

Le système de traduction automatique de l'Université de Montréal (T.A.U.M.)

Groupe de recherche pour la traduction automatique

Volume 18, Number 1-2, mars 1973

Actes du deuxième colloque international de linguistique et de traduction. Montréal, 4-7 octobre 1972

URI: <https://id.erudit.org/iderudit/002834ar>

DOI: <https://doi.org/10.7202/002834ar>

[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

Groupe de recherche pour la traduction automatique (1973). Le système de traduction automatique de l'Université de Montréal (T.A.U.M.). *Meta*, 18(1-2), 277–289. <https://doi.org/10.7202/002834ar>

Le système de traduction automatique de l'Université de Montréal (T.A.U.M.)

Le projet de traduction automatique (T.A.U.M.) est un projet de recherche visant à trouver un modèle mathématique, réalisable sur un ordinateur, du processus de traduction de l'anglais au français. Ce projet qui trouve sa place dans tout le courant des recherches actuelles en linguistique formelle est avant tout destiné à fournir une aide aux traducteurs. De plus en plus, les agences gouvernementales se trouvent débordées par le nombre de textes à traduire. Or, ces textes, le plus souvent dépourvus de métaphores et de tournures de style, se prêtent bien au traitement automatique. On pourrait se servir de l'ordinateur pour donner une première ébauche de traduction, cette ébauche, après révision humaine fournirait une traduction finale.

Dans les pages qui suivent, nous nous proposons de donner une présentation simple et claire de notre modèle de traduction. Nous écarterons les détails qui n'ajouteraient rien à la compréhension du problème linguistique. Parfois aussi, nous renoncerons, pour des raisons d'ordre pédagogique, à une description rigoureusement exacte.

Le projet de recherche T.A.U.M. est né en 1965. Ses premières années ont été consacrées à des travaux préliminaires comme le rassemblement de la documentation existante, l'examen de textes, l'étude théorique de certaines questions linguistiques et l'essai de divers formalismes pour représenter des grammaires.

En 1970, les Systèmes-Q (Québec) sont au point. Ils furent développés ici-même, au projet de traduction automatique, par monsieur Alain Colmerauer. Ce formalisme nous permet de *a*) traiter avec une flexibilité accrue les chaînes d'arbres¹; *b*) découper facilement en grammaires indépendantes les différentes phases du traitement linguistique que nous faisons subir à une phrase. Soulignons enfin, qu'il ne pose aucune difficulté d'apprentissage aux linguistes déjà habitués aux manipulations de grammaires transformationnelles.

I. LE MODÈLE

Le texte anglais à traduire est tapé soit sur cartes, soit à partir d'un télétype. Il doit satisfaire à certaines conventions, comme, par exemple, entre deux mots successifs, il ne doit y avoir qu'un seul blanc.

1. Pour un exemple de structure arborescente, voir p. 283.

L'entrée aura donc la forme suivante:

THE MAN LOVED THE BLOND GIRL.

Pour raisons d'espace, nous ne donnons ici qu'un exemple simple, mais il va sans dire qu'il est possible d'entrer un texte complet. Un programme de pré-édition transformera cette entrée en une chaîne acceptable par les Systèmes-Q. La sortie de ce programme se présente ainsi :

-01- \$(1)+THE+MAN+LOVED+THE+BLOND+GIRL+. -02-

Le traitement linguistique se fera à partir de cette chaîne par un ensemble de règles du type transformationnel, écrites en Système-Q. Nous détaillerons ce traitement plus loin. À la fin du traitement, la phrase se présentera ainsi :

-01- \$(1)+L'+HOMME+AIMAIT+LA+FILLE+BLONDE+. -02-

Puis, par un petit programme de postédition:

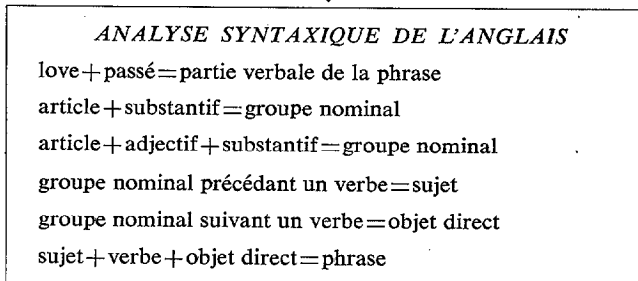
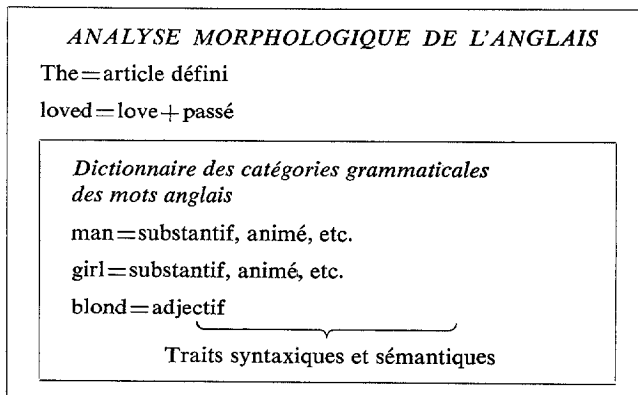
L'HOMME AIMAIT LA FILLE BLONDE.

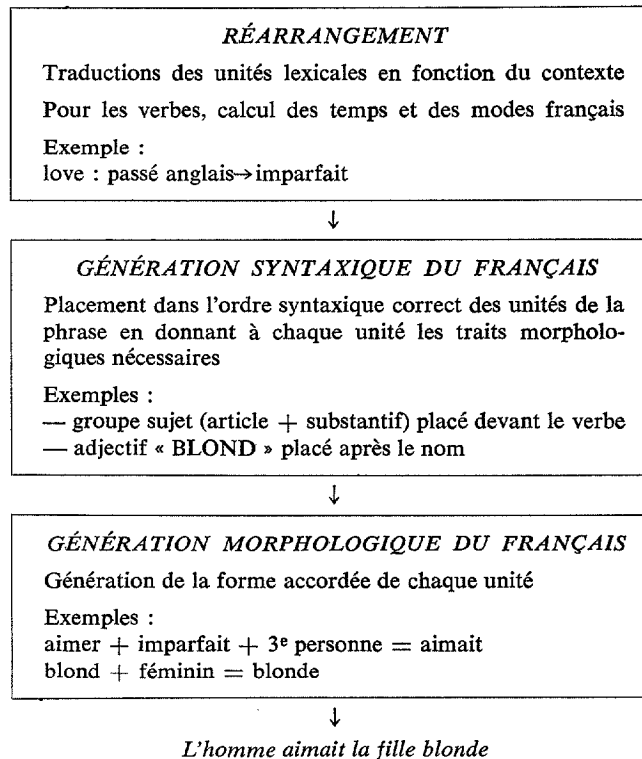
Si, au cours du traitement linguistique, on a trouvé deux ou plusieurs interprétations possibles pour une phrase anglaise, un numéro sera placé dans le texte et renverra à la fin de ce dernier, où, sous le numéro correspondant, on trouvera l'autre (ou les autres) interprétation(s). Le choix de l'interprétation qui sera placée dans le texte se fait arbitrairement.

II. LE TRAITEMENT LINGUISTIQUE

Le traitement linguistique d'une phrase se fait en cinq étapes consécutives :

The man loved the blond girl





A. L'ANALYSE MORPHOLOGIQUE DE L'ANGLAIS

Le but de l'analyse morphologique est de fournir de l'information sur la nature des morphèmes de la phrase d'entrée. Pour ce faire on remplace dans la phrase chacun des mots examinés par un ensemble de renseignements. Ceux-ci seront utilisés aux étapes subséquentes du traitement linguistique.

Des mots comme TENDING et TABLES seront remplacés respectivement par « nom concret TABLE, au pluriel » et « forme progressive de TEND. Ce verbe qui exprime une action qui dure possède deux arguments : le premier peut être un objet concret, abstrait, animé... Le second, obligatoire, peut être concret, abstrait... À l'occasion la préposition TOWARDS introduit ce dernier argument. »

Ces deux exemples montrent que pour faire une analyse morphologique, il faut un dictionnaire. Un dictionnaire qui donne pour chaque mot, la catégorie à laquelle il appartient (« verbe », « nom »...), certains traits sémantiques qui le décrivent (« qui exprime une action qui dure », « qui est concret »...). De plus il faudra pouvoir déterminer la valeur de certaines variables grammaticales comme les temps et la personne pour les verbes, le nombre pour les noms, etc.

Le dictionnaire dont se sert l'analyse morphologique de l'anglais ne doit pas être confondu avec le dictionnaire qui donne les traductions françaises des mots anglais. Nous parlerons de ce dernier à l'étape du réarrangement.

Par mesure d'économie, nous avons décidé d'établir un dictionnaire qui ne comprendrait que certaines formes entières comme les noms au singulier, les verbes à l'infinitif, les adjectifs, etc. Sont exclus les pluriels des noms, les dérivés de verbes en ED et en ING, les comparatifs et les superlatifs d'adjectifs, de même que les adverbes dérivables par adjonction de LY. Nous estimons que le facteur de réduction ainsi obtenu est de 2 ou 3.

L'économie réalisée au chapitre de l'espace et du temps se fait au détriment de la simplicité du programme d'analyse morphologique. Des mots comme TENDING et TABLES n'étant plus dans notre dictionnaire, il nous faudra découvrir qu'ils dérivent de TEND et de TABLE. L'analyse morphologique se fait en trois temps. Voici comment on procède :

1) Dans un premier temps, nous donnons la catégorie de certains mots invariables. Dans le programme d'analyse morphologique effectivement utilisé, on n'emploie pas d'étiquettes aussi explicites et aussi longues que celles employées dans les exemples ; on emploie plutôt des abréviations.

Ex. Prépositions

ABOUT → PRÉPOSITION (ABOUT)

Quand il y a ambiguïté quant à la catégorie d'un mot, on énumère toutes les possibilités. Au cours des étapes subséquentes on déterminera, selon le contexte, laquelle est la bonne. WILL peut être une formule modale, comme dans « HE WILL go ». Il nous faut donc une règle du genre WILL → FORME MODALE (WILL). Mais WILL peut être aussi un nom, comme dans « THIS IS MY WILL ». Pour couvrir cette possibilité on aura la règle WILL → NOM (WILL, /, ABSTRAIT, CONCRET).

2) Dans un deuxième temps, on normalise toutes les formes irrégulières.

Ex. Les verbes au passé

FORGOT → VERBE (FORGET, /, ARGUMENT 1 (ANIMÉ), ARGUMENT 2 (ABSTRAIT, CONCRET, ANIMÉ, PHRASE, PRÉPOSITION (ABOUT)), PERFECTIF, PARTICIPE PASSÉ UTILISABLE COMME ADJECTIF) + ED (PASSÉ)

Le premier argument du verbe est animé. Le deuxième argument peut être abstrait, concret, animé ou une phrase complète ; ce deuxième argument peut être introduit par la préposition ABOUT. FORGET quant à lui est un verbe perfectif et son participe passé est utilisable comme adjectif. La barre oblique sert à séparer l'élément lexical de ses traits.

3) Dans un troisième et dernier temps, on parcourt la phrase en notant tous les mots que l'on n'a pas encore réussi à réécrire sous une forme ou sous une autre. On examine ensuite chacun de ces mots en vue d'essayer de les découper en une base et une désinence. Les seules désinences considérées sont ED, ING, s, LY, ER et EST. Le découpage réussit si l'on trouve la base dans notre dictionnaire.

Ex. Le mot TENDED a sa base TEND dans le dictionnaire. Ce mot se découpera donc comme suit :

TENDED → TEND + ED

TEND + ED → VERBE (TEND, /, ARGUMENT 1 (ANIMÉ, ABSTRAIT, CONCRET, PHRASE), ARGUMENT 2 (ABSTRAIT, CONCRET, PHRASE, OBLIGATOIRE, PRÉPOSITION (TOWARD), DURÉE) + ED

Un mot comme STRING ne sera pas décomposé en VERBE (STR, /,...) + ING parce que STR n'existe pas dans le dictionnaire. Mais STRING au long s'y trouve, nous aurons donc : STRING → NOM (STRING,/,CONCRET). Certaines décompositions exigent un travail préliminaire avant de réussir. HUMBLY, par exemple, a pour base HUMBLE. Il ne faudra pas faire la recherche dans le dictionnaire de la base HUMBL. Pour résoudre ce problème on a recours à une série de règles heuristiques qui se chargent de construire les bonnes bases.

B. L'ANALYSE SYNTAXIQUE DE L'ANGLAIS

L'analyse syntaxique est, évidemment, le centre nerveux de tout le système. C'est à ce niveau que les plus gros problèmes se rencontrent et où le risque d'erreur ou d'échec est le plus grand. Nous allons voir comment fonctionne cette analyse, à l'aide d'un exemple simple, sans toutefois négliger de parler de certains problèmes plus complexes. Pour la phrase *The man loved the blond girl*, la morphologie aura donné la chaîne suivante :

ART (DEF) + N (MAN, /, *H *M) + V (LOVE, /, 1 (*H), 2 (*H, *C, *AB), *ST + ED + ART (DEF) + ADJ (BLOND, /) + N (GIRL, /, *H, *F)

soit une suite de petits arbres (écrits sous forme parenthésée).

*H = humain (animé)

*AB = abstrait

*M = masculin

*ST = *stative*

*C = concret

*F = féminin

Les traits sont précédés de l'astérisque pour les différencier des mots et des étiquettes.

On remarque qu'il est spécifié que LOVE ne peut prendre que 2 arguments et qu'il fait partie des *stative verbs*, c'est-à-dire qu'entre autres caractéristiques, il ne peut prendre la forme progressive : *he is loving* (cette forme peut se rencontrer, mais le sens de LOVE n'est pas le même).

Il s'agira maintenant, à partir de cette chaîne, de reconnaître la structure syntaxique de la phrase. Ceci se fera en regroupant les différents éléments de la phrase dans un arbre dont le sommet sera SENTENCE. La structure à laquelle on arrivera donnera l'information nécessaire à la traduction, avec les moyens que nous possédons actuellement.

L'analyse se fera en 2 étapes : la première, en gros, cherche à former les syntagmes nominaux (NP) ; on regroupera sous l'étiquette NP tout ce qui gravite autour du nom (l'article, les quantificateurs, les adjectifs, etc.) sauf la relative, qui devra, dans l'étape suivante, être analysée comme phrase avant d'être incorporée au NP. Grossièrement, le résultat de cette première étape, pour l'exemple choisi, sera :

NP(N(MAN),DET(ART(DEF)),/,*H,*M) + T(PST) + V(LOVE,/, 1(*H),2(*H,*C,*AB),*ST) + NP(N(GIRL),DET(ART(DEF)),ADJ(BLOND,/),/,*H,*F)

Comme on le voit, au fur et à mesure que l'analyse avance, on a de plus en plus d'information concernant la phrase. Les traits du nom ont été reportés au NP par commodité simplement.

La deuxième étape fera l'analyse du groupe verbal (GOV pour « gouverneur »), des phrases enchâssées (s'il y a lieu) et fera le regroupement final. Un seul bloc de règles s'applique récursivement pour les phrases enchâssées et pour la phrase dominante. Dans le cas de relatives, l'antécédent du pronom relatif sera retrouvé et recopié au bon endroit à l'intérieur de la relative (devant le verbe, s'il s'agit du verbe ; après le verbe, s'il s'agit de l'objet, etc.).

L'analyse du groupe verbal se fait de gauche à droite ; ainsi, dans l'exemple suivant: *He has come four times, has* sera d'abord pris comme verbe, puis en regardant le mot suivant, un réajustement se fera en donnant à *has* la fonction d'auxiliaire. Pour l'exemple *The man loved the blond girl*, le groupe verbal prendra la forme suivante :

GOV(T(PST),OPS,V(LOVE,/,1(*H),2(*H,*C,*AB),*ST))

T = temps

PST = *past*

OPS = opérateurs

Dans notre exemple, aucun opérateur n'agit sur le verbe, donc OPS ne domine rien. Sont rangés sous OPS, dans l'ordre d'apparition en surface, tous les adverbes modifiant le verbe (incluant la négation) ainsi que la marque du passif, la marque du conditionnel, la marque du progressif, etc.

Une fois le gouverneur analysé, on recherchera ses arguments un à la fois. Le NP précédant le verbe sera placé après le gouverneur, on lui donnera l'étiquette NP1 (sujet) et le tout sera dominé par PH. Cette opération sera permise seulement si le ou les traits du verbe sous le chiffre 1 correspondent aux traits du NP considéré ; ainsi, dans notre exemple, on a sous 1 le trait *H, qui exige que LOVE ait un sujet humain, et affectivement, ce trait se trouve parmi ceux de MAN, donc le regroupement a lieu. La même opération se fera pour les NP objets, en suivant la même procédure de correspondance de traits ; le NP objet direct sera marqué NP2 et les objets indirects NP3 et NP4 dans l'ordre d'apparition en surface. Le gouverneur ne peut avoir plus de quatre arguments. NP5 marque le groupe nominal enchâssé dans un autre NP ou contenu dans un circonstant :

NP2(N(DOG),DET(ART(DEF)),GP(P(OF),NP5(PAUL))... (*Paul's dog*)

CIRC(P(ON),NP5((N(WEDNESDAY),/,*PROP,*TM,*AB))... (*on Wednesday*)

DET = déterminant

GP = groupe prépositionnel

P = préposition

CIRC = circonstant

*PROP = propre

*TM = *time*

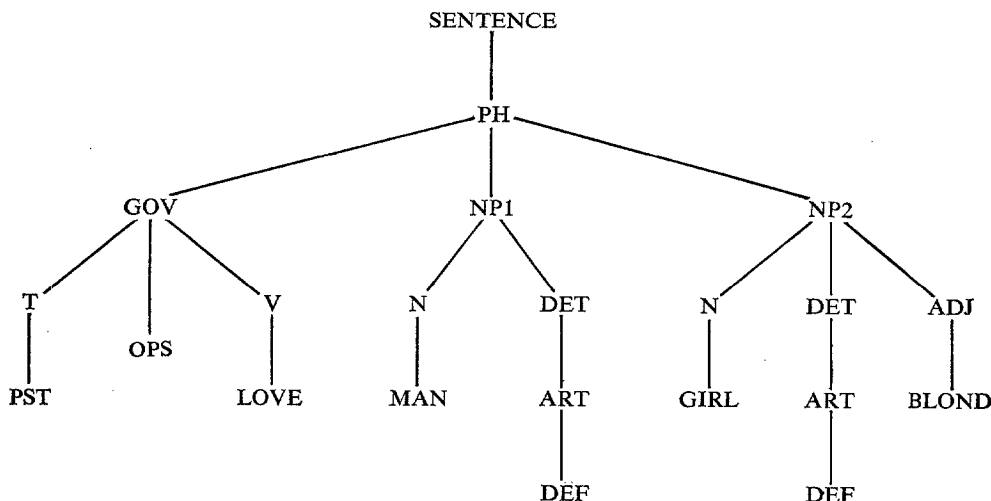
Une fois tous les arguments sous PH, on rattachera s'il y a lieu, les circonstants. La différenciation entre un NP3 ou NP4 et un circonstant est basée sur la préposition introduisant le groupe nominal. Le circonstant, qu'il ait été en tête de phrase, à la fin de la phrase ou à l'intérieur, sera placé, dans la structure normalisée, après les arguments, mais un trait spécial indiquera où il a été trouvé en surface.

La forme finale que prendra notre exemple est donc :

SENTENCE(PH(GOV(T(PST),OPS,V(LOVE,/,*ST)),NP1(N(MAN),DET(ART(DEF)),/,*H,*M),NP2(N(GIRL),DET(ART(DEF)),ADJ(BLOND,/,/,*H,*F)./))

On remarque que les traits du verbe concernant les arguments (ceux sous les numéros) ont été effacés lors du regroupement. La barre oblique à la fin de la

phrase et celle sous ADJ sont là même si dans l'exemple choisi, aucun trait ne fut donné à ces deux composants. L'adjectif peut avoir comme traits des marques du comparatif, superlatif, de temps, de degré, etc. La phrase peut contenir des marques de question, d'impératif, etc. Pour mieux visualiser la structure, nous en donnons la forme arborescente, en omettant les traits, pour simplifier :

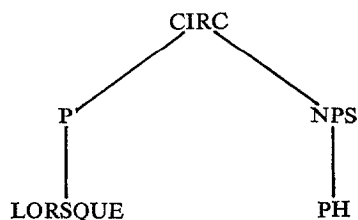


Nous allons maintenant regarder certains problèmes que l'exemple choisi ne montre pas.

Sous l'étiquette DET, on rangera les cardinaux, les ordinaux et les démonstratifs. La conjonction de cardinaux et d'ordinaux est acceptée par le système.

Les adverbes anglais en -LY régulièrement formés sont décomposés en deux parties : la racine adjectivale et un trait *LY qui indique qu'il s'agit d'un adverbe : CLEARLY → CLEAR, *LY. Cette procédure permet une grande économie au dictionnaire anglais-français, puisque seule l'entrée « CLEAR » est nécessaire. La génération du français recomposera l'adverbe en -MENT.

Sous NP, on peut, au lieu d'un nom, trouver un PH. Il s'agira alors d'une complétive sujet ou objet : *I think that Paul came* ; *that Paul came* sera dominé par NP2, puisque cette complétive joue le même rôle qu'un nom. À partir de la même idée, les propositions circonstancielles se verront donner la structure :



La conjonction sera alors seulement une coordination : AND, OR, BUT.

Les phrases passives sont remises sous forme active avec la marque *PSV qui permettra à la génération de redonner la forme passive. De même, les interrogatives sont remises sous forme affirmative avec la marque *QSTN.

Dans la structure normalisée, le verbe BE est inexistant, ceci conformément à une pratique courante en grammaire transformationnelle. Sous l'étiquette GOV, on pourra donc trouver NP, ADJ ou CIRC, si ceux-ci apparaissent en surface dans une structure avec BE.

C. LE RÉARRANGEMENT

Le réarrangement (ou transfert) a pour but d'effectuer le passage entre la structure anglaise normalisée et la structure française normalisée. Ce réarrangement se fera en deux parties :

1) Traduire les étiquettes et les mots

Pour ce faire, l'arbre doit être décomposé en une chaîne, de façon à avoir accès aux mots. Toutefois, toute l'information contenue dans l'arbre est conservée ; par exemple, le sous-arbre NP1 (N (MAN),/,*H) deviendra la chaîne

NP1 + [+ N + [+ MAN +] + / + *H +]

Après le passage du dictionnaire, l'arbre sera recomposé.

2) Modifier certaines parties de l'arbre pour simplifier la génération

La principale différence entre la structure anglaise et la structure française est la notation des fonctions :

NP1 devient P(SUJ) + SN

NP2 devient P(OBJ) + SN

Pour NP3 et NP4, en anglais, la préposition est contenue dans les traits du NP : NP (N (MAN).../, ... P (TO)). Le réarrangement place cette préposition devant le SN : P (A) + SN (N (HOMME)...). Ce changement est effectué pour simplifier la génération où il est fréquent de travailler sur SN sans se préoccuper de la fonction. Pour chaque opération, il aurait fallu écrire une règle pour SN1, une autre pour SN2, etc., alors que par ce procédé, une seule règle sur SN suffit.

En prévision du dictionnaire, certains traits seront transportés directement devant un mot, ce qui permettra une traduction du mot en tenant compte du contexte ; ainsi, les traits du nom sont placés devant la préposition, devant l'adjectif et devant le nom lui-même, et les traits des arguments sont placés devant le verbe. On aura donc, au dictionnaire, des règles du type :

P(U*) + BY = DE /*MSR -DANS- U*
= PAR /NOM --.

U* est la variable qui contient les traits du nom qui accompagnait BY. Le / sépare la règle de la condition. Cette règle se lit donc comme suit : traduire BY par DE si le nom possède le trait *MSR (*measure*) et par PAR si non.

V(U*) + BREAK = ROMPRE /NP2(*AB) -DANS- U*
= CASSER /NON

U* contient les traits des arguments du verbe, identifiés par l'étiquette qui domine l'argument d'où ils sont tirés. Donc les traits du sujet seront dominés par NP1,

ceux de l'objet par NP2, ainsi de suite pour NP3 et NP4. BREAK sera traduit par ROMPRE si l'objet direct (NP2) est un nom abstrait (*AB) et par CASSER si non.

Ces traits placés devant le mot, et dominés d'une étiquette P (préposition), V (verbe), A (adjectif), N (nom), servent aussi à identifier la catégorie syntaxique du mot à traduire ; ainsi

V(U*) + NAME = NOMMER

N(U*) + NAME = NOM

Il va sans dire qu'une bonne partie des règles du dictionnaire sont quand même très simples : FAMOUS = CÉLÈBRE.

Le dictionnaire offre une autre possibilité : celle de donner des traits nouveaux à un mot. Par exemple, on identifiera par la marque *ANTE, les adjectifs qui se placent, en français, devant le nom : GOOD = BON (*ANTE). La recombinaison de l'arbre, après le dictionnaire, placera ce nouveau trait au bon endroit :

ADJ(BON,/,*ANTE)

DREAM = RÊVER(*IND)

Le trait *IND (indicatif) veut dire que RÊVER demande l'indicatif dans une complétive objet : « J'ai rêvé que Paul est venu. » Les verbes auxquels on n'aura pas donné *IND auront leur complétive objet au subjonctif. Ce trait *IND sera transporté, par la recombinaison, à tous les arguments du verbe RÊVER.

Autre exemple : REMEMBER = SOUVENIR (*REF). *REF indique qu'en français, ce verbe est réfléchi.

Si le besoin s'en fait sentir, on peut même utiliser ce procédé pour changer la catégorie syntaxique d'un élément : A (U*) + COLD = AVOIR + FROID (V). La recombinaison donnera V (AVOIR, FROID), en remplacement du sous-arbre anglais ADJ (COLD).

Un changement de préposition sur un argument d'un verbe entre l'anglais et le français peut se faire selon cette procédure : WAIT FOR PAUL → ATTENDRE PAUL. L'entrée du dictionnaire sera : WAIT = ATTENDRE (2 (P (OBJ))). Ceci indique que le deuxième argument de ATTENDRE doit prendre la préposition vide OBJ. FOR sera traduit par le dictionnaire en POUR. La recombinaison effacera ce POUR et le remplacera par OBJ.

Ainsi, toute information nécessaire au français, mais non contenue dans la structure anglaise peut être ajoutée par le dictionnaire dans la structure française.

Pour conclure cette section, reprenons notre exemple :

SENTENCE(PH(GOV(T(PST),OPS,V(LOVE,/,*ST)),NP1(N(MAN),DET(ART(DEF)),/,*H,*M),NP2(N(GIRL),DET(ART(DEF)),ADJ(BLOND,/,/,*H,*F),/))

Cette structure deviendra, après le transfert :

PH(GOV(T(IMPF),OPS,V(AIMER,/,*ST)),P(SUJ),SN(N(HOMME),DET(ART(DEF)),/,*H,*M),P(OBJ),SN(N(FILLE),DET(ART(DEF)),ADJ(BLOND,/,/,*H,*F),/))

Le nœud SENTENCE, employé pour des raisons techniques en anglais, a disparu en français. On voit aussi que le transfert fait le calcul des temps pour le français.

Dans cet exemple, un verbe *stative* avec le temps PST (*past*) sera rendu par un imparfait, alors que les *non-statives* sont rendus par un passé composé.

D. LA GÉNÉRATION SYNTAXIQUE DU FRANÇAIS

La génération a pour tâche de décomposer l'arbre d'entrée pour arriver à une chaîne de mots, syntaxiquement bien formée, en donnant, à la phrase morphologique du français, toute l'information dont elle a besoin.

Il est clair que, pour un arbre à son entrée, la génération ne doit donner qu'une seule chaîne. Lorsqu'un choix syntaxique est possible, comme les formes interrogatives inversées ou avec « Est-ce que », une seule forme est donnée. Il est donc important de contrôler chaque arbre et chaque sous-arbre à chaque étape du parcours qu'il aura à suivre, de manière à ce que les règles ne produisent aucune chaîne ou partie de chaîne non-désirée. Ce contrôle se fait par la modification de l'étiquette dominant l'arbre (ou le sous-arbre) après chaque opération (ou presque) ; on passera donc de PH à PH1, PH2, PH3... PH30, de SN à SN1, SN2,..., de CIRC à CIRC1, etc. Un exemple simple : la règle qui inverse le sujet et l'objet d'un verbe dans le cas des phrases passives s'applique sur une étiquette PH et le résultat de cette règle donne l'étiquette PH1 ; donc la règle ne pourra plus s'appliquer sur ce même arbre puisqu'il ne sera plus jamais dominé par PH. En fait ce procédé permet d'ordonner les règles de Système-Q en génération (les règles de Système-Q n'ont pas d'ordre préétabli).

Essentiellement, la génération consiste en une décomposition progressive de l'arbre d'entrée, en extrayant de ce dernier les éléments dans l'ordre que la surface exige et en partant de la gauche.

En reprenant notre exemple :

PH(GOV(T(IMPF),OPS,V(AIMER,/,*ST)),P(SUJ),SN(N(HOMME),DET(ART(DEF)),/,*H,*M),P(OBJ),SN(N(FILLE),DET(ART(DEF)),ADJ(BLOND,/),/,*H,*F),/))

nous allons suivre, grossièrement, cet arbre à travers la génération :

1) Transmission des traits du sujet au gouverneur. Résultat :

PH6(GOV(T(IMPF),OPS,V(AIMER,/,*ST)),/,*H,*M),P(SUJ),...

2) Sortie du sujet. Résultat :

P(SUJ) + SN(N(HOMME),DET(ART(DEF)),/,*H,*M) + PH13(GOV(T(IMPF),OPS,V(AIMER,/,*ST)),/,*H,*M),P(OBJ),SN(N(FILLE))...

3) Suppression du nœud « PHRASE » (PH30). Résultat :

P(SUJ) + SN(N(HOMME),DET(ART(DEF)),/,*H,*M) + GOV(T(IMPF),OPS,V(AIMER,/,*ST)),/,*H,*M) + P(OBJ) + SN(N(FILLE),DET(ART(DEF)),ADJ(BLOND,/),/,*H,*F)

4) Traitement des SN. Nous suivrons le SN objet seulement ; le SN sujet suivra le même chemin, sauf en ce qui a trait à l'adjectif.

a) Sortie du déterminant. Résultat :

DET(ART(DEF)),/,*H,*F) + SN15(N(FILLE),ADJ(BLOND,/),/,*H,*F)

b) Sortie de l'adjectif. Résultat :

SN17(N(FILLE)),/,*H,*F) + ADJ(BLOND,/),/,*H,*F)

5) Traitement du gouverneur. Résultat :
VERBE(AIMER,/,*H,*M,T(IMPF))

6) Épuration des arbres et formation des déterminants. Résultat :
LE + NOM(HOMME,/,*H,*M) + VERBE(AIMER,/,*H,*M,T(IMPF)) + LA +
NOM(FILLE,/,*H,*F) + ADJECTIF(BLOND,/,*H,*F)

Évidemment, cet exemple est simple. Mais, faute d'espace, nous devons nous limiter ici à expliquer le procédé dans son ensemble. Il faut toutefois ajouter que des boucles ont été prévues, évidemment, pour les phrases enchâssées : une fois que le nœud SN qui les domine est effacé, on se retrouve avec l'étiquette PH et le processus recommence.

E. GÉNÉRATION MORPHOLOGIQUE DU FRANÇAIS

La génération morphologique du français est la dernière étape du traitement linguistique. Les principaux problèmes dont elle s'occupe sont la conjugaison du verbe, le féminin des adjectifs, le pluriel des noms et des adjectifs, l'élision et la contraction. De façon générale, les phrases d'entrée ne comportent que quatre types d'éléments :

- 1) Les étiquettes simples comme
LA LE DONT IL NOUS CETTE DE QUE LEURS ET LES PLUS : , () .
- 2) Les arbres coiffés de l'étiquette NOM comme
NOM(LIVRE,/,MASCULIN,TROISIÈME PERSONNE,PLURIEL)
- 3) Les arbres coiffés de l'étiquette ADJECTIF comme
ADJECTIF(INTELLECTUEL,/,MASCULIN,TROISIÈME PERSONNE,PLURIEL)
- 4) Les arbres coiffés de l'étiquette VERBE comme
VERBE(PARLER,/,ANIMÉ,MASCULIN,TROISIÈME PERSONNE,SINGULIER,
TEMPS(FUTUR SIMPLE))

Le résultat de la génération morphologique sera pour l'exemple :

L + ' + HOMME + AIMAIT + LA + FILLE + BLONDE + .

Il serait fastidieux et trop long de décrire ici toutes les étapes de la morphologie du français ; les opérations à faire sont en effet connues de tous ; contentons-nous de les rappeler : 1) le pluriel des noms et des adjectifs ; 2) le féminin des adjectifs ; 3) le féminin pluriel des adjectifs ; 4) les formes doubles de certains adjectifs (vieux/vieil) ; 5) accord du verbe et des participes (pour le participe passé, on se sert des mêmes règles que pour les adjectifs) ; 6) l'élision (L') et la contraction (DE + LE → DU, etc.). Signalons simplement que la morphologie du français est exhaustive, sauf en ce qui a trait au féminin des noms, qui reste encore à l'étude.

III. PERSPECTIVES D'AVENIR

Nous nous fixons pour but d'accroître la taille de notre dictionnaire jusqu'à 10 000 mots mais ce travail n'est pas pour l'instant prioritaire puisque plusieurs problèmes syntaxiques sont encore à l'étude. Les limitations théoriques du système actuel nous portent à penser que nous ne pourrions pas dépasser un taux de

70% de phrases correctement traduites². En vue de percer ce plafond, nous devons considérer des extensions à notre modèle.

APPENDICE 1

Exemples de traductions

(Le temps indiqué au-dessus des phrases donne la durée nécessaire, dans chaque cas, au traitement linguistique.)

A. Traitement de l'article

1)

5 SEC 149 MSEC

I DON'T WANT MILK. I DON'T WANT ANY MILK. I DON'T LIKE MILK.
JE NE VEUX PAS DE LAIT. JE NE VEUX PAS DE LAIT. JE N'AIME PAS LE LAIT.

2)

3 SEC 176 MSEC

I LIKE SOME TABLES. I WANT SOME TABLES.
J'AIME CERTAINES TABLES. JE VEUX DES TABLES.

B. Pronoms réfléchis

3 SEC 851 MSEC

HE IS WASHING HIS HANDS. HE BROKE HIS LEG.
IL SE LAVE LES MAINS. IL S'EST CASSÉ LA JAMBE.

C. Placement des pronoms devant le verbe

3 SEC 639 MSEC

SHE GAVE HIM THE BOOK. SHE GAVE IT TO HIM.
ELLE LUI A DONNÉ LE LIVRE. ELLE LE LUI A DONNÉ.

D. Traitement des verbes modaux

5 SEC 338 MSEC

THERE COULD BE CATS ON THE ROOF.
IL POURRAIT Y AVOIR DES CHATS SUR LE TOIT.

E. Traduction des mots selon la catégorie de leurs voisins syntaxiques

1)

5 SEC 26 MSEC

I KNOW FRED, BUT I KNOW HE DIDN'T COME YESTERDAY.
JE CONNAIS FRED MAIS JE SAIS QU'IL N'EST PAS VENU HIER.

2)

3 SEC 772 MSEC

JOHN BROUGHT MARY BUT HE DIDN'T BRING A BOTTLE.
JOHN A AMENÉ MARY MAIS IL N'A PAS APPORTÉ UNE BOUTEILLE.

2. C'est-à-dire en conservant l'essentiel de l'information dans une phrase syntaxiquement bien formée.

APPENDICE 2

Exemple d'une phrase donnant lieu à des difficultés

THE FEDERAL AND PROVINCIAL GOVERNMENTS HAVE BEEN INVESTIGATING COMPLETELY NEW WAYS OF LOCATING TIMBER AND MINERAL RESOURCES BY REMOTE MEANS.

1) Combien de gouvernements ?

Un fédéral et un provincial ou un fédéral et plusieurs provinciaux ou plusieurs fédéraux et plusieurs provinciaux ou plusieurs fédéraux et un provincial ?

2) S'agit-il d'« INVESTIGATING COMPLETELY » de « NEW WAYS OF... » ou bien d'« INVESTIGATING » de « COMPLETELY NEW WAYS OF... » ?

3) Que cherche-t-on ?

DES « TIMBER RESOURCES » et des « MINERAL RESOURCES » ou seulement du « TIMBER » et des « MINERAL RESOURCES » ?

4) Veut-on faire du « INVESTIGATING BY REMOTE MEANS » ou du « LOCATING BY REMOTE MEANS » ?

Le genre de problème que cette phrase soulève ne peut être résolu que par la sémantique. Pour bien traduire cette phrase il faut d'abord *comprendre* son sens, la *situer dans son contexte* et *déduire* de tout cela les conclusions appropriées; ce qu'un ordinateur, bien sûr, ne peut faire.

GROUPE DE RECHERCHE POUR
LA TRADUCTION AUTOMATIQUE