

QUEEN
QA
76.9
.N38
F47514
1989



Gouvernement du Canada
Ministère des Communications

Government of Canada
Department of Communications

Le Centre canadien de recherche sur l'informatisation du travail
Canadian Workplace Automation Research Centre

DE DE L'UTILISATEUR DU SYSTÈME SUN-TULI

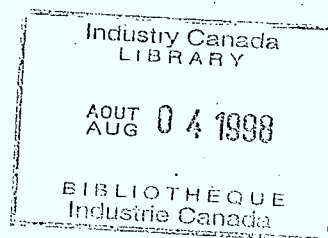
par

/ Ron Ferguson /

QA
76.9
N38
F47514
1989
C39

Canada

QA
76.9
N38
F475#1
1989
c.3



2. GUIDE DE L'UTILISATEUR DU SYSTÈME SUN-TULIP

par

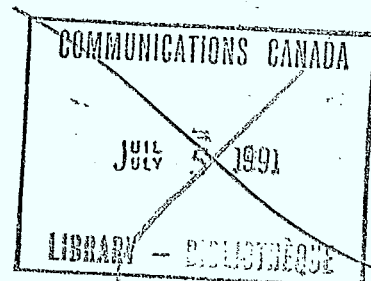
/ Ron Ferguson /

Ce rapport est le résultat de travaux de recherche effectués au Centre canadien de recherche sur l'informatisation du travail sous la direction de Joël Muzard, Groupe systèmes experts.

Ministère des Communications

Laval

Novembre 1989



Nº de cat. Co . 28-1/41-1989F

ISBN 0-662-95735-0

Les opinions émises dans ce rapport n'engagent que l'auteur.

* This report is also available in English

QA
76.9
N38
F475
1989
0.3

DD938272
DL1064098

Table des matières

<u>Introduction.....</u>	3
<u>Niveaux de compréhension.....</u>	4
<u>La compréhension conceptuelle.....</u>	5
<u>La compréhension syntaxique.....</u>	5
<u>L'association de mots.....</u>	6
<u>Limite de la compréhension conceptuelle du système Tulip.....</u>	7
<u>Procédures d'entrée et de sortie.....</u>	7
<u>Mise en application de la fonction d'aide du système Tulip.....</u>	8
<u>Suspension de l'exécution du programme Tulip.....</u>	9
<u>Données d'entrée pour le système Tulip.....</u>	10
<u>Les instructions.....</u>	13
<u>Forme des assertions en anglais admissibles.....</u>	14
<u>Forme des interrogations en anglais admissibles.....</u>	15
<u>Réponse aux interrogations par appariement de mots.....</u>	16
<u>Formulation des énoncés :</u>	
<u>Conjonction.....</u>	17
<u>Les syntagmes nominaux.....</u>	17
<u>Les déterminants.....</u>	19
<u>Les pronoms.....</u>	21
<u>Les verbes.....</u>	22
<u>Les noms.....</u>	23
<u>La date et l'heure.....</u>	24
<u>Les opérations arithmétiques.....</u>	24
<u>Les messages d'erreur.....</u>	24
<u>Les textes en français.....</u>	25
<u>L'expression des connaissances dans le système Tulip.....</u>	26

Introduction

Le système TULIP (Text Understanding and Logical Inferencing Processor) est une interface de langage naturel pour un bloc-notes électronique.

L'utilisateur peut se servir d'un sous-ensemble de la langue anglaise pour faire des assertions et poser des questions au système. Il peut également introduire dans le système des règles décrivant des relations entre des données d'entrée.

En plus des connaissances introduites par l'utilisateur, le système TULIP a une base de connaissances incorporée qui porte sur certains domaines : il peut interpréter les cartes professionnelles, la date et l'heure ainsi que certaines opérations arithmétiques.

Ce guide décrit le fonctionnement du système TULIP sur les postes de travail Sun et a été écrit en supposant que le lecteur connaît les rudiments des postes de travail Sun.

Dans ce guide, nous donnons des exemples d'informations admissibles et NON admissibles par le système TULIP. Par convention, les informations non admissibles par le système TULIP seront précédées de [*]. Exemple :

[*] Green ideas sleep furiously.

Niveaux de compréhension

On peut dire que le système TULIP comprend l'anglais en ce sens qu'il peut utiliser des assertions et des règles pour trouver des réponses à des interrogations. Toutefois, l'utilisateur ne peut s'attendre à ce que le système TULIP comprenne l'anglais aussi bien qu'un être humain. Pour qu'un programme machine comprenne un langage naturel (tel que l'anglais) aussi bien qu'un être humain, il faudrait qu'il ait l'expérience, les connaissances et les capacités d'inférence d'un être humain. A l'heure actuelle, ce n'est le cas d'aucun programme.

Le système TULIP a trois niveau de compréhension, ce qui permet d'obtenir une réduction progressive de la puissance de traitement. Si le système ne peut comprendre un énoncé au niveau de compréhension le plus élevé, il recommence l'opération au niveau inférieur immédiat. Par ordre de décroissance, les niveaux de compréhension sont les suivants :

- [1] la compréhension conceptuelle;
- [2] la compréhension syntaxique;
- [3] l'association de mots.

La compréhension conceptuelle

A ce niveau de compréhension, le système TULIP peut faire la relation entre des façons différentes de dire la même chose, même lorsqu'on utilise des mots différents. Par exemple, les assertions :

Mr. John Smith works for Kentek.

Kentek employs Mr. John Smith.

ont le même sens pour le système TULIP. Celui-ci sait qu'il y a une relation entre les expressions *works for* et *employs*.

La compréhension syntaxique

A ce niveau de compréhension, le système TULIP peut analyser une phrase et faire le lien entre des phrases qui ont le même sens, mais des structures syntaxiques différentes. Par exemple, les phrases :

Kentek employs Mr. John Smith.

Mr. John Smith is employed by Kentek.

fournissent la même information au système TULIP. La deuxième phrase est la forme passive de la première.

Le système comprend également des façons différentes d'exprimer une même relation de possession.

Par exemple :

the father of Jack

Jack's father

Le système TULIP peut également faire le lien entre une interrogation et une assertion lorsque celles-ci utilisent les mêmes expressions.

Exemple :

Kentek employs Mr. John Smith.

Who is Mr. John Smith employed by?

L'association de mots

Lorsqu'on pose au système TULIP une question qu'il ne peut analyser, le système trouve les assertions introduites précédemment qui contiennent le plus grand nombre de mots en commun avec la question. Ces assertions sont affichées comme réponses possibles à l'interrogation. Cette méthode simple permet assez souvent de trouver l'information désirée par l'utilisateur, bien qu'on puisse occasionnellement obtenir en même temps des informations non pertinentes.

Limites de la compréhension conceptuelle du système TULIP

Les fonctions de compréhension conceptuelle du système TULIP ont une capacité limitée. Au niveau syntaxique, le système peut comprendre la plupart des assertions introduites par l'utilisateur, mais non au niveau conceptuel. Par conséquent, il faut utiliser une terminologie cohérente. L'utilisateur peut accroître la compréhension conceptuelle du système en introduisant des règles qui définissent les relations entre les mots.

Procédures d'entrée et de sortie

Pour utiliser le système TULIP, il faut se trouver dans le répertoire :

/home/condor2/tulip

Charger TULIP en tapant :

tulip.bin

Vu la longueur du programme TULIP, le chargement du programme prend un certain temps. Lorsque le programme est chargé, le message de guidage Quintus Prolog suivant est affiché :

| ?-

Pour entrer dans le système TULIP, taper :

tulip.

Lorsque Tulip est prêt à recevoir les données de l'utilisateur, il affiche le message de guidage :

Enter a sentence :

Pour sortir du système Tulip, taper :

exit. (ou) quit.

Ceci fait afficher de nouveau le message de guidage Quintus Prolog :

| ?-

Pour sortir de Quintus Prolog, taper :

halt.

Mise en application de la fonction d'aide du système Tulip

Une fois à l'intérieur de Tulip, pour mettre en application la fonction d'aide du système Tulip, taper :

help.

Suspension de l'exécution du programme TULIP

L'utilisateur peut suspendre une exécution en cours du programme TULIP en tapant C (Control-C). Le message de guidage suivant est alors affiché :

Prolog interruption (h for help)?

Pour sortir du programme TULIP, l'utilisateur doit taper :

a

L'utilisateur revient alors au message de guidage Quintus Prolog. Il peut de nouveau entrer dans le programme Tulip en tapant :

tulip :

Données d'entrée pour le système Tulip

Les données d'entrée pour le système Tulip sont les suivantes :

- [1] Les instructions - Une instruction est une ligne qui commence par un mot-clé (par exemple *read*, *save*, *restart*) et se termine par un point suivi d'un retour du curseur. Les instructions commandent au programme Tulip d'exécuter certaines opérations spéciales.

Exemple :

restart.

- [2] Les assertions - Une assertion est une ligne en anglais constituant une déclaration de fait et se terminant par un point suivi d'un retour du curseur. Tulip conserve les assertions en mémoire et les utilise pour répondre aux interrogations. Exemple d'assertion :

Bill is the husband of Mary.

Si le premier mot de la phrase n'est pas un nom propre, il n'est pas nécessaire que la première lettre soit une majuscule. Par exemple :

a cat is on the mat.

- [3] Les interrogations - Une interrogation est une ligne en anglais ayant la forme d'une question terminée par un point d'interrogation suivi d'un retour du curseur. Lorsque l'utilisateur introduit une interrogation dans le système Tulip, celui-ci cherche une réponse en se basant sur l'information qui lui a été communiquée. Exemple d'interrogation :

who is Mary's husband?

Il n'est pas nécessaire de mettre une majuscule au premier mot d'une interrogation.

- [4] Les règles - Une règle est une ligne en langage évolué de la forme :

<Conclusion> si <Prémisse1> (et <Prémisse 2>...).

où la <Conclusion> et les <Prémisses> sont des énoncés simples pouvant contenir certaines *variables* servant de substituts à des objets. Une variable est une lettre majuscule autre que A ou I. Les règles sont utilisées pour établir les relations entre les concepts. Exemple de règle :

X is the wife of Y if Y is the husband of X.

En utilisant cette règle et l'assertion suivante :

Bill is the husband of Mary.

le système Tulip peut alors répondre à la question :

Who is Bill's wife?

Le sens des verbes *have* et *is* dépend du type de complément du verbe. Par conséquent, on ne peut PAS utiliser une variable comme complément des verbes *have* et *is* :

[*] X has Y if X owns Y.

[*] X is a Y if the type of X is Y.

Toutefois, on peut utiliser une variable comme sujet des verbes *have* et *is* :

X has a window office if X is a manager.

X is a mammal if X has fur.

Une règle ne comportant qu'une seule variable peut souvent être exprimée sous forme d'assertion. Par exemple, les règles précédentes peuvent être exprimées sous les formes suivantes :

Every manager has a window office.

Everything that has fur is mammal.

- [5) Les cartes professionnelles - Lorsqu'une ligne introduite ne se termine pas par un point ou un point d'interrogation, Tulip suppose qu'il s'agit de la première ligne d'une carte professionnelle. Si la

première ligne d'une carte professionnelle se termine par un point, il faut laisser un espace entre le point et le retour du curseur.

Tulip affiche le message de guidage suivant pour indiquer qu'il s'attend à ce qu'on introduise d'autres lignes pour la carte professionnelle :

I:

Une carte professionnelle doit se terminer par une ligne vierge.

Les instructions

save <nom> .

- sauvegarde des phrases tapées par l'utilisateur depuis la dernière instruction *Save* ou *Restart*. Les phrases sauvegardées sont identifiées par un <nom> , lequel est un mot alphanumérique. Si ce <nom> a été utilisé antérieurement avec une instruction *Save*, les phrases sont ajoutées à celles déjà identifiées par ce <nom>. On peut lire les phrases sauvegardées en utilisant l'instruction *Read*.

read <nom> .

- lecture des phrases sauvegardées à l'aide de l'instruction *Save* <nom> . Remarque : l'instruction *Save* <nom>. conserve les phrases dans un fichier (appelé <nom> .text) du répertoire **knowledge** de Tulip. On peut introduire un fichier de texte

dans le répertoire **knowledge** et le lire à l'aide de l'instruction *Read*. Ces fichiers peuvent être modifiés au moyen d'un programme d'édition.

restart.

- effacement du contenu de la mémoire de Tulip.debug.

debug.

- affichage du traitement d'une phrase par Tulip. Les données affichées sont très détaillées et ne sont probablement pas très utiles à un utilisateur moyen.

end debug.

- suppression de l'affichage détaillé du traitement des phrases.

help.

- assistance pour l'utilisation de Tulip.

Forme des assertions anglaises admissibles

En général, Tulip accepte les énoncés anglais qui sont des déclarations de fait.

Les énoncés sous forme négative ne sont pas admissibles. On ne peut introduire un énoncé comportant le mot *not*. Exemple :

[*] George is not a programmer.

Les énoncés comportant le mot *or* sont également inadmissibles.

Exemple :

[*] Helen or Mary is married to Jim.

Une assertion doit commencer par un syntagme nominal sujet de l'énoncé.

Forme des interrogations anglaises admissibles

Les types de question *Wh* et *yes/no* sont tous les deux admissibles. Une question *Wh* est une question qui commence par un mot comme : *who*, *what*, *which*, *where* ou *when*. L'expression *How many* est également considérée comme un mot *Wh*. Un mot *Wh* peut être précédé d'une préposition. Exemples d'interrogations *Wh* :

Who is Frank's wife?

To whom is Frank married?

How many people work for Kentek?

Which employees live in Laval?

Notons que le mot *Wh* doit être le premier mot de la phrase (ou le deuxième s'il est le complément d'une préposition). L'interrogation :

[*] Jack is married to whom?

n'est donc pas admissible.

Pour la rendre admissible, il faudrait l'exprimer sous la forme :

To whom is Jack married?

ou

Whom is Jack married to?

Le mot *how* est inadmissible s'il est utilisé seul :

Jack runs quickly.

[*] How does Jack run?

Une question *yes/no* est une question à laquelle on peut répondre par *yes* ou *no*. Ces questions commencent ordinairement par le verbe *is* ou *does*.

Exemples :

Does Frank work for Kentek?

Is Frank a manager?

Réponse aux interrogations par association de mots

Si l'on pose à Tulip une question qu'il ne peut analyser, il a recours à l'association de mots pour tenter d'obtenir des informations utiles. Par exemple, si l'on introduit des mots-clés suivis d'un point d'interrogation, Tulip affiche les assertions qui contiennent le plus grand nombre de mots-clés.

Les conjonctions

Le mot *and* ne peut être utilisé que de certaines façons. On peut l'utiliser entre des énoncés, par exemple :

Joe lives in Toronto and Henry lives in Montreal.

Si le sujet est le même dans les deux énoncés, il peut être omis dans le second :

Joe lives in Guelph and works in Toronto.

La conjonction *and* peut également être utilisée entre des clauses relatives (voir Syntagmes nominaux).

Les syntagmes nominaux

Un syntagme nominal peut être un syntagme nominal de nom commun, par exemple :

the tall man

ou un nom propre, par exemple :

John Doe

La première lettre d'un nom propre doit être une majuscule.

Un syntagme nominal de nom commun (mais non un nom propre) peut être modifié par une clause relative ou plusieurs clauses relatives jointes par des conjonctions. Par exemple, l'expression :

the man who lives in Halifax and is employed by Acme Inc.

constitue un syntagme nominal.

Les relatives réduites, comme par exemple :

[*] the man employed by Acme Inc.

ne sont pas admises par Tulip.

Celle-ci doit être présentée sous la forme :

the man who is employed by Acme Inc.

Un syntagme nominal de nom commun peut être singulier ou pluriel.

Exemples :

the cat

the cats

Toutefois, les syntagmes nominaux reliés par une conjonction ne sont pas admissibles.

Par exemple :

[*] the men and women

[*] Bill and Jane

ne sont pas admissibles.

Les mots compris entre guillemets sont traités comme un nom propre.

Exemple :

"The Young and the Restless" is a television program.

Les déterminants

Les syntagmes nominaux de noms communs sont généralement précédés d'un déterminant tel que *a*, *the*, *some*, *all* ou *every*. Les déterminants *most* et *few* ne sont pas admissibles.

Lorsqu'une assertion commence par le déterminant *a*, par exemple :

a bear has claws,

Tulip ne peut déterminer si le sens de l'assertion est :

some bear has claws ou every bear has claws.

Dans ce cas, Tulip affiche un menu où figurent ces deux options et demande à l'utilisateur de choisir le sens approprié.

Si le déterminant *a* est utilisé ailleurs qu'au début d'une phrase ou est utilisé dans le sens *some* au début d'une phrase, Tulip suppose qu'il est question d'un nouvel objet. Par contre, si le déterminant *the* est utilisé, Tulip suppose qu'il est question d'un objet mentionné dans une phrase précédente. Par exemple, si l'énoncé :

George has a car

a déjà été introduit et que l'utilisateur introduise l'énoncé :

Helen likes the car,

Tulip suppose que *the car* est la dernière voiture mentionnée, c'est-à-dire celle de George.

Par contre, si l'utilisateur introduit l'énoncé :

Helen likes a car,

Tulip ne suppose pas qu'Helen aime la voiture de George.

Le seul déterminant admissible devant un nom propre est *the*.

Exemples :

The Volvo

[*] a Volvo

Les pronoms

Tulip tente d'identifier les référents des pronoms singuliers à la troisième personne tels que *he, she, him, her* et *it*. Il tente également d'identifier le référent des pronoms possessifs singuliers à la troisième personne tels que *her* et *his*. Exemple :

Jack works for Acme Inc.

He owns a car.

Alice is his wife.

She gave him his coat.

who owns a car?

who is Jack's wife?

what did Alice give to Jack?

Dans l'exemple ci-dessus, Tulip suppose que le pronom *he* de la deuxième phrase désigne *Jack*, car il est la dernière personne de sexe masculin qui a été mentionnée.

Les référents des pronoms tels que *they, them* et *their* ne font PAS l'objet d'une tentative d'identification.

Les verbes

Les verbes au présent, au passé ou au futur sont admissibles. Toutefois, le temps des verbes n'est utilisé que pour interpréter les dates. Dans toute autre situation, Tulip ne tient pas compte du temps du verbe.

Les verbes modaux (ceux qui modifient d'autres verbes) tels que *can*, *should*, etc., sont également admissibles, mais sont négligés par le système.

Les phrases suivantes ont donc toutes le même sens pour Tulip :

Tweety can fly.

Tweety should fly.

Tweety could fly.

Tweety will fly.

Tweety flies.

À l'exception des verbes modaux, Tulip ne peut analyser les phrases contenant des verbes ayant d'autres verbes comme arguments (par exemple, *persuader*) :

[*] Tim persuaded Tom to hire Mary.

Les noms

Tulip peut parfois déterminer automatiquement quel type d'objet un nom propre désigne. Tulip connaît les prénoms masculins et féminins. Il utilise ses connaissances pour tenter de déterminer si les personnes mentionnées dans une phrase sont des hommes ou des femmes. De même, il suppose que les noms propres se terminant par *Limited*, *Incorporated* ou *Company* sont des sociétés. Il connaît également les noms de villes, de provinces et d'États.

Lorsqu'un nom apparaît dans une phrase, Tulip tente de l'associer à un objet qu'il connaît déjà. Il n'est pas nécessaire que ce nom soit absolument identique au nom déjà établi pour l'objet; toutefois, le nouveau nom doit être un sous-ensemble du nom établi. Par exemple, si Tulip connaît une personne appelée *Mr. Jack Smith*, lorsque le nom *Jack* ou *Mr. Smith* est utilisé, Tulip suppose qu'il s'agit de *Mr. Jack Smith*.

Lorsqu'un nom peut désigner plus d'une personne connue de Tulip, ce dernier choisit celle qui a été mentionnée en dernier lieu. Ainsi, si *Mr. Jack Jones* a été mentionné plus récemment que *Mr. Jack Smith*, l'utilisation du nom *Jack* renvoie le système à *Mr. Jack Jones*. Lorsqu'une personne est mentionnée pour la première fois, on doit utiliser la forme la plus complète de son nom car cette personne ne pourra être désignée que par ce nom ou ses sous-ensembles par la suite. Un objet ne peut avoir plus d'un nom dans le système Tulip.

La date et l'heure

Voir la démonstration *dates* pour des exemples d'utilisation de la date et de l'heure par le système Tulip.

Les opérations arithmétiques

Voir la démonstration *cars* pour des exemples d'utilisation des termes *average*, *maximum*, *total* et *minimum* par le système Tulip.

Le messages d'erreur

Lorsque Tulip est incapable d'analyser une assertion ou une interrogation, il affiche le message suivant :

Syntactic parser failed towards the end of this segment :

Il affiche ensuite la partie de la phrase qu'il n'a pu analyser.

Ordinairement, le dernier mot affiché dans la phrase est celui qui est la source du problème.

Lorsque Tulip peut analyser une interrogation, mais ne peut trouver une solution, il demande à l'utilisateur si celui-ci désire une liste des faits pouvant s'avérer utiles. Si l'utilisateur répond *Yes*, Tulip a recours à l'association de mots pour trouver les assertions ayant le plus grand nombre de mots en commun avec l'interrogation.

Les textes en français

Les capacités du système Tulip sont assez restreintes face aux textes en français. En général, Tulip ne peut analyser les phrases en français.

Toutefois il peut utiliser l'appariement de mots avec certains mots français. Son vocabulaire français est limité aux mots pouvant apparaître sur une carte professionnelle et aux termes utilisés dans le domaine de l'intelligence artificielle.

Pour utiliser les accents et autres signes graphiques propres au français, il faut inclure les lignes suivantes dans le fichier `.cshrc` :

```
setenv DEFAULT_FONT "/home/condor/critter/fonts/f_screen.r.14"  
setenv EMACS_FONT "/home/condor/critter/fonts/f_serif.r.14"
```

et avoir les lignes suivantes dans un fichier `.emacs_pro` dans le répertoire `home` :

```
(load "/home/condor/critter/.emacs_pro")  
(load " ../isabelle/emacs/dead_key_accents")
```

Les signes graphiques peuvent être ajoutés aux lettres en utilisant les touches R1 à R5 qui se trouvent à droite des touches principales sur une machine Sun 3. On ajoute un signe graphique en enfonceant la touche R correspondante avant d'enfoncer la touche de la lettre à laquelle le signe doit être ajouté.

Les signes graphiques des touches R sont :

R1 - accent aigu

R2 - accent grave

R3 - cédille

R4 - tréma

R5 - accent circonflexe

L'expression des connaissances dans le système Tulip

Tulip peut être utilisé pour exprimer la transmission de propriété entre des classes d'objets. Par exemple, on peut introduire dans le système les assertions :

every bird can fly.

every robin is a bird.

Ernie is a robin.

Si l'on pose alors la question

Can Ernie fly?

Tulip répond Yes. Étant donné que l'utilisateur a déclaré que chaque rouge-gorge est un oiseau, les propriétés associées aux oiseaux (par exemple, la capacité de voler) sont transmises à tous les rouges-gorges.

CACC / CCAC



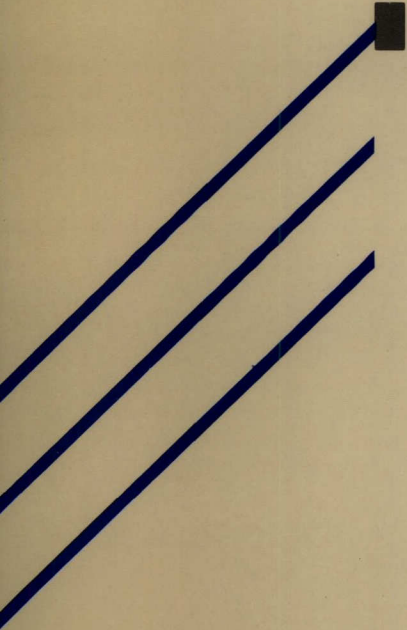
95336

QUEEN QA 76.9 .N38 F47514 19
Ferguson, Ronald John, 1951-
Guide de l'utilisateur du sy


QA
76.9
N38
F475f
1989
c.3

DATE DUE

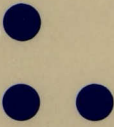
[illegible]



Pour plus de détails,
veuillez communiquer avec :



*Le Centre canadien de recherche
sur l'informatisation du travail*
1575, boulevard Chomedey
Laval (Québec)
H7V 2X2
(514) 682-3400



For more information,
please contact:

*Canadian Workplace
Automation Research Centre*
1575 Chomedey Blvd.
Laval, Quebec
H7V 2X2
(514) 682-3400