

TRADEX, un système de traduction de télex

Jean-Marc Aumaitre, Laurence Horel and Jean-Marie Lancel

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

Volume 37, Number 4, décembre 1992

URI: <https://id.erudit.org/iderudit/001924ar>

DOI: <https://doi.org/10.7202/001924ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Aumaitre, J.-M., Horel, L. & Lancel, J.-M. (1992). TRADEX, un système de traduction de télex. *Meta*, 37 (4), 624–634. <https://doi.org/10.7202/001924ar>

Article abstract

Current translation systems give satisfactory results in technical domains where both polysemy and syntactic complexity are limited. TRADEX is a translation system used to translate into French military telex messages written in English. It acts as an autonomous component in the telex transmission network, and can also be invoked from a translation environment. TRADEX was developed by CAP GEMINI INNOVATION and CAP SESA DEFENSE1 for the SEFT2. The article presents the context for the application, the computational characteristics of TRADEX, the linguistic data organization and the resulting performances.

TRADEX, UN SYSTÈME DE TRADUCTION DE TÉLEX

JEAN-MARC AUMAITRE,
LAURENCE HOREL, JEAN-MARIE LANCEL
CAP GEMINI INNOVATION, Paris, France

Résumé

Les systèmes de traduction donnent actuellement des résultats satisfaisants pour des domaines techniques où la polysémie et la complexité des structures syntaxiques utilisées sont limitées. TRADEX est un système de traduction utilisé pour traduire en français des télex militaires formulés en anglais. Il fonctionne comme un composant autonome dans le réseau de transmission de télex, et peut aussi être activé à partir d'un environnement de traduction. TRADEX a été réalisé par CAP GEMINI INNOVATION et CAP SESA DEFENSE¹ pour la SEFT². L'article présente le contexte de l'application, les caractéristiques logicielles de TRADEX, l'organisation des données linguistiques et les performances obtenues.

Abstract

Current translation systems give satisfactory results in technical domains where both polysemy and syntactic complexity are limited. TRADEX is a translation system used to translate into French military telex messages written in English. It acts as an autonomous component in the telex transmission network, and can also be invoked from a translation environment. TRADEX was developed by CAP GEMINI INNOVATION and CAP SESA DEFENSE¹ for the SEFT². The article presents the context for the application, the computational characteristics of TRADEX, the linguistic data organization and the resulting performances.

1. CONTEXTE DE L'APPLICATION

1.1. PRÉSENTATION

TRADEX est utilisé par l'armée de terre française pour consulter en français des messages transmis en anglais. Les utilisateurs ne sont pas des traducteurs mais des spécialistes qui souhaitent accéder rapidement à l'information transmise par le message.

Deux modes de fonctionnement sont fournis : un mode «à la demande» et un mode automatique. Dans le premier mode, l'utilisateur accède à un environnement qui lui permet de prendre connaissance des messages reçus et de demander la traduction spécifique d'un télex. L'utilisateur peut visualiser la traduction du télex et aussi le réémettre sur le réseau.

Dans le mode automatique, le système traduit les messages au fur et à mesure de leur réception. Les messages traduits sont stockés et l'accès est sur demande.

1.2. SPÉCIFICITÉS DES MESSAGES TÉLEX

Un message est composé d'un en-tête structuré et d'une partie en texte libre généralement organisée en paragraphes et qui seule est traduite.

Les structures rencontrées dans les phrases sont très variées : depuis une énumération de chiffres représentant des coordonnées géographiques jusqu'à des phrases complexes de vingt à trente mots.

Les phrases sont en général énonciatives. Elles décrivent la situation sur un champ de bataille, en termes de position et de mouvement des forces en présence. Elles rendent compte des opérations effectuées et de l'intention probable de l'adversaire. Elles peuvent également comporter des ordres sur la conduite à adopter.

Les structures syntaxiques sont assez variées ; elles incluent des constructions habituelles ainsi que des constructions spécifiques : phrases nominales, énumérations.

Les mots composés sont fréquents, ils représentent environ le tiers du vocabulaire du domaine. Les sigles ainsi que les abréviations de substantifs sont souvent utilisés, avec parfois des variantes d'écriture. Par exemple, *5GTA* est traduit par *5^e ABG*. Les messages comportent aussi un grand nombre de noms propres correspondant à des lieux géographiques et à des noms de personnes.

On trouve aussi dans ces messages des structures spécifiques du domaine telles que des horaires, des coordonnées géographiques (*NV 5473*), des indications de temps (durée, date) et de distance. On peut ainsi trouver *0530Z* qui signifie 5 heures 30 (heure zoulou) et *24 APR 88* qui correspond au 24 avril 1988. Les nombres sont habituellement écrits en chiffres arabes mais aussi parfois en chiffres romains.

2. CARACTÉRISTIQUES LOGICIELLES

2.1. CHOIX DE LA MÉTHODE

Les techniques utilisées pour la traduction peuvent être classées en trois catégories :

- les systèmes à représentation intermédiaire multilingue : on construit une représentation intermédiaire (langage «pivot») indépendante de la langue source et de la langue cible. Un composant de génération permet ensuite de produire la traduction dans la langue cible. Cette représentation interne peut être syntaxique ou conceptuelle.

Exemples :

- le système TITUS utilise une structure «pivot» syntaxique pour la traduction de textes en syntaxe contrainte. Les domaines d'application sont le textile et l'agriculture tropicale. Les langues sources et cibles sont l'allemand, l'anglais, l'espagnol et le français (Zingel 1978) ;
- le système SYSTRAN utilise également une structure intermédiaire pour la traduction (Toma 1978) ;
- CAP GEMINI INNOVATION a utilisé une représentation conceptuelle dans le cadre du projet GDM (Génération de Documents Multilingues) (Lancel, Rousselot, Simonin 1986).

Ce type d'approche n'est actuellement réalisable que pour des domaines restreints bien modélisables. Il a pour avantage un faible coût pour l'insertion d'une nouvelle langue (système en étoile), mais augmente les risques d'erreur du fait de la double transformation (de la langue source vers le pivot, puis du pivot vers la langue cible). Celle-ci est aussi limitée par la précision de la représentation «pivot» : l'objectif d'une structure intermédiaire complète et sans ambiguïté, valable pour toutes les langues, n'a pas encore été atteint.

- les systèmes basés sur le transfert : on construit une représentation syntaxique spécifique de la langue d'origine comprenant éventuellement des informations sémantiques. Pour traduire dans la langue cible, une phase de «transfert» intervient qui transforme la structure syntaxique dans la langue cible, avec une traduction des unités lexicales. La phrase de la langue cible est alors produite à partir de cette structure.

Exemples :

- GETA {russe-français, anglais-français, français-anglais, ...}, avec plusieurs domaines d'application dont l'aéronautique et les télécommunications (Vauquois et Boitet 1985);
- EUROTRA {allemand, anglais, danois, espagnol, grec, italien, français, néerlandais, portugais} (Danlos 1989) et (Danlos 1991);
- McCord {anglais-allemand} (McCord 1988); {anglais-français} (Bilange 1988), dans le domaine des manuels d'utilisation en informatique;
- les systèmes spécifiquement bilingues comme le système METEO (Thouin 1982): le processus de traduction est ici adapté à un couple de langues et à ses spécificités. Ceci permet de ne pas accumuler les difficultés de plusieurs langues, et ainsi d'obtenir de meilleurs résultats.
Le système TRADEX {allemand-français} (Cosserat, Lancel, Simonin 1988) et {anglais-français} comporte une phase de traduction spécifiquement bilingue. Il utilise comme structures intermédiaires le niveau morphologique canonique anglais et le niveau morphologique canonique français. La structure syntaxique française est construite par étapes lors de la reconnaissance des entités syntaxiques anglaises.

2.2. MODES DE FONCTIONNEMENT

Comme il n'existe pas actuellement de système informatique capable de traduire parfaitement un texte, les systèmes de traduction font souvent appel à une intervention extérieure :

- saisie contrainte : l'utilisation d'un pré-éditeur permet de détecter les structures non acceptées et de lever à l'avance certaines ambiguïtés en plaçant des marqueurs sur les structures à identifier (mots composés, groupes nominaux, groupe verbal...);
- traduction interactive : l'utilisateur peut intervenir à la demande du système pour le choix du vocabulaire et la levée d'ambiguïtés ;
- postédition : l'utilisateur peut affiner la traduction fournie par le système.

TRADEX {anglais-français} peut fonctionner de deux façons: en mode automatique, sans intervention extérieure ou en mode postédition, traduction d'un télex suivie d'une postédition. Le mode automatique ou «Opérationnel» correspond au besoin manifesté par l'armée de terre d'intégration de TRADEX dans un système de messagerie.

Ce choix implique une bonne qualité de traduction. L'approche a été prise de privilégier la qualité de traduction au détriment du nombre de phrases traduites. Dans environ 15 % des cas TRADEX ne fournit pas de traduction et conserve dans le texte la phrase initiale. Ceci correspond à des phrases ambiguës ou à des phrases dont la structure n'est pas reconnue.

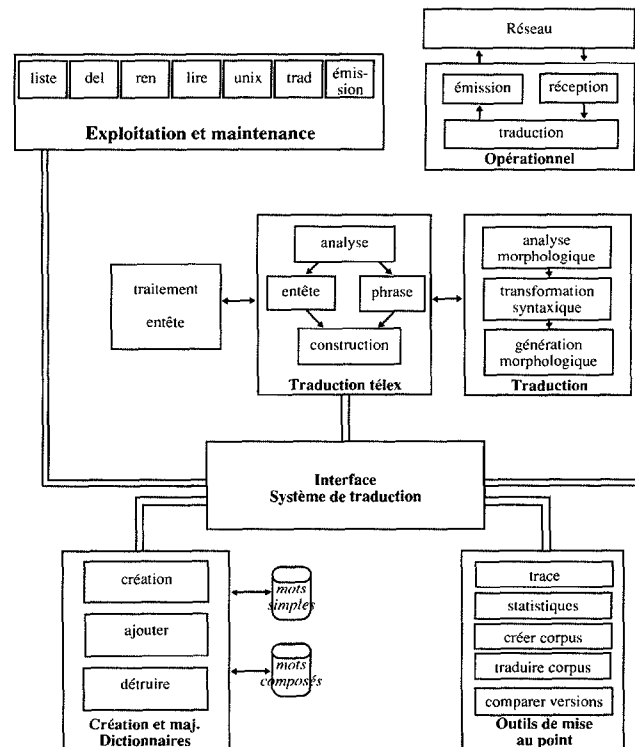
2.3. DESCRIPTION DES COMPOSANTS

TRADEX {anglais-français} a été conçu pour être évolutif, tant au niveau du vocabulaire qu'au niveau des types de phrases à traduire. Quatre modes permettent de réaliser ces fonctionnalités :

- Mode «Opérationnel»: il permet la traduction en continu des messages arrivant sur le réseau. Chaque message traduit est ré-émis par le système vers ses destinataires ;

- Mode «Exploitation et maintenance» : il permet la manipulation des télex par l'utilisateur (traduction d'un message donné, consultation, destruction, renommage, liste des messages présents sur le site, ré-émission de messages) ;
- Mode «Création et mise à jour des dictionnaires» : on peut ici assurer l'évolution des dictionnaires (création, ajout de mots nouveaux, suppression de mots obsolètes). Ces opérations peuvent être effectuées interactivement ou en différé à partir d'un fichier de mises à jour ;
- Mode «Outils de mise au point» : l'évolutivité de la grammaire de traduction est un atout important afin de conserver l'adéquation entre le système et le domaine d'application. Cette opération n'est toutefois pas sans risques et nécessite une attention particulière. C'est pourquoi des outils permettant le contrôle des modifications sont fournis :
 - traces : explications sur le cheminement du système lors de la traduction ;
 - statistiques : espace de recherche nécessaire au système pour effectuer la traduction. Plusieurs directions sont examinées (espace pour les mots, les groupes nominaux, les groupes verbaux...) ;
 - comparaison de versions : sur la base d'un échantillon («corpus»), on compare les performances (c'est-à-dire le nombre de phrases traduites, partiellement traduites, non traduites...) de deux versions de la grammaire. Cette fonctionnalité permet de contrôler la non-régression de la traduction lors de l'évolution du système.

Le schéma d'architecture ci-dessous présente les composants du système ainsi que les ressources utilisées :



2.4. MODULE DE TRADUCTION

La traduction est réalisée phrase par phrase. Pour chaque phrase anglaise à traiter, le processus est le suivant :

- analyse morphologique : analyse de chacun des mots de la phrase et analyse des mots composés ;
- transformation syntaxique : reconnaissance de la structure syntaxique de la phrase anglaise et transformation en la structure française correspondante ;
- génération morphologique : flexion des mots de la phrase française, élisions et contractions.

2.4.1. ANALYSE MORPHOLOGIQUE

La phase d'analyse morphologique permet la reconnaissance des mots simples, des mots composés («*mountain division*» : «*division alpine*») et des verbes à particule(s) («*to move forward*» : «*avancer*»).

L'analyse des mots simples fournit pour chaque mot son terme canonique et ses traits lexicaux. Ce traitement identifie par exemple le verbe «*to go*» à partir des différentes formes «*go*», «*goes*», «*going*», «*went*», «*gone*», ainsi que les renseignements nécessaires à la traduction (catégorie syntaxique, position de l'adjectif français, position des compléments de verbe, traductions...).

2.4.2. TRANSFORMATION SYNTAXIQUE

Pour effectuer la transformation syntaxique, on utilise un formalisme proche des DCG (Definite Clause Grammar) (Pereira et Warren 1980) permettant une écriture déclarative des règles de grammaire utilisées pour la reconnaissance des structures syntaxiques anglaises. La transformation en une structure française équivalente est effectuée simultanément. Ces règles peuvent être de type :

- reconnaissance de la structure syntaxique ;
- accords grammaticaux : accords en nombre, en temps, et en genre pour le français ;
- contrôle : regard en avant, choix stratégiques ;
- reconstruction : transformation de la structure anglaise en la structure française correspondante.

On donne ici quelques règles très simplifiées de reconnaissance de structures syntaxiques :

- PHRASE → PHRASE_COMPLÈTE | PHRASE_INCOMPLÈTE | PHRASE_NOMINALE [...] pas de traduction
- PHRASE_COMPLÈTE → STRUCT_COMPLÈTE SUITE_PHRASE
- PHRASE_INCOMPLÈTE → STRUCT_INCOMPLÈTE SUITE_PHRASE
- SUITE_PHRASE → vide | STRUCT_COMPLÈTE SUITE_PHRASE | STRUCT_INCOMPLÈTE SUITE_PHRASE

2.4.3. GÉNÉRATION MORPHOLOGIQUE

Une fois la structure française connue, il faut fléchir (conjuguer ou décliner) les termes de la phrase, suivant les caractéristiques déterminées lors de la phase de transformation syntaxique. On effectue enfin les opérations d'élision et de contraction propres au français.

3. ORGANISATION DES DONNÉES LINGUISTIQUES

Les dictionnaires du système de traduction sont composés de mots canoniques anglais auxquels sont rattachées les informations suivantes :

- traits lexicaux anglais : catégorie syntaxique (substantif, verbe, adjectif...), genre, nombre, personne ;
- syntaxe anglaise : structures attendues (principalement pour les verbes) ;
- syntaxe française : correspondance syntaxique entre l'anglais et le français ;
- traits lexicaux français : catégorie syntaxique, genre, nombre, personne.

Les dictionnaires couvrent à la fois le vocabulaire de l'anglais courant et le vocabulaire propre au domaine des télex militaires. Ils sont de deux types :

- mots simples (5 000 entrées) ;
- mots composés (2 000 entrées) : suite de mots simples dont le sens n'est pas équivalent à la suite de mots la composant.

Les mots simples anglais peuvent avoir un mot composé français pour traduction, de même un mot composé anglais peut se traduire par un mot simple français. Ainsi, on peut donner les exemples suivants :

- le mot simple *acknowledgement* se traduit par un mot composé en français : *accusé de réception*,
- et le mot composé *avoidance of action* est traduit en français par le mot simple : *décrochage*.

Le formalisme utilisé, ainsi que les informations attachées à un mot simple et à un mot composé, sont présentés ci-dessous.

3.1. LES MOTS SIMPLES

Le formalisme utilisé pour représenter un mot simple est inspiré des Descriptions Fonctionnelles (Kay 1985). Un mot est ainsi une suite de champs «attribut = valeur» entre crochets :

```
[ mot      = <générique>
  modèle  = <modèle de flexion>
  infos   = [<informations pertinentes>]
]
```

où :

- <générique> est la forme non fléchie (ou canonique) du mot,
- <modèle de flexion> permet de connaître la catégorie syntaxique du mot ainsi que les différentes formes sous lesquelles on le rencontre. Par exemple, le modèle *change* indique un verbe régulier dont le prétérit et le participe passé prennent un «d» et la 3^e personne du présent au singulier prend un «s»,
- <informations pertinentes> varie selon la catégorie syntaxique du mot.

Le contenu du champ <informations pertinentes> est détaillé ci-dessous en fonction des catégories syntaxiques des mots :

- adjectif : traduction de l'adjectif et indication de la position de l'adjectif français par rapport à celle du nom qu'il qualifie : avant, après ou indéterminé ;

- nom : traduction du nom et indication du genre de cette traduction. En effet, une traduction d'un nom anglais peut se composer de plusieurs mots français ; il est alors difficile de savoir lequel de ces mots donne ses traits morphologiques au groupe. L'indication du genre dans ce champ permet d'éviter cette indétermination sur le genre du groupe de mots français.
- verbe : une description détaillée du champ <informations pertinentes> du verbe est faite en 3.3.
- adverbe, préposition, déterminant et autres mots outils : le champ de ces mots contient la traduction du mot.

La partie traduction du champ <informations pertinentes> se présente sous la forme d'une liste de mots. À chaque mot est associé un code permettant de le fléchir correctement. Ce code est issu du DELAS, dictionnaire du français développé par le LADL (Courtois 1986).

Afin d'illustrer les différents champs d'information décrits ci-dessus, on donne ici une entrée d'un nom et une entrée d'un adjectif :

Un nom :

[mot	=	<i>stress</i>
	modèle	=	<i>box</i>
	infos	=	[[<i>tension(N21)</i>], <i>féminin</i>]
]			

On voit ici que le mot *stress* se fléchit comme le mot «*box*», c'est-à-dire qu'il prend «es» au pluriel (*boxes*). Sa traduction est «*tension*» qui suit la conjugaison N21 (mots féminins prenant un «s» au pluriel).

Un adjectif :

[mot	=	<i>cold</i>
	modèle	=	<i>tall</i>
	infos	=	[[<i>froid(A32)</i>], <i>après</i>]
]			

De la même façon, l'adjectif *cold* se fléchit comme *tall*, c'est-à-dire qu'au superlatif il prend «est» (*tallest*) et au comparatif «er» (*taller*). Sa traduction est «*froid*», adjectif qui prend un «s» au pluriel et un «e» au féminin. Enfin, il se place après le nom qu'il qualifie (*le temps froid*).

3.2. LES MOTS COMPOSÉS

Le formalisme utilisé est identique à celui des mots simples, à ceci près que pour les champs «mot» et «modèle», les valeurs sont des listes. On trouve pour les mots composés les champs suivants :

[mot	=	<générique1> ... <génériqueN>
	modèle	=	<modèle1> ... <modèleN>
	cat	=	<catégorie syntaxique>
	infos	=	[<informations pertinentes>]
]			

où :

- <générique1> ... <génériqueN> est la suite ordonnée des génériques constituant le mot composé ;

- <modèle1> ... <modèleN> est la suite des modèles associés aux génériques. Un signe «+» devant l'un des modèles indique que les traits morphologiques du mot composé seront ceux du mot simple ayant ce modèle ;
- <catégorie syntaxique> est la catégorie syntaxique du mot composé : nom, verbe, adjectif, adverbe ...
- <informations pertinentes> contient essentiellement la (ou les) traduction(s) du mot composé anglais.

La description du mot composé «casualty list» est donnée en exemple :

```
[ mot      =  casualty list
  modèle   =  baby + cat
  cat      =  nom
  infos    =  [[état(N1), des(INV), pertes(INV)], masculin]
]
```

On voit ainsi que le mot *list* est le mot directeur du mot composé. Le mot *casualty* se fléchit comme *baby* et *list* se fléchit comme *cat*. Ce mot composé a pour catégorie syntaxique «nom». Sa traduction est un mot composé français dont les mots *des* et *pertes* resteront invariables.

3.3. LE TRAITEMENT DES VERBES

En anglais, les verbes peuvent être associés à des particules, qui modifient alors leur sens premier. Ces particules sont soit des adverbes soit des prépositions. Pour un système de traduction, il est important de tenir compte de ces particules afin de rendre le sens exact des verbes. Ces informations facilitent aussi la reconnaissance de la structure de la phrase anglaise.

La modélisation linguistique utilisée s'inspire de la codification développée dans *Longman Dictionary of Contemporary English* (Longman 1981). Cette codification décrit les compléments que peut prendre un verbe, seul ou associé à une particule. Cette codification permet ainsi de réduire le nombre de structures possibles après un verbe. De plus, elle permet de donner la traduction appropriée à chaque construction du verbe.

Ces codes font partie du champ «infos» décrit ci-dessus. La description d'une partie du champ «infos» du verbe *to move* est donnée ici à titre d'exemple :

```
[
  [ [t(1)],    [déplacer(V3), C1],
    [i(0)],    [se(PRON6), déplacer(V3)]
    [v(3)],    [inciter(V3), C1, à(PREP), C2]
  ],
  [ [away],
    [i(0,a)],  [déménager(V3)]
  ],
  [ [off],
    [i(0,a)],  [partir(V28)]
  ]
]
```

Comme on peut le voir dans cet exemple, le code Longman est formé d'une lettre et d'un chiffre. La lettre indique le nombre de compléments et le type du verbe (transitif, ditransitif ...); le chiffre indique le contenu du (ou des) complément(s) (groupe nominal,

infinitive, that clause...). Lorsque le verbe est associé à une particule, un second argument indique la position de la particule par rapport au verbe :

- a : la particule est collée au verbe,
- b : la particule est séparée du verbe par un complément,
- _ : la place de la particule est indéterminée.

Ainsi, dans l'exemple, le code t(1) indique que le verbe *to move* peut être suivi d'un complément d'objet groupe nominal ou pronom. Dans la liste contenant la traduction, une variable indique où se trouvera ce complément.

L'utilisation de ces codes Longman facilite ainsi la reconnaissance de la structure et le réordonnement des mots de la phrase française.

4. ÉVALUATION

Le système de traduction Tradex a été évalué en plusieurs étapes, allant d'une évaluation théorique à l'évaluation finale en milieu opérationnel.

4.1. ÉVALUATION THÉORIQUE : PHASE D'APPRENTISSAGE

Lors de la phase de conception générale de Tradex, les opérationnels de l'armée de terre ont fourni un corpus de télex émanant de manœuvres, comportant environ 750 phrases. Ce corpus a permis d'une part d'étudier la structure des messages dans sa globalité, mais également d'extraire le vocabulaire spécifique au domaine, ainsi que les structures syntaxiques utilisées pour la rédaction des télex.

Pour la partie traduction, les 750 phrases ont été extraites afin de constituer un corpus de référence (corpus d'apprentissage). Le travail d'amélioration de la grammaire a ensuite pu commencer, avec les étapes classiques de production de versions successives, chaque version étant évaluée suivant les critères de :

- pourcentage de succès ;
- espace de recherche pour chacune des phrases à traduire (paramètre équivalent au temps de traduction) ;
- qualité des traductions (travail manuel en collaboration avec un interprète de l'armée de terre).

Cette méthode de travail a permis de traiter de manière semi-automatisée les problèmes de régression survenant en cours de développement (lors des modifications du système). On rencontre classiquement ces problèmes dans le domaine du traitement des langues naturelles.

4.2. ÉVALUATION THÉORIQUE : PHASE DE TESTS

La seconde étape d'évaluation se situe après la constitution des dictionnaires et le développement de la grammaire de traduction, lorsque la qualité de la traduction est jugée très satisfaisante sur le corpus d'apprentissage. Un corpus similaire au corpus d'apprentissage, nommé corpus de tests sert de base aux nouvelles évaluations du traducteur. Ce corpus provient également de manœuvres opérationnelles et comporte environ 750 phrases. Il n'a pas été utilisé lors de la conception ni lors du développement de la grammaire de traduction. Les enseignements que l'on en tire sont :

- le pourcentage de mots inconnus du système ;
- le pourcentage de phrases bien traduites, partiellement traduites, non traduites.

Cette phase permet d'évaluer le travail restant à effectuer pour rendre le système opérationnel et est suivie de «travaux» d'amélioration des dictionnaires et de la grammaire de traduction.

4.3. ÉVALUATION THÉORIQUE : PHASE DE RECETTE

À la suite de la phase de tests, un corpus d'environ 1 000 phrases a été fourni en vue d'effectuer la recette³ du système de traduction, une semaine avant la recette. Le but de cette phase est de déterminer la qualité du traducteur en ayant pris en compte le vocabulaire manquant (Tradex permet l'ajout interactif de mots dans les dictionnaires). La qualité des traductions est cette fois-ci évaluée par des opérationnels de l'armée de terre, afin de rendre compte du degré de satisfaction des utilisateurs finaux.

4.4. ÉVALUATION EN MILIEU OPÉRATIONNEL

Afin de tester le comportement du système de traduction en milieu opérationnel, celui-ci a été utilisé «en grandeur réelle» au cours de manœuvres militaires. Durant une semaine, des messages reçus par les Systèmes Informatiques de Commandement (SIC) ont été traduits à la fois par les opérationnels, et par le système Tradex, puis les traductions ont été comparées. Le comportement de Tradex lors de ces traductions a été jugé prometteur. On a pu noter lors de cette évaluation un nombre impressionnant de fautes de frappe dans les messages à traiter, nombre qui justifierait l'utilisation d'un correcteur. La conception modulaire de Tradex permet l'adjonction d'un tel outil, lequel permettrait l'augmentation du pourcentage de phrases traduites par le système de traduction Tradex.

5. CONCLUSION

TRADEX est destiné à des opérateurs qui ne sont pas des spécialistes de la langue anglaise. Il a pour but de permettre aux opérationnels d'avoir une connaissance rapide du contenu global d'un message. Une attention particulière a été apportée à la constitution des dictionnaires spécifiques, afin d'utiliser au mieux les expressions du domaine.

Le système a été évalué par des interprètes de l'armée de terre, ainsi que par des opérationnels utilisant le traducteur, grâce à un ensemble de 300 phrases prises au hasard parmi 1 500 phrases. L'objectif atteint est de 72 % de phrases bien traduites (phrases sans fautes de frappe). On entend par «bien traduites» les phrases qui n'ont pas subi d'altération de sens lors de la traduction. Dans 23 % des cas, le système décide de ne pas traduire la phrase (phrases ambiguës, structures inconnues de la grammaire) et, dans 5 % des cas, la traduction donnée par TRADEX n'est pas la bonne. Une phrase est traduite en moyenne en six secondes sur SUN 4/60.

Lors de la phase d'évaluation et de mise au point, les outils de trace, de statistiques et de non-régression ont été utilisés. Ils se sont avérés adaptés aux problèmes de l'évolution d'une grammaire de transformation et ont permis des améliorations sensibles de la qualité de la traduction avec une bonne sécurité. Le formalisme utilisé pour l'écriture des règles de grammaire assure leur lisibilité et permet donc des modifications aisées.

TRADEX est une des premières réalisations concrètes relevant du domaine de l'Intelligence Artificielle dans le domaine des Systèmes Informatiques de Commandement. Répondant à un besoin précis, il montre ce que peuvent apporter les techniques linguistiques dans les systèmes d'information et d'aide à la décision.

Notes

1. CAP GEMINI INNOVATION et CAP SESA DEFENSE font partie du groupe CAP GEMINI SOGETI.
2. Section d'Études et Fabrications des Télécommunications, rattachée à la Direction Générale de l'Armement.
3. Acceptation formelle du système par le client.

BIBLIOGRAPHIE

- BILANGE, Éric (1988) : *Un système de traduction automatique anglais-français*, Centre Scientifique IBM France, Étude F140, décembre 1988.
- COURTOIS, Blandine (1986) : *Les dictionnaires électroniques du LADL*, Rapport interne du Laboratoire d'Automatique Documentaire et Linguistique, Paris.
- DANLOS, Laurence (1989) : «La traduction automatique», *Annales des télécommunications*, 44, n° 1-2, pp. 101-110.
- DANLOS, L. et O. LAURENS (1991) : *Présentation du projet EUROTRA et des grammaires d'EUROTRA-France*, Rapport technique n° 1, TALANA — LISH, janvier 1991.
- COSSERAT, Laurent, LANCEL, Jean-Marie et Nathalie SIMONIN (1988) : *Une application de traduction allemand-français assistée par ordinateur*, Séminaire international de traduction assistée par ordinateur, Paris, mars 1988.
- KAY, Martin (1985) : «Parsing in Functional Unification Grammar», D. Dowty, L. Karttunen et A. Zwicky (Eds.), *Natural Language Parsing*, Cambridge, Cambridge University Press.
- LANCEL, Jean-Marie, ROUSSELOT, François et Nathalie SIMONIN (1986) : «A Grammar Used for Parsing and Generation», *COLING 86*, XIth International Conference on Computational Linguistics, Bonn, août 1986.
- Longman Dictionary of Contemporary English* (1981) : Longman Group Limited, 1978, Corrections 1981.
- MCCORD, Michael (1988) : «Design of LMT: A Prolog-Based Machine Translation System», *Computer Science*.
- PEREIRA, F. et D. WARREN (1980) : «Definite Clause Grammar for Language Analysis», *Artificial Intelligence*, 13, pp. 231-278.
- THOUIN, B. (1982) : *The METEO System, Practical Experience of Machine Translation*, Veronica Lawson (Ed.), Amsterdam, North Holland.
- TOMA, P. (1978) : «As a Multilingual Machine Translation System, Overcoming the Language Barrier», *Third European Congress on Information Systems and Networks*, Luxembourg, pp. 569-581.
- VAUQUOIS, Bernard et Christian BOITET (1985) : «Automated Translation at Grenoble University», *Computational Linguistics*, 11-1, pp. 28-36.
- ZINGEL, H. J. (1978) : «Experiences with TITUS II», *International Classification*, 5, pp. 33-37.