

CHAPITRE V : LA GENERATION AUTOMATIQUE DE TEXTES

5.1 INTRODUCTION :

La génération de textes consiste à produire des textes par un logiciel, de façon à exprimer automatiquement un contenu formel en langue naturelle. On partage traditionnellement la génération en deux étapes : le “ quoi dire ? ” qui consiste à déterminer le contenu du texte à engendrer, et le “ comment le dire ? ” qui est son expression en langue naturelle. De plus, l'objectif est l'élaboration de textes linguistiquement corrects et compréhensibles par le destinataire.

La qualité des textes engendrés, mais surtout leur compréhension sont fonction des informations dont le générateur dispose pour produire ces textes. Il devient alors essentiel de recenser toutes les connaissances nécessaires au générateur, afin de les lui fournir ou de lui donner les moyens de les calculer.

En effet, les textes élaborés dépendent d'abord des informations que l'on souhaite transmettre au destinataire du message. Ces éléments varient en fonction du contexte et donc de l'application visée. Ils sont représentés formellement (concepts, graphes ou autres) et constituent le “ quoi dire ? ”. D'autres informations plus contextuelles sont indispensables pour passer du “ quoi dire ? ” au “ comment le dire ? ”. Elles sont en partie données et en partie calculées au cours du traitement pour guider le générateur tout au long de sa tâche. Les informations manipulées par le processus de génération sont donc de types très divers :

- générales : on parle aussi de connaissances universelles,
- sur l'application : explicites ou implicites (on choisit de détailler plus ou moins une information),
- linguistiques : essentiellement morphologiques et syntaxiques, mais aussi sémantiques,
- sur le destinataire du message, avec le principe d'un “ modèle ” décrivant le destinataire.

5.2 LE GENERATEUR DE CRISTAL

Le générateur de CRISTAL est un générateur du français écrit, implémenté en Prolog. Cet outil est non contextuel (Balicco 1993, Ponton, 1996) c'est-à-dire indépendant de toute application. Il fonctionne sur la partie “ comment le dire ? ” et est donc fondamentalement un générateur de surface traduisant sous forme linguistique un contenu sémantique donné. Ce système est basé sur un modèle linguistique qui permet la production de plusieurs versions “ équivalentes ” d'un même contenu sémantique. Ce modèle organise le texte à l'aide d'opérations linguistiques qui servent à l'élaboration de toutes les versions différentes du texte.

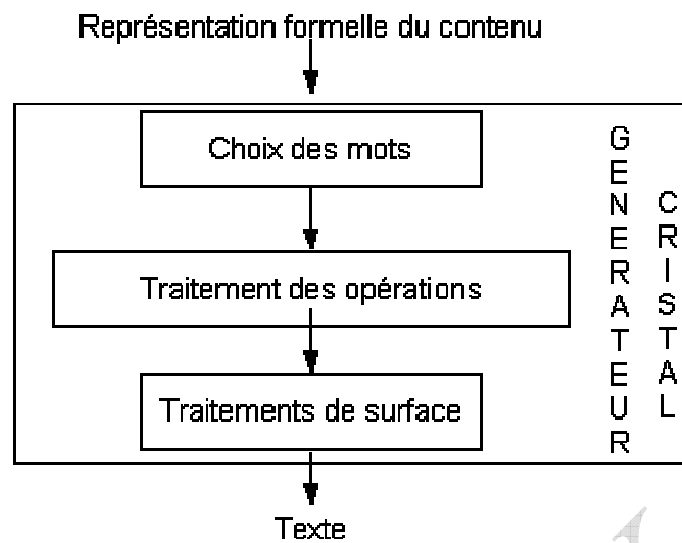


Fig. 1. Fonctionnement du générateur CRISTAL

Comme le montre le schéma ci-dessus, le processus de génération est décomposé en plusieurs tâches qui sont définies dans des modules distincts. Ils peuvent donc être modifiés indépendamment sans conséquence sur le fonctionnement général du générateur. Le système commence par le choix des mots (un message sous forme logique est transformé en unités lexicales), puis la détermination de la structure syntaxique à l'aide des opérations linguistiques. La génération se poursuit par une phase de génération morphologique (calcul des formes fléchies, conjuguées...). La génération se poursuit par une phase de génération morphologique (calcul des formes fléchies, conjuguées...) avant de terminer par des traitements de surface (majuscules, élisions...). Les opérations linguistiques sont concrètement réalisées lors de l'écriture du texte. Pour cela, des règles de transformation sont appliquées à la représentation formelle du contenu du texte. Le générateur utilise des ressources linguistiques indépendantes de l'application (un lexique et une grammaire) et un lexique propre à chaque application.

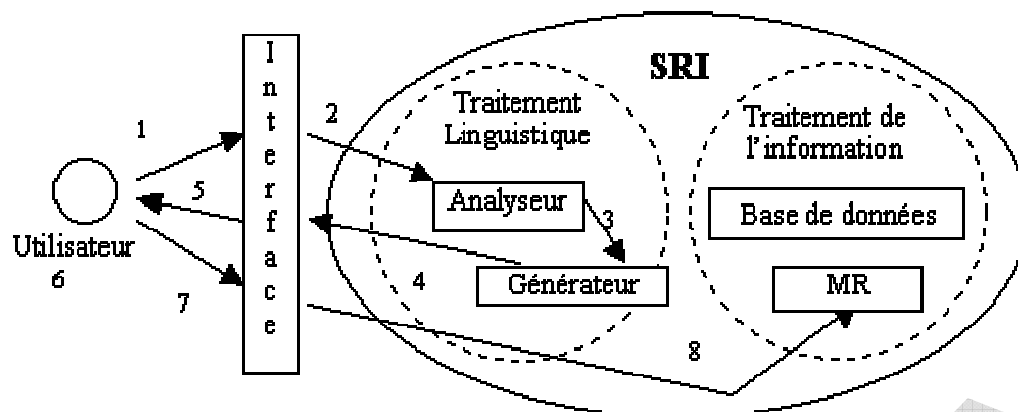
5.3 INTEGRATION D'UN GENERATEUR A UN SYSTEME DE RECHERCHE D'INFORMATIONS (SRI)

On considère qu'un SRI est constitué de quatre éléments : un moteur de recherche, une base de données, un analyseur de requête et un générateur. Ces éléments sont partagés en deux catégories de traitement : traitement de l'information (moteur de recherche et base de données) et traitement linguistique (analyseur et générateur). Une interface fait le lien entre l'utilisateur et le système de recherche d'information, et permet l'échange d'informations entre les deux.

Le processus d'interaction se décompose en deux étapes : de l'utilisateur vers le système et du système vers l'utilisateur. Nous présentons ici le rôle du générateur dans chacune de ces étapes.

5.3.1 De l'utilisateur vers le système : formulation de la requête

Ce processus est composé de huit étapes :

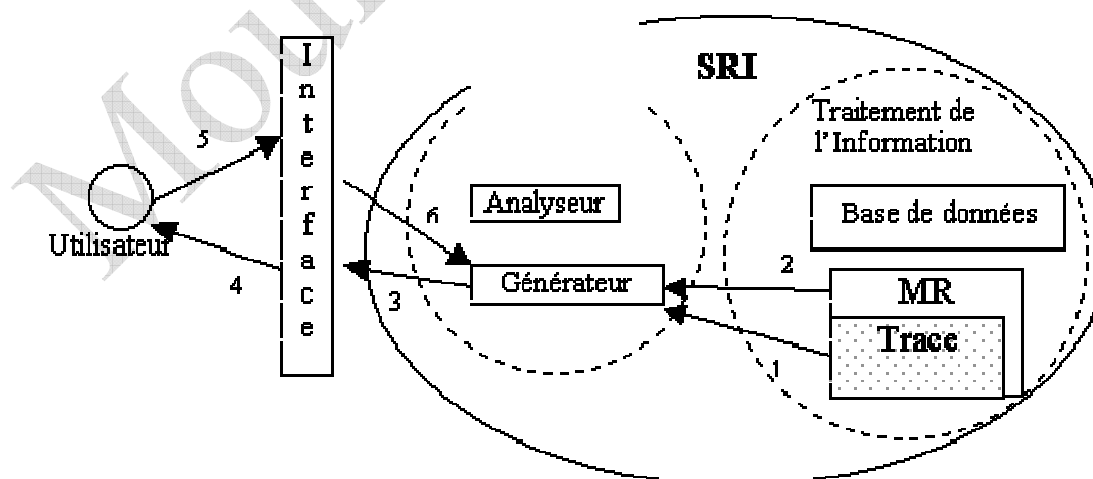


1. L'utilisateur exprime son besoin d'information en langue naturelle puis l'envoie à l'interface du système ;
2. L'interface achemine la demande à l'analyseur ;
3. La requête est analysée et la représentation formelle de son contenu est fournie en entrée au générateur ;
4. Le générateur produit des paraphrases de la requête et les renvoie à l'interface ;
5. L'interface propose les paraphrases du générateur à l'utilisateur ;
6. L'utilisateur sélectionne une ou plusieurs paraphrase(s), incluant ou non sa requête initiale ;
7. La sélection est retournée à l'interface ;
8. L'interface soumet la sélection de requêtes au moteur de recherche.

5.3.2 Du système vers l'utilisateur : donner des résultats

Ce processus est composé de six étapes :

Traitement
Linguistique



1. La trace est donnée en entrée de la tâche de génération ;

2. Dans le même temps, les résultats de la requête sont également fournis au générateur ;
3. Après le traitement, le générateur transmet des données (exprimées en langue naturelle) à l'interface ;
4. L'interface envoie les résultats à l'utilisateur ;
5. L'utilisateur peut demander d'autres affichages à l'interface ;
6. L'interface interroge le générateur selon les demandes de l'utilisateur.

Les étapes 5, 6, 3 et 4 sont répétées tant que l'utilisateur souhaite des présentations différentes.

Mourad LOUKAM