de la parole devra satisfaire d'autres exigences que la traduction de manuels techniques. L'ampleur du domaine d'application et la spécificité du langage utilisé auront une influence sur les aspects à représenter mais aussi sur la granularité de la représentation.

Un autre facteur à prendre en compte est la neutralisation plus ou moins forte des caractéristiques structurelles, qui aura un impact sur la similitude syntaxique entre les phrases source et cible. Si l'on veut une traduction plutôt littérale, il faudra conserver plus d'informations structurelles que si l'on veut une paraphrase, qui pourrait être plus idiomatique. Il n'y a toutefois pas de limite claire entre une traduction et une paraphrase dans une autre langue. Pour Nirenburg²² la traduction consiste à trouver des équivalences de sens et non de structures syntaxiques, ce qui équivaut presque à une paraphrase. Ce choix sera également influencé par le domaine d'application.

22 Nirenburg (1990)

Le choix des langues pour lesquelles l'interlangue sera l'intermédiaire aura aussi une influence sur celui-ci. Plus les langues seront proches, moins l'abstraction devra être conséquente.

Un dernier facteur à prendre en compte pour les projets plus vastes, dont le développement est réparti entre différents centres de recherche, est la complexité de l'interlangue. En effet, si l'interlangue est trop complexe, elle ne peut être employée de façon fiable par toutes les personnes impliquées dans le développement du système.²³

Dans les sections suivantes, nous allons parler tout d'abord de l'influence que le domaine d'application et les langues du système ont sur l'interlangue (2.2 et 2.3). Nous viendrons ensuite sur les difficultés rencontrées lors de la définition d'une interlangue. Ensuite nous nous intéresserons plus en détail à la représentation de plusieurs aspects, à savoir les relations temporelles, les structures et le contenu subjectif (2.4). Pour clore ce chapitre nous allons présenter quelques systèmes par interlangue existants (2.5).

2.2 Domaine et contexte d'application

Le choix d'une forme interlangue sera fortement influencé par le domaine d'application du système. Une interlangue destinée à traiter la langue générale devra représenter une quasi-infinité de concepts et de constructions. Une telle interlangue n'étant pratiquement pas faisable, la plupart des systèmes par interlangue actuels se concentrent sur un domaine ou sous-domaine précis. Le nombre de concepts à représenter est ainsi restreint. D'autres systèmes sont encore plus restrictifs en imposant un langage contrôlé pour les textes source, ce qui réduit le nombre de constructions et de formulations à représenter. C'est le cas du système KANT²⁴, développé à Carnegie Mellon pour la traduction de documentation technique.

Lors de la définition d'une interlangue pour un domaine précis, si on veut en assurer la portabilité entre domaines, il faudra faire une distinction claire entre les composantes générales et celles spécifiques à un domaine.²⁵

Les types de textes à traduire ont aussi une influence sur la définition de l'interlangue. Pour la

23 Levin *et al.* (2002)

24 Mitamura (1999)

25 Levin *et al.* (2002)

traduction de dialogues par exemple, on sera plus intéressés à avoir une paraphrase idiomatique qu'une traduction plus littérale conservant les caractéristiques syntaxiques de la phrase source.

2.3 Langues du système

Théoriquement, une interlangue devrait être indépendante des langues naturelles. Les langues du système auront toutefois une influence sur la définition de l'interlangue, pour plusieurs raisons.

D'une part, les personnes développant l'interlangue seront naturellement influencées par leur propre langue dans la façon de conceptualiser les choses énoncées.²⁶

D'autre part, les différences syntaxiques et lexicales entre les langues auront une influence sur le degré d'abstraction nécessaire. Moins les langues auront de différences au niveau lexical et syntaxique, moins la représentation devra être abstraite. Dans le contexte d'un système multilingue auquel de nouvelles langues seront ajoutées ultérieurement, une réadaptation de l'interlangue sera peut être nécessaire pour couvrir toutes les nuances.

Les problèmes rencontrés lors de la définition d'une interlangue sont très similaires à ceux observés dans une architecture de type transfert. Nous allons décrire ces difficultés dans le chapitre suivant.

2.4 Difficultés

2.4.1 Représentation des objets lexicaux

La non-isomorphie entre les langues, les trous lexicaux ou la non-compositionalité de certaines expressions, problèmes qui doivent être résolus dans les systèmes par transfert, influencent également la définition d'une interlangue.

2.4.1.1 Non-isomorphie

Une difficulté rencontrée lors de la traduction est le fait que les langues ne découpent pas la réalité de la même façon, c'est-à-dire les cas de non-isomorphie. Il se peut qu'une langue utilise le même mot pour décrire plusieurs concepts, tandis qu'une autre aura un mot par

26 Nirenburg (1990)

concept. Par exemple, en allemand, on fait la distinction entre *dasselbe*, qui désigne *le même*, dans le sens de *celui-ci et pas un autre* et *das gleiche*, qui signifie *le même, mais pas forcément celui-ci*. Si l'on se contentait donc d'une seule représentation du concept *même* dans l'interlangue, ce qui semblerait adéquat si l'on envisageait une traduction français-anglais, on aurait une perte de précision en cas de traduction vers l'allemand. On aura le même problème dans tous les cas où une langue dispose de plusieurs mots qui correspondent à un seul mot dans une autre langue. On peut donc choisir arbitrairement une langue dont on va adopter les nuances, ce qui est parfois qualifié d'« impérialisme culturel »²⁷. On peut aussi démultiplier les distinctions pour créer une interlangue maximaliste qui permettrait de représenter explicitement toutes les distinctions qui pourraient être exprimées dans une des langues du système. Cela ne semble toutefois que difficilement réalisable en pratique, à moins de ne traiter qu'un sous-domaine très limité.

2.4.1.2 Trous lexicaux

Une autre complication rencontrée lors du développement d'une interlangue est ce qu'on appelle les *trous lexicaux*, c'est à dire les cas où une langue doit employer une phrase pour exprimer ce qu'une autre exprime avec un seul mot. Il est alors difficile de choisir quelle variante représenter dans l'interlangue. Généralement, le seul moyen de représenter ces cas de façon universelle est de passer à un niveau d'abstraction supplémentaire, c'est à dire de se détacher des moyens lexicaux et syntaxiques utilisés dans les différentes langues.

2.4.1.3 Non-compositionnalité

Dans la langue parlée, on peut observer des constructions qui, tout en étant conforme aux règles syntaxiques d'une langue, ne peuvent être analysés de façon compositionnelle à partir des éléments qui la composent. D'autres locutions ne suivent même pas la syntaxe correcte d'une langue. Dans un dialogue, on est également souvent confronté à des expressions fixes, qui ne peuvent pas être traduites littéralement ou qui ont une signification trop indirecte pour que le sens littéral soit intéressant à traduire ('how about', 'would you mind')²⁸. Une phrase dans une langue A et sa traduction idiomatique en langue B peuvent n'avoir en commun rien d'autre que l'intention du locuteur (faire une suggestion, par exemple).

Ces expressions sont donc plus difficiles à traiter de façon lexico-sémantique. La stratégie

27 Nirenburg (1990)

28 Levin *et al.* (2002)

utilisant des 'domain action', que nous décrirons plus tard, permet de traiter ces expressions.

2.4.2 Représentation des relations temporelles et des aspects

Une interlangue doit également pouvoir représenter les conditions temporelles d'un énoncé, c'est-à-dire la relation temporelle entre le moment de l'énoncé et les événements décrits. Il faut donc développer une représentation des temps et aspects verbaux, qui sont d'une grande importance pour produire des traductions correctes. Cette tâche est complexe, puisque chaque langue naturelle emploie d'autres définitions des temps verbaux, et utilise d'autres moyens pour exprimer la continuité, la répétition etc. Un énoncé en français au présent de l'indicatif peut par exemple être traduit de différentes manières en anglais, selon la relation temporelle entre l'action décrite et le moment de l'énonciation:

Il mange.	He is eating.
	He eats.

Il faut donc pouvoir représenter ces nuances dans l'interlangue.

Chaque langue ayant des définitions de temps différentes, on ne peut pas utiliser les définitions des temps verbaux propres à une langue dans l'interlangue. Les distinctions existantes dans une langue ne seront pas forcément suffisantes pour décrire toutes les nuances de temporalité existantes dans une autre langue. On risque donc d'avoir un manque de précision au niveau de l'interlangue, et donc des traductions inexactes. Il est de plus en plus probable qu'il y ait une correspondance exacte entre les temps.

Une possibilité est d'utiliser dans l'interlangue des définitions générales communes aux différentes langues, telles que présent, passé etc. ainsi que les aspects verbaux généraux tels que perfect, continu, etc. C'est ainsi que les verbes sont représentés dans l'interlangue de MedSLT. Prenons par exemple la phrase suivante :

```
fr : 'la douleur dure-t-elle depuis plusieurs heures'
en : 'has the pain lasted for several hours'
interlangue : [[aspect,perfect], [prep,duration], [spec,several], [state,last], [symptom,pain],
[tense,present], [timeunit,hour], [utterance_type,ynq], [voice,active]]
```

Les conditions temporelles de la phrase sont représentées par les deux paires attribut-valeur [tense,present] et [aspect,perfect].

Il est aussi possible de représenter les conditions temporelles de façon plus abstraite, sans se référer aux définitions usuelles. Plutôt que d'utiliser des aspects verbaux, on décrit

explicitement la relation temporelle entre l'énoncé et le moment de l'énonciation. C'est le cas par exemple dans l'interlangue UNL²⁹. Les aspects temporels y sont représentés par des attributs associés aux prédicats. UNL dispose de deux attributs pour décrire les relations temporelles.

Un premier attribut décrit à quel moment se situe l'évènement décrit par rapport au moment du récit. Cet attribut est attaché au prédicat principal et peut prendre les valeurs @past, @present, et @future. Un deuxième attribut décrit le point de vue du locuteur sur le déroulement de l'évènement : le locuteur peut mettre l'accent sur le début de l'évènement, sur la fin, sur le progrès, sur l'incomplétude, etc. Cet attribut peut prendre les valeurs @complete, @continue, @repeat, @yet, etc.

Les exemples suivants³⁰ illustrent le fonctionnement de ces attributs :

He went on talking.
talk.@continue.@past

He has just come.
come.@complete.@just

It is so windy that the tree branches are knocking against the roof.
knock.@entry.@present.@repeat

Ce type de représentation présente l'avantage d'être absolument détaché des temps verbaux utilisés dans les langues naturelles. Il devrait donc pouvoir traiter tous les cas de temporalité possibles, dans toutes les langues.

2.4.3 Représentation des structures et fonctions

La plupart des interlangues utilisent des représentations structurées de façon similaire à celles utilisées dans les systèmes de transfert, avec un degré d'abstraction supplémentaire. En effet, les systèmes par interlangue utilisent généralement un processus d'analyse similaire aux systèmes de transfert, pour arriver à une représentation de la phrase source qui comporte les éléments lexicaux ainsi que les interrelations syntaxiques ou sémantiques. Plutôt que de passer directement de cette représentation source à une représentation cible par transfert lexical et structurel, les systèmes par interlangue transforment cette représentation en une forme plus abstraite qui neutralise les singularités inhérentes aux structures profondes des

29 Uchida *et al.* (2005)

30 Exemples trouvés sous :

<http://www.unl.org/unlsys/unl/unl2005/attribute.htm#3.2%20%20Attributes%20describing%20times%20with%20respect%20to%20the%20speaker> (15.07.08)

langues. Le degré d'abstraction sera influencé par le type de représentation choisie.

Par définition, une représentation interlingue ne devrait pas être influencée par la syntaxe d'une langue ou d'une autre. Elle doit toutefois comporter toutes les informations transmises par la syntaxe. Une représentation interlingue qui conserve des informations syntaxiques de la langue source permettra d'obtenir une phrase cible utilisant les mêmes moyens syntaxiques, ce qui pourrait être désirable dans certains cas, si l'on veut une traduction aussi proche que possible de la phrase source. Celle-ci risque toutefois d'être peu idiomatique en raison des divergences entre les moyens syntaxiques employés par les différentes langues. Une interlangue qui ne conserve aucune information syntaxique de la langue source se prêtera plutôt à la paraphrase multilingue. On obtiendra cependant des phrases plus idiomatiques.

Pour définir une interlangue détachée de la syntaxe des langues naturelles, plusieurs approches sont possibles.

Une possibilité est de renoncer à toute structuration, et de mettre tous les éléments au même niveau, dans une représentation suivant la logique des propositions, ne contenant aucune information portant sur la structure syntaxique. Selon Hutchins et Somers³¹ une telle représentation serait inadéquate pour la traduction, car il lui manque les informations nécessaires à la génération d'une phrase cible qui serait une traduction exacte. Dans ce type de représentation il n'est par exemple pas possible d'indiquer quel est le sujet de la phrase. L'exemple suivant³² illustre ce type de représentation :

source : Any government is dependent on its supporters.

interlangue : all(X), government(X), indefinite(Y), plural(Y), support(Y,X,T), depend-on(X,Y,T), timeless(T)

Les paraphrases générées à partir de cette représentation sont les suivantes :

Every government has supporters which it depends on.
 People who support all governments are depended on by them.
 All governments depend on people who support them.

Ces paraphrases n'expriment toutefois pas exactement le même sens que la phrase source, l'accent n'étant pas mis sur les mêmes éléments.

D'après Hutchins et Somers il est communément admis que les représentations interlangue pour la traduction devraient être des hiérarchies de sous-structures avec des interrelations

31 Hutchins&Somers (1992)

32 Exemple tiré de Hutchins&Somers (1992)

explicites.

Une solution souvent utilisée est celle reposant sur la grammaire des cas de Fillmore, dont nous avons parlé dans le premier chapitre. Selon la grammaire des cas, une phrase est constituée d'un verbe et de cas profonds, aussi appelés rôles sémantiques. Ces cas définissent par exemple l'agent, l'objet, l'instrument, etc. Une représentation utilisant les cas profonds permet de faire abstraction de la réalisation syntaxique de ces cas selon les langues.

2.4.4 Représentation du contenu subjectif

En plus du contenu propositionnel, les éléments de sens de type pragmatique devraient être préservés dans une traduction.³³ Il s'agit de tous les éléments exprimant l'attitude de l'auteur, le but de l'énoncé, les expressions de politesse, etc. Le contenu pragmatique est transmis par le style et les choix lexicaux, mais aussi par la structure argumentative ou les éléments rhétoriques. Si l'on veut préserver cette dimension pragmatique lors d'une traduction, il faut pouvoir représenter ces aspects dans l'interlangue.

Une interlangue destinée à traiter tous ces aspects est TAMERLAN, développée dans le cadre du projet DIONYSUS. Nous allons décrire cet interlangue plus en détail dans la section 2.5.3.

Nous allons présenter dans le chapitre suivant plusieurs approches employées dans différents systèmes par interlangue. Celles-ci illustrent des solutions apportées aux difficultés décrites précédemment.

2.5 Exemples d'approches interlangue

2.5.1 Rosetta – grammaire de Montague

Le projet Rosetta est un cas un peu particulier. Ce système, développé en 1980 aux Philips Research Laboratories à Eindhoven (Netherlands) utilise des représentations interlangue basées sur le principe de la grammaire de Montague, qui relie directement la syntaxe et la sémantique³⁴.

33 Defrise (1993)

34 Landsbergen (1989)

Les grammaires de Montague suivent le principe de compositionnalité, c'est à dire que le sens d'une expression est considéré comme étant une fonction du sens de ses constituants. Puisque la syntaxe définit les constituants, la sémantique est étroitement liée à la syntaxe. Une grammaire de Montague définit un ensemble d'expressions de base ainsi qu'un ensemble de règles syntaxiques.

En partant d'un arbre syntagmatique, on crée un arbre de dérivation sémantique qui présente la même géométrie que l'arbre de dérivation syntaxique, mais est étiqueté avec des 'basic meanings' et des 'meaning operation'. Dans le projet Rosetta, ces arbres sémantiques servent de représentation interlingue.

Le concept Rosetta est similaire à tous les autres systèmes à base linguistique: analyse morphologique, structures syntaxiques de surface, analyse sémantique en termes de relations structurelles profondes (deep structure relations) (valency, tense, voice, predicate-argument structures, etc). La seule différence essentielle est que là où d'autres systèmes mettent en parallèle des structures ou des objets non-égaux (par le transfert lexical et structurel), Rosetta met en parallèle les historiques dérivationnels des structures et objets.

Le projet Rosetta repose sur deux principes :

- grammaire unique: la même grammaire est utilisée pour la génération et pour l'analyse. Les règles syntaxiques sont réversibles.
- isomorphisme: deux phrases sont considérées comme étant des traductions l'une de l'autre si elles ont des arbres de dérivation sémantiques identiques. Les grammaires de deux langues doivent donc être développées en parallèle, afin que pour chaque expression de base dans une langue il y ait une expression de base correspondante dans l'autre langue, avec le même sens. De même pour les règles syntaxiques.

2.5.2 Interlangues restrictives

La quasi-infinité de concepts et de constructions possibles en langue naturelle pouvant difficilement être modélisée dans une interlangue, la plupart des interlangues sont restreintes à un domaine ou sous-langage précis. Cela présente l'avantage, en plus de restreindre le nombre de concepts et de constructions à représenter, qu'on aura moins de problèmes de polysémie.

Des exemples de projets de TA par interlangue développés pour des domaines limités sont

NESPOLE³⁵, que nous décrivons dans la section suivante, et MedSLT, dont nous parlerons dans le chapitre 3. Ces deux systèmes sont destinés à la traduction automatique de la parole, dans le contexte de dialogues. NESPOLE se caractérise par une approche basée sur les *domain actions*.

2.5.2.1 NESPOLE - *domain actions*

Une approche particulière de l'interlangue est celle reposant sur les *domain actions*. Ce schéma d'encodage a été développé pour la traduction automatique de la parole dans le contexte de dialogues orientés tâche (task-oriented dialogue). Un exemple sont les travaux du consortium C-STAR II³⁶. Cette approche repose sur l'hypothèse qu'il n'y a qu'un nombre limité d'actions ou d'évènements possibles dans un domaine donné. Dans le domaine d'application de C-STAR, les planifications de voyages, réservations et paiements d'hôtels, de moyens de transport etc., les actions se limitent à demander le prix d'une chambre, donner un numéro de carte de crédit, etc. Pour pouvoir représenter tous les énoncés possibles dans le contexte d'un tel dialogue, on définit un ensemble de *domain actions*. Tout énoncé pourra être classé comme une instance d'une de ces *domain actions*.

Les *domain actions* permettent de représenter des action ou évènements, plutôt que les moyens lexicaux et syntaxiques utilisés pour les décrire. Cette représentation fait abstraction des constructions du type prédicat-argument. Les *domain actions* sont construites de façon compositionnelle à partir d'un inventaire de *speech acts* et d'un inventaire de concepts, selon des

spécifications définissant les combinaisons permises. Un *speech act* peut être une requête, une interrogation, etc. Les concepts sont propres au domaine. Des paires attribut-valeur (qui ne font pas partie de la *domain action*) permettent d'ajouter des précisions (date, heure, nombre, etc).

NESPOLE

Le projet européen NESPOLE est une collaboration entre plusieurs groupes de recherche, IRST, Trento, Italie, ISL (Allemagne) et CLIPS (France), un groupe de recherche américain, ISL à Carnegie Mellon University, Pittsburgh, PA.

Le système s'intègre dans un environnement multimodal d'interaction (parole, vidéo et gestuel) en plus du visionnement de contenu web, du type e-commerce. NESPOLE propose un système de vidéo-conférence implémentant la traduction automatique de la parole. Deux interlocuteurs peuvent donc parler dans leur langue maternelle, et entendent les traductions synthétisées par le système. Des possibilités de communication multimodales sont également prévues: les deux parties peuvent envoyer et annoter des images ou diagrammes sur une "whiteboard application" partagée. Ce système est destiné à des utilisateurs non-spécialisés.

35 Lavie *et al.* (2001)

36 Levin *et al.* (1998)

Les interlangues basées sur les *domain actions*, par leur détachement des constructions syntaxiques, permettent de traiter élégamment les cas de divergences entre différentes langues, les divergences prédicatives par exemple. Celles-ci apparaissent souvent lors de l'expression de modalité ou d'éventualité, ainsi que dans certaines formes de discours indirect. Si l'on prend par exemple la phrase 'I just arrived - Je viens d'arriver', on voit que la notion de passé proche est exprimée en anglais par l'adverbe 'just' et en français par le verbe 'viens' combiné avec l'infinitif 'arriver'. Dans une interlangue basée sur les *domain actions*, plutôt que de représenter la structure prédicat-argument, on aura le concept 'arriver' associé au concept 'passé proche'. On obtient ainsi une interlangue indépendante des langues.³⁷

Un exemple concret d'application des *domain actions* est le projet NESPOLE! (NEgotiating through SPoken Language in E-commerce), destiné à la traduction automatique de la parole dans le contexte de l'e-commerce. L'interaction entre le client et le fournisseur implique la communication online entre personnes ne parlant pas la même langue. Le système est destiné à traiter des domaines vastes mais limités, et cela pour toutes les combinaisons possibles entre quatre langues, à savoir l'anglais, le français, l'allemand et l'italien. Le premier domaine développé est le tourisme et les informations de voyage. Le système utilise une interlangue relativement superficielle et orientée tâche. L'approche traductionnelle de NESPOLE repose sur les précédents travaux effectués dans le contexte du consortium C-STAR II.

Exemple de représentation interlingue NESPOLE³⁸

Source : I would like to make a hotel reservation for the fourth through the seventh of July

Nespole Interlingua : c:give-information+disposition+reservation+accommodation
 (disposition=(who=i, desire),
 reservation-spec=(reservation, identifiability=no),
 accommodation-spec=hotel,
 object-time=(start-time=(md=4), end-time=(md=7, month=7, incl-excl=inclusive)))}

La représentation interlingue est utilisée pour générer une paraphrase en langue source de l'énoncé, qui sera affichée à l'utilisateur afin qu'il puisse voir si son énoncé a été analysé correctement.

2.5.3 DIONYSUS – interlangue à base de connaissances

Le projet de traduction automatique DIONYSUS, développé à Carnegie Mellon, est un système interlangue à base de connaissances, qui a pour particularité d'avoir été développé

³⁷ Levin *et al.* (2002)

³⁸ exemple tiré de Levin *et al.* (2002)

pour la traduction de textes, et non de phrases isolées, entre l'anglais et le japonais³⁹.

DIONYSUS emploie l'interlangue TAMERLAN, résultant d'une modification du formalisme KBMT89⁴⁰.

Defrise départage les connaissances en trois niveaux: le contenu, les intentions de l'énonciateur et le contexte d'énonciation. Ces trois types de connaissances doivent d'une part être détectées et explicitées, d'autre part reproduites lors de la génération du texte cible. Sous contenu, Defrise entend le sens propositionnel de la phrase, les attitudes de l'énonciateur vis-à-vis de l'énoncé et les actes de langage. Les intentions représentent le but de l'énoncé. Finalement, le contexte décrit la situation d'énonciation.

Dans l'interlangue TAMERLAN le sens d'un texte est donc représenté par un triplet (G,S,T) :

- La composante G représente les buts et plans du locuteur, qui ne sont généralement pas explicités dans le texte. C'est ce que Defrise appelle l'agenda du locuteur. Ces buts sont modélisés dans l'ontologie qui est à la base du système. Ce modèle du monde définit les catégories sémantiques qui permettent de décrire les concepts et processus du domaine auquel appartient le texte à traduire.
- S, la situation d'énonciation, permet de représenter les éléments pragmatiques. Cette composante est représentée par deux paramètres, les *Indices Déictiques*, permettant de localiser l'énoncé dans le temps et l'espace et les *Facteurs Pragmatiques* modélisant les aspects stylistiques.
- La composante T est représentée par un quadruplet: *clauses, relations, attitudes et intentions de l'énonciateur*. La *clause* est l'ensemble des représentations du sens propositionnel, basées sur la grammaire des cas de Fillmore. Les *relations* servent à décrire l'ensemble des relations rhétoriques entre éléments de sens. Les *attitudes*, c'est-à-dire l'attitude du locuteur par rapport à son énoncé, reflète la structure argumentative du texte. Celle-ci repose sur l'emploi d'adverbes dits scalaires, de connecteurs et d'adjectifs. Cela permet de préserver lors de la traduction la distinction entre les expressions ayant la même valeur sémantique mais différentes forces argumentatives,

39 Defrise (1993)

40 Goodman&Nirenburg (1991)

par exemple *presque et pas tout à fait*⁴¹. Les *attitudes* servent également à différencier les mots dénotationnellement équivalents mais pas synonymes, dont l'usage reflète l'attitude du locuteur par rapport à l'objet décrit. La composante *intentions de l'énonciateur* représente l'ensemble des actes de langage, c'est à dire l'intention de l'énoncé (question, requête, etc.).

2.5.4 UNL – interlangue maximaliste

A l'opposé des interlangues restrictives, UNL⁴² (Universal Networking Language) n'est pas restreint, ni à un domaine, ni au niveau du langage. La représentation abstraite qui doit servir de pivot interlangue est destinée à encoder le sens d'énoncés en langage naturel sous forme d'hypergraphes 'sémantiques'.

La représentation d'un énoncé en UNL consiste en un hypergraphe avec des nœuds portant des UW (*Universal Words*, appelés *Unités de Vocabulaire Virtuel* en français) qui dénotent un sens de mot (ou acceptions) ainsi que des attributs (pluralité, modalité, aspects). Les nœuds sont reliés par des arcs portant des relations sémantiques telles que 'agent', 'conjonction', 'bénéficiaire', 'co-agent', etc. Au total, les spécifications de l'UNL prévoient 46 relations.

Les *Universal Words* comportent deux parties: le *headword* et une liste de restrictions. Les *headword* sont des mots ou composés anglais. Les restrictions sont indiquées sous forme de paires attribut-valeur, où les attributs sont des relations sémantiques et les valeurs d'autres UW.

La syntaxe d'un UW est la suivante:

<English word or compound> (<list of restrictions>)
 par exemple: look for (icl>action, agt>human, obj>thing)

Les UW représentent tous les sens du *headword* qui correspondent aux restrictions. Il existe

UNL

L'UNL n'est pas spécifiquement destiné à être utilisé dans le contexte de traduction automatique, mais plutôt pour servir de langage pour le stockage d'informations multilingues, c'est à dire de documents multilingues 'parfaitement alignés'. Ce projet de communication multilingue, développé à l'University of United Nations basée à Tokyo, a pour but de faciliter l'accès aux informations et services disponibles sur le web en surmontant la barrière linguistique. L'idée est de créer un 'HTML du contenu linguistique'. Une des principales applications est l'UNL Encyclopedia, dans laquelle les connaissances sont stockées en format UNL. UNL est actuellement intégré dans une structure HTML.

41 Exemple tiré de Defrise
 42 Sérasset *et al.* (2000)

une base de données⁴³ contenant plus de 200'000 UW.

Les concepts des UW sont répartis entre plusieurs catégories: concepts nominaux, concepts verbaux, concepts adjectivaux et concepts adverbiaux.

1. concepts nominaux : le facteur clé dans la définition des concepts nominaux est le rôle qu'ils peuvent jouer en relation avec d'autres concepts. Par exemple si un concept peut initier quelque chose, ou être impacté lors d'un événement. Les concepts nominaux sont classifiés en différentes catégories, arrangées dans une hiérarchie.

2. concepts verbaux : les concepts verbaux sont définis par les relations qu'ils peuvent avoir avec d'autres concepts. Ils sont divisés en 3 catégories: 'be', 'do' et 'occur'

3. concepts adjectivaux : sont divisés en 2 catégories: prédicatifs et attributifs; les concepts attributifs sont divisés en deux catégories supplémentaires, ceux qui expriment une notion quantitative et ceux qui qualifient un concept.

4. concepts adverbiaux : sont aussi divisés en plusieurs catégories, selon les concepts qu'ils modifient.

La définition des concepts doit répondre à deux attentes. D'une part, les concepts doivent être utiles pour la restriction d'autres concepts. 'mourir' dans le sens 's'arrêter de vivre' est défini comme prenant 'chose vivante' en tant qu'objet dans les restrictions, ce qui permet de le différencier des autres concepts de 'mourir', tels que 's'arrêter de fonctionner'. D'autre part, les concepts doivent pouvoir hériter les propriétés des concepts qui les précèdent dans la hiérarchie.

En plus des concepts ci-dessus, on peut aussi avoir des *compound concepts*, qui sont des concepts complexes qui doivent être compris/considérés comme un tout. Un scope permet de mettre en relation un tel *compound concept* avec d'autres UV. On attribue au scope un scope-ID, que l'on peut utiliser par la suite pour citer ou se référer à ce scope

Dans un graphe UNL, les UW portent des attributs décrivant les caractéristiques subjectives introduites par le locuteur, tels que l'attitude, la situation temporelle par rapport au moment du récit, le concept mis en avant, la référence, etc. Les attributs sont répartis en huit groupes. Nous avons vu deux de ces groupes d'attributs, ceux servant à la description temporelle, dans

43 <http://www.unl.fi.upm.es:8099/unlweb/index.jsp>

la section 2.4.2 de ce chapitre.

Dans les hypergraphes, les UW sont reliés par des relations. Les spécifications UNL définissent 46 relations, dont voici quelques exemples :

agt (agent) : définit quelque chose qui initie une action
 and (conjonction) : définit une relation de conjonction entre des concepts
 aoj (thing with attribute) : définit quelque chose qui est dans un certain état ou qui a un attribut

L'exemple⁴⁴ suivant montre une représentation interlingue UNL :

```
Source : culture takes diverse forms across time and space.
Interlangue UNL : {unl}
obj(take(icl>adopt(agt>thing,obj>thing)):08.@entry, form(icl>attribute):0M.@pl)
agt(take(icl>adopt(agt>thing,obj>thing)):08.@entry, culture(icl>abstract thing):00)
aoj(across(aoj>thing,obj>thing):0S,      form(icl>attribute):0M.@pl)
aoj(diverse(aoj>thing):0E,      form(icl>attribute):0M.@pl)
obj(across(aoj>thing,obj>thing):0S,      :01)
and:01(space(icl>universe):1D.@entry,   time(icl>abstract thing):14)
{/unl}
```

L'approche de UNL se différencie des autres systèmes par interlangue par le fait qu'elle ne fait aucune restriction au niveau du domaine ou du langage.

2.6 Conclusion

Nous avons vu dans ce chapitre les difficultés que présente le développement d'une interlangue pour la traduction automatique. Les problèmes à résoudre sont souvent les mêmes que pour les systèmes de transfert. Les différents systèmes interlangue existants emploient différentes stratégies de représentation, choisies en fonction de l'application du système.

44 Exemple tiré de : <http://www.undl.org/unesco/> (15.07.2008)

3 Application de la traduction automatique de la parole au domaine médical

3.1 Introduction

Dans certains pays ou certaines régions, il arrive souvent qu'un médecin et son patient n'aient aucune langue en commun. Cela arrive particulièrement souvent dans les pays ayant une importante population étrangère qui ne parle pas ou peu la langue locale, mais aussi dans les pays touristiques. On est alors obligé de recourir à un interprète pour que le médecin puisse poser les questions nécessaires au diagnostic. Ce qui pose plusieurs problèmes. D'une part, un interprète n'est pas toujours disponible, en particulier dans les situations d'urgence. D'autre part les frais engendrés par l'interprétation sont importants. Un système de traduction automatique de la parole pourrait résoudre ces problèmes.⁴⁵

La qualité la plus importante d'un système destiné au diagnostic médical est la fiabilité. Une erreur de traduction qui mènerait à une erreur de diagnostic peut avoir des conséquences graves pour le patient. Pour cette application, la précision est donc plus importante que la robustesse.

Le contexte du dialogue de diagnostic médecin-patient a plusieurs caractéristiques qui rendent possible le développement d'un système restrictif, et donc plus précis. D'une part, les énoncés sont relativement stéréotypés, le lexique est restreint, on a donc affaire à une sorte de langage contrôlé. D'autre part, dans le cas d'un système unidirectionnel où seuls les énoncés du médecin sont traduits (le patient répondant de façon non-verbale à des questions de type oui/non) on a un groupe restreint d'utilisateurs, qui sont susceptibles de se familiariser avec le système en l'utilisant à plusieurs reprises, et qui pourront donc adapter leurs énoncés à ce qui est couvert par le système. Les médecins appelés à utiliser régulièrement un tel système deviennent en quelque sorte naturellement des utilisateurs experts.

Un système bidirectionnel, qui traduirait aussi bien les questions du médecin que les réponses du patient sera beaucoup plus complexe, les patients n'ayant pas la possibilité de se familiariser avec le système et les énoncés couverts. C'est le cas par exemple du système NESPOLE! décrit précédemment, destiné à fournir des informations sur des destinations touristiques, où les utilisateurs ont peu de chances d'employer le système à répétition.

45 Bouillon *et al.* (2007)

Nous allons présenter dans la section suivante plusieurs systèmes de traduction automatique de la parole pour le domaine médical. Ensuite, nous présenterons plus en détail MedSLT, le système développé à l'Université de Genève, que nous utiliserons comme exemple dans la suite de ce travail.

3.2 Exemples de systèmes

3.2.1 Converser

Converser est un système commercial développé par Spoken Translation, Inc. à Berkeley⁴⁶. Il est destiné à aider des patients hispanophones à communiquer avec des médecins ou du personnel médical anglophone. Les buts recherchés sont la précision et une large couverture. Pour atteindre ces buts, le système inclut des fonctionnalités interactives. Ainsi l'utilisateur peut intervenir dans les deux processus qui sont le plus souvent à l'origine des erreurs, c'est-à-dire la reconnaissance de la parole et la traduction.

Après la reconnaissance de la parole, qui repose sur des méthodes statistiques, l'utilisateur peut corriger l'énoncé reconnu au moyen de commandes vocales, de la même façon que dans les systèmes de dictée. Lors de la traduction, l'utilisateur reçoit une paraphrase de la phrase source, résultat d'une re-traduction de la phrase cible vers la langue source, qui lui permet de juger de la qualité de la traduction. Dans les cas ambigus, l'utilisateur dispose d'une base de données de synonymes, définitions, etc. dans laquelle il peut choisir le sens des mots ambigus. Une fonctionnalité supplémentaire, les 'Translation Shortcuts™' sert à garder en mémoire les traductions des énoncés qui ont déjà été vérifiées et confirmées par l'utilisateur.

3.2.2 S-MINDS

S-MINDS⁴⁷ est un système de traduction automatique de la parole bidirectionnel développé par Sehda, Inc. L'accent est mis sur la possibilité de développer et de mettre en production rapidement des systèmes de traduction pour des tâches ou des langues pour lesquelles on n'a que peu ou pas de données à disposition. Le module de traduction de S-MINDS utilise un parseur sémantique pour extraire les mots et les expressions importantes de l'énoncé source résultant de la reconnaissance. Un moteur d'interprétation met ensuite ces éléments en

46 Dillinger *et al.* (2006)

47 Ehsani *et al.* (2006)

relation avec un ensemble de concepts spécifiques au domaine. On obtient finalement une paraphrase de l'énoncé source. Le système dispose d'une interface utilisateur vocale qui permet à l'utilisateur d'interagir avec le système tout en ayant les mains libres.

3.2.3 IBM MASTOR

IBM MASTOR⁴⁸ est un système conçu pour être utilisé sur des ordinateurs portables ou des PDA (Personal Digital Assistant, ordinateur de poche). La faible capacité de mémoire et de processeur de ce type d'ordinateur impose des contraintes conséquentes sur les algorithmes et programmes utilisables. Les programmes et ressources doivent non seulement être compacts, mais aussi peu gourmands en processeur, afin que l'application puisse tourner en temps réel sur un ordinateur de poche.

Ce système traite deux paires de langues, anglais-mandarin et anglais-arabe. Le manque de données et de ressources linguistiques disponibles pour ces langues rend le développement de la reconnaissance de la parole plus complexe.

La reconnaissance de la parole et le processus de traduction reposent sur des méthodes statistiques.

L'exigence principale formulée vis-à-vis des systèmes de traduction automatique de la parole pour le domaine médical est la précision, indispensable dans un domaine où la moindre erreur peut avoir des conséquences graves pour un patient. Selon les systèmes, d'autres exigences s'y ajoutent, telles que la facilité d'utilisation, la possibilité d'utilisation sur des ordinateurs de poche etc.

Dans la section suivante, nous allons décrire le système de traduction automatique de la parole pour le domaine médical développé à l'Université de Genève, MedSLT.

48 Gao *et al.* (2006)

3.3 MedSLT

3.3.1 Introduction

MedSLT est un système multilingue de traduction automatique de la parole en domaine limité pour le domaine médical, en développement à l'Université de Genève depuis 2003. Ce système est conçu pour traiter les dialogues de diagnostic entre médecin et patient, plus particulièrement dans le contexte du diagnostic d'urgence. MedSLT permet au médecin de poser des questions au patient dans sa propre langue, questions que le système va traduire dans la langue du patient. Le système supporte actuellement six langues: japonais, anglais, français, espagnol, arabe et catalan. Les traductions sont possibles entre n'importe quelle paire de ces langues. La traduction est unidirectionnelle, du médecin vers le patient. La plupart des questions sont donc du type oui-non, auxquelles le patient peut répondre de façon non-verbale. Une version bidirectionnelle, qui pourra également traduire les réponses verbales du patient est en cours de développement⁴⁹.

La thématique des dialogues que le système peut traiter est limitée à des sous-domaines médicaux, définis par des symptômes. Actuellement, MedSLT supporte trois sous domaines de diagnostic médical: les maux de tête, les douleurs abdominales et les douleurs thoracales.

Dans ce chapitre, nous consacrons une première section à l'architecture et la conception du système. Nous parlerons ensuite des outils utilisés par MedSLT, à savoir Nuance et Regulus. Finalement, nous reviendrons plus en détail sur chacune des étapes, de la reconnaissance de la parole jusqu'à la génération d'un énoncé cible.

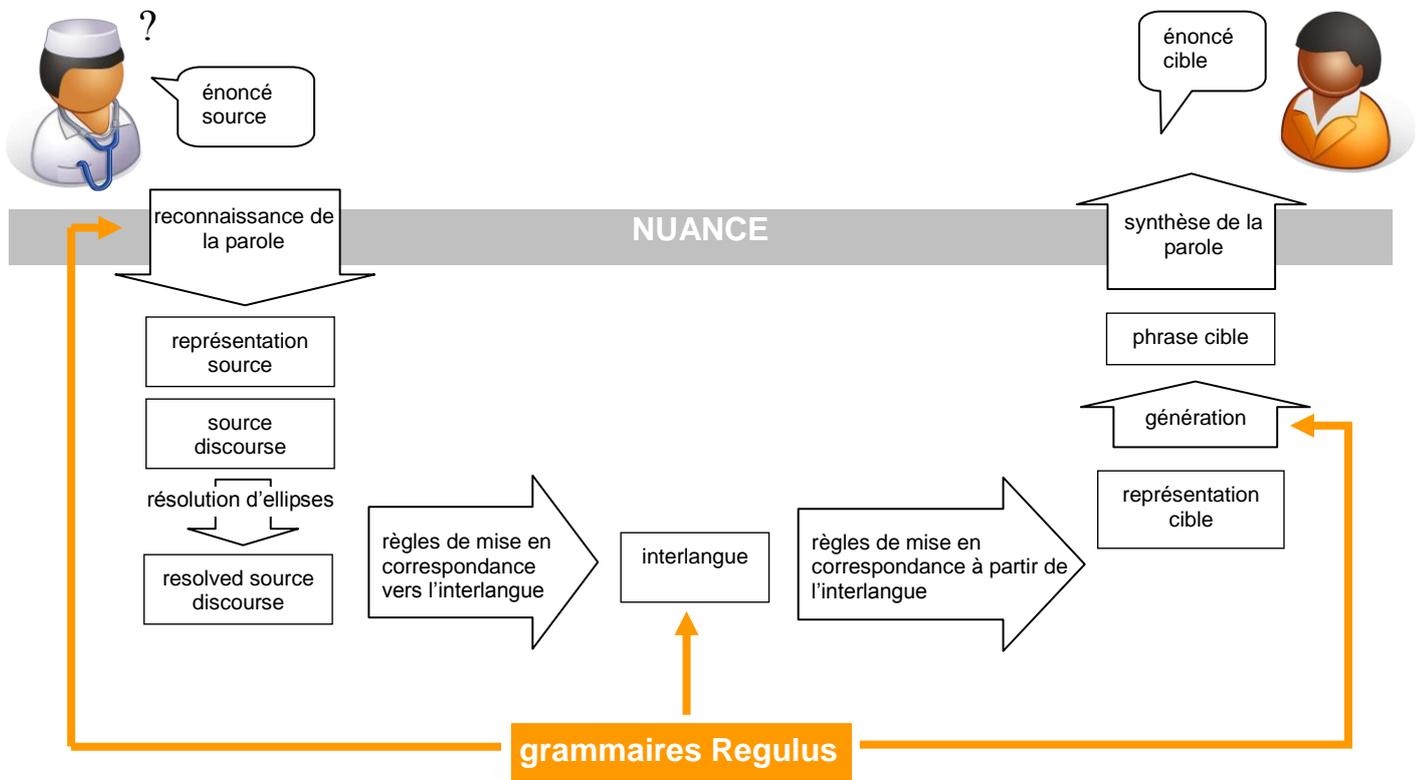
3.3.2 Architecture du système

Le choix de l'architecture de MedSLT a été influencé par plusieurs exigences. La première étant une exigence de précision, pour assurer la fiabilité du système, essentielle pour le domaine médical. Le système suit donc une approche du type langage contrôlé, avec une architecture basée sur des grammaires. Le problème usuel de ce type de système, le manque de robustesse, est compensé par l'ajout d'une fonction d'aide qui guide l'utilisateur vers les énoncés couverts par le système. Nous décrirons cette fonctionnalité plus en détail par la suite.

49 Bouillon *et al.* (2007)

Le choix de l'architecture a également été influencé par l'intention de développer un système dans lequel il serait raisonnablement simple et rapide d'implémenter une nouvelle langue. Une approche par interlangue a donc été adoptée.

Le schéma suivant illustre les composantes et processus de MedSLT :



Un module de reconnaissance de la parole linguistiquement motivé transforme les sons de l'énoncé source en chaînes de caractères. Les chaînes de caractères obtenues sont ensuite analysées pour obtenir une représentation source. Après la résolution d'ellipses, la représentation source est transposée en interlangue par application des règles de mise en correspondance. La représentation interlingue est ensuite transposée en une représentation cible, qui permet de générer la phrase cible en utilisant une grammaire de langue cible. Finalement, le module de synthèse de la parole va transformer l'énoncé traduit en son. Nous reviendrons sur ces étapes plus en détail dans la section 4.4.

Une des caractéristiques particulières de MedSLT est le fait que toutes les grammaires utilisées, que ce soit pour la reconnaissance, l'analyse ou la génération, sont compilées à partir de grammaires d'unification générales, motivées linguistiquement, en utilisant la plateforme

open-source Regulus 2⁵⁰.

Dans la section suivante, nous allons décrire plus en détail les outils utilisés dans MedSLT.

3.3.3 Outils

3.3.3.1 *Reconnaissance et synthèse de la parole: la plateforme Nuance*

Pour la reconnaissance et la synthèse vocale, MedSLT utilise la plateforme commerciale Nuance, qui emploie des modèles linguistiques basés sur des grammaires.

Nuance utilise des grammaires décrites dans le formalisme CFG (Context free grammar – grammaire indépendante de contexte) avec paramètres, correspondant au Nuance Grammar Specification Language (GSL)⁵¹. Ces grammaires sont compilées par le système Regulus 2 à partir d'une grammaire d'unification. Nuance compile ensuite les grammaires GSL pour en faire un module de reconnaissance.

3.3.3.2 *Développement et compilation des ressources linguistiques: Regulus 2*

MedSLT utilisant un reconnaiseur et un système de traduction linguistiques, il a besoin d'une grammaire par sous-tâche, c'est à dire pour la reconnaissance, l'analyse et la génération. Ces grammaires ne sont pas du même type, puisque la reconnaissance utilise une grammaire CFG, tandis que la génération repose sur des grammaires avec structures de traits. Afin de ne pas devoir écrire plusieurs grammaires, MedSLT utilise Regulus 2 pour créer les grammaires nécessaires à partir d'une grammaire d'unification.

Dans les sections suivantes, nous allons brièvement introduire les différents types de grammaires, puis décrire le processus de transformation des grammaires avec Regulus.

3.3.3.2.1 **Grammaires CFG (grammaires indépendantes de contexte) et grammaires d'unification**

Les grammaires CFG reposent sur la structure des constituants. Elles comportent un ensemble de règles qui décrivent comment les éléments de la langue peuvent être assemblés et un lexique contenant les mots de la langue. On aura par exemple une règle qui décrit un syntagme nominal (SN) comme étant composé d'un déterminant (Det) et d'un nom (Nom),

SN → Det Nom

⁵⁰ Rayner *et al.* (2006)

⁵¹ Bouillon, Rayner (2003)

ainsi que des entrées dans le lexique, définissant les mots 'une' et 'voiture'. Ces mots sont qualifiés de symboles terminaux :

Det → une
Nom → voiture

Tous les autres éléments, que ce soient des regroupements de mots tels que des syntagmes nominaux ou verbaux ou des généralisations tels que des catégories lexicales (déterminant, adjectif, etc.) sont qualifiées de symboles non-terminaux. Les règles ont toujours à gauche un seul non-terminal et à droite une liste ordonnée de terminaux et non-terminaux.

Une grammaire CFG peut servir à la fois à la génération de phrases conformes aux règles, mais aussi à l'analyse de phrases, pour leur assigner une structure⁵². En partant de la règle de l'exemple ci-dessus on peut obtenir le syntagme nominal 'une voiture', en remplaçant 'SN' par 'Det Nom', puis Det par 'une' et Nom par 'voiture'. Mais on peut aussi, en partant de la chaîne de caractères 'une voiture', utiliser la grammaire pour conclure qu'il s'agit d'un déterminant suivi d'un nom, regroupés pour former un syntagme nominal.

Les grammaires CFG permettent ainsi de décrire toutes les structures, en combinant une fonction descriptive avec une fonction représentative. Cependant, les grammaires CFG deviennent très vite très vastes. Si on veut par exemple que la grammaire couvre l'accord en genre, il faudra ajouter de nouvelles catégories atomiques pour couvrir les différents cas (Nom-f et Nom-m, Det-f et Det-m, SN-f et SN-m) ainsi que de nouvelles règles (SN-f → Det-f Nom-f et SN-m → Det-m Nom-m). Les règles sont ainsi démultipliées, les grammaires deviennent vite très complexes et difficiles à maintenir.

Une possibilité de réduire cette démultiplication est d'ajouter des paramètres pour décrire les catégories, plutôt que d'utiliser des catégories atomiques. On parle alors de grammaire CFG avec paramètres. Pour reprendre l'exemple précédent, plutôt que d'avoir une règle pour chacun des genres, on aura une seule règle : SN[nbre, genre] → Det[nbre, genre] Nom[nbre, genre]. Le nombre et l'ordre des paramètres décrivant une catégorie doivent être prédéfinis. Toute catégorie doit donc toujours avoir tous les paramètres dans le bon ordre et ceci dans toutes les règles. Ces grammaires sont donc plus compactes que les grammaires CFG de base, mais restent très lourdes.

Un formalisme encore plus compact est celui des grammaires d'unification. Celles-ci reposent

52 Jurafsky&Martin (2006)

sur le même principe de compositionnalité que les grammaires CFG, mais plutôt que d'utiliser des catégories atomiques ou des catégories avec paramètres, elles utilisent ce qu'on appelle des structures de traits, un ensemble d'attributs/valeurs. Un objet lexical féminin pluriel aura par exemple les traits 'genre=f' et 'nbre=pl'. Dans les règles, l'ordre des traits n'a aucune importance et il n'est pas nécessaire d'utiliser des descriptions complètes, on peut donc n'indiquer que les traits utiles pour la règle en question. Dans le formalisme d'unification, on manipule ces structures de traits, par comparaison ou par combinaison. L'unification sert d'une part à la vérification de la compatibilité des valeurs, par exemple le nombre d'un déterminant et celui d'un nom. La règle ci-dessous détermine que le trait 'nbre' de l'objet 'Det' doit être unifié avec le trait 'nbre' de l'objet 'Nom' :

$$\text{SN : [genre=G, nbre=N]} \rightarrow \text{Det : [genre=G, nbre=N], Nom: [genre=G, nbre=N]}$$

L'unification permet d'autre part la combinaison d'informations, par exemple de traits se trouvent de part et d'autre d'une règle. Dans la règle ci-dessus, le trait 'genre' du SN à gauche est unifié avec les traits 'genre' des éléments de droite.

On peut ainsi décrire les mêmes structures qu'avec une grammaire CFG, tout en ayant beaucoup moins de règles. Le formalisme des grammaires d'unification présente l'avantage d'être plus compact et donc plus simple à écrire que les grammaires CFG.

Le formalisme d'unification, par sa déclarativité, permet à une description d'être bidirectionnelle.⁵³ Les grammaires d'unification peuvent donc être utilisées aussi bien pour l'analyse que pour la génération.

3.3.3.2.2 Regulus

Regulus 2 est un ensemble de logiciels libres développé par un consortium open-source avec comme partenaires principaux la NASA et l'Université de Genève. Il est utilisé dans MedSLT pour la génération de la langue cible et la transformation des grammaires d'unification en grammaires GSL pour la reconnaissance.

Regulus permet de développer une seule grammaire d'unification pour chaque langue, à partir de laquelle peuvent ensuite être compilées les différentes grammaires requises pour la reconnaissance, la génération, etc. L'intérêt de l'architecture Regulus réside dans la simplification du processus d'écriture et de mise à jour d'un grand nombre de grammaires

53 Estival (1993)

similaires, tout en maintenant une cohérence entre ces grammaires. La cohérence entre les grammaires de reconnaissance et d'analyse est tout particulièrement utile, puisqu'elle assure que toute phrase reconnue sera également analysable. Le développement des ressources linguistiques pour un système de traduction automatique est ainsi simplifié, toutes les grammaires du système étant dérivées à partir de la même grammaire générale.

Les versions plus anciennes de MedSLT avaient une grammaire de base unique pour chaque langue. Pour simplifier encore le développement, dans les versions plus récentes, les grammaires de langues proches ont été fusionnées⁵⁴.

3.3.3.2.3 Spécialisation des grammaires

Les grammaires de base de MedSLT sont des grammaires générales, c'est-à-dire qu'elles ne sont pas spécifiques à un domaine. Il y a généralement une part non négligeable de recoupements entre des grammaires développées pour différents domaines, conséquence inévitable du fait que toutes les grammaires doivent modéliser les mêmes concepts linguistiques généraux des langues naturelles. Il est donc intéressant de concevoir une seule grammaire générale, applicable à de nombreux domaines⁵⁵. Une telle grammaire, de par sa large couverture, risque cependant d'être moins restrictive, et donc de ne pas satisfaire l'exigence de précision du domaine médical. Pour contrer ce problème, on utilise cette grammaire générale comme base pour créer des grammaires spécifiques à chaque sous-domaine. Ces grammaires sont spécialisées en utilisant des corpus d'entraînement propres aux sous-domaines, en appliquant des méthodes reposant sur l'EBL (Explanation based learning)⁵⁶.

La spécialisation se fait en plusieurs étapes. Le lexique de la grammaire est complété par l'ajout d'entrées lexicales spécifiques au domaine. Le corpus d'entraînement est converti en un « treebank » de représentations analysées. Ce treebank est utilisé pour produire une grammaire spécialisée « brute » en format REGULUS en utilisant l'algorithme EBL⁵⁷. La grammaire spécialisée « brute » est ensuite post-traitée pour obtenir la grammaire spécialisée finale.

Les grammaires d'unification spécialisées résultantes sont ensuite soumises à une nouvelle

54 Bouillon *et al.* (2007)

55 Rayner *et al.* (2004)

56 Rayner *et al.* (2002)

57 van Harmelen&Bundy (1988), Rayner (1988)

compilation pour les convertir en format CFG pour la reconnaissance. Cette compilation consiste essentiellement à étendre de façon énumérative la grammaire d'unification en instanciant chaque trait de chaque règle pour chacune de ses valeurs possibles⁵⁸.

Dans le chapitre suivant, nous allons décrire les différentes étapes de la traduction dans MedSLT, à savoir la reconnaissance et l'analyse, la transposition vers l'interlangue, la transposition vers la représentation cible, la génération de la phrase cible et finalement la synthèse vocale.

3.3.4 Etapes de la traduction dans MedSLT

L'exemple suivant représente les différentes étapes, de la phrase source à la phrase cible, avec les représentations produites à chaque niveau :

Source : Avez vous toujours mal sur le même côté ?
Source representation: [[frequency,toujours], [locative_prep,sur], [part,côté], [path_proc,avoir], [pronoun,vous], [spec,le_même], [symptom,mal], [tense,present], [utterance_type,sentence], [voice,active]]
Source discourse: [[utterance_type,sentence], [voice,active], [pronoun,vous], [tense,present], [path_proc,avoir], [frequency,toujours], [symptom,mal], [locative_prep,sur], [spec,le_même], [part,côté]]
Resolved source discourse: [[utterance_type,sentence], [voice,active], [pronoun,vous], [tense,present], [path_proc,avoir], [frequency,toujours], [symptom,mal], [locative_prep,sur], [spec,le_même], [part,côté]]
Interlingua: [[freq,always], [part,side], [prep,in_loc], [spec,the_same], [symptom,pain], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]
Target representation: [[freq,always], [part,side], [prep,in_loc], [secondary_symptom,pain], [spec,the_same], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]
Target: is the pain always in the same side

3.3.4.1 Reconnaissance et analyse

Les étapes de reconnaissance et d'analyse sont toutes les deux effectuées par la plateforme Nuance. Le reconnaiseur vocal, qui repose sur un modèle linguistique, permet de transformer les sons enregistrés en une chaîne de caractères. Il utilise la grammaire CFG compilée à partir de la grammaire de langue source pour trouver une séquence de mots correcte qui corresponde aux sons enregistrés. Si la reconnaissance produit un résultat, celui-ci sera donc forcément une phrase grammaticalement correcte et couverte par les grammaires du système. La qualité de la reconnaissance est améliorée en utilisant l'interlangue comme source de

⁵⁸ Rayner *et al.* (2001)

connaissances. Nous reviendrons sur ce sujet dans le chapitre 4.

Lors de la reconnaissance, une représentation sémantique est générée en utilisant les éléments sémantiques associés aux mots dans la grammaire de langue source. Les représentations sémantiques de MedSLT suivent le principe de la sémantique compositionnelle. La représentation des phrases est ainsi basée sur le fait que le sens de la phrase est une combinaison des sens des éléments qui la composent.

La représentation d'une phrase sera donc la concaténation de la sémantique des éléments qu'elle comporte. La représentation est plate, sauf lorsqu'il y a une relation temporelle/causale, où un niveau d'enchâssement est prévu. Les objets lexicaux sont représentés par des paires attribut-valeur. Les mots (valeur) sont liés à leur type ontologique (attribut).

exemple:

	<u>représentation sémantique</u>
mal	[symptom, mal]
toujours	[frequency,toujours]
le même	[spec,le_même]

L'ordre des éléments n'ayant aucune signification dans ces représentations, les éléments sont assemblés dans l'ordre alphabétique de l'attribut. La phrase de l'exemple ci-dessus 'Avez-vous toujours mal sur le même côté ?' aura donc comme représentation source :

[[frequency,toujours], [locative_prep,sur], [part,côté], [path_proc,avoir], [pronoun,vous], [spec,le_même], [symptom,mal], [tense,present], [utterance_type,sentence], [voice,active]]

3.3.4.1.1 Résolution d'ellipses

Dans la langue parlée, et plus particulièrement dans les dialogues, on est souvent confronté à ce qu'on appelle des phrases elliptiques. Il s'agit en quelque sorte de raccourcis, où le locuteur utilise des phrases incomplètes, qui se rapportent aux phrases précédentes du dialogue. Ce phénomène est illustré dans l'exemple suivant :

- (1) médecin : Avez-vous toujours mal sur le même côté ?
 - (2) médecin : le devant de la tête ?
 - (3) médecin : à l'arrière de la tête ?
- etc.

Ici le médecin utilise une formulation elliptique pour éviter de répéter la question 'avez-vous toujours mal [...]'

Les ellipses sont en quelque sorte des phrases auxquelles il manque certains constituants. La plupart des dialogues contenant des phrases elliptiques, un système tel que MedSLT doit pouvoir les traiter. Ceci est particulièrement vrai dans le contexte d'un dialogue unidirectionnel médecin-patient où le médecin est amené à poser principalement des questions du type oui-non qui peuvent être très répétitives.

Bouillon *et al.*⁵⁹ estiment que les ellipses ne peuvent pas être simplement traduites mot à mot. Une traduction mot-à-mot entraînerait des erreurs d'accord ou de cas. Si l'ellipse comporte un adjectif par exemple, il devrait être accordé correctement. Dans certains cas, l'aspect verbal de la phrase complète devra également être utilisé pour la phrase elliptique. Il est donc nécessaire de traduire les phrases elliptiques dans le contexte de la phrase complète qui les précède.

La résolution des ellipses dans MedSLT repose sur une manipulation de listes d'attributs⁶⁰. Elle est effectuée avant le passage vers l'interlangue, dans une version simplifiée de la représentation source, appelée 'source discourse'. Celle-ci est obtenue en ôtant certains éléments, le possessif par exemple.

Les phrases de notre exemple ci-dessus auront les représentations suivantes :

(1)
[[utterance_type,sentence], [voice,active], [pronoun,vous], [tense,present], [path_proc,avoir],
[frequency,toujours], [symptom,mal], [locative_prep,sur], [**spec,le_même**], [**part,côté**]]

(2)
[[**body_part,tête**], [**part,devant**], [utterance_type,phrase]

La phrase elliptique (2) est résolue en utilisant la phrase complète qui la précède (1). Dans la liste d'attributs-valeurs, on substitue la sous-représentation dans la phrase complète ('[spec, le_même], [part, côté]') par celle correspondante de la phrase elliptique ('[body_part, tête], [part,devant]').

Une fois l'ellipse résolue, donc la substitution effectuée, la phrase (2) aura la représentation suivante :

[[utterance_type,sentence], [voice,active], [pronoun,vous], [tense,present], [path_proc,avoir],
[frequency,toujours], [symptom,mal], [locative_prep,sur], [**body_part,tête**], [**part,devant**]]

Cette représentation est appelée 'résolved source discourse' et sera transposée par la suite dans la représentation interlangue.

59 Bouillon *et al.* (2007)

60 *ibid.*

Pour définir quels éléments sont substituables, on utilise des classes. Celles-ci regroupent des patrons des sous-représentations qui peuvent être substituées. On aura par exemple une classe 'place_of_pain_np' qui contiendra des patrons tels que '[[body_part, _], [part, _]]', '[[adj, _], [body_part, _]]', etc. Les patrons de cette classe permettront de décrire les sous-représentations définissant l'endroit où se trouve la douleur, tels que 'the front of the head', 'the left side', 'the side', etc.

Les phrases elliptiques ainsi résolues peuvent ensuite être traduites correctement, comme s'il s'agissait de phrases complètes.

3.3.4.1.2 Traitement des énoncés hors couverture

Puisque la reconnaissance de la parole dans MedSLT est linguistique, le système ne peut reconnaître et traduire que les énoncés couverts par ses ressources linguistiques. MedSLT est destiné à traiter des phrases d'un sous domaine très limité et les questions posées lors d'un examen médical sont raisonnablement stéréotypées, voire conditionnée par des guides⁶¹. L'occurrence d'énoncés hors couverture sera donc beaucoup plus faible que dans un contexte plus général. De plus, si le médecin utilise régulièrement le système, il saura quels sont les énoncés couverts. On ne peut toutefois pas exclure l'occurrence d'énoncés non couverts.

Pour résoudre le problème de ces énoncés, MedSLT dispose d'un module d'aide.⁶² Ce module prend comme input le résultat d'un reconnaiseur statistique plus robuste que le reconnaiseur linguistique, qui donne un meilleur résultat sur les phrases qui ne sont pas couvertes par la grammaire. Ce reconnaiseur reconnaît des paires de mots plutôt que des phrases complètes. Il produit ensuite une série de phrases exemples couvertes par le système, sélectionnées dans une liste en privilégiant les phrases ayant un maximum de N-grams en commun avec les phrases résultantes de la reconnaissance. Le médecin peut alors voir quelles phrases similaires à son énoncé sont couvertes.

L'exemple suivant illustre le fonctionnement du module d'aide. Prenons un dialogue de diagnostic dans le domaine des maux de gorge :

médecin : Where are your pains ?
patient : I have a sore throat.
médecin : For how long ?

La question du médecin 'for how long ?' est une phrase incomplète, énoncé qui n'est pas

61 pour les Etats-Unis, www.guidelines.gov, National Guideline Clearinghouse

62 Starlander *et al.* (2005)

couvert par la grammaire anglaise. Comme nous l'avons vu dans la section précédente, certaines phrases incomplètes, les phrases elliptiques, sont traitées par le système. Dans le cas présent, il s'agit d'un type d'ellipse qui n'est pas traité par le système.

La reconnaissance linguistique ne donne aucun résultat pour cette phrase, mais le reconnaiseur statistique identifie les mots 'for how long'. Le module d'aide propose alors au médecin les phrases suivantes, couvertes par la grammaire et similaires à son énoncé :

```
<help>
<helpitem>for how long have you had your sore throat</helpitem>
<helpitem>for how long have you had a sore throat</helpitem>
<helpitem>how long have you had a sore throat</helpitem>
<helpitem>how long have you had congestion</helpitem>
<helpitem>how long have you had a stomachache</helpitem>
<helpitem>how long have you had a pain in your throat</helpitem>
<helpitem>how long have you had a headache</helpitem>
<helpitem>how long has your throat hurt</helpitem>
<helpitem>how long does your sore throat last</helpitem>
</help>
```

Le médecin peut choisir parmi celles-ci la phrase qui lui convient, qui sera ensuite traduite :

```
<helpSelection>for how long have you had your sore throat</helpSelection>
```

Le module d'aide permet non seulement de rendre le système plus robuste, mais cela permet aussi au médecin d'apprendre rapidement quels types d'énoncés et quel lexique sont couverts par le système⁶³.

3.3.4.2 Traduction

La traduction repose sur une approche par interlangue. Le système MedSLT se différencie d'une bonne partie des autres systèmes de traduction automatique par le fait qu'il est plus focalisé sur le transfert correct du sens que de la forme, et que dans certains cas il choisira délibérément une paraphrase plutôt qu'une traduction littérale.

La traduction se fait en deux étapes: la transposition depuis la représentation de la langue source vers l'interlangue et la transposition depuis l'interlangue vers la représentation de la langue cible. Ces transpositions sont faites par des règles de mise en correspondance, que nous décrirons plus en détail dans la suite de ce chapitre.

En reprenant notre exemple, 'Avez-vous toujours mal sur le même côté ?', on aura donc :

63 Chatzichrisafis *et al.* (2005)

Les interlangues en TA – l'exemple de MedSLT

représentation source : [[frequency,toujours], [locative_prep,sur], [part,côté], [path_proc,avoir], [pronoun,vous], [spec,le_même], [symptom,mal], [tense,present], [utterance_type,sentence], [voice,active]]

règles de mise en correspondance vers
l'interlangue

interlangue : [[freq,always], [part,side], [prep,in_loc], [spec,the_same], [symptom,pain], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]

règles de mise en correspondance à partir
de l'interlangue

représentation cible : [[freq,always], [part,side], [prep,in_loc], [secondary_symptom,pain], [spec,the_same], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]

Le formalisme de représentation de l'interlangue est le même que celui utilisé pour les représentations sémantiques produites par les grammaires du système. Nous allons décrire celui-ci plus en détail dans les sections suivantes

3.3.4.3 Description de l'interlangue

3.3.4.3.1 Éléments de l'interlangue

La forme structurée de l'interlangue MedSLT est une liste de primitives sémantiques (*interlingua_constants*), qui sont des paires attribut/valeur comportant un concept sémantique (p.ex. 'body_part', 'symptom', 'action', etc) et une valeur qui est une expression en anglais qui correspond au concept (p.ex. 'eye', 'pain', 'read', etc).

Les primitives sémantiques (*interlingua_constants*) de l'interlangue de MedSLT sont déclarées dans un fichier *interlingua_declarations*, sous la forme *interlingua_constant* (*[<key>, <value>]*), par exemple:

```
interlingua_constant([symptom, pain])
```

Dans MedSLT, on s'intéresse au sens des mots dans un domaine particulier plutôt qu'au sens dans la langue générale. Cela présente l'avantage qu'il n'est pas nécessaire de prendre en compte tous les aspects du sens. On a donc un ensemble limité de sens possibles, ce qui simplifie la résolution d'ambiguïtés lexicales.

L'attribution des types sémantiques n'est toutefois pas toujours évidente, puisqu'un objet lexical ne sera pas forcément du même type dans tous les contextes. Un élément peut être une

cause dans un contexte, et un symptôme secondaire dans un autre. Prenons par exemple la lumière forte, souvent associée aux maux de tête.

La lumière forte ('bright light') peut être la cause de la douleur, comme dans l'exemple suivant :

'can bright light give you headaches' / 'vos maux de tête sont-ils causés par une lumière forte'

La représentation interlingue de ces phrases est la suivante :

```
[[clause,
  [cause,bright_light], [pronoun,you], [state,experience],
  [tense,present], [utterance_type,dcl], [voice,active]],
 [pronoun,you], [sc,when], [state,have_symptom],
 [symptom,headache], [tense,present], [utterance_type,ynq],
 [voice,active]]
```

L'objet lexical 'bright light' est donc représentée par [cause,bright_light].

Cependant, la lumière forte, ou plus particulièrement la sensibilité à la lumière forte, est également un symptôme secondaire des maux de tête, comme illustré dans l'exemple suivant :

'are you sensitive to bright light when you experience the pain' / 'êtes-vous sensible à la lumière quand vous avez mal'

La représentation interlingue de ces phrases est la suivante :

```
[[clause,
  [[pronoun,you], [state,have_symptom], [symptom,pain],
  [tense,present], [utterance_type,dcl], [voice,active]],
 [pronoun,you], [sc,when], [secondary_symptom,photophobia],
 [state,experience], [tense,present], [utterance_type,ynq],
 [voice,active]]
```

Dans ce cas, l'objet lexical 'bright light' est représentée par le symptôme secondaire qui lui est associé, la photophobie, [secondary_symptom,photophobia].

3.3.4.3.2 Ordre des concepts dans la représentation

Dans l'interlangue, les éléments sont dans l'ordre alphabétique des concepts. Prenons par exemple la représentation interlingue de la phrase suivante:

la douleur commence t elle l'après-midi

```
[[prep,in_time], [state,start], [symptom,pain], [tense,present],
 [time,afternoon], [utterance_type,ynq], [voice,active]]
```

3.3.4.3.3 Une représentation interlingue pour plusieurs phrases

Comme mentionné dans le chapitre précédent, MedSLT est focalisé sur le transfert du sens

Les interlangues en TA – l'exemple de MedSLT

plutôt que de la forme. Des phrases exprimant le même sens seront donc transposées dans la même représentation interlingue, même si elles recourent à des moyens syntaxiques différents pour exprimer le sens.

Les trois phrases suivantes :

la douleur débute t elle le soir
est-ce que la douleur commence le soir
vos douleurs commencent elles le soir

auront la même représentation interlingue:

```
[[prep,in_time], [state,start], [symptom,pain], [tense,present],  
[time,evening], [utterance_type,ynq], [voice,active]]
```

et donc la même traduction:

does the pain start in the evening

On peut aussi regrouper une expression verbale avec son équivalent nominalisé.

Les deux phrases suivantes:

la douleur **débute** t elle de manière graduelle
est-ce que **le début** est graduel

auront la même représentation interlingue:

```
[[manner,gradually], [state,start], [symptom,pain],  
[tense,present], [utterance_type,ynq], [voice,active]]
```

et donc la même traduction

does the pain start gradually

3.3.4.3.4 Mots composés, expressions figées

Une autre question à résoudre lors du développement d'une interlangue est le traitement des mots composés, ou plutôt la définition des critères selon lesquels des mots peuvent être regroupés en expressions figées (compounds). Dans l'exemple suivant, on voit par exemple que l'expression « difficult to localize » est regroupée, et considérée comme un adjectif.

'is the pain difficult to localize'
'la douleur est elle difficile à localiser'

```
[[adj,difficult_to_localize], [symptom,pain], [tense,present],  
[utterance_type,ynq], [verb,be], [voice,active]]
```

Ce regroupement rend la représentation beaucoup plus simple, puisque l'on n'a pas besoin de traiter l'adjectif, la préposition et l'infinitif séparément, mais que toute l'expression est considérée de la même façon que les autres adjectifs décrivant la douleur. Un tel

regroupement n'est toutefois envisageable que parce que l'on est dans un contexte très limité.

On a un regroupement du même genre pour un élément 'cause', « sudden temperature change »

'are the headaches worse when you experience sudden temperature changes'
'vos maux de tête sont-ils aggravés par des changements de température brusques'

```
[[clause,
[[cause,sudden_temperature_change], [pronoun,you],
[state,experience], [tense,present], [utterance_type,dcl],
[voice,active]]],
[event,become_worse], [sc,when], [symptom,headache],
[tense,present], [utterance_type,ynq], [voice,active]]
```

3.3.4.3.5 Représentation des déterminants

Les déterminants étant difficiles à reconnaître, ils ne sont pas représentés, ni dans les représentations source, ni dans l'interlangue. Ils sont ajoutés au moment de la génération de la phrase cible. Le générateur choisit les articles en se basant sur des corpus⁶⁴.

3.3.4.3.6 Représentation des relations temporelles ou causales

Pour exprimer les relations temporelles ou causales, une représentation enchâssée est nécessaire. Dans l'interlangue MedSLT, un seul niveau d'enchâssement pour les phrases subordonnées est utilisé. Les relations causales et temporelles sont toutes les deux représentées comme des structures enchâssées à relation temporelle.

Bien qu'à priori il y ait une différence de sens entre les relations temporelles et causales, dans le contexte d'utilisation de MedSLT, c'est à dire dans un contexte de diagnostic médical, les deux relations ont le même résultat. En effet, le patient ne peut que décrire les symptômes secondaires qu'il a pu observer en corrélation avec le symptôme au sujet duquel il vient consulter. Le patient ne dispose à priori pas des connaissances nécessaires pour déterminer s'il y a effectivement corrélation entre les différents symptômes, donc relation de cause à effet. C'est au médecin de déduire les relations exactes. Il est donc suffisant, pour le diagnostic, de parler en termes de relations temporelles.⁶⁵

Trois types de relations sont prévus dans le formalisme de l'interlangue: 'when', 'before' et 'after'.

64 Bouillon *et al.* (2006)

65 Rayner *et al.* (2005)

L'exemple suivant illustre la relation 'when'. Dans l'interlangue, cette relation est indiquée par l'élément 'sc-when' ('sc' étant l'abréviation de 'subclause'). L'exemple suivant montre la représentation interlingue de la phrase 'la douleur diminue-t-elle **quand** vous buvez du café',

```
[[clause,  
[[action,drink], [cause,coffee], [pronoun,you], [tense,present],  
[utterance_type,dcl], [voice,active]]],  
[event,become_better], [sc,when], [symptom,pain],  
[tense,present], [utterance_type,ynq], [voice,active]],
```

3.3.4.4 Règles de mise en correspondance

Les règles de mise en correspondance servent à la transposition de la représentation source en interlangue et de l'interlangue en représentation cible.

Dans MedSLT, ces règles sont développées à partir d'un corpus interlangue combinant toutes les expressions interlangues qui peuvent être créées en traduisant tous les corpus de langues source vers l'interlangue. L'exemple suivant montre une entrée de ce corpus :

```
interlingua_item('WH-QUESTION attacks last how-long usually PRESENT ACTIVE',  
[[duration,how_long], [freq,usually], [itemizer,attacks], [state,last], [tense,present], [utterance_type,whq],  
[voice,active]],  
[from(eng)-'how long do the attacks usually last',  
from(fre)-'combien de temps ces crises durent elles habituellement',  
to(cat)-'quant temps duren els atacs habitualment',  
to(eng)-'how long do the attacks usually last',  
to(fre)-'combien de temps ces crises durent-elles habituellement',  
to(jap)-'daitai dorekurai no aida tsuzuki masu ka',  
to(spa)-'cuánto tiempo duran normalmente los ataques']).
```

Cette méthode permet de développer des règles de mise en correspondance qui couvriront toutes les expressions interlangues possibles dans le système.

3.3.4.5 Retrotraduction

Avant d'être traduite vers la langue cible, la représentation interlingue est d'abord retraduite vers la langue source. Le résultat de cette retrotraduction est affiché à l'utilisateur, afin qu'il puisse confirmer que son énoncé a été compris correctement. Si l'utilisateur approuve la retrotraduction, l'interlangue sera transposée en langue cible.

3.3.4.6 Génération

La génération est le processus inverse de l'analyse. La phrase cible est générée à partir de la représentation cible en utilisant une grammaire de la langue cible.

représentation cible : [[freq,always], [part,side], [prep,in_loc], [secondary_symptom,pain],
[spec,the_same], [tense,present], [utterance_type,ynq], [verb,be],
[voice,active]]

phrase cible : is the pain always in the same side

Lors de la génération, la grammaire de génération ajoute les composantes qui ne sont pas représentées ni dans l'interlangue, ni dans la représentation cible, les déterminants par exemple⁶⁶.

3.3.4.7 Synthèse de la parole

La chaîne de caractères ainsi obtenue est finalement transformée en son par la plateforme NUANCE et transmise au patient.

3.4 Conclusion

Dans MedSLT, aussi bien la reconnaissance que l'analyse et la génération reposent sur des grammaires, compilées par Regulus à partir d'une grammaire unique, ce qui permet d'assurer la cohérence entre les différentes étapes de traduction. Nous verrons par la suite que la représentation interlingue repose également sur une grammaire.

Pour améliorer l'utilisabilité du système, le manque de robustesse de la reconnaissance basée sur les grammaires est compensé par un module d'aide reposant sur des méthodes de reconnaissance statistiques. La résolution d'ellipses permet de traiter une forme de discours très fréquente dans les dialogues de diagnostic médical. La traduction dans MedSLT repose sur une stratégie par interlangue, afin de pouvoir ajouter plus facilement de nouvelles langues au système. L'interlangue est également utilisée pour produire une retrotraduction de l'énoncé reconnu vers la langue source, ce qui permet au médecin de confirmer que sa question a bien été comprise, assurant ainsi la fiabilité du système.

Dans le chapitre suivant, nous allons décrire plus en détail l'interlangue utilisée par MedSLT.

66 Bouillon *et al.* (2006)

4 L'interlangue utilisée par MedSLT

4.1 Introduction

Nous avons vu dans le chapitre précédant la forme structurée de l'interlangue utilisée par MedSLT. Cette forme permet de représenter le contenu sémantique des énoncés, en satisfaisant les exigences de précision et de fiabilité nécessaires dans le domaine d'application, tout en permettant une portabilité vers d'autres sous-domaines médicaux.

L'interlangue MedSLT doit cependant répondre à deux conditions supplémentaires. D'une part, elle doit être lisible pour des humains, d'autre part elle doit être vérifiable. Dans les sections suivantes, nous allons parler de l'intérêt de ces caractéristiques dans le contexte d'un système multilingue, puis nous décrirons l'approche utilisée dans MedSLT pour satisfaire ces exigences.

Dans la suite de ce chapitre, nous allons nous intéresser de plus près à l'une de ces caractéristiques, la lisibilité. Nous présenterons plusieurs expériences destinées à évaluer cette qualité, en utilisant des phrases extraites du corpus interlangue de MedSLT.

4.2 Caractéristiques de l'interlangue

4.2.1 Une interlangue lisible

La lisibilité de l'interlangue, donc la facilité qu'ont des humains à la comprendre, a plusieurs intérêts. D'une part, une interlangue compréhensible permet le développement monolingue, d'autre part, elle permet une évaluation monolingue des traductions produites.

4.2.1.1 Développement monolingue

Lors du développement d'un système par interlangue, il sera plus simple de développer les règles de mise en correspondance vers et à partir de l'interlangue si la représentation interlingue est claire et explicite. Ces règles sont spécifiques à chacune des langues du système. Un développeur ne se chargera généralement que des règles pour une langue, donc pour passer de la représentation propre à cette langue vers l'interlangue et inversement. Pour pouvoir tester le fonctionnement de ces règles, une possibilité serait d'effectuer une traduction

depuis la langue source vers une langue cible, et d'évaluer cette traduction. Cela présente plusieurs désavantages. Premièrement, un tel test implique deux jeux de règles : les règles de mise en correspondance de la langue source (donc celles en développement) et un autre jeu de règles, spécifique à une autre langue, probablement développé par une autre personne. En cas d'erreur de traduction, il est donc difficile de déterminer dans quel jeu se trouvent les règles incorrectes. Le développement devient ainsi très complexe. Une deuxième difficulté réside dans le fait qu'il n'est pas toujours possible, selon les langues, de trouver des développeurs suffisamment compétents dans deux langues pour que ce genre de test soit fiable. L'intérêt de pouvoir développer les règles de mise en correspondance indépendamment d'une autre langue est donc évident.

Une interlangue claire et compréhensible permettrait de développer les règles de mise en correspondance en utilisant uniquement une langue source et l'interlangue, sans devoir faire recours à une langue cible. L'interlangue prend alors la place de la langue cible. La recherche d'erreurs serait simplifiée, et il ne serait plus nécessaire pour les développeurs de connaître plusieurs langues du système.

4.2.1.2 Évaluation monolingue

Lors de l'évaluation des traductions produites par un système de TA, on évalue généralement des paires langue source – langue cible. Cela devient délicat lorsqu'on a un système traitant des paires de langues pour lesquelles on a peu de chance de trouver des évaluateurs maîtrisant les deux langues, telles que le japonais et l'arabe par exemple⁶⁷. Il serait donc intéressant de pouvoir effectuer les évaluations de façon monolingue, en utilisant la représentation interlangue plutôt qu'une langue cible.

Si un locuteur de la langue source peut facilement comprendre l'interlangue, il pourra évaluer la qualité de la représentation interlingue qui servira de source à la génération de la langue cible (ou des langues cibles). On peut ainsi évaluer une moitié du processus de traduction, à savoir la traduction langue source-interlangue, sans qu'il soit nécessaire de connaître la ou les langues cibles. De plus, si l'on peut évaluer séparément la traduction langue source – interlangue et la traduction interlangue – langue cible, on pourra aussi beaucoup plus aisément localiser l'erreur dans le système si l'on a une traduction langue source – langue cible fautive.

⁶⁷ Bouillon *et al.* (2008)

4.2.2 Une interlangue vérifiable

Sous interlangue vérifiable, on entend qu'il y a moyen de valider les phrases interlingues, donc qu'il existe une définition de ce qu'est une expression interlingue bien formée. La possibilité de vérifier les expressions interlangue présente clairement un avantage pour le développement des règles de mise en correspondance, puisqu'elle permet de voir si les expressions sont bien complètes et cohérentes.

4.2.2.1 L'interlangue comme source de connaissances

Une interlangue qui définit un ensemble de représentations sémantiques correctes permet également d'encoder les contraintes du domaine indépendamment des langues. On peut ainsi définir des contraintes sémantiques, qui seront communes à toutes les langues, directement dans l'interlangue plutôt que dans les grammaires de chacune des langues. L'interlangue peut alors servir de source de connaissances pour l'analyse ou la reconnaissance. Nous verrons plus loin comment cet aspect de l'interlangue est exploité dans MedSLT.

Dans la section suivante, nous allons décrire par quel moyen ces caractéristiques sont réalisées dans MedSLT.

4.2.3 Conception de l'interlangue MedSLT

4.2.3.1 Une grammaire *Regulus*

La forme structurée de l'interlangue MedSLT, que nous avons décrite dans le chapitre précédent (section 3.3.4.4), composée de paires attributs-valeurs, ne remplit pas les conditions de lisibilité, du moins pour les personnes pour qui ce type de représentation n'est pas familier. Il s'est donc avéré nécessaire de créer une forme de surface de l'interlangue, plus facilement lisible par des humains⁶⁸. Cette forme de surface est générée à partir de la forme structurée en utilisant une grammaire. L'utilisation d'une grammaire pour la définition de la forme de surface (et donc parallèlement de la forme structurée) permet également de remplir la deuxième condition, que l'interlangue soit vérifiable.

La grammaire définissant l'interlangue est une grammaire *Regulus*, semblable aux grammaires des différentes langues du système, mais associée à une syntaxe artificielle et simplifiée. Les non-terminaux de cette grammaire sont des concepts sémantiques, tels que

68 Rayner *et al.* (2008)

'symptom', 'location', etc. L'exemple suivant montre une règle de cette grammaire :

```
location : [sem=concat(BodyPart, Part, Side)] -->
body_part : [sem=BodyPart],
?part: [sem=Part],
?side: [sem=Side].
```

La règle ci-dessus décrit une réalisation possible de la représentation 'location', à savoir la concaténation d'un élément 'body part' avec des éléments optionnels 'part' et 'side'. Ces règles permettent d'encoder des contraintes du domaine, donc de définir les combinaisons sémantiques possibles.

Cette grammaire permet de transposer la forme structurée de l'interlangue en forme de surface, en remplaçant et en réarrangeant les éléments. Le procédé est similaire à celui utilisé pour générer les phrases cible, à la différence que la phrase obtenue est en quelque sorte une version simplifiée d'une phrase cible, dans laquelle certains éléments sont conservés tels quels plutôt que d'être réalisés avec les moyens syntaxiques et lexicaux d'une langue cible. Dans la section suivante, nous décrirons plus en détail cette forme de l'interlangue, destinée à être plus aisément lisible pour des humains.

4.2.3.2 Forme de surface de l'interlangue

La forme de surface actuelle utilise un lexique anglais et suit une syntaxe relativement semblable à l'anglais. Elle n'utilise pas les paires attribut-valeur, mais seulement les valeurs : [state,stay_still] sera remplacé par 'stay-still'. La fonction des éléments, plutôt que d'être indiquée par l'attribut, est transmise par la position de l'élément dans la phrase. L'ordre des éléments correspond à l'ordre qui serait usuel en anglais.

Certains éléments, qui ne peuvent être remplacés par des objets lexicaux, sont conservés sans modification majeure. C'est notamment le cas de l'élément [utterance_type,ynq], qui indique qu'il s'agit d'une question oui/non. Il est remplacé par 'YN-QUESTION' positionné en début de phrase. De même, les verbes ne sont pas conjugués, mais représentés par leur forme de base et d'une indication de temps, par exemple 'stay-still PRESENT ACTIVE'.

L'exemple suivant montre une phrase, sa représentation interlingue et la forme de surface correspondante.

phrase source :	'is the pain better when you stay still'
forme structurée :	[[clause, [[pronoun,you], [state,stay_still], [tense,present], [utterance_type,dcl], [voice,active]]],

Les interlangues en TA – l'exemple de MedSLT

[event,become_better], [sc,when], [symptom,pain], [tense,present],
[utterance_type,ynq], [voice,active]],

forme de surface : 'YN-QUESTION pain become-better sc-when [you stay-still PRESENT ACTIVE]
PRESENT ACTIVE'

L'ordre des éléments est destiné à faciliter la lecture et donc la compréhension, surtout pour les phrases longues. Les clauses enchâssées devraient également être plus simples à identifier. Celles-ci sont séparées du reste de la phrase par des parenthèses carrées, tout comme dans la forme structurée, et précédées de l'élément indiquant le type de subordonnée, 'sc-when' dans l'exemple ci-dessus.

Il serait envisageable de créer d'autres jeux de règles pour obtenir des formes de surface adaptées à des locuteurs d'autres langues que l'anglais. Il faudrait alors remplacer les éléments par des objets lexicaux d'une autre langue et les agencer dans un autre ordre.

Comme nous l'avons mentionné précédemment, le fait d'utiliser une grammaire pour la transformation de la forme structurée en forme de surface rend l'interlangue vérifiable. En effet, la grammaire garantit que toute expression en forme de surface sera bien-formée. De même, toute expression en forme structurée à partir de laquelle la grammaire permet de générer une forme de surface sera elle aussi bien formée. La grammaire définit donc toutes les expressions bien-formées. Cet aspect est exploité dans MedSLT à plusieurs fins, que nous décrirons dans les sections suivantes.

4.2.3.3 L'interlangue comme source de connaissances

La grammaire Regulus définissant l'interlangue est utilisée comme composante du modèle du langage, pour encoder les contraintes du domaine indépendantes des langues. Lors de la reconnaissance de la parole, le système produit plusieurs hypothèses de reconnaissance, toutes couvertes par la grammaire de la langue source. On utilise alors l'interlangue comme source de connaissances pour choisir la phrase la plus probable qui produise une représentation interlingue valide. L'interlangue permet ainsi d'améliorer la reconnaissance. Cette solution présente l'avantage que les contraintes sémantiques spécifiques au domaine sont contenues dans l'interlangue et peuvent être appliquées à la reconnaissance pour toutes les langues.

4.2.3.4 Interlingua checker

Un outil destiné à vérifier la validité d'une expression interlangue (interlingua checker) a été

créée en appliquant le générateur-compileur de Regulus à la grammaire de l'interlangue⁶⁹. En effet, l'interlangue étant une langue spécifiée par une grammaire au même titre que toutes les langues naturelles traitées par MedSLT, elle peut être manipulée par les mêmes outils que ceux utilisés pour les langues naturelles.

L'interlingua checker permet de vérifier si une expression interlingue est bien conforme à la grammaire et donc valide. Les expressions en forme structurée sont soumises au checker, qui va essayer de les transposer en forme de surface. S'il est possible de générer une forme de surface, l'expression est considérée comme étant valide. A l'opposé, si l'expression ne correspond à aucune des constructions encodées par la grammaire, elle sera rejetée. Pour faciliter le développement, le checker va proposer à l'utilisateur des propositions de substitution, d'ajout ou de suppression d'éléments qui rendraient l'expression valide. L'exemple suivant illustre le fonctionnement du checker :

Source : avez vous mal plus de deux heures ?

Source representation : [[comparative,plus_de], [number,2], [path_proc,avoir], [pronoun,vous], [symptom,mal], [tense,present], [timeunit,heure], [utterance_type,sentence], [voice,active]]

Source discourse: [[utterance_type,sentence], [voice,active], [pronoun,vous], [tense,present], [path_proc,avoir], [symptom,mal], [comparative,plus_de], [number,2], [timeunit,heure]]

Resolved source discourse: [[utterance_type,sentence], [voice,active], [pronoun,vous], [tense,present], [path_proc,avoir], [symptom,mal], [comparative,plus_de], [number,2], [timeunit,heure]]

Interlingua: [[pronoun,you], [spec,[more_than,2]], [state,have_symptom], [symptom,pain], [tense,present], [timeunit,hour], [utterance_type,ynq], [voice,active]]

Interlingua surface: WARNING: INTERLINGUA REPRESENTATION FAILED STRUCTURE CHECK. APPLYING MODIFICATION (null-->[prep,ago]) GIVES "YN-QUESTION you have pain more-than two hour ago PRESENT ACTIVE"

Pour cette phrase, il n'est pas possible de générer la forme de surface, la forme structurée étant incomplète. Il manque une indication de préposition. Cette erreur est probablement due au fait qu'il n'y a pas de préposition dans la phrase française. La préposition doit toutefois être représentée, puisque elle sera nécessaire dans d'autres langues, telles que l'anglais par exemple ('do you experience the pain for more than two hours'). Le checker propose dans ce cas d'ajouter l'élément [prep,ago], qui n'est pas la bonne. Il faudrait ici ajouter [prep,duration], qui indique que l'on parle de la durée de la douleur.

4.2.4 Conclusion

Pour rendre l'interlangue plus aisément lisible, et donc pour faciliter le développement et

69 Rayner *et al.* (2008)

l'évaluation, MedSLT en génère une forme de surface. Celle-ci est définie par une grammaire Regulus. Cette grammaire permet à la fois de transformer la forme structurée en forme de surface et de vérifier sa bonne formation. La vérifiabilité de l'interlangue est exploitée pour améliorer la reconnaissance, ainsi que dans un outil permettant de vérifier la bonne formation de la forme structurée, résultat de la transposition depuis les représentations source.

Dans la suite de ce travail, nous allons évaluer la compréhensibilité de l'interlangue de MedSLT, cet aspect étant particulièrement utile pour le développement et l'évaluation du système, comme nous l'avons vu dans ce chapitre.

4.3 Evaluation de l'interlangue MedSLT

4.3.1 Expériences

Pour évaluer la compréhensibilité de l'interlangue utilisée dans MedSLT, nous avons prévu de faire plusieurs expériences. Dans celles-ci, nous allons soumettre à des sujets humains des phrases en interlangue (aussi bien sous la forme structurée de l'interlangue que sous la forme de surface). Les phrases utilisées sont issues d'un corpus extrait de la version unidirectionnelle de MedSLT. Il s'agit donc de phrases interrogatives dont l'énonciateur est le médecin. Les phrases appartiennent au sous-domaine des maux de tête.

Les langues utilisées pour ces expériences sont l'anglais et le français

Nous avons conçu trois expériences, que nous introduisons brièvement ci-dessous :

- Première expérience : l'interlangue peut-il servir de source pour reconstituer une phrase en langue naturelle ?

Dans cette expérience, nous voulons voir si les expressions interlangue comportent toutes les informations sémantiques d'une phrase sous une forme suffit suffisamment explicite pour qu'un humain puisse les utiliser comme base pour produire une phrase en français ou en anglais. Dans une première partie, nous utilisons la forme structurée de l'interlangue ; dans une deuxième partie, la forme de surface. Cela nous permettra de comparer les deux formes.

- Deuxième expérience : l'interlangue permet-elle de repérer les erreurs de traduction langue source- interlangue?

Comme mentionné précédemment, l'une des exigences formulées à l'égard de l'interlangue est de faciliter le développement des règles de mise en correspondance, ou plus particulièrement de simplifier l'identification des erreurs dans celles-ci. Afin d'évaluer cet aspect, nous avons créé un corpus avec des phrases sources ainsi que leurs représentations interlingues. Nous avons délibérément introduit des erreurs de traduction dans certaines paires, et avons demandé aux évaluateurs de les trouver.

- Troisième expérience : évaluation source-interlangue, interlangue-cible et source-cible.

Afin d'évaluer l'utilité de l'interlangue comme étape intermédiaire pour évaluer les traductions, ainsi que pour trouver la source d'erreur en cas de mauvaise traduction, nous avons créé une expérience avec des triplets phrase source – représentation interlingue – phrase cible, dans lesquels nous avons introduit des erreurs de traduction. Nous avons demandé aux évaluateurs d'indiquer s'il y avait des erreurs et à quel niveau elles étaient (source-interlangue, interlangue-cible ou source-cible).

Il aurait été souhaitable de faire une quatrième expérience, avec des paires de phrases français–anglais, comportant des erreurs similaires à celles de la troisième expérience. Cela aurait permis de comparer le nombre d'erreurs repérées dans des paires de phrases source-cible avec le nombre d'erreurs repérées avec l'interlangue comme information supplémentaire.

Une telle expérience a cependant déjà été réalisée avec des corpus extraits de MedSLT, pour la paire de langues français-japonais⁷⁰. Cette expérience consistait à comparer les résultats obtenus avec une évaluation partagée français-interlangue / interlangue-japonais avec ceux obtenus lors d'une évaluation directe français-japonais. Sur 496 phrases évaluées, seules 9 ont donné des résultats d'évaluation jugés significativement différents. Dans 8 de ces cas, il s'agissait de problèmes au niveau de la granularité de l'interlangue. En conclusion, l'évaluation partagée a donné presque les mêmes résultats que l'évaluation directe.

Dans la suite de ce chapitre, nous allons tout d'abord décrire le déroulement pratique des expériences (4.3.1.1). Ensuite, nous aurons une section par expérience (4.3.2, 4.3.3 et 4.3.4), comportant la conception de l'expérience ainsi que les résultats obtenus. Finalement nous

⁷⁰ Bouillon *et al.* (2008)

récapitulerons les résultats de ces expériences.

4.3.1.1 Déroulement des expériences

Les trois expériences ont été soumises en format électronique. Les expériences telles que soumises au sujet se trouvent dans l'annexe 1. Les sujets avaient un temps illimité à disposition, il leur a cependant été demandé d'indiquer le temps nécessaire pour chaque expérience.

Les trois expériences ont été soumises à dix personnes maîtrisant l'anglais et le français, de formations différentes (informaticiens, linguistes, chimistes, traducteurs). Les explications et instructions fournies avec les expériences étaient les mêmes pour tous les sujets (également dans l'annexe 1). Une partie des sujets est impliquée dans le projet MedSLT, mais aucune n'a de connaissances particulières de l'interlangue.

4.3.2 Première expérience

L'objectif de cette expérience est de déterminer si l'interlangue de MedSLT comporte toutes les informations sémantiques sous une forme suffisamment explicite pour que des humains puissent l'utiliser pour recomposer une phrase en langue naturelle. Nous avons donc demandé aux sujets de traduire des phrases depuis l'interlangue au choix vers l'anglais ou vers le français.

Cette expérience est divisée en deux parties, la première portant sur la forme structurée de l'interlangue, la deuxième sur la forme de surface. Dans chacune des parties, les sujets ont dû traduire 13 expressions interlangue. Les phrases ont été choisies afin de représenter tous les types de phrases possibles du système, plus ou moins complexes. On a donc des phrases utilisant tous les temps et aspects verbaux possibles, différentes clauses enchâssées, etc. Les phrases ont été sélectionnées de sorte à ce que les deux parties de l'expérience comportent un assortiment semblable, de difficulté équivalente.

Nous avons demandé aux sujets d'indiquer le temps nécessaire à chacune des parties de cette expérience.

4.3.2.1 Description de l'expérience

4.3.2.1.1 Première partie

Comme nous l'avons vu dans la section 5.1.2, la forme structurée de l'interlangue de MedSLT est composée de paires attribut-valeur, concaténées dans l'ordre alphabétique des attributs. L'ordre des concepts ne correspond donc à aucun ordre usuel en langue naturelle. La seule source d'information qu'un lecteur humain aura sur la fonction des différents éléments vient des concepts eux-mêmes. Dans la première partie de cette expérience, nous voulons voir si les concepts sont suffisamment explicites pour en quelque sorte compenser le 'désordre', et donc permettre à un humain d'identifier le sens transmis par les éléments.

Les concepts nous semblent plus ou moins intuitifs, en tout cas pour des personnes comprenant l'anglais. Certains concepts tels que « symptom » ou « body-part » sont raisonnablement simples à comprendre, d'autant plus que les valeurs qui leur sont associées sont également intuitives, par exemple « pain » ou « neck ». D'autres concepts, tels que « utterance_type », qui peut avoir comme valeur « ynq » ou « dcl » sont moins évidents, bien que standards dans un contexte linguistique. Ces concepts donnent une idée de la fonction de la valeur qui leur est associée.

Pour l'évaluation, nous avons choisi de ne pas fournir d'explications par rapport à la signification des différents concepts, le but de l'expérience étant de voir si ces concepts sont intuitifs. Nous avons toutefois fait deux exceptions, pour les concepts suivants, qui nous ont semblé plus compliqués :

- utterance_type

Puisque ce concept prend des abréviations comme valeur, nous avons jugé utile de fournir leur signification aux évaluateurs.

valeur du concept	définition
utterance_type	
dcl	phrase déclarative
whq	WH-QUESTION : phrase interrogative qui en anglais commencerait par un adverbe en 'wh', par exemple 'what', 'where', 'when', etc.
ynq	YN-QUESTION : phrase interrogative demandant une réponse de type oui/non (non verbale)

- aspect

Le concept aspect peut prendre deux valeurs dans MedSLT: *perfect* et *continuous*. Ces notions ne nous semblant pas forcément intuitives pour les personnes n'ayant pas suivi d'enseignement grammatical anglais, il nous a semblé nécessaire de les expliciter.

Le *perfect* est un aspect qui reflète le point de vue de l'énonciateur, dénotant une action achevée au moment de l'énonciation ou un état ou action ayant commencé dans le passé et se poursuivant dans le présent. Le *present perfect* est formé de l'auxiliaire *have/has* et d'un participe passé. Par exemple:

'how long have you had the headaches'
'depuis combien de temps avez-vous vos maux de tête'

Le *present perfect* s'oppose au *simple present*, qui décrit un état de fait, par exemple une action qui se répète ou qui est habituelle:

'how long do you have the headaches'
'pendant combien de temps avez-vous vos maux de tête'

La notion de *continuous* décrit une action en progrès. Cet aspect peut être exprimé de différentes manières :

'is the headache often caused by reading'
'do you often have the headaches when you read'
'avez-vous souvent vos maux de tête quand vous lisez'

Dans MedSLT, l'aspect *continuous* apparaît principalement dans les clauses enchâssées, typiquement introduites par 'when'.

4.3.2.1.2 Deuxième partie

Comme décrit dans la section 4.2.3.2, MedSLT peut générer une forme de surface de l'interlangue, destinée à la rendre plus compréhensible pour des humains. Nous avons donc ajouté une deuxième partie à cette expérience, afin de comparer la forme structurée et la forme de surface. L'idée est de voir si la forme de surface améliore effectivement la compréhension. Les phrases de cette partie ont été choisies de sorte à comporter les mêmes difficultés que celles de la première partie de l'expérience.

4.3.2.2 Résultats

Nous avons comparé les phrases produites par les sujets avec les phrases du corpus, extraites de MedSLT, c'est-à-dire les phrases anglaises ou françaises que le système peut analyser ou générer qui correspondent aux phrases interlangue proposées. L'évaluation des résultats de

cette expérience a une composante subjective, puisqu'il s'agit d'une évaluation de qualité de traduction. Dans beaucoup de cas, le corpus contient plusieurs phrases en langue naturelle correspondant à la même expression interlangue, employant des moyens syntaxiques ou lexicaux différents. Nous avons donc considéré comme correctes les phrases exprimant le même sens, même si elles n'étaient pas exactement les mêmes que les phrases du corpus.

Les différences observées sont souvent au niveau des articles (défini/indéfini) ainsi qu'au niveau des pronoms possessifs (vos maux de tête/les maux de tête). L'interlangue ne comportant aucune indication permettant de choisir une de ces variantes, nous avons considéré ces variations comme correctes. En voici quelques exemples :

phrase du corpus : 'vos maux de tête sont-ils sur l'un des côtés'
phrase obtenue : Les maux de tête sont-ils situés d'un seul côté ?

phrase du corpus : 'vos maux de tête ont-ils commencé progressivement'
phrase obtenue : est-ce que les maux de tête ont commencé peu à peu ?

phrase du corpus : 'avez-vous vos maux de tête une fois par mois'
phrase obtenue : Avez-vous mal à la tête une fois par mois ?

Les erreurs grammaticales ou orthographiques ont été ignorées, du moment qu'elles n'apportaient pas d'ambiguïté au niveau du sens de la phrase.

Afin de ne pas alourdir cette discussion des résultats, nous n'avons indiqué pour chaque phrase que les solutions incorrectes. Les résultats complets de la première expérience se trouvent dans l'Annexe 3.

Nous indiquons pour chaque phrase en interlangue/forme de surface (en gras) les phrases correspondantes du corpus. Les phrases précédées par from() sont celles que le système peut analyser, et qui sont donc transposées dans l'interlangue en question. Les phrases précédées par to() sont celles qui peuvent être générées à partir de l'interlangue en question. Nous indiquons ensuite les phrases obtenues (en italique).

Pour chaque phrase, nous commentons brièvement les erreurs dans les phrases obtenues. Celles-ci seront ensuite récapitulées dans la section suivante.

4.3.2.2.1 Première partie

1. **[[adj,diffuse], [symptom,pain], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]**

from(fr)-'la douleur est elle diffuse'

to(eng)-'is the pain diffuse'

to(fr)-'la douleur est-elle diffuse'

Bien que 'diffuse' soit la valeur de *adj*, deux personnes l'ont utilisé comme verbe :

Est-ce que la douleur se propage?

Does the pain diffuse?

Ici nous avons une traduction incorrecte de 'diffuse' par 'générale' :

Avez-vous une douleur générale?

2. **[[part,side], [prep,in_loc], [spec,both], [symptom,headache], [tense,present],[utterance_type,ynq], [verb,be], [voice,active]]**

to(eng)-'are the headaches in both sides'

to(fre)-'vos maux de tête sont-ils des deux côtés'

Dans cette phrase, une seule personne a omis le [spec,both] :

Est-ce que le mal de tête est de côté ?

3. **[[adj,left], [body_part,head], [freq,always], [part,side], [prep,in_loc], [pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]**,

to(eng)-'do you always have the headaches in the left side of the head'

to(fre)-'avez-vous toujours vos maux de tête sur le côté gauche'

L'interlangue comporte l'information [body_part,head] ainsi que le symptôme [symptom, headache]. Il y a surabondance d'informations, puisqu'un 'headache' ne peut pas concerner d'autre parties du corps que la tête. Une partie des sujets ont omis cette redondance dans leurs phrases, soit en utilisant le symptôme plus général et en gardant une localisation précise :

Est-ce que vous avez toujours mal au côté gauche de la tête ?

soit en gardant le symptôme précis et en omettant l'indication de la partie du corps concernée :

Is your headache always on the left side?

Does the headache always occur on the left side?

Est-ce que vous avez toujours mal à la tête du côté gauche ?

Ce qui est d'ailleurs également le cas pour la phrase française du corpus. Nous avons donc considéré ces phrases comme correctes

4. **[[loc,where], [symptom,pain], [tense,present], [utterance_type,whq], [verb,be], [voice,active]]**

from(eng)-'where does it hurt', from(eng)-'where is the pain',

from(eng)-'where is your pain',

from(fre)-'où avez vous mal

Tous les sujets ont traduit cette phrase correctement.

5. **[[adj,right], [body_part,arm], [state,radiate], [symptom,pain], [tense,present],**

[utterance_type,ynq], [voice,active]]

from(fre)-'la douleur irradie t elle le bras droit'
 to(eng)-'does the pain radiate to the right arm'
 to(fre)-'la douleur irradie-t-elle le bras droit'

L'élément [state, radiate] a causé des difficultés à la plupart des sujets. Le concept 'state' prête à confusion, puisqu'il peut être interprété comme décrivant l'état d'un objet, mais aussi comme un prédicat, donc l'action associée à un objet. On peut toutefois déduire la fonction de 'state' à partir des autres valeurs qui lui sont associées, à savoir 'experience', 'have_symptom', 'start' etc.

Un sujet n'a pas utilisé 'radiate' comme une action de la douleur, mais comme un état décrivant le bras droit :

Votre bras droit irradié, fait-il mal?

'radiate' a aussi été omis complètement dans une phrase :

Avez-vous mal au bras droit?

Un autre problème rencontré est le manque d'indication de direction pour le concept 'radiate', on ne peut donc pas savoir dans quel direction l'irradiation a lieu. Plusieurs sujets ont d'ailleurs commenté sur ce manque d'information. On a donc plusieurs phrases où la radiation ne se fait pas dans la bonne direction, ou où la direction de la radiation n'est pas précisée :

est-ce que la douleur irradie du bras droit?
Does the pain radiate from the right arm?
Is the pain radiating?
La douleur irradie-t-elle du bras gauche ?

Dans la dernière phrase nous avons en plus une inversion de la droite et de la gauche, ce qui est certainement une faute d'inattention.

6. [[event,become_worse], [itemizer,attacks], [symptom,pain], [tense,present], [utterance_type,ynq], [voice,active]]

from(fre)-'l intensité de ces crises augmente t elle'
 to(eng)-'is the pain worse'
 to(fre)-'ces crises empirent-elles'

L'interlangue comporte ici deux éléments, [itemizer, attacks] et [symptom, pain], qui sont tous les deux susceptibles de servir de sujet à la phrase, à savoir [event, become worse]. Dans les phrases françaises du corpus, on parle de crises ; dans la phrase anglaise, de douleur. Aucune n'utilise donc à la fois la notion de crise et la notion de douleur. Les phrases françaises n'expriment ainsi pas exactement le même sens que les phrases anglaises.

Les phrases du corpus n'étant pas tout à fait satisfaisantes, l'évaluation des phrases obtenues n'est pas évidente. Nous avons obtenu deux phrases comportant à la fois la notion de 'pain' et la notion de 'attacks' :

Do the pain attacks become worse?
Are the pain attacks getting worse?

L'élément [itemizer, attacks] est problématique, dans le sens que son utilité n'est pas intuitive.

Itemizer sert à indiquer que l'on veut parler de la forme quantifiable d'un objet. Si l'on a [itemizer,attacks], [symptom,pain] on parle d'attaques de douleur, qui peuvent être quantifiées (une attaque, deux attaques, etc.), par opposition à la douleur en tant que telle à laquelle on ne peut pas associer de quantité. Il s'agit ici de la même distinction que celle entre 'du riz' et 'un grain de riz'. Il s'agit à priori du même objet, sauf que lorsqu'on parle de 'riz' en général, on ne s'intéresse pas aux grains qui le composent.

Deux phrases obtenues sont totalement différentes des phrases du corpus :

Lors d'une crise, la douleur s'aggrave t'elle?
Avez-vous plus mal lors des crises?

7. [[manner,gradually], [state,start], [symptom,headache], [tense,past], [utterance_type,ynq], [voice,active]]

to(eng)-'did the headaches start gradually'
to(fre)-'vos maux de tête ont-ils commencé progressivement'

Cette phrase est la première de l'expérience qui soit au passé et non au présent. Plusieurs sujets ont écrit des phrases au présent :

vos maux de tête commencent graduellement?
Vos maux de tête commencent-ils progressivement?

Encore une fois, le concept 'state' n'est pas interprété correctement. Dans la phrase suivante [state,start] est interprété comme un état :

Depuis que vos maux de têtes ont commencé, est ce que ça c'est empiré?

8. [[history,family_history], [pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

to(eng)-'do you have a family history of the headaches',
to(fre)-'avez-vous des antécédents familiaux de maux de tête',

Tous les sujets ont traduit cette phrase correctement.

9. **[[frequency,time], [prep,duration], [prep,frequency], [pronoun,you], [spec,1],[state,have_symptom], [symptom,headache], [tense,present], [timeunit,month], [utterance_type,ynq], [voice,active]]**

from(eng)-'does the headache happen once a month',
 to(eng)-'do you have the headaches once a month',
 to(fre)-'avez-vous vos maux de tête une fois par mois'

Nous n'avons obtenu que très peu de bonnes réponses pour cette phrase, qui comporte plusieurs éléments peu clairs.

La plupart des sujets ont été perturbés par les deux éléments 'prep': [prep,duration] et [prep,frequency]. Un sujet propose deux phrases :

vos maux de tête durent-t-il un mois / avez-vous mal à la tête une fois par mois

D'autres choisissent [prep,duration] :

Avez-vous eu ces symptomes pendant 1 mois?

Combien de mois ces maux de tête durent-ils?

ou [prep,frequency] :

Combien de fois par mois est-ce que vous avez vos maux de tête?

Do you experience headache more than once per mounth in average?

Do you have headaches every month?

Est-ce que vous avez plus d'une fois par mois ces symptômes de maux de têtes?

How often do you have headaches each month?

L'élément [spec,1] est aussi un problème puisqu'on ne sait pas à quoi il se rapporte. Il n'y a donc pas moyen de savoir si on parle d'une durée de un mois ou d'une fréquence d'une fois par mois.

Un autre élément problématique est [frequency,time] : il n'a pas d'utilité apparente.

10. **[[aspect,perfect], [prep,duration], [pronoun,you], [spec,[more_than,2]], [state,have_symptom], [symptom,headache], [tense,present], [timeunit,week], [utterance_type,ynq], [voice,active]]**

from(eng)-'have you had them for more than two weeks'
 from(fre)-'avez vous vos maux de tête depuis plus de deux semaines'
 to(eng)-'have you had the headaches for more than two weeks'
 to(fre)-'avez-vous vos maux de tête depuis plus de deux semaines'

Cette phrase est la première de l'expérience à utiliser l'aspect perfect, dont nous supposons qu'il causerait quelque difficulté aux personnes n'ayant pas de connaissances grammaticales approfondies de l'anglais (ce qui est le cas de la plupart de nos sujets).

La phrase décrit ici une action qui a commencé il y a plus de deux semaines, et qui est encore en cours. Les phrases suivantes sont donc fausses, puisqu'elles impliquent que le patient n'a

plus mal à la tête. Sur ces cinq phrases, quatre sont au passé, une seule est au présent :

Est-ce que vous avez eu mal à la tête pendant plus de 2 semaines ?
Avez-vous eu ces symptômes pendant plus de 2 semaines?
Avez-vous eu ces maux de tête pendant plus de deux semaines?
avez-vous mal à la tête pendant plus de deux semaines?
Have you experienced headache for more than two weeks?

Il y a également une incertitude au niveau de l'élément [spec,[more_than_2], puisqu'il pourrait aussi s'appliquer au symptôme, plutôt qu'à l'unité temporelle 'week':

Est-ce que vous avez toujours plus de deux fois par semaine ces symptômes de maux de tête?
Avez-vous mal à la tête plus de 2 fois par semaine ?
Did you have more than two headaches this week?

11. **[[action,cough],
[clause,
[[pronoun,you], [state,have_symptom], [symptom,headache], [tense,present],
[utterance_type,dcl], [voice,active]]],
[pronoun,you], [sc,when], [tense,present], [utterance_type,ynq],[voice,active]]**
to(eng)-'do you cough when you have the headaches'
to(fre)-'toussez-vous quand vous avez vos maux de tête'

Cette phrase est la première de l'expérience avec une clause subordonnée. Le fait que la subordonnée soit la partie introduite par [clause,...] ne semble pas très intuitif. Nous avons obtenu quatre phrases où les clauses sont inversées :

Lorsque vous toussiez, avez-vous des symptômes de maux de têtes ?
avez-vous mal à la tête quand vous toussiez
avez- vous vos maux de tête quand vous toussiez
Do you have a headache when you cough?

Nous avons également obtenu une phrase sans clause :

When do you experience headache?

12. **[[clause,
[[pronoun,you], [secondary_symptom,trauma], [state,experience], [tense,present],
[utterance_type,dcl], [voice,active]]],
[pronoun,you], [sc,after], [state,have_symptom], [symptom,pain],
[tense,past],[utterance_type,ynq], [voice,active]]**
from(fre)-'avez vous eu mal après un traumatisme'
to(eng)-'did you experience the pain after you experienced trauma'
to(fre)-'avez-vous eu mal après un traumatisme'

Le fait que 'trauma' soit défini comme un 'secondary symptom' prête à confusion, puisqu'un symptôme secondaire n'apparaît généralement pas avant le symptôme principal, et ne peut pas être la cause du symptôme principal. Comme pour la phrase précédente, nous avons certains cas d'inversion des deux clauses :

*Avez-vous ressenti un traumatisme après avoir eu mal ?
Vous êtes traumatisé après avoir eu ces douleurs?
Avez-vous d'autres symptômes suite aux maux de tête?*

En plus de la difficulté liée à la clause enchâssée, nous avons dans cette phrase des indications de temps potentiellement conflictuelles. En effet, la phrase principale est au passé, la phrase subordonnée est au présent.

Dans les deux phrases suivantes, la problématique du temps de la phrase subordonnée est élégamment contournée en supprimant le verbe. La phrase principale est toutefois au présent, et non au passé :

*avez-vous mal suite à un traumatisme?
avez-vous la douleur après une traumatisation?*

Une autre phrase est juste, mis à part le temps (présent) :

Lorsque vous avez un trauma, ressentez vous des symptômes de douleurs après ?

La phrase suivante est presque juste, mis à part le 'any' qui a été ajouté :

Have you experienced any pain after the trauma?

13. **[[clause, [[pronoun,you], [state,have_symptom], [symptom,pain], [tense,present], [utterance_type,dcl], [voice,active]]], [pronoun,you], [sc,before], [secondary_symptom,muscle_aches], [state,experience], [tense,present], [utterance_type,ynq], [voice,active]]**

to(eng)-'do you experience muscle aches before you experience the pain'
to(fre)-'la douleur est-elle précédée par des douleurs musculaires'

Nous avons ici une clause du type [sc,before], donc un évènement qui en précède un autre. L'ordre a parfois été inversé :

Do you have pain before you experience muscle aches?

La notion temporelle 'avant' a été remplacée par 'en même temps' :

*avez-vous des douleurs musculaires quand vous avez mal à la tête
Avez-vous des maux de tête quand vous avez des douleurs musculaires?
Resentez-vous des douleurs musculaires avec ces douleurs?*

Deux phrases sont correctes au niveau de la relation temporelle (douleurs musculaires avant le mal de tête), mais il y a une légère distorsion du sens :

*Après avoir eu des courbatures, ressentez vous des symptômes de douleurs ?
Do you experience pain after muscles aches ?*

Et une phrase presque correcte mis à part l'ajout de 'déjà' :

avez-vous déjà eu des douleurs musculaires avant la douleur

Nous récapitulerons les résultats de la première partie de cette expérience dans la section

5.3.3, ensemble avec les résultats de la deuxième partie, afin de comparer les deux formes de l'interlangue.

4.3.2.2.2 Deuxième partie

14. 'YN-QUESTION continuous pain be PRESENT ACTIVE'

from(fre)-'la douleur est elle continue'

to(eng)-'is the pain continuous'

to(fre)-'la douleur est-elle continue'

Nous avons un cas d'ommission d'adjectif :

avez-vous mal

Et un mauvais choix d'adjectif :

vos douleurs sont-elles persistantes

15. 'YN-QUESTION headache be in-loc one side-part PRESENT ACTIVE',

to(eng)-'are the headaches in one side',

to(fre)-'vos maux de tête sont-ils sur l'un des côtés'

Le mot 'one' semble être ambigu. Deux sujets l'ont considéré comme un article :

Est-ce que les maux de tête sont de côté ?

Is the headache on the side?

Prenons le cas d'un francophone, qui traduirait 'one' par 'un'. Il en résulte une ambiguïté, puisqu'en français 'un' peut être le nombre '1' mais aussi un article indéfini. Cette ambiguïté n'existe a priori pas en anglais, et donc pas dans l'interlangue, mais pourrait apparaître lors de la traduction. Pour éviter ce problème, une possibilité serait de remplacer dans la forme de surface le mot 'one' par le chiffre '1', ce qui est d'ailleurs le cas dans la forme structurée, avec laquelle nous n'avons pas observé cette difficulté.

La phrase suivante n'est pas entièrement juste, puisqu'elle est moins précise que l'interlangue. 'côté' est remplacé par 'partie' :

Avez-vous les maux de tête dans une seule partie de la tête?

16. 'YN-QUESTION you have pain in-loc head back-part ever PRESENT ACTIVE',

to(eng)-'do you ever experience the pain in the back of the head',

to(fre)-'avez-vous déjà eu mal à l'arrière de la tête'

Cette phrase présente deux difficultés. D'une part, 'ever' semble poser problème aux non-anglophones. Dans plusieurs phrases françaises il a été omis :

Avez-vous eu des douleurs à l'arrière tête ?

Est-ce que vous avez mal à l'arrière de la tête ?

avez-vous mal sur l'arrière de la tête

Il a aussi été traduit par 'toujours' :

Avez-vous toujours les douleurs dans la partie arrière de la tête?

D'autre part, le temps de la phrase est difficile à déterminer. L'interlangue indique le présent, mais la phrase française du corpus est au passé. "Déjà" ne peut pas être utilisé avec le présent.

La phrase suivante n'est pas assez précise, puisque, hormis le contexte du dialogue, rien ne permet de savoir que l'on parle de la tête :

Do you ever have pain in the back?

17. 'WH-QUESTION pain last how-long PRESENT ACTIVE',

from(eng)-'what is the duration of your pain'

from(fre)-'quelle est la durée de la douleur'

to(fre)-'combien de temps la douleur dure-t-elle'

Aucune erreur n'a été faite pour cette phrase.

18. 'YN-QUESTION pain radiate shoulder left PRESENT ACTIVE',

from(fre)-'la douleur irradie t elle l'épaule gauche'

to(eng)-'does the pain radiate to the left shoulder'

to(fre)-'la douleur irradie-t-elle l'épaule gauche'

Tout comme pour la phrase no.5, 'radiate' et le manque d'indication de direction cause problème :

L'épaule gauche irradiée, fait t'elle mal ?

La douleur irradie-t-elle de l'épaule gauche

La douleur irradie-t-elle de l'épaule gauche ?

Resentez-vous de la douleur à l'épaule gauche irradiée

Does the pain radiate in your left shoulder?

Does the pain radiate to/from the left shoulder? (same problem as mentioned in the first part)

19. 'YN-QUESTION attacks become-better PRESENT ACTIVE',

from(eng)-'are the attacks becoming less severe'

to(eng)-'are the attacks better'

to(fre)-'ces crises diminuent-elles'

Cette phrase est délicate, puisque l'interlangue peut être interprétée de deux façons: si les crises s'améliorent, cela peut d'une part être par une diminution de la fréquence des crises, d'autre part par une diminution de l'intensité des crises. Ces deux possibilités sont reflétées par les phrases obtenues :

Avez-vous moins de crises?

Do the attacks become less frequent?

Est-ce que les crises s'améliorent ?
Est-ce que les crises ont diminué d'ampleur?
Do the attacks become less frequent?
La crise va mieux ?
Are the attacks easing?
Les crises s'atténuent –elles ?
est-ce que les crises s'améliorent
Are the attacks getting better?
vos crises s'améliorent?

Toutes ces phrases ont été considérées comme correctes.

20. 'YN-QUESTION you experience nausea PAST ACTIVE',

to(eng)-'did you feel nauseous',
to(fre)-'avez-vous eu des nausées'

Nous avons ici une erreur au niveau du temps verbal :

avez-vous de la nausée

Et un ajout de 'déjà' :

Avez vous déjà eu des nausées?

21. 'YN-QUESTION you have family-history of high-blood-pressure PRESENT ACTIVE',

from(fre)-'avez vous des antécédents familiaux de hypertension',
to(eng)-'do you have a family history of high blood pressure',

Aucune erreur n'a été faite pour cette phrase.

22. 'YN-QUESTION you have attacks two times per month PRESENT ACTIVE'

from(eng)-'do you have the attacks twice a month'
to(eng)-'do you have the attacks twice a month'
to(fre)-'avez-vous ces crises deux fois par mois'

Aucune erreur n'a été faite pour cette phrase.

23. 'YN-QUESTION you have attacks at-least three times per month PRESENT ACTIVE'

from(eng)-'do you have the attacks at least three times a month'
to(eng)-'do you have the attacks at least three times a month'
to(fre)-'avez-vous ces crises au moins trois fois par mois'

Une seule nuance: 'at-least' est remplacé par 'more than', ce qui ne change pas vraiment le sens de la phrase :

Do you have attacks more than three times per month?

24. 'YN-QUESTION you experience pain sc-when [you move-suddenly head PRESENT ACTIVE] PRESENT ACTIVE'

to(eng)-'do you experience the pain when you move the head suddenly'
to(fre)-'avez-vous la douleur quand vous bougez soudainement la tête'

Aucune erreur n'a été faite pour cette phrase.

25. 'YN-QUESTION you have pain sc-after [you eat large-meal PRESENT ACTIVE] PRESENT ACTIVE'

from(eng)-'does the pain occur after you eat a large meal'
to(eng)-'do you experience the pain after you eat a large meal'
to(fre)-'avez-vous mal après un gros repas'

Une seule erreur, où la relation temporelle 'after' est remplacée par 'when' :

Avez-vous de la douleur quand vous mangez beaucoup?

26. 'YN-QUESTION you experience nausea sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE',

from(fre)-'vos maux de tête sont ils précédés par des nausées'
to(eng)-'do you feel nauseous before you have the headaches'
to(fre)-'vos maux de tête sont-ils précédés par des nausées'

Inversion de la principale et de la subordonnée :

Avez-vous mal à la tête avant les nausées ?

Relation temporelle fautive: 'when' au lieu de 'before' :

Est-ce que vous ressentez des nausées quand vous avez mal à la tête?

Ajout de 'déjà' :

avez-vous déjà eu des nausées avant les maux de tête?

4.3.2.3 Récapitulation des difficultés

Dans cette expérience, nous avons observé plusieurs types de difficultés : celles spécifiques à la forme structurée, celles spécifiques à la forme de surface, et des difficultés générales communes aux deux formes. Nous allons donc discuter de ces difficultés dans trois sections distinctes.

4.3.2.3.1 Difficultés liées à la forme structurée

Concepts peu intuitifs : certains concepts ne semblent pas indiquer clairement leur fonction.

- 'state'

La fonction de ce concept n'est pas toujours intuitive, tout particulièrement lorsqu'il est associé à la valeur 'radiate'. Il n'est pas clair qu'il remplit une fonction prédicative.

- 'itemizer'

Le concept d'itemizer n'est pas intuitif.

- 'frequency, time'

Ce concept a perturbé les sujets dans le sens que son utilité n'est pas évidente : il ne semble pas apporter d'informations utiles à un humain pour écrire une phrase en anglais ou français.

- 'secondary_symptom'

Ce concept est ambigu, puisqu'un symptôme secondaire est généralement une manifestation clinique accompagnant une maladie, mais qui n'en est pas le symptôme principal. Un symptôme secondaire ainsi défini ne peut donc pas être la cause du symptôme principal. (voir phrase 12)

Informations potentiellement conflictuelles :

- Deux prépositions dans une même phrase ([prep,duration], [prep,frequency]), phrase no. 9.

Ordre des éléments problématique :

- Dans l'ensemble, les phrases plus courtes sont à l'origine de moins d'erreurs que les phrases plus longues.
- Les phrases comportant des clauses subordonnées sont à l'origine de beaucoup d'erreurs, en particulier dans la première partie utilisant l'interlangue structurée. Il semble difficile d'identifier quelle est la phrase principale et quelle est la clause enchâssée. De plus, l'indication du type de clause (before, when, after) venant après la clause, celle-ci n'est pas toujours utilisée correctement.

« Trop » d'informations dans l'interlangue :

- Dans certains cas, l'interlangue comporte des informations dont les humains ne savent que faire, par exemple [frequency,time]. L'interlangue n'étant pas destinée aux humains, mais servant d'étape intermédiaire à la traduction, elle doit comporter toutes les composantes sémantiques nécessaires au système pour la génération d'une phrase cible complète. Il se peut donc que l'interlangue comporte sous une forme explicite des informations nécessaires à une machine, mais dont l'humain n'a pas besoin car elles

sont pour lui implicites, c'est-à-dire que l'humain les déduit instinctivement grâce à sa connaissance du monde.

4.3.2.3.2 Difficultés liées à la forme de surface

Ambiguïté du mot 'one' :

- Dans la forme structurée, 'one' est représenté par [spec,1]. Sa fonction de nombre n'y est pas donc pas ambiguë. Dans la forme de surface, le nombre 'un' est représenté par 'one'. Nous avons vu que cette représentation, bien que non-ambigüe en tant que telle, peut être une source d'erreurs pour les francophones. En effet, si l'on considère 'one' comme équivalent de 'un', on risque de le considérer comme un article indéfini, et non comme un nombre.

4.3.2.3.3 Difficultés communes aux deux formes

Manque d'informations dans l'interlangue :

- 'become-better' et 'become-worse', lorsqu'ils s'appliquent à des objets ayant plusieurs propriétés, dans notre exemple 'attacks', ne sont pas suffisamment précis. En effet, 'become-better' ne nous indique pas s'il s'agit d'une amélioration quantitative, donc d'une diminution du nombre de crises, ou d'une amélioration qualitative, donc d'une diminution de l'intensité des crises.
- L'élément [state, radiate], qui ne comporte pas d'indication de direction.
- Il n'y a aucun lien entre les spécificateurs de quantité et l'élément qu'ils quantifient, ce qui est problématique dans les phrases comportant deux éléments quantifiables.

Le problème des indications temporelles :

- L'interlangue utilise une représentation basée sur les aspects verbaux, qui permet de décrire toutes les distinctions nécessaires dans le domaine du diagnostic médical. La notion d'aspect, usuelle pour les anglophones, n'est cependant pas toujours intuitive pour les francophones. La distinction entre le present et le present-perfect est à l'origine de beaucoup d'erreurs. Une représentation plus abstraite, qui indiquerait les relations temporelles entre le fait décrit et l'énoncé serait plus explicite. On pourrait par exemple envisager de remplacer [aspect, perfect], [tense, present] par une

représentation qui indiquerait que l'action a commencé dans le passé et est encore en cours.

- Un autre problème rencontré avec les temps verbaux est celui des clauses enchâssées, dont on ne peut savoir si l'indication de temps est à considérer comme absolue ou comme relative au temps de la phrase principale.

4.3.2.4 Comparaison de l'interlangue structurée avec sa forme de surface

Nous avons observé beaucoup moins d'erreurs dans les phrases de la deuxième partie, traduites à partir de la forme de surface. La plus grande différence concerne les phrases comportant des clauses enchâssées. L'identification de la phrase principale et de la clause est plus simple dans la forme de surface. Dans l'interlangue structurée, le fait que l'ordre des éléments ne soit pas similaire à celui d'une langue naturelle semble nuire à la compréhension. De plus, le positionnement de l'élément 'sc-*' (indiquant la relation temporelle) après la clause ne facilite pas la compréhension. Avec les phrases en forme de surface, on observe également moins d'erreurs au niveau du type de relation (before, when ou after). Le type de relation (sc-before, sc-when et sc-after) étant indiqué juste avant la clause, il est moins facile à oublier.

Bien que le nombre de sujets ne permette pas de donner des valeurs statistiquement valables, il nous semble intéressant d'indiquer quelques données quantitatives. Dans la première partie de cette expérience, nous avons au total 49 erreurs sur 130 phrases traitées, dans la deuxième partie nous avons 23 erreurs sur 130 phrases traitées. La forme de surface semble donc être beaucoup plus claire pour des humains.

Un autre facteur intéressant à considérer est le temps de travail indiqué par nos sujets pour ces deux parties. Le temps est très variable selon les sujets. Certains ont pris 9min pour la première partie quand d'autres en ont pris 30min. De même pour la deuxième partie, que les plus rapides ont complétée en 2min, contre les moins rapides qui ont mis 12 ou 20min. Par contre, presque tous les sujets ont fait la deuxième partie en moins de la moitié du temps nécessaire à la première. Tout en faisant moins d'erreurs. Cela renforce notre conclusion que la forme de surface est plus aisément compréhensible.

Cependant, l'ordre des parties a pu influencer les résultats obtenus. Tous les sujets ayant d'abord traité les phrases en forme structurée puis ensuite celles en forme de surface, il est

possible que les résultats de la deuxième partie aient été influencés par un certain effet d'apprentissage. Après avoir traité 13 phrases, il est probable que les sujets se soient familiarisés avec les concepts et le type de phrase utilisé, et aient eu plus de facilité à résoudre les 13 phrases suivantes. Idéalement, il aurait fallu soumettre les deux parties dans l'ordre inverse à la moitié des participants.

Il serait également intéressant de voir si les anglophones font en moyenne moins d'erreurs que les non-anglophones, l'interlangue utilisant des éléments lexicaux anglais. Cependant, comme nous n'avons eu que trois sujets anglophones, nous n'avons pas assez de valeurs pour effectuer cette comparaison.

4.3.2.5 Conclusion

Cette première expérience tend à montrer que la forme de surface est plus facilement compréhensible que la forme structurée. Les phrases longues et complexes, en particulier les phrases avec des clauses enchâssées, sont plus claires dans la forme de surface. Les deux expériences suivantes ne portent plus que sur la forme de surface. Nous allons voir si elle présente les mêmes difficultés dans un contexte d'évaluation que dans le contexte de cette première expérience.

4.3.3 Deuxième expérience

Afin de voir si l'interlangue est suffisamment compréhensible pour servir à l'évaluation de la traduction langue source – interlangue, nous avons préparé une expérience simulant cette situation. Pour des paires de phrases langue source – interlangue de surface, l'évaluateur doit indiquer si le sens est bien traduit. Nous avons introduit des erreurs plus ou moins conséquentes dans certaines paires de phrases. Pour cette expérience, nous avons utilisé la forme de surface de l'interlangue, puisque celle-ci s'avère plus intuitive.

Deux versions de la deuxième expérience ont été préparées, une interlangue – français et une interlangue – anglais, afin que les évaluateurs puissent employer la même langue que celle choisie pour la première expérience.

Dans un souci de réalisme, les erreurs introduites dans l'interlangue ont été choisies pour ne pas être détectables par l'interlingua checker de MedSLT dont nous avons parlé au début de ce chapitre. Il n'y a donc pas omission d'éléments nécessaires à une expression interlangue

bien formée, ni d'éléments inexistant dans le formalisme d'interlangue, ni de combinaisons non-conformes aux spécifications de l'interlangue. Les énoncés interlingues sont donc tous corrects en tant que tels⁷¹, mais certains ne traduisent pas le sens des énoncés en langue naturelle.

Pour créer ces paires de phrases, nous avons sélectionné des paires de phrases similaires dans le corpus extrait de MedSLT. Ces phrases ne diffèrent que sur un détail, par exemple un adverbe, comme dans l'exemple suivant:

(1)
'YN-QUESTION pain be in-loc head front-part **sometimes** PRESENT ACTIVE'
from(fre)-'avez vous **parfois** mal sur le devant de la tête'
to(eng)-'is the pain **sometimes** in the front of the head'
to(fre)-'avez-vous **parfois** mal sur le devant de la tête'

(2)
'YN-QUESTION pain be in-loc head front-part **usually** PRESENT ACTIVE'
from(eng)-'is the pain **usually** frontal'
from(fre)-'avez vous **fréquemment** mal sur le devant de la tête'
from(fre)-'avez vous **habituellement** mal sur le devant de la tête'
from(fre)-'la douleur est elle **habituellement** sur le devant de la tête'
to(eng)-'is the pain **usually** in the front of the head'
to(fre)-'avez-vous **habituellement** mal sur le devant de la tête'

L'adverbe 'sometimes' dans la première phrase est remplacé par l'adverbe 'usually' dans la deuxième. Le reste de la phrase est exactement le même. Pour composer les paires de phrases de notre expérience, nous avons combiné l'interlangue de la première phrase avec la phrase en langue source/cible de la deuxième phrase. L'évaluateur aura donc les deux phrases suivantes:

interlangue : 'YN-QUESTION pain be in-loc head front-part **sometimes** PRESENT ACTIVE'
from(fre)-'la douleur est elle **habituellement** sur le devant de la tête'

et devra indiquer si elles expriment exactement le même sens ou pas.

Cette façon de procéder nous permet de n'utiliser que des phrases source/cible qui sont effectivement traitées par le système, donc couvertes par les grammaires des langues source/cible, ainsi que des expressions interlingues correctes.

Nous avons introduit plusieurs types d'erreurs, que nous allons décrire dans la section suivante.

71 sauf un type d'énoncé, mentionné plus loin.

4.3.3.1 Erreurs introduites

4.3.3.1.1 Adverbes

Les adverbes utilisés dans le corpus MedSLT servent souvent à décrire la fréquence d'apparition des douleurs : 'souvent', 'habituellement', 'parfois', 'déjà', 'toujours'. Les erreurs introduites sont:

- remplacement d'un adverbe par un autre
- omission d'un adverbe
- ajout d'un adverbe

4.3.3.1.2 Adjectifs

Les adjectifs servent principalement à préciser les indications de temps, 'quelques jours', 'plusieurs jours', 'tôt le matin'. De même que pour les adverbes, les adjectifs ont été remplacés, supprimés ou ajoutés.

4.3.3.1.3 Nombre cardinal 'one'

Comme nous l'avons vu dans l'expérience précédente, le nombre 'one' tel qu'il apparaît dans l'interlangue peut causer certaines difficultés

'YN-QUESTION headache be in-loc head one side-part PRESENT ACTIVE'
to(eng)-'are the headaches in one side of the head'
to(fre)-'vos maux de tête sont-ils sur l'un des côtés'

'YN-QUESTION headache be in-loc head side-part PRESENT ACTIVE'
to(eng)-'are the headaches in the side of the head'
to(fre)-'vos maux de tête sont-ils sur le côté'

Les deux phrases ci-dessus ne décrivent pas le même concept, puisque la première précise que la douleur est d'un côté (donc d'un seul côté), tandis que la deuxième indique que la douleur est sur le côté (par opposition à l'avant ou l'arrière de la tête).

Nous avons donc ajouté comme erreur l'omission ou l'ajout de 'one' dans l'interlangue, afin de voir si cet élément est bien pris en compte comme un nombre.

4.3.3.1.4 Clauses subordonnées

Les clauses subordonnées sont toujours introduites par 'sc-when', 'sc-before' et 'sc-after', ce qui indique la relation entre le concept décrit dans la phrase principale et celui de la phrase subordonnée.

Nous avons introduit des phrases avec le mauvais 'sc-*'. Il n'était pas possible de le supprimer

entièrement, car cela aurait rendu les expressions interlingues invalides.

4.3.3.1.5 Modification des temps verbaux

Le système permet de traiter les temps verbaux 'present' et 'past' ainsi que les aspects 'present' et 'perfect'. Nous avons introduit des erreurs en modifiant ces temps verbaux, dans des phrases simples et dans des phrases enchâssées.

4.3.3.1.6 Inversion de la phrase principale avec la phrase subordonnée

Nous avons introduit quelques phrases où la principale est inversée avec la subordonnée, comme dans l'exemple suivant:

'YN-QUESTION you experience fatigue sc-when [you have pain PRESENT ACTIVE] PRESENT ACTIVE'

to(eng)-'do you feel tired when you experience the pain'

to(fre)-'êtes-vous fatigué quand vous avez mal'

'YN-QUESTION you have pain sc-when [you experience fatigue PRESENT ACTIVE] PRESENT ACTIVE'

from(eng)-'does the pain appear when you are tired'

from(fre)-'avez vous mal quand vous êtes fatigué'

to(eng)-'do you experience the pain when you feel tired'

to(fre)-'avez-vous mal quand vous êtes fatigué'

Bien que ces deux phrases comportent les mêmes éléments, elles n'ont pas le même sens, puisque la relation de cause à effet n'est pas la même.

4.3.3.1.7 Modification de la préposition

'YN-QUESTION pain be in-time side-part always PRESENT ACTIVE'

Nous avons remplacé ici 'in-loc' par 'in-time'. A priori l'interlingua checker ne devrait pas laisser passer une telle erreur, des restrictions sur les concepts empêchant les associations erronées. Nous avons toutefois ajouté cette erreur, puisqu'il paraît probable que lors de l'extension du système à de nouveaux sous-domaines du diagnostic médical, de nouveaux concepts seront ajoutés, et ce type d'erreur pourrait arriver.

Pour un total de 40 paires de phrases à évaluer, nous avons introduit 20 paires incorrectes. Les paires de phrases ont été soumises aux sujets dans un ordre aléatoire. Nous avons demandé aux sujets d'indiquer pour chaque paire langue-interlangue si les deux phrases exprimaient le même sens. Les sujets ont été invités à ajouter des commentaires sur les phrases qui leur

paraissaient ambiguës ou les cas qui leur semblaient compliqués. Il aurait été plus judicieux de demander aux sujets d'indiquer pour chaque phrase la raison de leur choix. En effet, l'appréciation de certaines nuances de sens varie d'une personne à une autre. Savoir les raisons des choix aurait permis de différencier les cas où le sujet n'a pas observé la différence de sens de ceux où il l'a observée mais jugée négligeable. Les commentaires ajoutés par une partie des sujets nous ont toutefois permis de reconnaître les cas sujets à cette problématique.

4.3.3.2 Résultats obtenus

Les phrases sont ici triées par type d'erreur. Pour les expériences, elles ont été soumises dans un ordre aléatoire. Pour chaque phrase, nous indiquons le nombre de sujets qui ont jugé que l'interlangue et l'anglais/le français⁷² ont le même sens (oui) ou un sens différent (non).

Nous indiquons ici pour chaque phrase interlangue la phrase anglaise et la phrase française. Les sujets n'ont bien entendu eu qu'une seule phrase en langue naturelle, dans la même langue que celle employée pour la première expérience.

4.3.3.2.1 Erreurs introduites au niveau des adverbes

1. interlangue : 'YN-QUESTION deep pain be often PRESENT ACTIVE'

en : 'is the pain usually deep'

fr : 'la douleur est-elle habituellement profonde'

erreur introduite : l'adverbe 'often' dans l'interlangue est remplacé par 'usually' dans les phrases en/fr

même sens : oui : 6 non : 4

Dans cette phrase, la différence de sens entre 'souvent/often' et 'habituellement/usually' semble à première vue assez faible. Dans le contexte d'un patient décrivant les symptômes qu'il a observés sur sa propre personne, la réalité décrite pourrait être la même : le patient a fréquemment une douleur profonde.

Les définitions⁷³ des deux mots ne sont toutefois pas les mêmes :

often : frequently, in many instances

usually : 1. in a manner that is usual or normal. 2. generally speaking, as a rule.

La distinction des deux qualificatifs dans l'interlangue a donc son utilité, puisque 'habituellement' sous-entend qu'il s'agit d'une condition qui est devenu une habitude, en quelque sorte un état « normal », mais n'implique pas forcément que cet état est encore

⁷² Les sujets ont eu au choix les phrases anglaises ou les phrases françaises.

⁷³ Toutes les définitions sont tirées du Concise Oxford English Dictionary, Eleventh Edition (revised) 2006

valable au moment de l'énoncé. On pourrait imaginer un patient qui aurait habituellement des douleurs profondes, mais qui aurait depuis quelques jours des douleurs suraigües. Par opposition, 'souvent' décrit un état de fait qui est encore valable au moment de l'énoncé (pour autant qu'il soit utilisé avec un verbe au présent, comme dans notre phrase).

Il s'agit ici non pas d'un problème de compréhension, mais plutôt d'un problème d'appréciation.

2. interlangue : 'YN-QUESTION pain be in-loc head front-part sometimes PRESENT ACTIVE'

en : 'is the pain usually in the front of the head'

fr : 'la douleur est elle habituellement sur le devant de la tête'

erreur introduite : l'adverbe 'sometimes' dans l'interlangue est remplacé par 'usually' dans les phrases en/fr

même sens : oui : 0 non : 10

Tous les sujets ont observé la différence de sens entre les deux phrases. La distinction entre 'sometimes' et 'usually' est très nette.

3. interlangue : 'YN-QUESTION headache last duration a-few day PRESENT ACTIVE'

en : 'does the pain last for several days'

fr : 'la douleur dure-t-elle plusieurs jours'

erreur introduite : 'a few' dans l'interlangue est remplacé par 'several/plusieurs' dans les phrases en/fr

même sens : oui : 4 non : 6

Comme pour la première phrase, la différence de sens entre 'a few' et 'several' est très faible.

Les définitions de ces deux mots sont les suivantes :

a few : a small number of, not many

several : more than two but not many

La signification est donc très similaire. La différence réside au niveau de l'accentuation. L'expression 'a few' a un effet plutôt diminutif, dans l'idée de 'juste quelques uns' par opposition à 'beaucoup'. L'expression 'several' accentue plutôt le fait qu'il y a plusieurs occurrences, donc que le phénomène n'est pas négligeable.

Comme pour la phrase no.2, les quatre sujets ayant considéré les deux phrases comme équivalentes au niveau sémantique n'ont probablement pas rencontré de problème de compréhension, mais ont une appréciation différente de la distinction sémantique entre 'a few' et 'several'.

4. interlangue : 'YN-QUESTION pain be in-loc head front-part ever PRESENT ACTIVE'

en : 'is the pain sometimes in the front of the head'

fr : 'avez-vous parfois mal sur le devant de la tête'

erreur introduite 'ever' dans l'interlangue a été remplacé par 'souvent' dans les phrases en/fr
même sens : oui : 7 non : 3

Les phrases correctes associées à cette interlangue sont :

to(eng)-'is the pain ever in the front of the head'

to(fre)-'avez-vous déjà eu mal sur le devant de la tête'

Plusieurs sujets ont émis des doutes sur l'équivalence sémantique de l'interlangue et des phrases en/fr proposées. La différence entre 'ever/déjà' et 'souvent/parfois' est assez subtile. D'après les définitions, le sens est clairement différent:

ever : [usu. with neg. or in questions] : at any time

souvent : occasionally

Cependant, dans notre question, le choix de l'une ou de l'autre variante n'aurait pas d'impact sur la réponse du patient : soit il a eu des douleurs à l'avant de la tête et il répondra oui, soit il n'en a pas eu et il répondra non. La question avec 'ever' met plus d'accent sur la volonté d'éliminer avec sûreté une possibilité. Encore une fois, les résultats obtenus dépendent de l'appréciation des sujets.

5. interlangue : 'WH-QUESTION you have headache when usually PRESENT ACTIVE'

en : 'when do you have the headaches'

fr : 'quand avez-vous vos maux de tête'

erreur introduite : l'adverbe 'usually' dans l'interlangue est supprimé dans les phrases en/fr
même sens : oui : 5 non : 5

Cinq sujets ont considéré qu'il n'y a pas de différence au niveau sémantique. Soit ces sujets n'ont pas vu que 'usually' a été supprimé, soit ils ont jugé que cette suppression n'avait pas d'impact sur le sens. La conception de cette expérience ne nous permet pas de distinguer ces deux cas.

6. interlangue : 'YN-QUESTION headache be in-loc head side-part left always PRESENT ACTIVE'

en : 'are the headaches in the left side'

fr : 'vos maux de tête sont-ils sur le côté gauche'

erreur introduite : l'adverbe 'always' dans l'interlangue est supprimé dans les phrases en/fr
même sens : oui : 1 non : 9

Un seul sujet n'a pas observé l'omission de 'always'.

7. interlangue : 'YN-QUESTION headache become-better sc-when [you experience massage PRESENT ACTIVE] PRESENT ACTIVE'

en : 'are the headaches usually better when you receive massage'

fr : 'vos maux de tête sont-ils habituellement soulagés par des massages'

erreur introduite : l'adverbe 'usually' est ajouté aux phrases en/fr, il n'est pas présent dans l'interlangue

même sens : oui : 5 non : 5

De même que pour la phrase 5, une moitié des sujets a jugé les phrases comme équivalentes d'un point de vue sémantique, bien que les phrases en anglais/français ont un adverbe 'usually' qui n'est pas présent dans l'interlangue. De même que pour le no. 5, on ne peut s'avoir s'il s'agit d'une erreur de compréhension ou d'appréciation.

4.3.3.2.2 Erreurs introduites au niveau des adjectifs

8. interlangue : 'YN-QUESTION you have headache in-time morning PRESENT ACTIVE'

en : 'do you have the headaches in the early morning'

fr : 'vos maux de tête surviennent-ils tôt le matin'

erreur introduite : l'adjectif 'early' est ajouté aux phrases en/fr, il n'est pas présent dans l'interlangue

même sens : oui : 3 non : 7

Trois sujets n'ont pas observé l'ajout de 'early'.

4.3.3.2.3 Erreurs introduites au niveau du nombre cardinal 'one'

9. interlangue : 'YN-QUESTION you have headache in-loc side-part PRESENT ACTIVE'

en : 'do you have the headaches in one side of the head'

fr : 'avez-vous vos maux de tête sur l'un des côtés'

erreur introduite : ajout de 'one' ; plutôt que de désigner le côté en général (par opposition à l'avant ou à l'arrière) comme le fait l'interlangue, on désigne spécifiquement un des côtés de la tête.

même sens : oui : 8 non : 2

Huit sujets n'ont pas observé que l'interlangue est moins précis que le français/l'anglais.

10. interlangue : 'YN-QUESTION headache be in-loc head one side-part PRESENT ACTIVE'

en : 'are the headaches in the side of the head'

fr : 'vos maux de tête sont-ils sur le côté'

erreur introduite : le nombre cardinal 'one', présent dans l'interlangue, est supprimé dans les phrases en/fr

même sens : oui : 6 non : 4

Cette paire de phrases est l'inverse de la précédente, on supprime la précision plutôt que de l'ajouter. Il y a moins d'erreurs que pour la phrase précédente (à noter que ces phrases ne se suivaient pas dans les expériences soumises aux sujets).

4.3.3.2.4 Erreurs introduites au niveau des clauses subordonnées

11. interlangue : 'YN-QUESTION you experience blurred-vision sc-when [you have pain PRESENT ACTIVE] PRESENT ACTIVE'

en : 'do you experience blurred vision before you experience the pain'

fr : 'la douleur est-elle précédée par des troubles de vision'

erreur introduite : la clause en 'when' de l'interlangue est remplacée par une clause en 'before' dans les phrases en/fr

même sens : oui : 0 non : 10

Tous les sujets ont observé cette substitution.

12. interlangue : 'YN-QUESTION you have pain sc-when [you experience fever PRESENT ACTIVE] PRESENT ACTIVE'

en : 'do you have fever when you experience the pain'

fr : 'avez-vous de la fièvre quand vous avez mal'

erreur introduite : inversion de la relation, la subordonnée devient phrase principale.

même sens : oui : 4 non : 6

phrases correctes :

to(eng)-'do you experience the pain when you have fever'

to(fre)-'avez-vous mal quand vous avez de la fièvre'

Comme nous l'avons déjà observé avec la première expérience, les clauses enchâssées ne sont pas forcément intuitives, d'autant plus si les deux possibilités sont cliniquement réalistes (douleur lorsque fièvre, ou fièvre lorsque douleur).

On ne peut malheureusement pas savoir si les sujets n'ont pas observé l'inversion des clauses, ou s'ils ont jugé que celle-ci n'avait aucun impact sur le sens.

13. interlangue : 'YN-QUESTION you experience depression sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE'

en : 'do you have the headaches when you feel depressed'

fr : 'avez-vous vos maux de tête quand vous êtes dépressif'

erreur introduite : inversion de la phrase principale et de la phrase enchâssée ; remplacement de la clause en 'before' par une clause en 'when'

même sens : oui : 3 non : 7

phrases correctes :

from(fre)-'vos maux de tête sont ils précédés d'un état dépressif'

to(eng)-'do you feel depressed before you have the headaches'

La nuance est ici plus complexe que pour la phrase précédente, puisque l'impact de l'inversion des clauses est partiellement compensé par le remplacement de 'sc-before' par 'sc-when'. Plutôt que d'avoir une phrase qui exprime l'inverse de l'interlangue ('avez-vous vos

maux de tête avant de ressentir un état dépressif?), les deux clauses sont mises au même niveau temporel. La différence au niveau sémantique n'est donc pas aussi importante.

4.3.3.2.5 Erreurs introduites au niveau des temps verbaux

14. interlangue : 'YN-QUESTION intense-pain be PRESENT ACTIVE'

en : 'was the pain severe'

fr : 'la douleur a-t-elle été intense'

erreur introduite : le temps verbal passe du présent dans l'interlangue au passé dans les phrases en/fr

même sens : oui : 1 non : 9

Un seul sujet n'a pas observé la différence de temps verbal.

15. interlangue : 'YN-QUESTION you have pain duration several day PRESENT ACTIVE'

en : 'have you had the pain for several days'

fr : 'avez vous mal depuis plusieurs jours'

erreur introduite : le temps verbal passe du présent dans l'interlangue au present perfect dans les phrases en/fr

même sens : oui : 3 non : 7

phrases correctes :

to(eng)-'do the headaches last for a few days'

to(fre)-'vos maux de tête durent-ils quelques jours'

Trois sujets n'ont pas observé le passage du présent au present-perfect. La différence sémantique est cependant importante, puisque l'on passe d'un patient qui n'a pas de douleur au moment du dialogue à un patient qui a des maux de tête au moment du dialogue.

16. interlangue : 'WH-QUESTION you have attacks how-long PRESENT-PERFECT ACTIVE'

en : 'how long do you have the attacks'

fr : 'pendant combien de temps avez-vous ces crises'

erreur introduite : le temps verbal passe du présent perfect dans l'interlangue au present dans les phrases en/fr

même sens : oui : 5 non : 5

Cette paire de phrases est l'inverse de la précédente, le present-perfect est remplacé par le présent.

Les phrases correctes correspondant à l'interlangue sont les suivantes :

en : 'how long have you had attacks'

en : 'how long have you had the attacks'

fr : 'depuis combien de temps avez vous ces crises'

On passe donc d'une question portant sur le moment du début des crises, qui implique que le

patient souffre encore de crises au moment du dialogue, à une question portant sur la durée des crises. La moitié des sujets n'ont pas observé la différence de sens.

17. interlangue : 'YN-QUESTION you smoke PAST ACTIVE'

en : 'do you smoke'

fr : 'fumez-vous'

erreur introduite : le temps verbal passe de past dans l'interlangue au present dans les phrases en/fr

même sens : oui : 1 non : 9

Un sujet n'a pas observé le changement de temps.

18. interlangue : 'WH-QUESTION you have headache how-long PAST ACTIVE'

en : 'how long do you have the headaches'

fr : 'pendant combien de temps avez-vous vos maux de tête'

erreur introduite : le temps verbal passe de past dans l'interlangue au present dans les phrases en/fr

même sens : oui : 1 non : 9

Un sujet n'a pas observé le changement de temps.

19. interlangue : 'YN-QUESTION you have pain sc-when [you lie-down PRESENT ACTIVE] PAST ACTIVE'

en : 'do you experience the pain when you lie down'

fr : 'avez-vous mal quand vous vous allongez'

erreur introduite : passage de la phrase principale du passé au présent

même sens : oui : 4 non : 6

De même que pour la phrase 12 de la première expérience, plusieurs sujets ont commenté sur la difficulté de concevoir une phrase principale au passé avec une clause enchâssée au présent.

Les phrases correctes correspondant à l'interlangue sont les suivantes :

from(fre)-'avez vous eu mal en vous allongeant'

to(eng)-'did you experience the pain when you lay down'

to(fre)-'avez-vous eu mal en vous allongeant'

4.3.3.2.6 Autres erreurs introduites

20. interlangue : 'YN-QUESTION pain be in-time side-part always PRESENT ACTIVE'

en : 'is the pain always in the side'

fr : 'avez vous toujours mal sur le côté'

erreur introduite : ici une erreur a été introduite spécifiquement dans l'interlangue. 'in-loc' a été remplacé par 'in-time'. Cette expression interlangue n'est pas valide, elle n'existe pas dans le corpus.

même sens : oui : 9 non : 1

Cette erreur est difficile à voir, un seul sujet sur dix s'en est aperçu.

4.3.3.2.7 Phrases sans erreurs

21. interlangue : 'WH-QUESTION you have headache how-long PRESENT-PERFECT ACTIVE'

en : 'how long have you had your headaches'

fr : 'depuis combien de temps avez vous mal à la tête'

même sens : oui : 8 non : 2

Encore une fois, l'aspect 'perfect' est à l'origine de difficultés de compréhension.

22. interlangue : 'YN-QUESTION headache be in-loc both side-part PRESENT ACTIVE'

en : 'are the headaches in both sides'

fr : 'vos maux de tête sont-ils des deux côtés'

même sens : oui : 8 non : 2

Un sujet a ajouté ici un commentaire sur la différence en nombre : dans l'interlangue 'headache' est toujours au singulier, dans les phrases il est parfois au singulier (voir phrase 31, 'is the headache on the left side of the head'), parfois au pluriel (comme dans cette phrase). En français, il est plus idiomatique de parler de 'maux de tête' que de 'mal de tête'.

L'interlangue ne contient pas d'informations sur le nombre du symptôme 'headache', celui-ci n'ayant pas vraiment d'importance pour le sens de la phrase. En effet, si l'on parle de 'headaches/maux de tête' au pluriel, cela exprime plutôt l'aspect récurrent du phénomène, et non le fait que le patient ait plusieurs douleurs différentes à la tête de façon simultanée.

23. interlangue : 'WH-QUESTION attacks last how-long usually PRESENT ACTIVE'

en : 'how long do the attacks usually last'

fr : 'combien de temps ces crises durent elles habituellement'

même sens : oui : 9 non : 1

24. interlangue : 'WH-QUESTION you take what-medication regularly PRESENT ACTIVE'

en : 'what medications do you regularly take'

fr : 'quels médicaments prenez vous régulièrement'

même sens : oui : 10 non : 0

25. interlangue : 'WH-QUESTION what is daily-physical-activity PRESENT ACTIVE'

fr : 'quelle est votre activité physique quotidienne'

en : 'what is your daily physical activity'

même sens : oui : 10 non : 0

Nous voulions voir ici si le fait qu'il n'y a pas de pronom possessif dans l'interlangue

permettant de conclure que l'activité physique est celle du patient pourrait induire les sujets en erreur.

26. interlangue : 'YN-QUESTION you have headache in-time morning ever PRESENT ACTIVE'

en : 'do you ever have headaches in the morning'

fr : 'avez vous déjà eu vos maux de tête le matin'

même sens : oui : 7 non : 3

De même que dans la première expérience, le concept de 'ever' est à l'origine de difficultés de compréhension. Les trois sujets ayant indiqué que les phrases n'ont pas le même sens ne sont pas de langue maternelle anglaise.

27. interlangue : 'YN-QUESTION headache last duration a-few week PRESENT ACTIVE'

en : 'do the headaches last for a few weeks'

fr : 'vos maux de tête durent-ils quelques semaines'

même sens : oui : 10 non : 0

28. interlangue : 'YN-QUESTION bursting pain be PAST ACTIVE'

en : 'was the pain bursting'

fr : 'la douleur a-t-elle été éclatante'

même sens : oui : 10 non : 0

29. interlangue : 'YN-QUESTION headache last duration a-few minute PRESENT ACTIVE'

en : 'do the headaches last for a few minutes'

fr : 'vos maux de tête durent-ils quelques minutes'

même sens : oui : 10 non : 0

30. interlangue : 'YN-QUESTION gradual pain start PAST ACTIVE'

fr : 'la douleur a t elle commencé progressivement'

en : 'did the pain start gradually'

même sens : oui : 9 non : 1

31. interlangue : 'YN-QUESTION headache be in-loc head side-part left PRESENT ACTIVE'

en : 'is the headache on the left side of the head'

fr : 'vos maux de tête sont-ils sur le côté gauche'

même sens : oui : 10 non : 0

32. interlangue : 'WH-QUESTION pain last how-long PAST ACTIVE'

en : 'how long did the pain last'

fr : 'combien de temps la douleur a-t-elle duré'

même sens : oui : 10 non : 0

33. interlangue : 'YN-QUESTION pain be in-loc head side-part PRESENT ACTIVE'

en : 'is the pain temporal'

fr : 'la douleur est elle temporale'

même sens : oui : 5 non : 5

Plusieurs sujets ont émis doutes sur l'utilisation de l'adjectif 'temporal' en tant que synonyme de l'expression 'head side-part'. Le terme 'temporal' ne faisant pas partie du langage commun, il aurait du être expliqué dans le lexique. Il ne nous semble toutefois pas certain qu'il soit admissible, d'un point de vue anatomique, d'utiliser l'adjectif temporal de façon synonyme avec 'head side-part', puisque 'temporal' est plus précis et définit une zone distincte située sur le côté de la tête. Si le patient ressent une douleur sur le côté de la tête, derrière l'oreille par exemple, il ne la décrirait probablement pas comme étant temporale.

Nous avons donc ici une situation où l'interlangue est moins précise que les phrases source/cible anglaises ou françaises.

34. interlangue : 'YN-QUESTION you have headache duration more-than two week PRESENT-PERFECT ACTIVE'

en : 'have you had the headaches for more than two weeks'

fr : 'avez vous vos maux de tête depuis plus de deux semaines'

même sens : oui : 10 non : 0

35. interlangue : 'WH-QUESTION you have attacks how-often PRESENT ACTIVE'

en : 'how frequently do you have attacks'

fr : 'quelle est la fréquence de ces crises'

même sens : oui : 9 non : 1

Le sujet ayant affirmé que les phrases n'ont pas le même sens a fait l'expérience en français. Il y a plusieurs différences entre l'interlangue et le français, qui ont pu mener à ce résultat. L'interlangue donne l'idée que le sujet est le patient, la phrase française a pour sujet les crises. Il ne semble pas possible d'exprimer cette phrase de façon idiomatique en français en utilisant le patient comme sujet. Une autre différence est le fait que la phrase française ne se réfère pas directement au patient, puisqu'elle parle de 'ces crises' et non de 'vos crises'. Le sens reste toutefois le même, puisque les crises dont on parle sont implicitement celles du patient.

36. interlangue : 'WH-QUESTION you have pain when usually PRESENT ACTIVE'

en : 'when does your pain usually appear'

fr : 'quand la douleur survient elle habituellement'

même sens : oui : 10 non : 0

37. interlangue : 'YN-QUESTION you have pain sc-when [you experience anger PRESENT ACTIVE] PAST

en : 'did you experience the pain when you felt angry'

fr : 'la douleur a t elle été causée par la colère'

même sens : oui : 5 non : 5

La version française de cette phrase n'exprime pas tout à fait le même sens que l'interlangue: l'interlangue exprime une relation purement temporelle, le français exprime une relation clairement causale. Il est donc justifié que les sujets ayant fait l'expérience en français aient jugé que les deux phrases n'ont pas le même sens.

38. interlangue : 'YN-QUESTION you experience drowsiness sc-before [you have headache PRESENT ACTIVE] PAST ACTIVE'

en : 'did you feel drowsy before you had the headaches'

fr : 'vos maux de tête ont-ils été précédés d'un état de somnolence'

même sens : oui : 8 non : 2

Les trois phrases expriment la même relation temporelle. La phrase française utilise les maux de tête comme sujet, l'interlangue le patient, ce qui peut prêter à confusion.

Un autre problème mentionné par un des sujets est la définition des temps. En effet, sans avoir de spécification de l'interlangue, on ne peut savoir comment interpréter les temps verbaux des clauses subordonnées. Si la clause est au présent, cela implique t'il que le verbe de la clause sera au présent? Ou cela signifie t'il que la clause sera au même temps que la phrase principale? La phrase française contourne élégamment ce problème en utilisant une construction sans clause enchâssée.

39. interlangue : 'YN-QUESTION headache last duration a-few minute usually PRESENT ACTIVE'

en : 'do the headaches usually last for a few minutes'

fr : 'vos maux de tête durent-ils habituellement quelques minutes'

même sens : oui : 10 non : 0

40. interlangue : 'YN-QUESTION headache become-worse sc-when [read CONTINUOUS ACTIVE] PRESENT ACTIVE'

en : 'are the headaches worse when you read'

fr : 'vos maux de tête empirent-ils quand vous lisez'

même sens : oui : 9 non : 1

Un sujet a relevé ici la différence sémantique entre 'are worse' et 'become worse'. D'un point de vue purement diagnostique, il s'agit de découvrir si la lecture a un effet négatif sur le patient. Les deux variantes impliquent que le patient a mal à la tête. La variante avec 'are worse' sous-entend que le patient a plus ou moins mal (de façon constante) en fonction de ses activités, et qu'en l'occurrence il a plus mal lorsqu'il lit. La variante avec 'become worse' implique que les maux de tête deviennent plus forts lorsque le patient se met à lire. Cette distinction nous semble toutefois très faible, et négligeable dans le contexte du diagnostic.

Il s'agit donc ici d'un problème d'appréciation et non de compréhension de l'interlangue.

4.3.3.3 Conclusion

Sur les 400 paires de phrases traitées, traductions correctes et incorrectes confondues, nous avons observé 99 mauvaises réponses. 76 de ces erreurs sont des cas où une différence de sens introduite entre l'interlangue et la phrase française/anglaise n'a pas été observée, 23 sont des cas où une différence de sens incorrecte a été observée entre une expression interlangue et son équivalent français/anglais dans le corpus MedSLT.

Les difficultés rencontrées étant principalement les mêmes que pour la troisième expérience, nous les discuterons en commun, après les résultats de la troisième expérience.

4.3.4 Troisième expérience

4.3.4.1 Description de l'expérience

La troisième expérience repose sur le même concept que la deuxième, c'est-à-dire l'introduction d'erreurs de traduction entre des phrases en langue naturelle et leur représentation interlingue. A la différence de la deuxième expérience, en plus de la phrase source et de sa représentation interlingue, on a une phrase cible. Il s'agira donc d'évaluer chaque paire possible de ces trois phrases.

Cette expérience simule une situation d'évaluation bilingue des traductions, où l'interlangue sert d'information supplémentaire pour déterminer, en cas de mauvaise traduction source cible, si l'erreur de traduction est causée par une mauvaise traduction source – interlangue ou par une mauvaise traduction interlangue - cible.

Voici un exemple d'un tel triplet de phrases : ici la paire anglais - interlangue est correcte, tandis que la paire interlangue – français est incorrecte, puisque 'gradually' est traduit par

'soudainement' et non par graduellement. La paire anglais - français est donc également incorrecte.

en : 'did the headaches start gradually'

interlangue : 'YN-QUESTION headache start gradually PAST ACTIVE'

fr : 'vos maux de tête ont-ils commencé soudainement'

Les erreurs introduites sont du même type que pour la deuxième expérience (changement de temps, omissions, etc.). Dans certains cas, la traduction source – cible est correcte, mais l'interlangue contient une erreur. Les triplets de phrases ont été constitués de la même façon que les paires de la deuxième expérience, en utilisant des paires de phrases similaires du corpus extrait de MedSLT. Pour composer le premier triplet par exemple, nous avons extrait les deux phrases suivantes du corpus :

(1)
 'YN-QUESTION headache start **gradually** PAST ACTIVE'
 to(eng)-'did the headaches start **gradually**'
 to(fre)-'vos maux de tête ont-ils commencé **progressivement**'

(2)
 'YN-QUESTION headache start **suddenly** PAST ACTIVE'
 [from(eng)-'did your headache start **suddenly**'
 to(eng)-'did the headaches start **suddenly**'
 to(fre)-'vos maux de tête ont-ils commencé **soudainement**'

Ces phrases ne se différencient qu'au niveau de l'adverbe, 'gradually' dans la première et 'suddenly' dans la deuxième. Pour former le triplet de l'expérience, nous avons associé la phrase anglaise (1) à l'interlangue (1) et à la phrase française (2). On obtient ainsi une paire anglais-interlangue correcte, une paire interlangue-français incorrecte, et, forcément, une paire anglais-français incorrecte.

Nous avons reporté ci-dessous les phrases proposées aux sujets, suivies des erreurs introduites et des réponses des sujets.

4.3.4.2 Résultats obtenus

4.3.4.2.1 Phrases avec des erreurs introduites au niveau des adverbes

1. en : 'did the headaches start gradually'

interlangue : 'YN-QUESTION headache start gradually PAST ACTIVE'

fr : 'vos maux de tête ont-ils commencé soudainement'

en-int : correct

int-fr : 'gradually' remplacé par 'suddenly'

Résultats :

en-int, même sens : oui : 10 non : 0
int-fr, même sens : oui : 0 non : 10
en-fr, même sens : oui : 0 non : 10

Ces phrases n'ont posé aucune difficulté aux sujets.

2. en : 'do you ever have the headaches at night'

interlangue : 'YN-QUESTION you have headache in-time night ever PRESENT ACTIVE'

fr : 'vos maux de tête surviennent-ils souvent la nuit'

en-int : correct

int-fr : ever -> often

Résultats :

en-int, même sens : oui : 9 non : 1
int-fr, même sens : oui : 0 non : 10
en-fr, même sens : oui : 0 non : 10

Pour ces phrases, un sujet a jugé la phrase anglaise comme étant non-idiomatique, et donc fausse.

3. en : 'does bright light ever make your headaches worse'

interlangue : 'YN-QUESTION headache become-worse sc-when [you experience bright-light PRESENT ACTIVE] PRESENT ACTIVE'

fr : 'vos maux de tête sont ils aggravés par une lumière forte'

en-int : omission de 'ever'

int-fr : correct

Résultats :

en-int, même sens : oui : 9 non : 1
int-fr, même sens : oui : 9 non : 1
en-fr, même sens : oui : 7 non : 3

Encore une fois, le concept de 'ever' pose problème aux non-anglophones. La phrase anglaise correspondant à l'interlangue (donc sans la notion de 'ever') est la suivante :

'does bright light make your headaches worse'

Cette phrase exprime une influence systématique de la lumière forte sur l'intensité des maux de tête. La phrase avec 'ever', qui se traduit en français de la façon suivante :

'vos maux de tête ont ils été déjà aggravés par une lumière forte'

exprime que la lumière a déjà eu une influence sur les maux de tête, sans pour autant que la lumière ait toujours une influence. La différence est donc relativement faible.

4. en : 'do the headaches sometimes last for more than two hours'

interlangue : 'YN-QUESTION headache last duration more-than two hour PRESENT ACTIVE'

fr : 'vos maux de tête durent-ils parfois plus de deux heures'

'sometimes' manque dans l'interlangue
en-fr : correct

Résultats :

en-int, même sens : **oui : 5** non : 5
int-fr, même sens : **oui : 5** non : 5
en-fr, même sens : oui : 9 **non : 1**

Presque tous les évaluateurs ont jugé la phrase anglaise et la phrase française équivalente, mais seulement une moitié a vu que l'interlangue est incomplète.

4.3.4.2.2 Phrases avec des erreurs introduites au niveau du nombre cardinal 'one'

5. en : 'do you experience the pain in one side'

interlangue : 'YN-QUESTION you have pain in-loc side-part PRESENT ACTIVE'

fr : 'avez-vous mal sur le côté'

en-int : omission de 'one'
int-fr : correct

Résultats :

en-int, même sens : **oui : 5** non : 5
int-fr, même sens : oui : 8 **non : 2**
en-fr, même sens : **oui : 2** non : 8

La phrase anglaise est ici plus précise que l'interlangue, puisqu'elle spécifie que la douleur est sur un des côtés, et non sur le côté (par opposition à l'avant ou à l'arrière). Comme nous l'avons déjà observé dans les expériences précédentes, la notion de 'one' est source de difficultés.

4.3.4.2.3 Phrases avec des erreurs introduites au niveau des clauses subordonnées

6. en : 'does the pain appear when you are tired'

interlangue : 'YN-QUESTION you experience fatigue sc-when [you have pain PRESENT ACTIVE] PRESENT ACTIVE'

fr : 'avez vous mal quand vous êtes fatigué'

en-int : inversion phrase principale et subordonnée
int-fr : réinversion
en-fr correct

Résultats :

en-int, même sens : **oui : 1** non : 9
int-fr, même sens : **oui : 3** non : 7
en-fr, même sens : oui : 6 **non : 4**

Les phrases anglaises et françaises correspondant à l'interlangue sont les suivantes :

to(eng)-'do you feel tired when you experience the pain'

to(fre)-'êtes-vous fatigué quand vous avez mal'

Trois sujets n'ont pas observé l'inversion des clauses. Deux de ces sujets jugé que l'interlangue et le anglais n'avaient pas le même sens à cause du verbe 'appear', qui ne correspond pas au 'have' de l'interlangue.

Six sujets ont jugé que la phrase anglaise correspond au français, mais que l'interlangue ne correspond pas.

Il semble intéressant de noter ici que lorsqu'il y a plusieurs différences, impactant le sens ou non, certains sujets semblent se focaliser sur une des différences (en l'occurrence, pour deux sujets, le changement de verbe), ce qui les empêche de voir l'autre (l'inversion des clauses).

7. en : 'is the pain accompanied by nausea'

interlangue : 'YN-QUESTION you have pain sc-when [you experience nausea PRESENT ACTIVE] PRESENT ACTIVE'

fr : 'avez-vous mal quand vous avez des nausées'

en-int : inversion phrase principale et subordonnée

int -fr : correct

Résultats :

en-int, même sens : **oui : 3** non : 7

int-fr, même sens : oui : 9 **non : 1**

en-fr, même sens : **oui : 3** non : 7

La paire anglais-interlangue est difficile, puisque les éléments lexicaux 'pain' et 'nausea' s'y trouvent dans le même ordre, bien que les clauses soient inversées. Dans la phrase anglaise, la douleur est accompagnée de nausées tandis que dans l'interlangue, des nausées sont accompagnées par la douleur. D'un point de vue médical, il s'agit de deux cas clairement distincts.

8. en : 'do you feel stressed before you have the headaches'

interlangue : 'YN-QUESTION you experience stress sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE'

fr : 'avez vous vos maux de tête quand vous êtes stressé'

en-int : correct

int-fr : inversion principale et subordonnée, 'sc-before' devient 'sc-when'

Résultats :

en-int, même sens : oui : 10 **non : 0**

int-fr, même sens : **oui : 1** non : 9

en-fr, même sens : **oui : 1** non : 9

Un sujet n'a pas observé l'inversion.

Ces phrases, où l'anglais et le français ne correspondent pas, ont posé moins de problèmes que l'autre cas d'inversion (no. 6), où les deux phrases en langue naturelle étaient des traductions correctes l'une de l'autre.

9. en : 'are you sensitive to bright light when you experience the pain'

interlangue : 'YN-QUESTION you experience photophobia sc-when [you have pain PRESENT ACTIVE] PRESENT ACTIVE'

fr : 'la douleur est-elle précédée d'une photophobie'

en-int : correct

int-fr : 'sc-when' remplacé par 'sc-before'

Résultats :

en-int, même sens : oui : 8 non : 2

int-fr, même sens : oui : 0 non : 10

en-fr, même sens : oui : 0 non : 10

Tous les sujets ont observé la différence entre l'interlangue et le français. Deux sujets ont estimé que l'anglais ne correspond pas à l'interlangue, 'sensitive to bright light' n'étant pas équivalent à 'photophobia'. Ce terme aurait peut-être dû être défini dans le glossaire fourni avec les expériences.

10. en : 'are your headaches usually accompanied by drowsiness'

interlangue : 'YN-QUESTION you experience drowsiness sc-before [you have headache PRESENT ACTIVE] PAST ACTIVE'

fr : 'vos maux de tête ont-ils été précédés d'un état de somnolence'

en-int : 'sc-when' devient 'sc-before', usually

int-fr : correct

Résultats :

en-int, même sens : oui : 0 non : 10

int-fr, même sens : oui : 10 non : 0

en-fr, même sens : oui : 0 non : 10

Aucune erreur n'a été faite pour ces phrases.

4.3.4.2.4 Phrases avec des modifications des temps verbaux

11. en : 'did you experience the pain for more than ten days'

interlangue : 'YN-QUESTION you have pain duration more-than ten day PAST ACTIVE'

fr : 'avez-vous mal depuis plus de dix jours'

en-int : correct

int-fr : past -> present perfect

Résultats :

en-int, même sens : oui : 9 non : 1

int-fr, même sens : oui : 7 non : 2

en-fr, même sens : **oui : 7** non : 2

Seuls deux sujets ont observé le changement de temps entre l'interlangue et le français. La phrase française correspondant à l'interlangue est la suivante :

'avez-vous eu mal pendant plus de dix jours'

La différence entre le past et le present-perfect est difficile à cerner, puisqu'il s'agit dans les deux cas de situations qui ont commencé dans le passé, l'une étant terminée (past), l'autre étant encore en cours (present-perfect). En français cela se traduit par un verbe au passé et 'pendant' (past) où un verbe au présent et 'depuis' (present perfect). Un des sujets a émis un doute quant au 'depuis', qui aurait du être 'pendant'.

12. en : 'did you feel weak before you had the headaches'

interlangue : 'YN-QUESTION you experience weakness sc-before [you have headache PRESENT ACTIVE] PAST ACTIVE'

fr : 'vos maux de tête sont ils précédés d'une impression de faiblesse'

en-int : correct

int-fr : phrase principale : past -> present

Résultats :

en-int, même sens : **oui : 7** **non : 3**

int-fr, même sens : **oui : 7** non : 3

en-fr, même sens : **oui : 5** non : 5

La différence entre l'interlangue et le français n'a été observée que par trois sujets. Cinq sujets ont estimé que les trois phrases sont équivalentes.

Comme pour la phrase no. 38 de la deuxième expérience, les indications de temps pour les phrases avec clauses ne sont pas claires.

La phrase française correspondant à l'interlangue est la suivante :

'vos maux de tête ont-ils été précédés d'une impression de faiblesse'

Le sens transmis par cette phrase nous semble être le même que celui de la phrase française proposée dans l'expérience. La seule différence réside dans le fait que cette deuxième phrase serait également applicable lorsque le patient ne souffre plus de maux de tête. Auquel cas il n'irait probablement pas consulter un médecin. Dans le contexte du dialogue de diagnostic, ces deux phrases nous paraissent donc équivalentes.

13. en : 'did you have an upper respiratory infection'

interlangue : 'YN-QUESTION you have upper-respiratory-infection PRESENT ACTIVE'

fr : 'avez-vous eu une infection des voies respiratoires supérieures'

en-int : past -> present
int-fr : present -> past
en-fr : correct

Résultats :

en-int, même sens : **oui : 0** non : 10
int-fr, même sens : **oui : 2** non : 8
en-fr, même sens : **oui : 8** **non : 2**

Deux sujets n'ont pas observé le changement de temps entre l'interlangue et le français.

14. en : 'does the pain last for several hours'

interlangue : 'YN-QUESTION pain last duration several hour PRESENT-PERFECT ACTIVE'

fr : 'la douleur dure-t-elle depuis plusieurs heures'

en-int : present -> present perfect
int-fr : correct

Résultats :

en-int, même sens : **oui : 5** non : 5
int-fr, même sens : **oui : 6** **non : 4**
en-fr, même sens : **oui : 4** non : 6

Nous avons ici encore un exemple utilisant le present-perfect, causant des difficultés aux non-anglophones. Étonnement, quatre sujets n'ont pas observé la différence de sens entre l'anglais et le français, qui semble ici assez évidente, contrairement au no. 11 de cette expérience.

15. en : 'did you have a sudden headache'

interlangue : 'YN-QUESTION you have sudden headache PAST ACTIVE'

fr : 'avez-vous soudainement vos maux de tête'

en-int : correct
int-fr : past -> present

Résultats :

en-int, même sens : **oui : 10** **non : 0**
int-fr, même sens : **oui : 2** non : 8
en-fr, même sens : **oui : 2** non : 8

Deux sujets n'ont pas observé le passage du passé au présent entre l'interlangue et le français.

16. en : 'did you experience the pain more often'

interlangue : 'YN-QUESTION you have pain more-often PRESENT ACTIVE'

fr : 'avez-vous plus fréquemment la douleur'

en-int : past -> present
int-fr : correct

Résultats :

en-int, même sens : **oui : 2** non : 8

int-fr, même sens : oui : 10 non : 0
en-fr, même sens : oui : 2 non : 8

Deux sujets n'ont pas observé le passage du passé au présent entre l'anglais et l'interlangue.

4.3.4.2.5 Phrases avec d'autres modifications

17. en : 'are the headaches becoming less severe'

interlangue : 'YN-QUESTION headache become-better PRESENT ACTIVE'

fr : 'l'intensité de vos maux de tête augmente t elle'

en-int : correct

int- fr : become better -> become worse

Résultats :

en-int, même sens : oui : 8 non : 2
int-fr, même sens : oui : 0 non : 10
en-fr, même sens : oui : 0 non : 10

Ces phrases ont été choisies à cause de la potentielle ambiguïté des paires : 'becoming less severe' avec 'become better' est juste, tandis que 'become better' et 'augmente' est fausse. La difficulté repose sur l'association de mots à connotation positive ('better', 'augmenter') avec des mots à connotation négative ('less').

Seuls deux sujets se sont trompés pour la première paire.

18. en : 'do you experience the pain when you feel stressed'

interlangue : 'YN-QUESTION you have pain sc-when [you experience stress PRESENT ACTIVE] PRESENT ACTIVE'

fr : 'la douleur empire-t-elle quand vous êtes stressé'

en-int :correct

int-fr: have -> become worse

Résultats :

en-int, même sens : oui : 10 non : 0
int-fr, même sens : oui : 1 non : 9
en-fr, même sens : oui : 1 non : 9

Seul un sujet n'a pas observé le changement de prédicat dans la paire interlangue-français.

19. en : 'how frequently do you have attacks'

interlangue : 'WH-QUESTION you have attacks how-often PRESENT ACTIVE'

fr : 'quelle est la fréquence de ces crises par jour'

en-int : correct

int-fr : ajout de 'per day'

Résultats :

en-int, même sens : oui : 10 non : 0

int-fr, même sens : **oui : 2** non : 8
 en-fr, même sens : **oui : 2** non : 8

Deux sujets ont jugé que l'interlangue correspond au français. Un de ces sujets a ajouté un commentaire précisant qu'il avait observé l'ajout de 'par jour', mais qu'il l'a jugé comme n'affectant pas le sens.

20. en : 'does the pain usually last for more than one hour'

interlangue : 'YN-QUESTION pain last duration more-than one hour usually PRESENT ACTIVE'

fr : 'la douleur dure-t-elle habituellement une heure'

en-int : correct

int-fr : omission de 'more than'

Résultats :

en-int, même sens : **oui : 10** **non : 0**

int-fr, même sens : **oui : 1** non : 9

en-fr, même sens : **oui : 1** non : 9

Un seul sujet n'a pas observé l'omission de 'plus que'.

21. en : 'are the attacks becoming more frequent'

interlangue : 'YN-QUESTION attacks of pain become-worse PRESENT ACTIVE'

fr : 'ces crises empirent-elles'

en-int : augmentation de la fréquence -> augmentation de la douleur

int-fr : correcte

Résultats :

en-int, même sens : **oui : 1** non : 9

int-fr, même sens : oui : 9 **non : 1**

en-fr, même sens : **oui : 1** non : 9

L'expression 'become-worse' associée dans l'interlangue à 'attacks of pain' est ambiguë. Les crises peuvent en effet empirer de deux façons, en devenant plus fréquentes ou en devenant plus douloureuses. Le sens de la phrase anglaise pourrait donc être inclus dans le sens de la phrase interlangue. Cependant l'inverse n'est pas correct, la phrase anglaise apportant plus de précisions que l'interlangue. Plusieurs sujets ont commenté sur l'ambiguïté du concept 'become worse'

22. en : 'do you take medication'

interlangue : 'WH-QUESTION you take what-medication PRESENT ACTIVE'

fr : 'quels médicaments prenez-vous'

en-int : ajout de 'what'

int-fr : correct

Résultats :

en-int, même sens : **oui : 0** non : 10

int-fr, même sens : oui : 10 **non : 0**

en-fr, même sens : **oui : 0** non : 10

Tous les sujets ont observé la différence entre l'anglais et l'interlangue, ainsi que l'équivalence entre l'interlangue et le français.

23. en : 'do you sometimes have the headaches in the afternoon'

interlangue : 'YN-QUESTION you have headache in-loc afternoon sometimes PRESENT ACTIVE'

fr : 'avez vous parfois vos maux de tête l'après-midi'

int : remplacement de 'in-time' par 'in-loc' (n'existe pas dans le corpus)

en-fr : correct

Résultats :

en-int, même sens : **oui : 7** non : 3

int-fr, même sens : **oui : 7** non : 3

en-fr, même sens : oui : 9 **non : 1**

Seuls trois sujets ont observé l'erreur dans l'interlangue. Cette phrase a été ajoutée pour voir si les sujets faisaient attention au complément attaché la préposition. Manifestement, il s'agit d'une erreur difficilement visible.

4.3.4.2.6 Phrases sans erreurs

24. en : 'are the headaches in the back of the head'

interlangue : 'YN-QUESTION headache be in-loc head back-part PRESENT ACTIVE'

fr : 'vos maux de tête sont-ils à l'arrière de la tête'

Résultats :

en-int, même sens : oui : 10 **non : 0**

int-fr, même sens : oui : 10 **non : 0**

en-fr, même sens : oui : 10 **non : 0**

Aucune erreur pour ces phrases.

25. en : 'did you experience the pain after you experienced unusual activity'

interlangue : 'YN-QUESTION you have pain sc-after [you experience unusual-activity PRESENT ACTIVE] PAST ACTIVE'

fr : 'avez vous eu mal après une activité inhabituelle'

Résultats :

en-int, même sens : oui : 6 **non : 4**

int-fr, même sens : oui : 7 **non : 3**

en-fr, même sens : oui : 7 **non : 3**

Encore une fois, les temps verbaux des clauses causent des difficultés. Plusieurs commentaires des sujets mentionnent une erreur de temps au niveau de l'interlangue.

26. en : 'is the pain always on the same side of the head'

interlangue : 'YN-QUESTION you have pain in-loc head the-same side-part always PRESENT ACTIVE'

fr : 'avez-vous toujours mal du même côté'

Résultats :

en-int, même sens : oui : 10 non : 0

int-fr, même sens : oui : 6 non : 4

en-fr, même sens : oui : 6 non : 4

Pour ces phrases, quatre sujets ont très justement remarqué qu'il manque 'de la tête' dans la phrase française. Dans le contexte d'un dialogue portant exclusivement sur les maux de tête cette précision n'est pas réellement nécessaire. Cependant, il ne paraît pas très cohérent que la phrase anglaise et la phrase française (qui sont toutes les deux des phrases générées à partir de l'interlangue) ne contiennent pas les mêmes informations.

27. en : 'did you experience the pain at night'

interlangue : 'YN-QUESTION you have pain in-time night PAST ACTIVE'

fr : 'la douleur est-elle survenue la nuit'

Résultats :

en-int, même sens : oui : 10 non : 0

int-fr, même sens : oui : 10 non : 0

en-fr, même sens : oui : 10 non : 0

Aucune erreur pour ces phrases.

28. en : 'do you have the headaches after you drank red wine'

interlangue : 'YN-QUESTION you have headache sc-after [you drink red-wine PAST ACTIVE] PRESENT ACTIVE'

fr : 'avez-vous vos maux de tête quand vous avez bu du vin rouge'

Résultats :

en-int, même sens : oui : 10 non : 0

int-fr, même sens : oui : 7 non : 3

en-fr, même sens : oui : 7 non : 3

Plusieurs sujets ont ici commenté que la phrase anglaise est agrammaticale. Il paraît toutefois difficile de formuler une phrase au présent avec une clause spécifiquement au passé. La clause étant introduite par 'after', il est implicite que ce qui y est décrit s'est déroulé avant l'évènement décrit dans la phrase principale.

29. en : 'have you had the headaches for more than one week'

interlangue : 'YN-QUESTION you have headache duration more-than one week PRESENT-PERFECT ACTIVE'

fr : 'avez vous vos maux de tête depuis plus d'une semaine'

Résultats :

en-int, même sens : oui : 8 non : 2

int-fr, même sens : oui : 8 non : 2

en-fr, même sens : oui : 6 non : 4

Le present perfect est encore une fois source d'erreurs.

30. en : 'do you have fever'

interlangue : 'YN-QUESTION you have fever PRESENT ACTIVE'

fr : 'est-ce que vous êtes fiévreux'

Résultats :

en-int, même sens : oui : 10 non : 0

int-fr, même sens : oui : 10 non : 0

en-fr, même sens : oui : 10 non : 0

Ces phrases n'ont causé aucune difficulté.

31. en : 'were the headaches better'

interlangue : 'YN-QUESTION headache become-better PAST ACTIVE'

fr : 'vos maux de tête ont-ils diminué'

Résultats :

en-int, même sens : oui : 7 non : 3

int-fr, même sens : oui : 8 non : 2

en-fr, même sens : oui : 5 non : 5

La difficulté réside ici dans le fait que l'interlangue et l'anglais ont les deux l'expression 'better', à connotation positive, tandis que le français a l'expression 'diminué', à connotation négative.

Un sujet a remarqué sur la différence sémantique entre l'expression 'be better' de la phrase anglaise et 'become better' de l'interlangue.

32. en : 'do the headaches usually last for a few seconds'

interlangue : 'YN-QUESTION headache last duration a-few second usually PRESENT ACTIVE'

fr : 'vos maux de tête durent-ils habituellement quelques secondes'

Résultats :

en-int, même sens : oui : 10 non : 0

int-fr, même sens : oui : 10 non : 0

en-fr, même sens : oui : 10 non : 0

Aucune erreur pour ces phrases.

4.3.4.3 Récapitulation des erreurs

Les résultats obtenus avec ces expériences ne permettent pas de conclure directement sur la compréhensibilité de l'interlangue, puisqu'ils sont influencés par d'autres facteurs. En effet, lorsque les sujets estiment qu'une paire de phrases a le même sens, mais que cette paire de phrases fait partie de celles où nous avons délibérément introduit une erreur, deux cas sont

possibles.

Le premier cas est celui où le sujet n'a pas remarqué de différence et où l'on peut donc remettre en question la compréhensibilité de l'interlangue. Dans le deuxième cas, le sujet a remarqué une différence, mais a jugé que celle-ci n'avait pas d'impact sur le contenu sémantique. Nos expériences ne permettent pas de différencier ces deux cas. Il faudrait pour cela développer d'autres expériences, où l'on demanderait au sujet d'expliquer pourquoi il juge que deux phrases ont le même sens ou pas. Dans certains cas, les sujets ont ajouté des commentaires indiquant les raisons de leurs choix.

Les réponses que nous considérons ici comme des erreurs peuvent donc très bien être des cas où le sujet n'a simplement pas la même appréciation de la notion de 'même sens' que celle appliquée lors du développement de MedSLT.

Sur les 960 paires de phrases (en-int, int-fr et en-fr) traitées, nous observons 182 erreurs (taux de réussite de 81%). 56 de ces erreurs portent sur les paires anglais-interlangue, 61 sur les paires interlangue-français, et 65 sur les paires anglais-français. Il y a donc un nombre relativement similaire d'erreurs pour les paires langue-interlangue que pour les paires langue-langue.

Le jugement des trois paires d'un triplet n'étant pas dissociables (si l'on décide que $en \neq int$ et que $int = fr$, on aura forcément $en \neq fr$) et les choix faits pour l'une ou l'autre des paires risquant d'avoir un impact sur les autres choix du triplet, il vaudrait peut-être mieux compter le nombre de triplets incorrects. Sur un total de 320 triplets traités, 92 étaient faux (taux de réussite de 71%). Un triplet est considéré comme faux à partir du moment où l'une des paires est jugée incorrectement.

Les chiffres ci-dessus ne sont cependant pas statistiquement représentatifs, d'une part à cause de l'erreur de conception de l'expérience mentionnée ci-dessus, d'autre part par le nombre restreint de sujets.

Les deuxième et troisième expériences nous ont toutefois permis d'observer quelques difficultés de compréhension liées à la forme de surface de l'interlangue de MedSLT. Nous les récapitulons dans les sections suivantes.

4.3.4.3.1 Erreurs d'adverbes

Beaucoup des erreurs d'adverbes n'entraînent que de faibles différences sémantiques, et sont donc sujettes à la problématique ci-dessus. Celle-ci apparaît aussi bien lorsqu'un adverbe est remplacé par un autre très similaire, que lorsqu'un adverbe changeant peu l'énoncé, 'usually' par exemple est ajouté ou supprimé. Par contre, lorsque les différences sont frappantes, par exemple entre 'gradually' et 'suddenly', tous les sujets ont observé que les phrases n'ont pas le même sens.

4.3.4.3.2 Nombre 'one'

Les phrases comportant le nombre 'one' sont toutes problématiques, le mot 'one' de l'interlangue étant souvent confondu avec un article indéfini, bien que 'one' en anglais ne soit jamais un article. Comme nous l'avons mentionné pour la première expérience, il est possible que les francophones fassent cette erreur en interprétant 'one' par analogie avec 'un', qui lui est ambigu.

4.3.4.3.3 Clauses

Les inversions de clauses où la phrase subordonnée devient la phrase principale sont difficiles à déceler, tout particulièrement lorsque le sujet n'a qu'une seule phrase en langue naturelle à disposition (deuxième expérience) ou lorsque les deux phrases en langue naturelle sont des traductions correctes l'une de l'autre (troisième expérience). Les deux cas de la troisième expérience où l'anglais ne correspond pas au français ont produit moins d'erreurs. Ce type d'erreur semble donc être plus visible en langues naturelles que dans l'interlangue de surface.

Les erreurs introduites au niveau des identificateurs de clauses (sc-when, etc), ont été repérées par presque tous les sujets dans tous les cas.

4.3.4.3.4 Temps verbaux

Dans les phrases simples (sans clause), les changements de temps entre le présent et le passé ont presque toujours été repérés. Les phrases impliquant le present- perfect sont source de beaucoup plus d'erreurs, l'aspect 'perfect' ne semblant pas intuitif pour des francophones.

Les phrases avec clauses dont on a modifié le temps verbal ont causé plus de difficultés. En effet, il n'est pas clair si le temps verbal indiqué pour la clause est à considérer comme absolu ou comme relatif au temps verbal de la phrase principale.

4.3.4.3.5 Autres erreurs

Les autres erreurs introduites, ajout de précision ('more than'), modification de prédicat ('become-better' / 'become-worse) ont presque toujours été identifiées. Le remplacement de 'in-time' par 'in-loc' n'a par contre que rarement été remarqué.

4.3.4.4 Conclusion

Dans l'ensemble, il n'y a pas plus d'erreurs d'évaluation pour les paires langue-interlangue que pour les paires anglais-français. La forme de surface de l'interlangue semble donc être à peu près aussi compréhensible que le français ou l'anglais.

Les principales sources d'erreurs sont les phrases comportant des clauses subordonnées. Il est cependant difficile de dire si cela est dû à la non-intuitivité de la représentation ou à la complexité inhérente à ce type de phrase. Une autre difficulté récurrente est liée aux aspects verbaux, particulièrement à l'aspect perfect, qui n'est pas intuitif pour des francophones.

Pour ces deux expériences, une cause d'incertitude non négligeable est la différence d'appréciation de certaines nuances de sens par les sujets. Il aurait été préférable de demander aux sujets d'indiquer les raisons de leurs choix (même sens/pas le même sens) afin d'éliminer ce problème.

Comme mentionné précédemment, il aurait été intéressant d'ajouter une quatrième expérience comportant seulement des paires de phrases anglaises et françaises, dans lesquelles on aurait introduit le même type d'erreurs. On aurait ainsi pu voir si l'interlangue apporte ou non un avantage pour l'évaluation.

5 Conclusion

Dans la partie théorique, nous avons vu l'intérêt de l'approche par interlangue pour les systèmes de traduction automatique multilingues. D'une part, cette approche facilite l'ajout de nouvelles langues au système ; d'autre part, elle permet de produire des paraphrases en langue source, ce qui est particulièrement utile pour la traduction automatique de la parole, où la reconnaissance est potentiellement sujette à erreurs.

Nous avons également vu les avantages apportés par la possibilité de vérifier la bonne formation de l'interlangue pour le développement des règles de mise en correspondance. Puis nous avons parlé de l'utilité de la lisibilité de cette représentation pour l'évaluation des traductions produites par un système, en permettant l'évaluation de paires langue source – interlangue et interlangue – langue cible. Finalement, nous avons décrit comment ces différentes fonctions de l'interlangue sont réalisées dans le système MedSLT.

La partie pratique de ce travail avait pour but de tester la compréhensibilité de l'interlangue MedSLT pour des humains non-spécialistes. La première expérience montre que la forme de surface de l'interlangue est bien plus lisible que la forme structurée. La deuxième expérience nous a permis de voir que la forme de surface de l'interlangue est suffisamment compréhensible pour permettre aux sujets de repérer la plupart des différences de sens. Enfin, la troisième expérience a permis d'observer que l'évaluation langue – interlangue ne semble pas plus sujette à erreurs que l'évaluation langue source – langue cible. Les résultats obtenus sont toutefois faussés par des erreurs dans la conception de ces expériences.

En effet, ce travail nous a permis de voir que notre processus expérimental n'était pas tout à fait bien conçu. D'une part, l'ordre dans lequel les expériences sont soumises aux sujets aura une influence sur les résultats obtenus, puisqu'il est probable que les sujets s'habituent aux concepts et types de phrases proposés, ce dont nous n'avons pas tenu compte. D'autre part, pour les deuxième et troisième expériences, nous avons demandé aux sujets d'indiquer s'ils jugeaient deux phrases comme ayant le même sens. La notion de 'même sens' n'étant pas absolue, certaines nuances semblant importantes pour certains mais négligeables pour d'autres, nous aurions dû demander aux sujets d'indiquer les raisons de leurs choix. Cela aurait permis de différencier les cas réellement faux des cas où il ne s'agit que d'une question d'appréciation.

Les expériences effectuées dans ce travail sont des tests pilotes, les résultats obtenus ne sont pas statistiquement représentatifs. Pour réellement évaluer la compréhensibilité de l'interlangue utilisée dans MedSLT, il faudrait effectuer des expériences à grande échelle.

6 Bibliographie

Arnolds, Douglas ; Balkan, Lorna ; Meijer, Siety ; Humphreys, R. Lee ; Sadler, Louise (1994). *Machine Translation: an Introductory Guide*. NCC Blackwell, London

Bouillon, Pierette ; Clas, André (Eds.) (1993). *La traductique: études et recherches de traduction par ordinateur*. Presses de l'université de Montréal.

Bouillon, Pierette ; Rayner, Manny ; Novellas, Bruna ; Starlander, Marianne ; Santaholma, Marianne ; Nakao, Yukie ; Chatzichrisafis, Nikos (2006). *Une grammaire partagé multi-tâche pour le traitement de la parole : application aux langues romanes*. TAL. Volume 47 – n°3/2006, pages 1 à25

Bouillon, Pierette ; Rayner, Manny ; Chatzichrisafis, Nikos ; Hockey, Beth Ann ; Santaholma, Marianne ; Starlander, Marianne ; Nakao, Yukie ; Kanzaki, Kyoko ; Isahara, Hitoshi (2006). *A Generic multilingual Open Source Platform for Limited-Domain Medical Speech Translation*. Proceedings of the tenth Conference on European Association of Machine Translation, 30-31, May, 2005, Budapest, Hungary, pp. 5-58.

Bouillon, Pierette ; Rayner, Manny ; Novellas, Bruna ; Nakao, Yukie ; Santaholma, Marianne ; Starlander, Marianne ; Chatzichrisafis, Nikos (2006). *Une grammaire multilingue partagée pour la traduction automatique de la parole*. TALN 2006, Leuven

Bouillon, Pierette ; Rayner, Manny ; Starlander, Marianne ; Santaholma, Marianne (2007), *Les ellipses dans un système de Traduction Automatique de la Parole*. TALN 2007, Toulouse.

Bouillon, Pierette ; Flores, Glen ; Starlander, Marianne ; Chatzichrisafis, Nikos ; Santaholma, Marianne ; Tsourakis, Nikos ; Rayner, Manny ; Hockey, Beth Ann (2007). *A Bidirectional Grammar-Based Medical Speech Translator*. Proceedings of workshop on Grammar-based approaches to spoken language processing, ACL 2007, June 29, Prague, Czech Republic, 2007, pp. 41-48.

Bouillon, Pierette ; Chatzichrisafis, Nikos ; Halimi, Sonia ; Hockey, Beth Ann ; Isahara, Hitoshi ; Kanzaki, Kyoko ; Nakao, Yukie ; Novellas Vall, Bruna ; Rayner, Manny ; Santaholma, Marianne ; Starlander, Marianne (2007). *MedSLT : A Multi-Lingual Grammar-Based Medical Speech Translator*. Proceedings of First International Workshop on Intercultural Collaboration, IWIC2007, January 25-26, Kyoto, Japan.

Bouillon, Pierette ; Halimi, Sonia ; Nakao, Yukie ; Kanzaki, Kyoko ; Hitoshi, Isahara ; Tsourakis, Nikos ; Starlander, Marianne ; Hockey, Beth Ann ; Rayner, Manny (2008). *Developing Non-European Translation Pairs in a Medium-Vocabulary Medical Speech Translation System*. Proceedings the Sixth ELRA International Conference on Language Resources and Evaluation (LREC), 2008, Marrakech, Morocco.

Chatzichrisafis, Nikos ; Bouillon, Pierette ; Ranyer, Manny ; Santaholma, Marianne ; Starlander, Marianne ; Hockey, Beth Ann (2006). *Evaluating Task Performance for a Unidirectional Controlled Language Medical Speech Translation System*. Proceedings of First International Workshop on Medical Speech Translation, HLT-NAACL, June 9, 2006, New York.

Defrise, Christine (1993). *Discours et traduction automatique : une approche interlangue basée sur les connaissances*. La Traductique – Etudes et recherches de traduction par ordinateur, eds. P. Bouillon & A. Clas. Montréal: Les Presses de l'Université de Montréal. pp.392-407.

Dillinger, Mike ; Seligman, Mark (2006). *Converser™: Highly Interactive Speech-to-Speech Translation for Healthcare*. Proceedings of the Workshop : Medical Speech Translation, New York

Dutoit, Thierry ; Couvreur, Laurent ; Malfrère, Fabrice ; Pagel, Vincent ; Ris, Christophe (2002). *Synthèse Vocale et Reconnaissance de la Parole : Droites Gauches et Mondes Parallèles*. Actes du 6^{ème} Congrès Français d'Acoustique, Lille.

Ehsani, Farzad ; Kinzey, Jim ; Master, Demetrios ; Sudre, Karen ; Domingo, David ; Park, Hunil (2006). *S-MINDS 2-Way Speech-to-Speech Translation System*. Proceedings of the Workshop : Medical Speech Translation, New York

Estival, Dominique (1993). *Grammaires d'unification et traduction automatique*. La Traductique – Etudes et recherches de traduction par ordinateur, eds. P. Bouillon & A. Clas. Montréal: Les Presses de l'Université de Montréal. pp.212-227.

Gao, Yuqing ; Zhou, Bowen ; Sarika, Ruhi ; Afify, Mohamed ; Kuo, Hong-Kwang ; Zhu, Weizhong ; Deng, Yong-gang ; Prosser, Charles ; Zhang, Wei ; Besacier, Laurent (2006). *IBM MASTOR SYSTEM: Multilingual Automatic Speech-to-Speech Translator*. Proceedings of the Workshop : Medical Speech Translation, New York

Goodman, Kenneth ; Nirenburg, Sergei (eds.) (1991). *The KBMT Project: A Case Study in Knowledge Based Machine Translation*. Morgan Kaufmann, San Mateo, California

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

Jurafsky, Daniel ; Martin, James (2000). *Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition*

Kitano, Hiroaki (1993). *La traduction de la langue parlée*. La Traductique – Etudes et recherches de traduction par ordinateur, eds. P. Bouillon & A. Clas. Montréal: Les Presses de l'Université de Montréal. pp.408-422.

Kittredge, Richard I. (1987). *The significance of Sublanguage for Automatic Translation*. S. Nirenburg (ed.), Machine Translation. Cambridge: Cambridge University Press

Kittredge, Richard I. (2003). *Sublanguages and Controlled Languages*. Chapitre 23 de "The Oxford Handbook of Computational Linguistics". Ruslan Mitkov (ed.), Oxford University Press

Landsbergen, Jan (1989). *The Rosetta Project*. MT Summit II, August 16-18, 1989, Munich, Germany

Lavie, Alon ; Langley, Chad ; Waibel, Alex ; Pianese, Fabio ; Lazzarini Gianni ; Coletti, Paolo ; Taddei, Loredana ; Balducci, Franco (2001). *Architecture and Design Considerations in NESPOLE!: a Speech Translation System for E-commerce Applications*. in Proc. of the HLT2001

L'Homme, Marie-Claude (2001). *Initiation à la traductique*. Montréal : Lingatech

Levin Lori ; Langley, Chad ; Lavie, Alon ; Gates, Donna ; Wallace, Dorcas ; Peterson, Kay (2003). *Domain Specific Speech Acts for Spoken Language Translation*

Levin, Lori ; Gates, Donna ; Wallace, Dorcas ; Peterson, Kay ; Lavie, Alon ; Pianesi, Fabio ; Pianta, Emanuele ; Cattoni, Roldano ; Mana, Nadia (2002). *Balancing Expressiveness and Simplicity in an Interlingua for Task Based Dialogue*.

Mitamura, Teruko ; Nyberg, Eric H. 3rd ; Carbonell, Jaime G. (1991), *An Efficient Interlingua Translation System for Multi-lingual Document Production*. Proceedings of the Third Machine Translation Summit, 1991

Mitamura, Teruko (1999). *Controlled Language for Multilingual Machine Translation*. Proceedings of MT Summit, 1999

Rayner, Manny ; Bouillon, Pierette ; Hockey, Beth Ann ; Chatzichrisafis, Nikos ; Starlander, Marianne (2004). *Comparing Rule-Based and Statistical Approaches to Speech Understanding in a Limited Domain Speech Translation System*. TMI-2004: proceedings of the Tenth Conference on Theoretical and Methodological Issues in Machine Translation, October 4-6, 2004, Baltimore, Maryland, USA; pp.21-29.

Rayner, Manny ; Bouillon, Pierette ; Chatzichrisafis, Nikos ; Hockey, Beth Ann ; Santaholma, Marianne ; Starlander, Marianne ; Isahara, Hitoshi ; Kanzaki, Kyoko ; Nakao, Yukie (2005). *A Methodology for Comparing Grammar-Based and Robust Approaches to Speech Understanding*. Proceedings of Eurospeech-Interspeech, 4-8, September, 2005, Lisboa, Portugal.

Rayner, Manny ; Bouillon, Pierette ; Santaholma, Marianne ; Nakao, Yukie (2005). *Representational and architectural issues in a limited-domain medical speech translator*. Proceedings of TALN/RECITAL, 6-10, June, 2005, Dourdan, France, pp. 163-172

Rayner, Manny ; Bouillon, Pierette ; Hockey, Beth Ann ; Nakao, Yukie ; Starlander, Marianne ; Tsourakis, Nikos ; Kanzaki, Kyoko ; Isahara, Hitoshi (2008). *Using a Semantic Grammar as an Interlingua in a Multilingual Speech Translation System*, TALN 2008, Avignon.

Sérasset, Gilles ; Boitet, Christian (2000) *On UNL as the future « html of the linguistic content » & the reuse of existing NLP components in UNL-related applications with the example of a UNL-French deconverter*. <http://www-clips.imag.fr/projets/unl/> (11.03.08), <http://www.unl.org/unlsys/unl/unl2005/attribute.htm> (11.03.08) <http://www.unl.org/unlsys/unl/unl2005/relation.htm> (11.03.08)

Somers, Harold L. (1992). *La traduction automatique basée sur l'exemple ou sur les corpus*. La Traductique – Etudes et recherches de traduction par ordinateur, eds. P. Bouillon & A. Clas. Montréal: Les Presses de l'Université de Montréal. pp.149-166.

Starlander, Marianne ; Bouillon, Pierrette ; Chatzichrisafis, Nikos ; Santaholma, Marianne ; Rayner, Manny ; Hockey, Beth Ann ; Isahara, Hitoshi ; Kanzaki, Kyoko ; Nakao, Yukie (2005). *Practising Controlled Language through a Help System integrated into the Medical Speech Translation System (MedSLT)* In: Proceedings of the MT Summit X, 12-16 September, 2005, Phuket, Thailand.

Styger, Thomas ; Keller, Eric (1994). *Fundamentals of speech synthesis and speech recognition: basic concepts, state of the art and future challenges*. J.Wiley, Chichester.

Uchida, Hiroshi ; Zhu, Meiying ; Della Senta, Tarcisio G (2005). *UNL Universal Networking Language*. UNDL Foundation

Official Website of the ASD Simplified Technical English Maintenance Group (STEMG), <http://www.asd-ste100.org/INDEX.HTM>, 21.07.08

Annexe 1

Expériences sur l'interlangue MedSLT . explications

Introduction

MedSLT est un système multilingue de traduction automatique de la parole pour le domaine médical conçu pour traiter les dialogues de diagnostic entre médecin et patient. Il permet au médecin de poser des questions dans sa propre langue, questions que le système va traduire dans la langue du patient. La plupart des questions sont du type oui-non, auxquelles le patient peut répondre de façon non-verbale. La thématique des dialogues est limitée à des sous-domaines médicaux, définis par des symptômes tels que les maux de tête, les douleurs abdominales ou les douleurs thoracales.

Pour la traduction, MedSLT repose sur une approche de type interlangue. Le résultat de l'analyse des phrases source n'est donc pas directement transposé dans la langue cible, mais d'abord transposé en interlangue à l'aide de règles de mise en correspondance. Cette représentation interlangue sert de source pour la génération de toutes les langues cibles. L'approche par l'interlangue se prête tout particulièrement aux systèmes multilingues, présentant l'avantage de simplifier l'ajout de nouvelles langues, puisqu'il n'y a pas besoin de modules de transfert pour chaque paire de langues possible dans le système, mais seulement pour chaque paire langue-interlangue.

Évaluation de la compréhensibilité de l'interlangue

La compréhensibilité d'une interlangue à plusieurs intérêts. D'une part, plus l'interlangue sera compréhensible, plus il sera simple de développer les règles de mise en correspondance. D'autre part, une interlangue compréhensible par un humain ouvre de nouvelles perspectives pour l'évaluation de la traduction automatique. En effet, si un locuteur de la langue source peut comprendre l'interlangue, il pourra évaluer la qualité de la représentation qui servira de source à la génération de la langue cible (ou des langues cibles), ce qui permet déjà d'évaluer une moitié du processus de traduction, sans pour autant que ce locuteur ait nécessairement besoin de connaître la ou les langues cibles.

L'interlangue et sa forme de surface

Les expressions interlangue dans MedSLT sont des listes d'attributs-valeurs, avec un niveau d'enchâssement pour le traitement des clauses subordonnées. Les attributs sont des concepts sémantiques (p.ex. 'body_part', 'symptom', 'action', etc) et les valeurs sont des mots anglais (p.ex. 'eye', 'pain', 'read', etc). Pour former des phrases, ces éléments sont concaténés dans l'ordre alphabétique des attributs.

La phrase « où la douleur irradie t elle ? »
aura la représentation interlangue suivante:

```
[[loc,where], [state,radiate], [symptom,pain], [tense,present], [utterance_type,whq], [voice,active]]
```

Afin de rendre l'interlangue plus compréhensible, une forme de surface a été créée. Cette forme de surface ne comporte que les valeurs, arrangées dans un ordre destiné à simplifier la compréhension:

```
'WH-QUESTION pain radiate where PRESENT ACTIVE'
```

La première expérience utilise les deux formes de l'interlangue, la deuxième et la troisième expérience ne portent que sur la forme de surface.

Première expérience

Dans cette expérience, nous voulons dans une première partie voir si les concepts sont intuitifs pour des personnes comprenant l'anglais, donc si les expressions interlangue sont suffisamment explicites pour qu'un humain puisse les utiliser comme base pour produire une phrase en français ou en anglais. Dans une deuxième partie, nous allons voir si la forme de surface est plus compréhensible que l'interlangue. Il s'agira donc de traduire deux fois 13 expressions vers l'anglais ou le français.

Pour la première partie de cette expérience nous avons choisi de ne pas fournir d'explications par rapport à la signification des différents concepts, un des but de l'expérience étant de voir si ces concepts sont intuitifs. Nous avons toutefois fait deux exceptions, pour les concepts suivants, qui ne nous semblaient pas explicites:

- aspect

Le concept 'aspect' peut prendre la valeur *perfect*. Le *perfect* est un aspect qui reflète le point de vue de l'énonciateur, dénotant une action achevée au moment de l'énonciation ou un état ou action ayant commencé dans le passé et se poursuivant dans le présent. Le *present perfect* est formé de l'auxiliaire *have/has* et d'un participe passé. Par exemple:

'how long have you had the headaches'
'depuis combien de temps avez-vous vos maux de tête'

Le *present perfect* s'oppose au *simple present*, qui décrit un état de fait, par exemple une action qui se répète ou qui est habituelle:

'how long do you have the headaches'
'pendant combien de temps avez-vous vos maux de tête'

- utterance_type

Ce concept définit le type de phrase et peut avoir comme valeur les abréviations suivantes:

valeur	définition
dcl	phrase déclarative
whq	WH-QUESTION : phrase interrogative qui en anglais commencerait par un adverbe en 'wh', par exemple 'what', 'where', 'when', etc.
ynq	YN-QUESTION : phrase interrogative demandant une réponse de type oui/non (non verbale)

Deuxième expérience

Dans cette expérience, nous voulons voir s'il est facile de déceler des erreurs de traduction dans des paires de phrases langue naturelle – interlangue. Il s'agira donc d'indiquer pour une paire de phrases si elles expriment le même sens ou pas.

Troisième expérience

Comme pour la deuxième expérience, il s'agira de trouver les erreurs de sens, cette fois entre trois phrases : une phrase en anglais, une phrase en interlangue et une phrase en français.

Exemples interlangue – français/anglais

Voici quelques exemples d'expressions interlangue avec les phrases correspondantes en anglais et en français. A noter que des phrases exprimant le même sens sont transposées dans la même représentation interlangue, même si elles recourent à des moyens syntaxiques différents pour exprimer le sens. Pour chaque expression interlangue, plusieurs phrases sont donc possibles en langue naturelle :

interlangue: [[adj,crushing], [symptom,headache], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]

surface: 'YN-QUESTION crushing headache be PRESENT ACTIVE'

en: 'are the headaches crushing'

fr: 'vos maux de tête sont-ils constrictifs',

interlangue: [[action,decrease], [degree,frequency], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

surface: 'YN-QUESTION frequency decrease headache PRESENT ACTIVE'

en: 'are the headaches becoming less frequent'

en: 'are the headaches occurring less frequently'

en: 'are the headaches occurring less often'

fr: 'la fréquence de vos maux de tête diminue t elle'

fr: 'la fréquence des maux de tête diminue t elle'

interlangue : [[frequency,time], [prep,duration], [prep,frequency], [pronoun,you], [spec,[at_least,1]], [state,have_symptom], [symptom,pain], [tense,present], [timeunit,day], [utterance_type,ynq], [voice,active]]

surface: 'YN-QUESTION you have pain at-least one times per day PRESENT ACTIVE'

en: 'does the pain occur at least once a day'

en: 'do you experience the pain at least once a day'

fr: 'avez-vous mal au moins une fois par jour'

interlangue: [[clause, [[action,stand_up], [pronoun,you], [tense,present], [utterance_type,dcl], [voice,active]]], [pronoun,you], [sc,when], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

surface: 'YN-QUESTION you have headache sc-when [you stand-up PRESENT ACTIVE] PRESENT ACTIVE'

en: 'do you have the headaches when you stand up'

fr: 'avez-vous vos maux de tête quand vous vous relevez'

Glossaire

Puisque l'évaluation ne devrait pas être influencée par des problèmes de type terminologique, nous ajoutons ici un petit glossaire des termes utilisés dans les phrases d'évaluation.

anglais	français
to radiate	irradier
attack	crise
family history	antécédent familial
high blood pressure	hypertension
muscle ache	douleur musculaire
drowsiness	somnolence

Interlangue MedSLT . première expérience

Informations sur l'évaluateur

Nom et prénom :

Langue maternelle :

Niveau en français :

Pour chaque expression, écrire la phrase correspondante en français. Le but est de faire des phrases simples, sans accorder particulièrement d'importance à la forme (donc pas besoin de chercher longtemps pour trouver une formulation particulièrement idiomatique ou élégante). Puisque nous cherchons à évaluer la compréhensibilité, l'évaluateur ne devrait pas passer beaucoup de temps à trouver le sens des expressions. Si une expression interlangue est incompréhensible, indiquer quel concept / quelle combinaison de concepts est source de difficultés et passer à l'expression suivante.

Dans une première partie, les expressions seront sous la forme brute (listes d'attributs-valeurs), dans une deuxième partie les expressions seront sous la forme de surface. Pour que nous puissions comparer ces deux parties, merci d'indiquer le temps passé pour chaque partie.

Première partie

1.

[[adj,diffuse], [symptom,pain], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]

2.

[[part,side], [prep,in_loc], [spec,both], [symptom,headache], [tense,present],[utterance_type,ynq], [verb,be], [voice,active]]

3.

[[adj,left], [body_part,head], [freq,always], [part,side], [prep,in_loc], [pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

4.

[[loc,where], [symptom,pain], [tense,present], [utterance_type,whq], [verb,be], [voice,active]]

5.

[[adj,right], [body_part,arm], [state,radiate], [symptom,pain], [tense,present], [utterance_type,ynq], [voice,active]]

6.

[[event,become_worse], [itemizer,attacks], [symptom,pain], [tense,present], [utterance_type,ynq], [voice,active]]

7.

[[manner,gradually], [state,start], [symptom,headache], [tense,past], [utterance_type,ynq], [voice,active]]

8.

[[history,family_history], [pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

9.

[[frequency,time], [prep,duration], [prep,frequency], [pronoun,you], [spec,1],[state,have_symptom], [symptom,headache], [tense,present], [timeunit,month], [utterance_type,ynq], [voice,active]]

10.

[[aspect,perfect], [prep,duration], [pronoun,you], [spec,[more_than,2]], [state,have_symptom], [symptom,headache], [tense,present], [timeunit,week], [utterance_type,ynq], [voice,active]]

11.

[[action,cough],
[clause,
[[pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,dcl],
[voice,active]]],
[pronoun,you], [sc,when], [tense,present], [utterance_type,ynq],[voice,active]]

12.

[[clause,
[[pronoun,you], [secondary_symptom,trauma], [state,experience], [tense,present], [utterance_type,dcl],
[voice,active]]],
[pronoun,you], [sc,after], [state,have_symptom], [symptom,pain], [tense,past],[utterance_type,ynq],
[voice,active]]

13.

[[clause,
[[pronoun,you], [state,have_symptom], [symptom,pain], [tense,present], [utterance_type,dcl], [voice,active]]],
[pronoun,you], [sc,before], [secondary_symptom,muscle_aches], [state,experience], [tense,present],
[utterance_type,ynq], [voice,active]]

Temps nécessaire au traitement des 13 expressions interlangue :

Deuxième partie

14.

'YN-QUESTION continuous pain be PRESENT ACTIVE'

Annexe 1

15.

'YN-QUESTION headache be in-loc one side-part PRESENT ACTIVE'

16.

'YN-QUESTION you have pain in-loc head back-part ever PRESENT ACTIVE'

17.

'WH-QUESTION pain last how-long PRESENT ACTIVE'

18.

'YN-QUESTION pain radiate shoulder left PRESENT ACTIVE'

19.

'YN-QUESTION attacks become-better PRESENT ACTIVE'

20.

'YN-QUESTION you experience nausea PAST ACTIVE'

21.

'YN-QUESTION you have family-history of high-blood-pressure PRESENT ACTIVE'

22.

'YN-QUESTION you have attacks two times per month PRESENT ACTIVE'

23.

'YN-QUESTION you have attacks at-least three times per month PRESENT ACTIVE'

24.

'YN-QUESTION you experience pain sc-when [you move-suddenly head PRESENT ACTIVE] PRESENT ACTIVE'

25.

'YN-QUESTION you have pain sc-after [you eat large-meal PRESENT ACTIVE] PRESENT ACTIVE'

26.

'YN-QUESTION you experience nausea sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE'

Temps nécessaire au traitement des 13 expressions surface :

Interlangue MedSLT . deuxième expérience . français

Informations sur l'évaluateur

Nom :

Prénom :

Dans cette expérience, il s'agira d'indiquer pour une paire de phrases interlangue-français si les deux phrases expriment bien le même sens. Des erreurs ont délibérément été introduites dans certaines des paires.

N.B. Sous "même sens" on entend "exactement le même sens", c'est à dire que s'il y a une nuance, même faible (pas le même temps verbal, omission d'une précision, etc.), le sens n'est pas considéré comme étant le même.

Comme pour la première expérience, merci d'indiquer le temps nécessaire au traitement de ces phrases. N'hésitez pas à ajouter des commentaires pour les cas ambigus.

1 interlangue : 'WH-QUESTION you have headache when usually PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'quand avez-vous vos maux de tête'

commentaire :

2 interlangue : 'YN-QUESTION deep pain be often PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'la douleur est-elle habituellement profonde'

commentaire :

3 interlangue : 'WH-QUESTION you have headache how-long PRESENT-PERFECT ACTIVE'

↑
 même sens
 ↓

fr : 'depuis combien de temps avez vous mal à la tête'

commentaire :

4 interlangue : 'YN-QUESTION intense-pain be PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'la douleur a-t-elle été intense'

commentaire :

5 interlangue : 'YN-QUESTION headache be in-loc both side-part PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'vos maux de tête sont-ils des deux côtés'

commentaire :

6 interlangue : 'YN-QUESTION headache last duration a-few day PRESENT ACTIVE'

↑
 même sens
↓

fr : 'la douleur dure-t-elle plusieurs jours'

commentaire :

7 interlangue : 'YN-QUESTION you have pain sc-when [you experience fever PRESENT ACTIVE] PRESENT ACTIVE'

↑
 même sens
↓

fr : 'avez-vous de la fièvre quand vous avez mal'

commentaire :

8 interlangue : 'WH-QUESTION attacks last how-long usually PRESENT ACTIVE'

↑
 même sens
↓

fr : 'combien de temps ces crises durent elles habituellement'

commentaire :

9 interlangue : 'WH-QUESTION you take what-medication regularly PRESENT ACTIVE'

↑
 même sens
↓

fr : 'quels médicaments prenez vous régulièrement'

commentaire :

10 interlangue : 'YN-QUESTION you have pain duration several day PRESENT ACTIVE'

↑
 même sens
↓

fr : 'avez vous mal depuis plusieurs jours'

commentaire :

11 interlangue : 'YN-QUESTION you have headache in-loc side-part PRESENT ACTIVE'

↑
 même sens
↓

fr : 'avez-vous vos maux de tête sur l'un des côtés'

commentaire :

12 interlangue : 'YN-QUESTION you experience blurred-vision sc-when [you have pain PRESENT ACTIVE] PRESENT ACTIVE'

↑
 même sens
↓

fr : 'la douleur est-elle précédée par des troubles de vision'

commentaire :

13 interlangue : 'YN-QUESTION headache be in-loc head side-part left always PRESENT ACTIVE'

↑
 même sens
↓

fr : 'vos maux de tête sont-ils sur le côté gauche'

commentaire :

14 interlangue : 'WH-QUESTION what is daily-physical-activity PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'quelle est votre activité physique quotidienne'

commentaire :

15 interlangue : 'YN-QUESTION you have headache in-time morning ever PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'avez vous déjà eu vos maux de tête le matin'

commentaire :

16 interlangue : 'WH-QUESTION you have attacks how-long PRESENT-PERFECT ACTIVE'

↑
 même sens
 ↓

fr : 'pendant combien de temps avez-vous ces crises'

commentaire :

17 interlangue : 'YN-QUESTION headache be in-loc head one side-part PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'vos maux de tête sont-ils sur le côté'

commentaire :

18 interlangue : 'YN-QUESTION you smoke PAST ACTIVE'

↑
 même sens
 ↓

fr : 'fumez-vous'

commentaire :

19 interlangue : 'YN-QUESTION you have headache in-time morning PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'vos maux de tête surviennent-ils tôt le matin'

commentaire :

20 interlangue : 'WH-QUESTION you have headache how-long PAST ACTIVE'

↑
 même sens
 ↓

fr : 'pendant combien de temps avez-vous vos maux de tête'

commentaire :

21 interlangue : 'YN-QUESTION headache last duration a-few week PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'vos maux de tête durent-ils quelques semaines'

commentaire :

22 interlangue : 'YN-QUESTION bursting pain be PAST ACTIVE'

↑
 même sens
 ↓

fr : 'la douleur a-t-elle été éclatante'

commentaire :

23 interlangue : 'YN-QUESTION headache last duration a-few minute PRESENT ACTIVE'

↑
 même sens
↓

fr : 'vos maux de tête durent-ils quelques minutes'

commentaire :

24 interlangue : 'YN-QUESTION you have pain sc-when [you lie-down PRESENT ACTIVE] PAST ACTIVE'

↑
 même sens
↓

fr : 'avez-vous mal quand vous vous allongez'

commentaire :

25 interlangue : 'YN-QUESTION gradual pain start PAST ACTIVE'

↑
 même sens
↓

fr : 'la douleur a t elle commencé progressivement'

commentaire :

26 interlangue : 'YN-QUESTION you experience depression sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE'

↑
 même sens
↓

fr : 'avez-vous vos maux de tête quand vous êtes dépressif'

commentaire :

27 interlangue : 'YN-QUESTION pain be in-loc head front-part sometimes PRESENT ACTIVE'

↑
 même sens
↓

fr : 'la douleur est elle habituellement sur le devant de la tête'

commentaire :

28 interlangue : 'YN-QUESTION headache be in-loc head side-part left PRESENT ACTIVE'

↑
 même sens
↓

fr : 'vos maux de tête sont-ils sur le côté gauche'

commentaire :

29 interlangue : 'WH-QUESTION pain last how-long PAST ACTIVE'

↑
 même sens
↓

fr : 'combien de temps la douleur a-t-elle duré'

commentaire :

30 interlangue : 'YN-QUESTION pain be in-loc head side-part PRESENT ACTIVE'

↑
 même sens
↓

fr : 'la douleur est elle temporaire'

commentaire :

31 interlangue : 'YN-QUESTION you have headache duration more-than two week PRESENT-PERFECT ACTIVE'

↑
 même sens
 ↓

fr : 'avez vous vos maux de tête depuis plus de deux semaines'

commentaire :

32 interlangue : 'YN-QUESTION pain be in-loc head front-part ever PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'avez-vous parfois mal sur le devant de la tête'

commentaire :

33 interlangue : 'WH-QUESTION you have attacks how-often PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'quelle est la fréquence de ces crises'

commentaire :

34 interlangue : 'WH-QUESTION you have pain when usually PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'quand la douleur survient elle habituellement'

commentaire :

35 interlangue : 'YN-QUESTION you have pain sc-when [you experience anger PRESENT ACTIVE] PAST'

↑
 même sens
 ↓

fr : 'la douleur a t elle été causée par la colère'

commentaire :

36 interlangue : 'YN-QUESTION you experience drowsiness sc-before [you have headache PRESENT ACTIVE] PAST ACTIVE'

↑
 même sens
 ↓

fr : 'vos maux de tête ont-ils été précédés d'un état de somnolence'

commentaire :

37 interlangue : 'YN-QUESTION headache last duration a-few minute usually PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'vos maux de tête durent-ils habituellement quelques minutes'

commentaire :

38 interlangue : 'YN-QUESTION pain be in-time side-part always PRESENT ACTIVE'

↑
 même sens
 ↓

fr : 'avez vous toujours mal sur le côté'

commentaire :

39 interlangue : 'YN-QUESTION you experience drowsiness sc-before [you have headache PRESENT ACTIVE] PAST ACTIVE'

↑
 même sens
↓

fr : 'vos maux de tête ont-ils été précédés d'un état de somnolence'

commentaire :

40 interlangue : 'YN-QUESTION headache become-worse sc-when [read CONTINUOUS ACTIVE] PRESENT ACTIVE'

↑
 même sens
↓

fr : 'vos maux de tête empirent-ils quand vous lisez'

commentaire :

41 interlangue : 'YN-QUESTION headache become-better sc-when [you experience massage PRESENT ACTIVE] PRESENT ACTIVE'

↑
 même sens
↓

fr : 'vos maux de tête sont-ils habituellement soulagés par des massages'

commentaire :

temps nécessaire :

Interlangue MedSLT . troisième expérience

Informations sur l'évaluateur

Nom :

Prénom :

Dans cette expérience, des erreurs ont été introduites dans des triplets anglais – interlangue – français. Ces erreurs peuvent être entre l'anglais et l'interlangue et entre l'interlangue et le français. Il est aussi possible que les deux étapes soient fausses. Il s'agira d'indiquer si les paires correspondent (anglais – interlangue, interlangue – français et anglais – français). Lorsque deux phrases ont le même sens, cocher la case.

N.B. Sous “même sens” on entend “exactement le même sens”, c’est à dire que s’il y a une nuance, même faible (pas le même temps verbal, omission d’une précision, etc.), le sens n’est pas considéré comme étant le même.

Comme pour les autres expériences, merci d'indiquer le temps nécessaire au traitement de ces phrases. N’hésitez pas à ajouter des commentaires pour les cas ambigus.

1 en : 'did the headaches start gradually'

↑
 même sens

↓
 même sens

interlangue : 'YN-QUESTION headache start gradually PAST ACTIVE'

↖
 même sens

fr : 'vos maux de tête ont-ils commencé soudainement'

commentaire :

2 en : 'did you experience the pain for more than ten days'

↑
 même sens

↓
 même sens

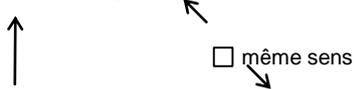
interlangue : 'YN-QUESTION you have pain duration more-than ten day PAST ACTIVE'

↖
 même sens

fr : 'avez-vous mal depuis plus de dix jours'

commentaire :

3 en : 'did you feel weak before you had the headaches'



même sens

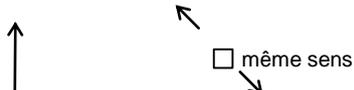
interlangue : 'YN-QUESTION you experience weakness sc-before [you have headache PRESENT ACTIVE] PAST ACTIVE'



fr : 'vos maux de tête sont ils précédés d'une impression de faiblesse'

commentaire :

4 en : 'are the headaches in the back of the head'



même sens

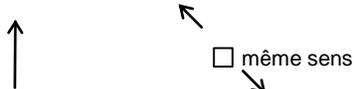
interlangue : 'YN-QUESTION headache be in-loc head back-part PRESENT ACTIVE'



fr : 'vos maux de tête sont-ils à l'arrière de la tête'

commentaire :

5 en : 'are the headaches becoming less severe'



même sens

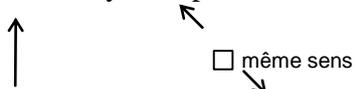
interlangue : 'YN-QUESTION headache become-better PRESENT ACTIVE'



fr : 'l'intensité de vos maux de tête augmente t elle'

commentaire :

6 en : 'do you experience the pain when you feel stressed'



même sens

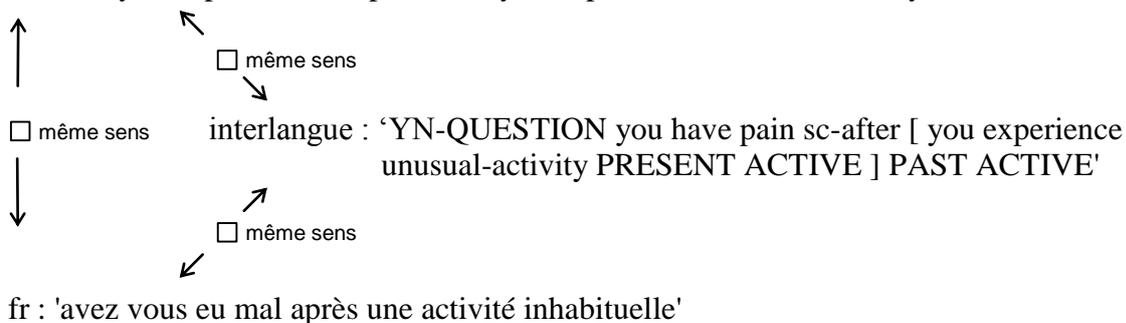
interlangue : 'YN-QUESTION you have pain sc-when [you experience stress PRESENT ACTIVE] PRESENT ACTIVE'



fr : 'la douleur empire-t-elle quand vous êtes stressé'

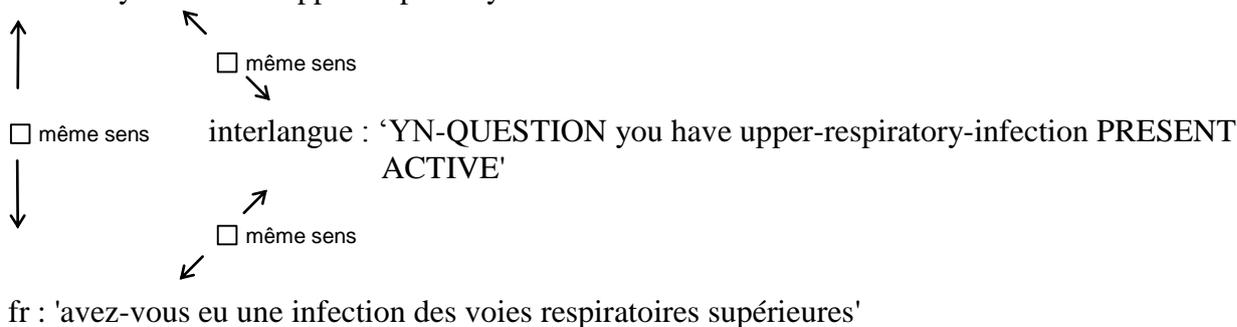
commentaire :

7 en : 'did you experience the pain after you experienced unusual activity'



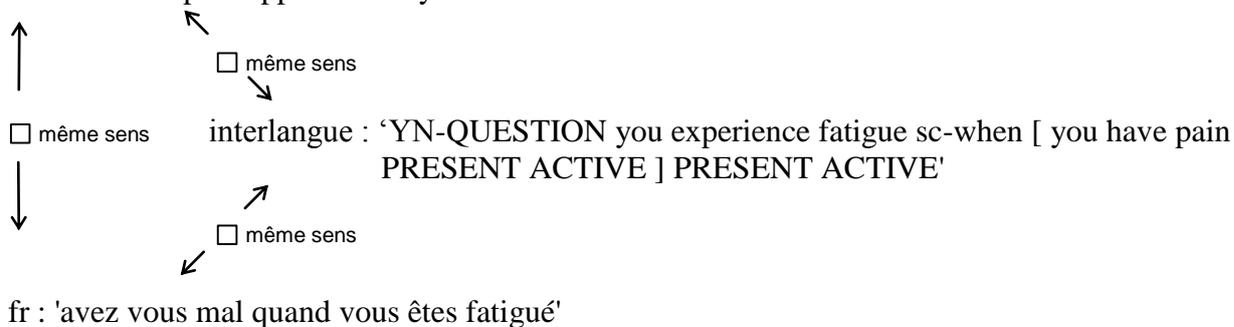
commentaire :

8 en : 'did you have an upper respiratory infection'



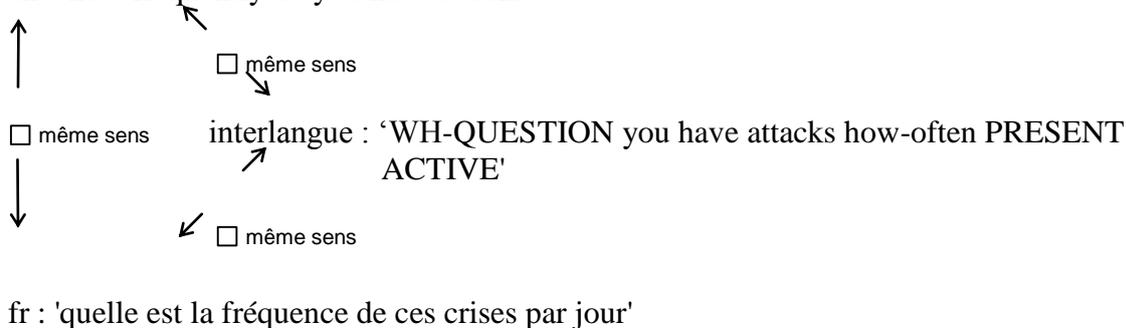
commentaire :

9 en : 'does the pain appear when you are tired'



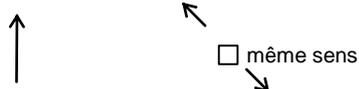
commentaire :

10 en : 'how frequently do you have attacks'



commentaire :

11 en : 'does the pain usually last for more than one hour'



même sens

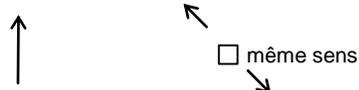
interlangue : 'YN-QUESTION pain last duration more-than one hour usually PRESENT ACTIVE'



fr : 'la douleur dure-t-elle habituellement une heure'

commentaire :

12 en : 'does bright light ever make your headaches worse'



même sens

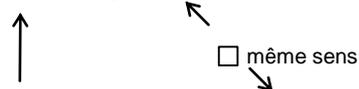
interlangue : 'YN-QUESTION headache become-worse sc-when [you experience bright-light PRESENT ACTIVE] PRESENT ACTIVE'



fr : 'vos maux de tête sont ils aggravés par une lumière forte'

commentaire :

13 en : 'do you ever have the headaches at night'



même sens

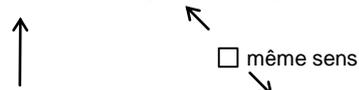
interlangue : 'YN-QUESTION you have headache in-time night ever PRESENT ACTIVE'



fr : 'vos maux de tête surviennent-ils souvent la nuit'

commentaire :

14 en : 'is the pain always on the same side of the head'



même sens

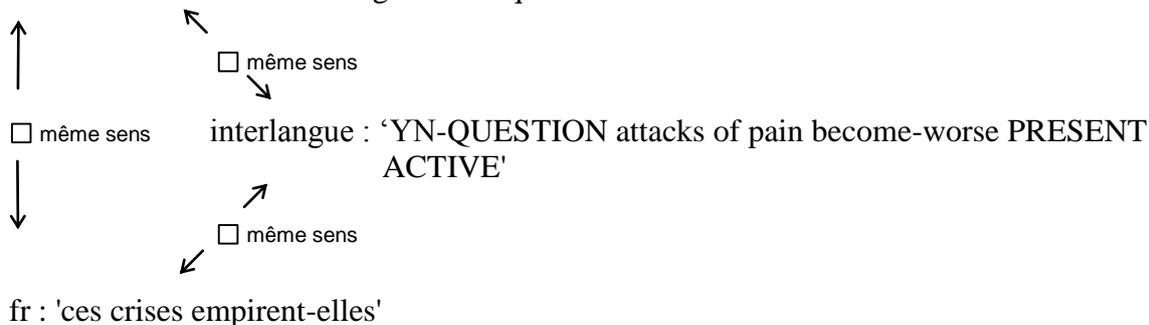
interlangue : 'YN-QUESTION you have pain in-loc head the-same side-part always PRESENT ACTIVE'



fr : 'avez-vous toujours mal du même côté'

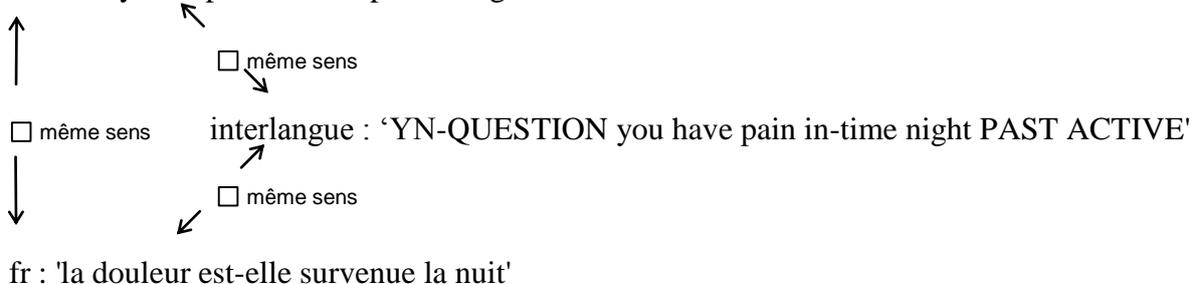
commentaire :

15 en : 'are the attacks becoming more frequent'



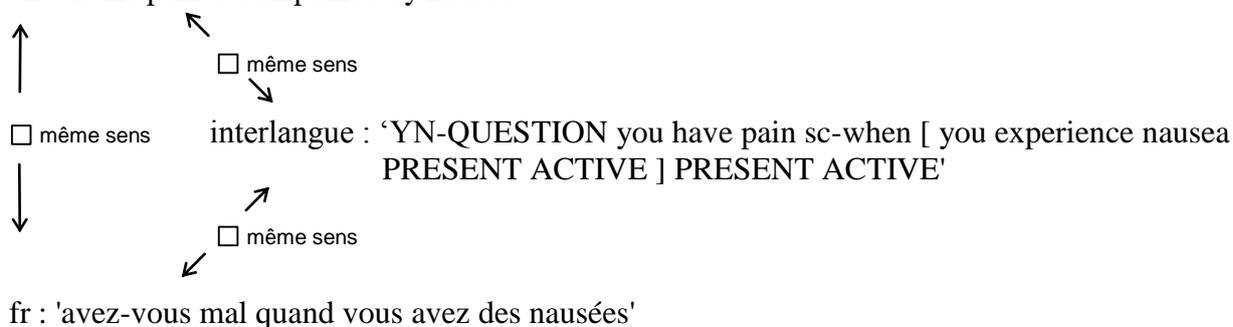
commentaire :

16 en : 'did you experience the pain at night'



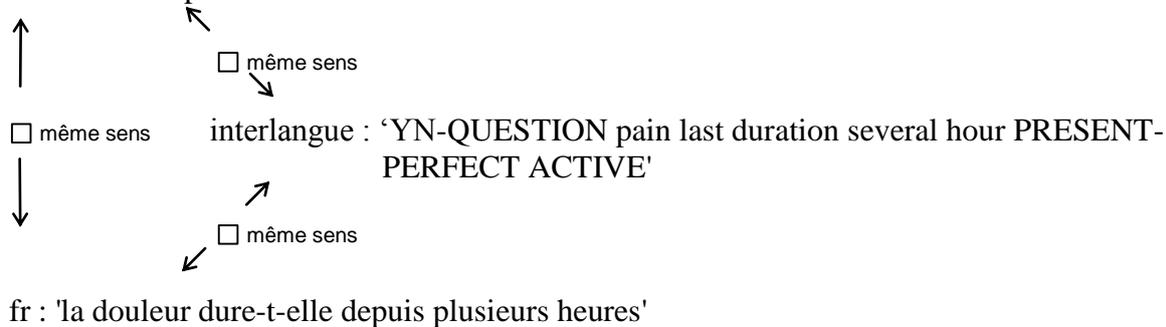
commentaire :

17 en : 'is the pain accompanied by nausea'



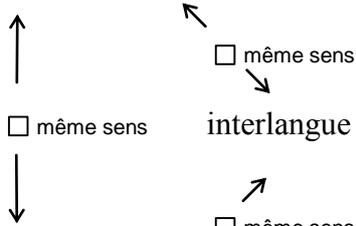
commentaire :

18 en : 'does the pain last for several hours'



commentaire :

19 en : 'do you have the headaches after you drank red wine'

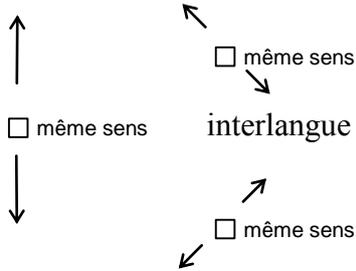


interlangue : 'YN-QUESTION you have headache sc-after [you drink red-wine PAST ACTIVE] PRESENT ACTIVE'

fr : 'avez-vous vos maux de tête quand vous avez bu du vin rouge'

commentaire :

20 en : 'do the headaches sometimes last for more than two hours'

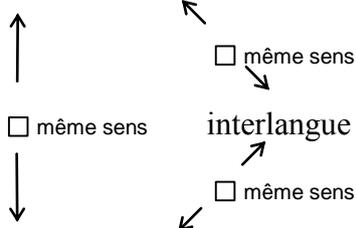


interlangue : 'YN-QUESTION headache last duration more-than two hour PRESENT ACTIVE'

fr : 'vos maux de tête durent-ils parfois plus de deux heures'

commentaire :

21 en : 'do you take medication'

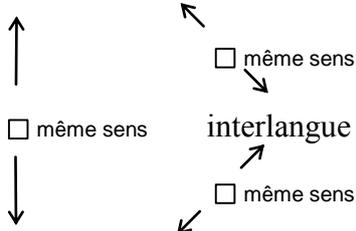


interlangue : 'WH-QUESTION you take what-medication PRESENT ACTIVE'

fr : 'quels médicaments prenez-vous'

commentaire :

22 en : 'did you have a sudden headache'



interlangue : 'YN-QUESTION you have sudden headache PAST ACTIVE'

fr : 'avez-vous soudainement vos maux de tête'

commentaire :

23 en : 'have you had the headaches for more than one week'

↑
□ même sens

□ même sens
↓

↖ □ même sens
↘

interlangue : 'YN-QUESTION you have headache duration more-than one week PRESENT-PERFECT ACTIVE'

↗ □ même sens
↙

fr : 'avez vous vos maux de tête depuis plus d'une semaine'

commentaire :

24 en : 'do you feel stressed before you have the headaches'

↑
□ même sens

□ même sens
↓

↖ □ même sens
↘

interlangue : 'YN-QUESTION you experience stress sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE'

↗ □ même sens
↙

fr : 'avez vous vos maux de tête quand vous êtes stressé'

commentaire :

25 en : 'do the headaches usually last for a few seconds'

↑
□ même sens

□ même sens
↓

↖ □ même sens
↘

interlangue : 'YN-QUESTION headache last duration a-few second usually PRESENT ACTIVE'

↗ □ même sens
↙

fr : 'vos maux de tête durent-ils habituellement quelques secondes'

commentaire :

26 en : 'do you sometimes have the headaches in the afternoon'

↑
□ même sens

□ même sens
↓

↖ □ même sens
↘

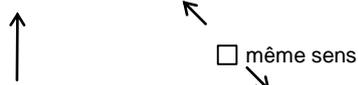
interlangue : 'YN-QUESTION you have headache in-loc afternoon sometimes PRESENT ACTIVE'

↗ □ même sens
↙

fr : 'avez vous parfois vos maux de tête l'après-midi'

commentaire :

27 en : 'are you sensitive to bright light when you experience the pain'



même sens



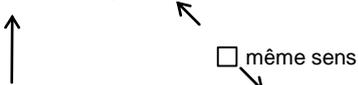
interlangue : 'YN-QUESTION you experience photophobia sc-when [you have pain PRESENT ACTIVE] PRESENT ACTIVE'



fr : 'la douleur est-elle précédée d'une photophobie'

commentaire :

28 en : 'do you have fever'



même sens



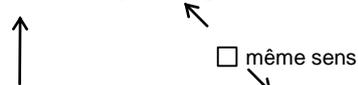
interlangue : 'YN-QUESTION you have fever PRESENT ACTIVE'



fr : 'est-ce que vous êtes fiévreux'

commentaire :

29 en : 'did you experience the pain more often'



même sens



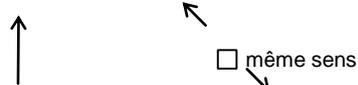
interlangue : 'YN-QUESTION you have pain more-often PRESENT ACTIVE'



fr : 'avez-vous plus fréquemment la douleur'

commentaire :

30 en : 'were the headaches better'



même sens



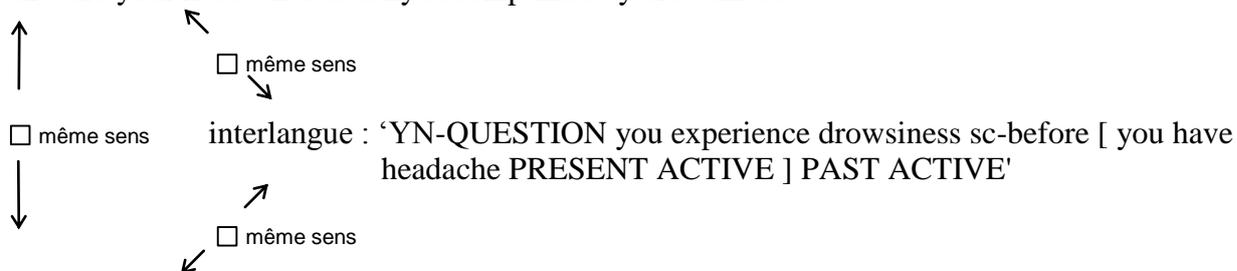
interlangue : 'YN-QUESTION headache become-better PAST ACTIVE'



fr : 'vos maux de tête ont-ils diminué'

commentaire :

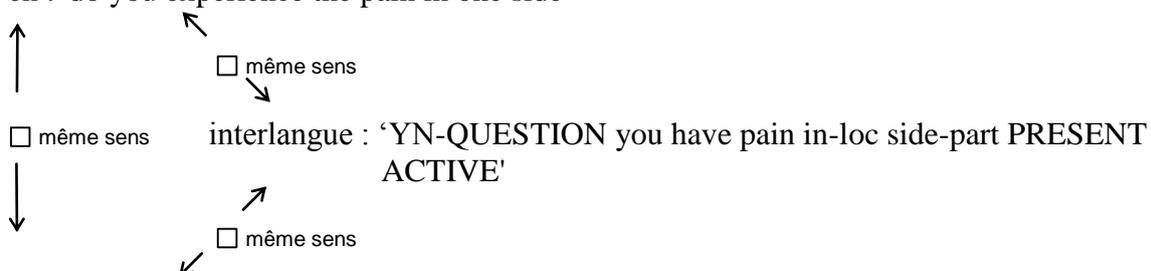
31 en : 'are your headaches usually accompanied by drowsiness'



fr : 'vos maux de tête ont-ils été précédés d'un état de somnolence'

commentaire :

32 en : 'do you experience the pain in one side'



fr : 'avez-vous mal sur le côté'

commentaire :

temps nécessaire au traitement des 32 phrases :

Annexe 2 : résultats complets de la première expérience

Les commentaires des sujets sont reproduits entre parenthèses carrées.

Les phrases jugées fausses sont en italique.

première partie

1. **[[adj,diffuse], [symptom,pain], [tense,present], [utterance_type,ynq], [verb,be], [voice,active]]**

from(fre)-'la douleur est elle diffuse'

to(eng)-'is the pain diffuse'

to(fre)-'la douleur est-elle diffuse'

- (a) *Avez-vous une douleur générale?*
 (b) Est-ce que la douleur est diffuse ?
 (c) La douleur est-elle diffuse?
 (d) *Does the pain diffuse?*
 (e) *Est-ce que la douleur ce propage?*
 (f) Is the pain diffuse?
 (g) Est-ce que la douleur est diffuse ?
 (h) est-ce qu'il y a une diffuse douleur ?
 (i) is the pain diffuse?
 (j) Est-ce que la douleur est difuse ?

erreurs : 3

2. **[[part,side], [prep,in_loc], [spec,both], [symptom,headache], [tense,present],[utterance_type,ynq], [verb,be], [voice,active]]**

to(eng)-'are the headaches in both sides'

to(fre)-'vos maux de tête sont-ils des deux côtés'

- (a) Avez-vous des maux de tête des deux côtés?
 (b) *Est-ce que le mal de tête est de côté ?*
 (c) Est-ce que les maux de tête sont présents des deux côtés?
 (d) Is the headache present in both side?
 (e) Est-ce que vous avez les maux de têtes des deux côtés ?
 (f) Is the headache on both sides?
 (g) Est-ce que vous avez mal des deux côtés de la tête ?
 (h) est-ce que les maux de tête se trouvent sur les deux côtés ?
 (i) Is the headache present on both sides?
 (j) Avez-vous des maux de tête des deux côtés ?

erreurs : 1

3. **[[adj,left], [body_part,head], [freq,always], [part,side], [prep,in_loc], [pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]],**

to(eng)-'do you always have the headaches in the left side of the head'

to(fre)-'avez-vous toujours vos maux de tête sur le côté gauche'

- (a) les maux de tête sont-ils toujours à la partie gauche de la tête?

- (b) Est-ce que vous avez toujours mal au côté gauche de la tête ?
 - (c) Est-ce que vous avez toujours mal du côté gauche de la tête?
 - (d) Do you always experience headache in the left side of your head?
 - (e) Est-ce que vos symptômes de maux de têtes sont toujours du côté gauche de votre tête?
 - (f) Is your headache always on the left side?
 - (g) Est-ce que vous avez toujours mal à la tête du côté gauche ?
 - (h) est-ce que vous avez les maux de tête toujours à gauche de la tête ?
 - (i) Do you always have the headache on the left side of the head? [Do you always experience..., Does the headache always occur on the left side?]
 - (j) Avez-vous toujours mal à la tête du côté gauche ?
- erreurs : 0

4. [[loc,where], [symptom,pain], [tense,present], [utterance_type,whq], [verb,be], [voice,active]]

from(eng)-'where does it hurt', from(eng)-'where is the pain',
from(eng)-'where is your pain',
from(fre)-'où avez vous mal

- (a) Où avez-vous mal?
 - (b) Où est la douleur ?
 - (c) Où avez-vous mal?
 - (d) Where do you feel pain?
 - (e) Où est-ce que vous avez mal?
 - (f) Where is the pain?
 - (g) Où avez-vous mal ?
 - (h) la douleur se trouve où ?
 - (i) Where is the pain?
 - (j) Où avez vous mal ?
- erreurs : 0

5. [[adj,right], [body_part,arm], [state,radiate], [symptom,pain], [tense,present], [utterance_type,ynq], [voice,active]]

from(fre)-'la douleur irradie t elle le bras droit'
to(eng)-'does the pain radiate to the right arm'
to(fre)-'la douleur irradie-t-elle le bras droit'

- (a) *Avez-vous mal au bras droit?*
 - (b) Est-ce que la douleur irradie dans le bras droite ?
 - (c) La douleur irradie-t-elle le bras droit?
 - (d) *Does the pain radiate from the right arm?* [il manque des infos pour savoir si ça radiate to ou from]
 - (e) *Votre bras droit irradié, fait-il mal?*
 - (f) Does the pain radiate to your right arm?
 - (g) *La douleur irradie-t-elle du bras gauche ?*
 - (h) *est-ce que la douleur irradie du bras droit?*
 - (i) *Is the pain radiating* [unclear: from? down? up? the right arm?]
 - (j) la douleur irradie-t-elle votre bras droit ?
- erreurs : 6

6. [[event,become_worse], [itemizer,attacks], [symptom,pain], [tense,present],

[utterance_type,ynq], [voice,active]]

from(fre)-'l intensité de ces crises augmente t elle'

to(eng)-'is the pain worse'

to(fre)-'ces crises empirent-elles'

- (a) *Avez-vous plus mal lors des crises?*
 (b) Est-ce que les crises de douleur s'aggravent ? [mais je ne sais pas ce qu'est un itemizer]
 (c) Avez-vous des crises de douleur plus aïgues?
 (d) Do the pain attacks become worse?
 (e) *Lors d'une crise, la douleur s'aggrave t'elle?*
 (f) Is the pain worsening?
 (g) Les crises deviennent-elle de plus en plus douloureuse ?
 (h) est-ce que les crises deviennent pires?
 (i) Are the pain attacks getting worse?
 (j) vos crises de douleur empire-t-elle?
 erreurs : 2

7. [[manner,gradually], [state,start], [symptom,headache], [tense,past], [utterance_type,ynq], [voice,active]]

to(eng)-'did the headaches start gradually'

to(fre)-'vos maux de tête ont-ils commencé progressivement'

- (a) *Vos maux de tête commencent-ils progressivement?*
 (b) Est-ce que le mal de tête a commencé graduellement ?
 (c) Les maux de tête sont-ils apparus graduellement?
 (d) Did the headache started gradually?
 (e) *Depuis que vos maux de têtes ont commencés, est ce que ça c'est empiré?*
 (f) Did the headache start gradually?
 (g) Vos maux de têtes ont-ils débuté graduellement ?
 (h) est-ce que les maux de tête ont commencé peu à peu ?
 (i) Did the headaches begin gradually?
 (j) *vos maux de tête commencent graduellement ?*
 erreurs : 3

8. [[history,family_history], [pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,ynq], [voice,active]]

to(eng)-'do you have a family history of the headaches',

to(fre)-'avez-vous des antécédents familiaux de maux de tête',

- (a) Ya-t-il dans votre famille des personnes qui ont des maux de tête?
 (b) Est-ce que vous avez des maux de tête dans vos antécédents familiaux ?
 (c) Avez-vous des cas de maux de tête dans l'historique familial?
 (d) Do other members of your family experience headache as well?
 (e) Y a t'il d'autres personnes dans vos antécédent familial qui présentes les même symptômes de maux de têtes?
 (f) Do you have a family history of headaches?
 (g) Avez-vous des antécédents familiaux de maux de têtes ?
 (h) est-ce que vous avez eu les symptomes des maux de tête dans antécédent familial?
 (i) Do you have a family history of headaches? [Does your family have history of headaches?]
 (j) avez-vous des antécédents familiaux de maux de tête ?
 erreurs : 0

9. [[frequency,time], [prep,duration], [prep,frequency], [pronoun,you], [spec,1],[state,have_symptom], [symptom,headache], [tense,present], [timeunit,month], [utterance_type,ynq], [voice,active]]

from(eng)-'does the headache happen once a month',

to(eng)-'do you have the headaches once a month',

to(fre)-'avez-vous vos maux de tête une fois par mois'

- (a) *Combien de mois ces maux de tête durent-ils?*
 (b) *Est-ce que vous avez des maux de tête une fois par mois ?*
 (c) *Avez-vous eu ces symptômes pendant 1 mois?*
 (d) *Do you experience headache more than once per month in average? [je vois un conflit entre duration et frequency, et donc j'ai droppé duration]*
 (e) *Est-ce que vous avez plus d'une fois par mois ces symptômes de maux de têtes?*
 (f) *Do you have a headache once a month?*
 (g) *Avez-vous mal à la tête une fois par mois ?*
 (h) *combien des fois par mois est-ce que vous avez vos maux de tête?*
 (i) *Do you have headaches every month? [difficult to understand: spec, 1; what's intended by preposition of duration and of frequency, and how to deal with "frequency, time", hmm... How often do you have headaches each month?]*
 (j) *vos maux de tête durent-t-il un mois ou avez-vous mal à la tête une fois par mois ?*
 erreurs : 7

10. [[aspect,perfect], [prep,duration], [pronoun,you], [spec,[more_than,2]], [state,have_symptom], [symptom,headache], [tense,present], [timeunit,week], [utterance_type,ynq], [voice,active]]

from(eng)-'have you had them for more than two weeks'

from(fre)-'avez vous vos maux de tête depuis plus de deux semaines'

to(eng)-'have you had the headaches for more than two weeks'

to(fre)-'avez-vous vos maux de tête depuis plus de deux semaines'

- (a) *Avez-vous eu ces maux de tête pendant plus de deux semaines?*
 (b) *Est-ce que vous avez eu mal à la tête pendant plus de 2 semaines ?*
 (c) *Avez-vous eu ces symptômes pendant plus de 2 semaines?*
 (d) *Have you experienced headache for more than two weeks?*
 (e) *Est-ce que vous avez toujours plus de deux fois par semaine ces symptômes de maux de têtes?*
 (f) *Have you had the headache for more than two weeks?*
 (g) *Avez-vous mal à la tête plus de 2 fois par semaines ?*
 (h) *avez-vous vos maux de tête depuis plus que 2 semaines ?*
 (i) *Did you have more than two headaches this week?*
 (j) *avez-vous mal à la tête pendant plus de deux semaines ?*
 erreurs : 8

11. [[action,cough], [clause, [[pronoun,you], [state,have_symptom], [symptom,headache], [tense,present], [utterance_type,dcl], [voice,active]]], [pronoun,you], [sc,when], [tense,present], [utterance_type,ynq],[voice,active]]

to(eng)-'do you cough when you have the headaches'

to(fre)-'toussez-vous quand vous avez vos maux de tête'

- (a) Avez-vous de la toux aux moments de ces maux de tête?
 (b) Est-ce que vous toussiez quand vous avez mal à la tête ?
 (c) Toussiez-vous quand vous avez des maux de tête?
 (d) *When do you experience headache?* [là y a un truc que je comprends pas, il y a une clause interne mais la clause externe n'a pas de verbe ; de plus, je suis troublé par l'usage du présent...]
 (e) *Lorsque vous toussiez, avez-vous des symptômes de maux de têtes ?*
 (f) Do you cough when you have a headache?
 (g) Est-ce que vous toussiez lorsque vous avez mal à la tête ?
 (h) *avez- vous vos maux de tête quand vous toussiez?*
 (i) *Do you have a headache when you cough?*
 (j) *avez-vous mal à la tête quand vous toussiez ?*
 erreurs : 5

12. **[[clause,**
[[pronoun,you], [secondary_symptom,trauma], [state,experience], [tense,present],
[utterance_type,dcl], [voice,active]]],
[pronoun,you], [sc,after], [state,have_symptom], [symptom,pain],
[tense,past],[utterance_type,ynq], [voice,active]]
 from(fre)-'avez vous eu mal après un traumatisme'
 to(eng)-'did you experience the pain after you experienced trauma'
 to(fre)-'avez-vous eu mal après un traumatisme'

- (a) *Vous êtes traumatisé après avoir eu ces douleurs?*
 (b) Est-ce que vous avez eu mal après que vous avez subi un traumatisme ?
 (c) *Avez-vous d'autres symptômes suite aux maux de tête?*
 (d) Did you begin to experience pain after any trauma?
 (e) *Lorsque vous avez un trauma, ressentez vous des symptômes de douleurs après ?*
 (f)?
 (g) *Avez-vous ressenti un traumatisme après avoir eu mal ?*
 (h) *avez -vous la douleur après une traumatisation?*
 (i) *Have you experienced any pain after the trauma?* [I'm guessing here, just stringing together the words in a sentence that would make sense.]
 (j) *avez-vous mal suite à un traumatisme ?*
 erreurs : 7

13. **[[clause,**
[[pronoun,you], [state,have_symptom], [symptom,pain], [tense,present],
[utterance_type,dcl], [voice,active]]],
[pronoun,you], [sc,before], [secondary_symptom,muscle_aches], [state,experience],
[tense,present], [utterance_type,ynq], [voice,active]]
 to(eng)-'do you experience muscle aches before you experience the pain'
 to(fre)-'la douleur est-elle précédée par des douleurs musculaires'

- (a) *Resentez-vous des douleurs musculaires avec ces douleurs?*
 (b) Est-ce que vous avez des douleurs musculaires avant d'avoir mal ?
 (c) *Avez-vous des maux de tête quand vous avez des douleurs musculaires?*
 (d) *Do you experience pain after muscles aches ?* [sans définition plus formelle de MedSLT, je ne peux pas désambiguïser l'ordre de précedence des clauses]
 (e) *Après avoir eu des courbatures, ressentez vous des symptômes de douleurs ?*

(f)?

(g) Ressentez-vous des douleurs musculaire avant d'avoir mal ?

(h) *avez-vous déjà eu des douleurs musculaires avant la douleur?*

(i) *Do you have pain before you experience muscle aches?*

(j) *avez-vous des douleurs musculaires quand vous avez mal à la tête ?*

erreurs : 7

nombre total d'erreurs pour la première partie : 49

temps utilisé (en minutes) :

(a) 30

(b) 15

(c) 9

(d) 10

(e) 28

(f) 14

(g) 30

(h) ?

(i) 12

(j) 11

deuxième partie

14. 'YN-QUESTION continuous pain be PRESENT ACTIVE'

from(fre)-'la douleur est elle continue'

to(eng)-'is the pain continuous'

to(fre)-'la douleur est-elle continue'

(a) Avez-vous ces douleurs sans arrêt?

(b) Est-ce que la douleur est continue ?

(c) La douleur est-elle continue?

(d) Is the pain continuous?

(e) La douleur est t'elle toujours présente ?

(f) Is the pain continuous?

(g) La douleur est-elle continue ?

(h) *vos douleurs sont-elles persistantes?*

(i) Is the pain continuous?

(j) *avez-vous mal ?*

erreurs : 2

15. 'YN-QUESTION headache be in-loc one side-part PRESENT ACTIVE',

to(eng)-'are the headaches in one side',

to(fre)-'vos maux de tête sont-ils sur l'un des côtés'

(a) *Avez-vous les maux de tête dans une seule partie de la tête?*

(b) *Est-ce que les maux de tête sont de côté ?*

(c) Les maux de tête sont-ils présents que d'un seul côté?

(d) Is the headache in one side?

(e) Les maux de têtes sont seulement d'un côté ?

- (f) *Is the headache on the side?*
 (g) Les maux de tête sont-ils situé d'un seul côté ?
 (h) vos maux de tête sont-ils sur un côté?
 (i) Is the headache on one side?
 (j) avez-vous mal sur un côté de la tête ?

erreurs :3

16. 'YN-QUESTION you have pain in-loc head back-part ever PRESENT ACTIVE',

to(eng)-'do you ever experience the pain in the back of the head',
 to(fre)-'avez-vous déjà eu mal à l'arrière de la tête'

- (a) *Avez-vous toujours les douleurs dans la partie arrière de la tête?*
 (b) *Est-ce que vous avez mal à l'arrière de la tête ?*
 (c) *Avez-vous déjà eu des douleurs dans la partie arrière de la tête?*
 (d) *Do you ever have pain in the back of your head?*
 (e) *Avez-vous eu des douleurs à l'arrière tête ?*
 (f) *Do you ever have pain in the back?*
 (g) *Avez-vous toujours mal à l'arrière de la tête ?*
 (h) *vous n'avez jamais eu la douleur dans l'arrière tête?*
 (i) *Do you ever have pain in the back of the head?*
 (j) *avez-vous mal sur l'arrière de la tête ?*

erreurs : 5

17. 'WH-QUESTION pain last how-long PRESENT ACTIVE',

from(eng)-'what is the duration of your pain'
 from(fre)-'quelle est la durée de la douleur'
 to(fre)-'combien de temps la douleur dure-t-elle'

- (a) *Combien de temps dure la douleur?*
 (b) *Combien de temps dure la douleur ?*
 (c) *Combien de temps dure la douleur?*
 (d) *How long does the pain last?*
 (e) *Combien de temps dure la douleur ?*
 (f) *How long does the pain last?*
 (g) *Combien de temps dure la douleur ?*
 (h) *combien de temps la douleur reste –elle?*
 (i) *How long does the pain last?*
 (j) *combien de temps dure la douleur ?*

erreurs : 0

18. 'YN-QUESTION pain radiate shoulder left PRESENT ACTIVE',

from(fre)-'la douleur irradie t elle l'épaule gauche'
 to(eng)-'does the pain radiate to the left shoulder'
 to(fre)-'la douleur irradie-t-elle l'épaule gauche'

- (a) *Ressentez-vous de la douleur à l'épaule gauche irradié*
 (b) *Est-ce que la douleur irradie dans l'épaule gauche ?*
 (c) *Est-ce que la douleur irradie l'épaule gauche?*
 (d) *Does the pain radiate in your left shoulder?*
 (e) *L'épaule gauche irradiée, fait t'elle mal ?*

- (f) Does the pain radiate to the left shoulder?
(g) *La douleur irradie-t-elle de l'épaule gauche ?*
(h) *La douleur irradie-t-elle de l'épaule gauche?*
(i) *Does the pain radiate to/from the left shoulder?* [same problem as mentioned in the first part]
(j) *la douleur irradie-t-elle l'épaule gauche ?*
erreurs : 6

19. 'YN-QUESTION attacks become-better PRESENT ACTIVE',

from(eng)-'are the attacks becoming less severe'
to(eng)-'are the attacks better'
to(fre)-'ces crises diminuent-elles'

- (a) *Avez-vous moins de crises?*
(b) *Est-ce que les crises s'améliorent ?*
(c) *Est-ce que les crises ont diminué d'ampleur?*
(d) *Do the attacks become less frequent?*
(e) *La crise va mieux ?*
(f) *Are the attacks easing?*
(g) *Les crises s'atténuent –elle ?*
(h) *est-ce que les crises s'améliorent?*
(i) *Are the attacks getting better?*
(j) *vos crisent s'améliorent ?*
erreurs : 0

20. 'YN-QUESTION you experience nausea PAST ACTIVE',

to(eng)-'did you feel nauseous',
to(fre)-'avez-vous eu des nausées'

- (a) *Avez-vous eu des nausées?*
(b) *Est-ce que vous avez eu la nausée ?*
(c) *Avez vous déjà eu des nausées?*
(d) *Did you experienced nausea?*
(e) *Avez-vous eu des nausées ?*
(f) *did you experience nausea?*
(g) *Avez-vous eu des nausées ?*
(h) *avez –vous eu des nausées?*
(i) *Did you experience nausea?*
(j) *avez-vous de la nausée ?*
erreurs : 2

21. 'YN-QUESTION you have family-history of high-blood-pressure PRESENT ACTIVE',

[from(fre)-'avez vous des antécédents familiaux de hypertension',
to(eng)-'do you have a family history of high blood pressure',

- (a) *Y a-t-il des antécédants d'hypertension au sein de votre famille?*
(b) *Est-ce que vous avez un antécédent familial d'hypertension ?*
(c) *Est-ce que votre famille a un historique de pression sanguine élevée?*
(d) *Do you have a family history of high-blood pressure?*
(e) *Vos antécédent familiaux ont t'ils de hypertension ?*
(f) *Do you have a family history of high blood pressure?*

- (g) Avez-vous des antécédents familiaux d'hypertension ?
 (h) dans votre antécédent familiale avez -vous des membres avec hypertension?
 (i) Do you have a family history of high blood pressure?
 (j) avez-vous des antécédents familiaux d'hypertension ?
 erreurs : 0

22. 'YN-QUESTION you have attacks two times per month PRESENT ACTIVE'

from(eng)-'do you have the attacks twice a month'
 to(eng)-'do you have the attacks twice a month'
 to(fre)-'avez-vous ces crises deux fois par mois'

- (a) Avez-vous des crises deux fois par mois?
 (b) Est-ce que vous avez des crises deux fois par mois ?
 (c) Avez-vous des crises deux fois par moi?
 (d) Do you have attacks two times per mounth?
 (e) Avez-vous deux fois par mois une crise ?
 (f) Do you have teh attacks twice a month?
 (g) Avez-vous des crises deux fois par mois ?
 (h) avez –vous des crises deux fois par mois?
 (i) Do you have attacks twice per month?
 (j) avez-vous des crises deux fois par mois ?
 erreurs : 0

23. 'YN-QUESTION you have attacks at-least three times per month PRESENT ACTIVE'

from(eng)-'do you have the attacks at least three times a month'
 to(eng)-'do you have the attacks at least three times a month'
 to(fre)-'avez-vous ces crises au moins trois fois par mois'

- (a) Avez-vous au moins trois crises par mois?
 (b) Est-ce que vous avez des crises au moins trois fois par mois ?
 (c) Avez vous des crises au moins trois fois par mois?
 (d) *Do you have attacks more than three times per mounth?*
 (e) Vous avez au moins trois fois par mois une crise ?
 (f) Do you have the attacks at least 3 times a month?
 (g) Avez-vous des crises au moins trois fois par mois ?
 (h) avez-vous des crises au minimum trois fois par mois?
 (i) Do you have attacks at least three times per month?
 (j) avez-vous des crises au moins trois fois par mois ?
 erreurs : 1

24. 'YN-QUESTION you experience pain sc-when [you move-suddenly head PRESENT ACTIVE] PRESENT ACTIVE'

to(eng)-'do you experience the pain when you move the head suddenly'
 to(fre)-'avez-vous la douleur quand vous bougez soudainement la tête'

- (a) Resentez-vous de la douleur quand vous bougez brusquement la tête?
 (b) Est-ce que vous avez mal quand vous bougez soudainement la tête ?
 (c) Est-ce que vous ressentez des douleurs quand vous déplacez rapidement la tête?
 (d) Do you experience pain when you suddenly move your head?
 (e) Lorsque vous bougez brusquement la tête, vous ressentez de la douleur ?
 (f) Do you experience pain when you move your head suddenly?

- (g) Ressentez-vous de la douleur lorsque vous bougez brusquement la tête ?
 - (h) avez-vous une douleur quand vous bougez la tête brusquement?
 - (i) Do you experience pain when you move your head suddenly?
 - (j) avez-vous mal quand vous bougez la tête brusquement ?
- erreurs : 0

25. 'YN-QUESTION you have pain sc-after [you eat large-meal PRESENT ACTIVE] PRESENT ACTIVE'

from(eng)-'does the pain occur after you eat a large meal'
to(eng)-'do you experience the pain after you eat a large meal'
to(fre)-'avez-vous mal après un gros repas'

- (a) *Avez-vous de la douleur quand vous mangez beaucoup?*
 - (b) Est-ce que vous avez mal après que vous avez mangé un grand repas ?
 - (c) Est-ce que vous ressentez des douleurs après un repas copieux?
 - (d) Do you have pain after you eat a large meal?
 - (e) Après un grand repas, vous ressentez de la douleur ?
 - (f) Do you have pain after you eat a large meal?
 - (g) Avez-vous mal après avoir beaucoup mangé ?
 - (h) avez-vous une douleur après un repas copieux?
 - (i) Do you have pain after you eat a large meal?
 - (j) avez vous mal après un grand repas ?
- erreurs : 1

26. 'YN-QUESTION you experience nausea sc-before [you have headache PRESENT ACTIVE] PRESENT ACTIVE'

from(fre)-'vos maux de tête sont ils précédés par des nausées'
to(eng)-'do you feel nauseous before you have the headaches'
to(fre)-'vos maux de tête sont-ils précédés par des nausées'

- (a) Avez-vous des nausées avant les maux de tête?
 - (b) Est-ce que vous avez la nausée avant que vous ayiez mal à la tête ?
 - (c) *Est-ce que vous ressentez des nausées quand vous avez mal à la tête?*
 - (d) Do you experience nausea before you have headache?
 - (e) *Avez-vous mal à la tête avant les nausées ?*
 - (f) Do you get nausea before your headaches?
 - (g) Ressentez-vous des nausées avant vos migraines ?
 - (h) *avez-vous déjà eu des nausées avant les maux de tête?*
 - (i) Do you experience nausea before you have a headache?
 - (j) avez-vous de la nausée avant vos maux de tête ?
- erreurs : 3

nombre total d'erreurs pour la deuxième partie : 23

temps utilisé (en minutes) :

- (a) 20
- (b) 5
- (c) 4
- (d) 2
- (e) 12
- (f) 6

- (g) 12
- (h) ?
- (i) 4
- (j) 5