

METEO : un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public

John Chandioux

Volume 21, numéro 2, juin 1976

URI : <https://id.erudit.org/iderudit/002214ar>

DOI : <https://doi.org/10.7202/002214ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)

1492-1421 (numérique)

[Découvrir la revue](#)

Citer cet article

Chandioux, J. (1976). METEO : un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public. *Meta*, 21(2), 127–133. <https://doi.org/10.7202/002214ar>

METEO :

un système opérationnel pour la traduction automatique des bulletins météorologiques destinés au grand public

INTRODUCTION

Le projet T.A.U.M. (Traduction automatique à l'Université de Montréal) travaille depuis près de six ans à la mise au point d'un modèle expérimental pour la traduction de l'anglais vers le français. Un tel système est étroitement tributaire des progrès de la linguistique — plus particulièrement de la sémantique — et suppose la compilation de quantités considérables de données syntaxiques et lexicales. Pour cette raison, il n'est pas possible d'envisager actuellement, ni même à court terme, un système opérationnel pour la traduction de textes quelconques.

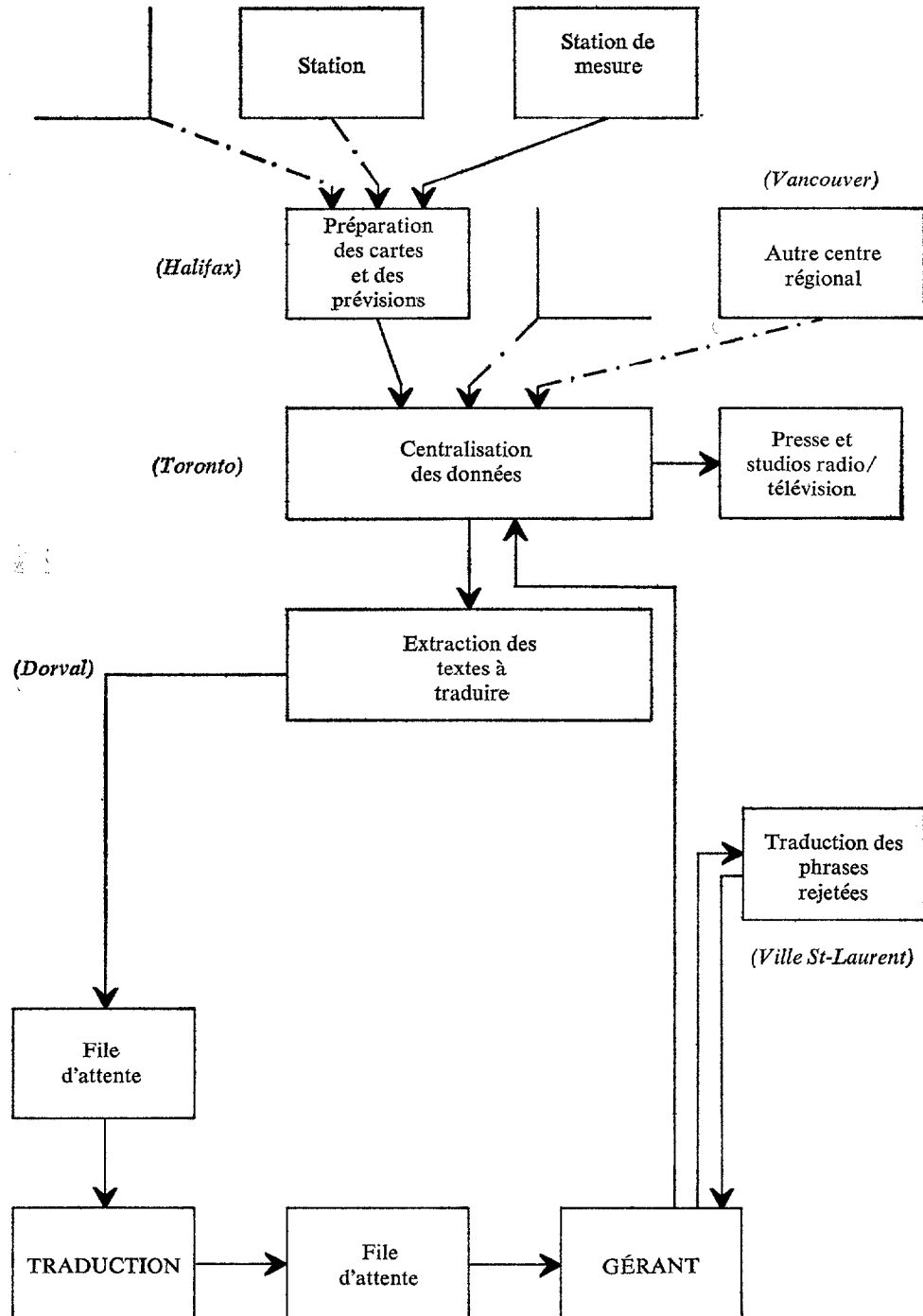
Il est par contre envisageable de réaliser des systèmes répondant à des besoins spécifiques. Le système METEO en est une illustration. De plus, nous venons de terminer avec le Bureau des traductions (gouvernement fédéral) une expérience de deux mois visant à démontrer la supériorité d'un traitement automatique dans le domaine de la traduction technique. Il faut dire que la terminologie technique se développe à une rapidité telle que les organismes ayant besoin de traduction ont souvent plus confiance en un bon technicien partiellement bilingue qu'en un bon traducteur. Il ne s'agit pas ici de faire du style mais uniquement de transmettre le message avec la plus grande fidélité possible.

DESCRIPTION DU SYSTÈME

METEO est un système entièrement automatique pour la traduction des bulletins météorologiques destinés au grand public à travers tout le Canada. Il est implanté au Centre météorologique canadien (C.M.C.) sur une base expérimentale depuis le mois de décembre dernier et doit être livré au ministère de l'Environnement le 15 mai 1976.

Les bulletins météorologiques sont préparés dans des centres régionaux (figure 1) à partir des données envoyées par des stations de mesure et acheminées sur le réseau de communication CN/CP. À Toronto, plaque tournante du réseau CN/CP, les communications concernant la météorologie sont extraites du flot et canalisées vers l'ordinateur du C.M.C. à Dorval.

FIGURE 1



Là, un nouveau tri est effectué car le système METEO a été conçu pour la traduction de l'anglais au français des bulletins météorologiques destinés au grand public ; les bulletins rédigés en français et les situations générales sont envoyés directement au Bureau des traductions pour traduction humaine, et les données numériques sont prises en charge par le programme de simulation du C.M.C.

Les communications retenues sont placées dans un fichier structuré en file d'attente où elles sont récupérées à intervalles réguliers par le système METEO proprement dit. Il n'y a aucune intervention humaine avant la traduction, chaque communication est acceptée telle quelle et livrée après traitement à un moniteur interactif.

Ce dernier parcourt la communication afin de voir si des phrases ont été rejetées par le programme de traduction et, le cas échéant, affiche les phrases correspondantes sur un terminal à écran au Bureau des traductions de Ville Saint-Laurent. Là, un traducteur humain les traduit par la machine. Une fois une communication ainsi complétée, le même réseau de communication CN/CP assure sa distribution aux différents organismes concernés. On remarquera qu'il n'y a aucune révision manuelle des phrases acceptées et traduites par le système ; à notre connaissance, c'est la première fois que le produit d'un système de traduction automatique sera distribué directement au public (les gros systèmes russe/anglais fonctionnant actuellement aux États-Unis ne sont pratiquement utilisés que par des techniciens désireux de voir si un texte étranger est susceptible de les intéresser ou non).

Les phrases rejetées par le système le sont pour les raisons suivantes : faute d'orthographe, erreur de transmission, mot inconnu du dictionnaire, anglais incorrect, structure syntaxique non reconnue par l'analyseur, etc. Une fois totalement opérationnel, le système traduira 30 000 mots par jour à la vitesse de 1 000 mots/minute. Les phrases rejetées représentent actuellement moins de 20 % des entrées et le coût total est évalué au tiers de celui d'une traduction humaine.

L'APPROCHE LINGUISTIQUE

Le système METEO se divise en deux programmes, le programme de traduction et le moniteur interactif. La partie traduction comprend quatre grammaires écrites dans un langage informatique de haut niveau spécialement conçu pour les linguistes :

1. le dictionnaire d'idiomes
2. le dictionnaire principal
3. l'analyseur
4. le générateur

1. *Le dictionnaire d'idiomes*

Le dictionnaire d'idiomes comprend environ 300 entrées réparties en trois catégories :

a) Plusieurs idiomes :

clear period → éclaircie

clouding over → se couvrant

b) Quelques chaînes de mots qui ne sont pas analysées pour des raisons de performance car elles figurent obligatoirement dans toutes les communications :

forecast issued by the atmospheric environment service ; end par 1 of 2

c) Une majorité de noms de lieu devant être traduits ou ayant une traduction imprévisible :

Lake St Claire → Lac Ste Claire

Indian lake → Lac *des* Indiens

Thunder Bay → Thunder Bay (et non « baie »)

2. *Le dictionnaire principal*

Le dictionnaire principal contient environ 1 200 entrées et variantes morphologiques. Il contient toute l'information lexicale nécessaire tant pour l'analyse que pour la génération et associe à chaque mot les catégories syntaxiques potentielles, à chaque catégorie la ou les traductions et à chaque traduction les traits sémantiques correspondants. Les variantes morphologiques sont entrées directement au dictionnaire pour raison d'économie car elles sont comparative-ment peu nombreuses.

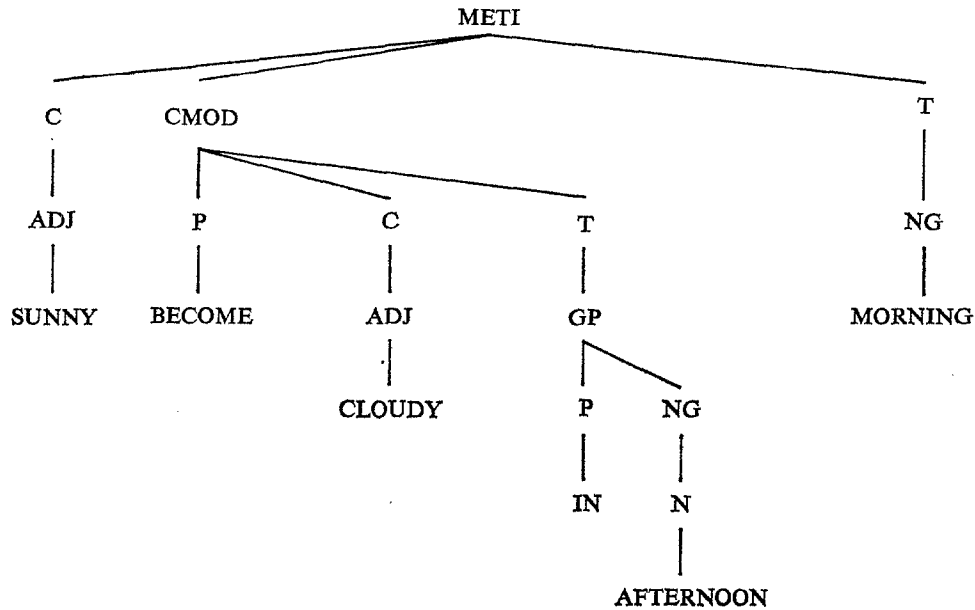
3. *L'analyseur*

L'originalité du système METEO réside en grande partie dans la stratégie d'analyse utilisée. Les textes météorologiques ont leurs caractéristiques propres. Sur le plan syntaxique, les phrases sont simples, pas de relatives ni de passives par exemple. Le problème essentiel est le découpage de la phrase en syntagmes en raison du style essentiellement télégraphique et de la prolifération de conjonctions. Dès le départ, il s'est avéré évident qu'une analyse syntaxique conventionnelle serait insuffisante car cette dernière repose en grande partie sur la présence de mots fonctionnels souvent escamotés en style télégraphique.

Nous avons donc opté pour une approche plus sémantique et construit un analyseur effectuant l'analyse en plusieurs phases.

Le but de l'analyseur est de faire correspondre à chaque phrase d'entrée une structure unique donnant pour chaque mot la catégorie et la traduction réalisées (figure 2).

FIGURE 2



« Sunny this morning clouding over in the afternoon »

— Dans un premier temps, les sous-chaînes de mots contenant des nombres sont reconnues comme étant des dates, des heures ou des températures.

— Ensuite on passe à la reconnaissance des unités de temps ou de lieu. Dans le cas des unités de temps, il est nécessaire de distinguer le point de vue duratif du point de vue ponctuel :

in the morning → dans la matinée
 this morning → ce matin ;

il faut aussi distribuer l'article lorsque nécessaire :

this afternoon or evening → cet après-midi ou ce soir ;

et faire converger différentes formes vers une même traduction :

near dawn tomorrow } → demain à l'aube.
 tomorrow near dawn }

Pour ce qui est des unités de lieu, le principal problème est l'identification correcte des mots non reconnus par le dictionnaire. En effet, les noms de lieu ne devant pas être traduits ne figurent pas au dictionnaire en raison de leur très grand nombre.

— Lors d'un troisième passage, les sous-chaînes restantes sont analysées. Les règles correspondantes s'appuient surtout sur les sous-catégorisations sémantiques définies au dictionnaire pour choisir la traduction correcte d'un mot :

heavy fog → brouillard généralisé

heavy rain → forte pluie

ou la portée des conjonctions :

snowflurries or rainshowers becoming intermittent tonight →

(snowflurries or rainshowers) becoming...

Il est par exemple nécessaire de répartir les conditions atmosphériques en phénomènes stationnaires, précipitations et vents pour guider l'analyse.

— Les suites de conditions, d'unités de temps et d'unités de lieu sont alors vérifiées et chaque fois qu'une structure unique et bien formée peut être construite cette dernière est retenue et passée au générateur.

— Enfin, les structures partielles ou analyses incorrectes sont purgées et on en profite pour effectuer quelques ajustements « stylistiques ». Le mot « vallée », lorsqu'il figure comme nom de lieu dans une communication en provenance d'Halifax, doit par exemple être transformé en « vallée d'Annapolis ». Un cas intéressant est celui de « occasional » entré au dictionnaire comme signifiant « passer » lorsqu'il entre dans des instructions comme :

occasional cloudy periods → * *passages nuageux passagers*

il faut alors supposer que le météorologue fait allusion à une distribution dans l'espace :

occasional cloudy periods } → passages nuageux isolés
isolated cloudy periods }

4. *Le générateur*

Le générateur effectue la décomposition de l'arborescence construite par l'analyseur en introduisant les articles là où nécessaire et en tenant compte de l'ordre des mots ou des syntagmes en français.

La traduction est alors terminée et on peut juger de la qualité des sorties avec l'exemple suivant :

LOWER ST JOHN RIVER VALLEY
UPPER ST JOHN RIVER.
WIND WARNING ENDED BOTH REGIONS.
SNOW AND BLOWING SNOW TONIGHT
BECOMING INTERMITTENT NEAR DAWN
FRIDAY. CLOUDY WITH PERIODS OF
LIGHT SNOW FRIDAY. STRONG
GUSTY NORTHEASTERLY WINDS TONIGHT

BECOMING NORTHWESTERLY WINDS
FRIDAY AFTERNOON. LOWS TONIGHT
MID 20S. HIGHS ON FRIDAY 30
TO 35. OUTLOOK FOR SATURDAY
SUNNY.

VALLÉE DU BAS ST JEAN
HAUT ST JEAN.
FIN DE L'AVIS DE VENT POUR LES
DEUX RÉGIONS. CETTE NUIT NEIGE
ET POWDRERIE DEVENANT PASSAGÈRES
VENDREDI À L'AUBE. VENDREDI
NUAGEUX AVEC FAIBLES CHUTES DE
NEIGE PASSAGÈRES. CETTE NUIT
VENTS FORTS DU NORD OUEST SOUFFLANT
EN RAFALES DEVENANT VENTS FORTS
DU NORD OUEST VENDREDI APRÈS-MIDI.
MINIMUM CETTE NUIT 24 À 26.
MAXIMUM VENDREDI 30 À 35. APERÇU
POUR SAMEDI ENSOLEILLÉ.

CONCLUSION

Le système METEO ne saurait être utilisé pour la traduction de textes autres que des bulletins météorologiques, puisqu'il repose sur une sémantique très restreinte. Cependant la stratégie décrite est certainement applicable à d'autres domaines où le volume de textes à traduire permet d'envisager l'investissement dans un système spécialisé.

JOHN CHANDIOUX