

# Arbres et lettres, de l'information au langage.

## **Philippe Ameline**

philippe.ameline@free.fr

Rendre compte n'est jamais une finalité. C'est à la fois raconter ce qui a été fait et passer le témoin à un autre ; c'est, idéalement, enrichir sa propre base d'information et celle du correspondant avec des données actives à court, moyen et long terme.

Dans ce cadre, il est stratégique de bien distinguer les données structurées (qui peuvent potentiellement permettre à l'ordinateur de réaliser des traitements à valeur ajoutée – y compris en terme de génération automatique de documents) et le langage naturel, compréhensible par les humains, mais hermétiques pour la machine.

Trouver le juste équilibre, c'est déterminer la « narration structurée » qui offre le meilleur potentiel de valeur ajoutée et établir les règles optimales de transformation de ces informations en textes ou fragments de textes au sein de divers documents transmis d'humain à humain.

## **Table des matières**

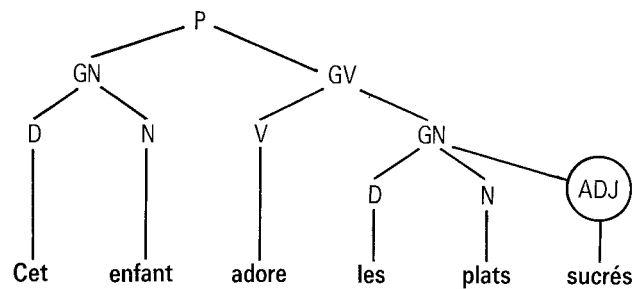
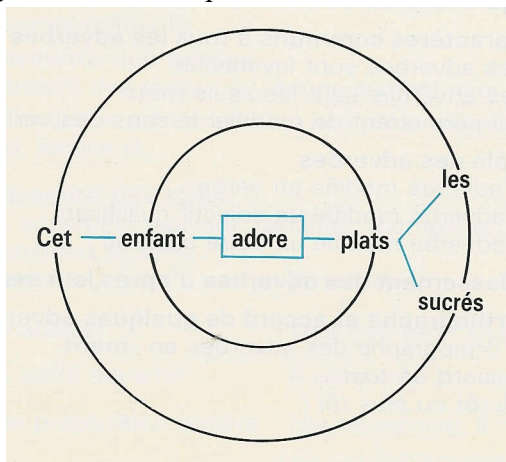
Soyons pragmatique.....	2
Génération automatique de langage naturel.....	2
Génération automatique.....	2
Génération guidée.....	3
Texte à compléter.....	3
Conclusion.....	4

## Soyons pragmatique

Généralement, l'informatique est conçue comme la cohabitation d'une base de données (un ensemble de cases à remplir) et d'une gestion documentaire (un ensemble de textes et d'images). Le lien entre les deux « aspects » de l'information (structurée et libre) est généralement ténu, car le langage naturel réalise une réelle narration, alors que les cases d'une base de données semblent dans un tout autre espace... on imagine mal raconter quelque chose en remplissant les cases d'un formulaire !

Au sein d'Episodus, les règles sont fort différentes car les données structurées qui y sont stockées sont des **arbres** constitués de **concepts de l'ontologie**. Comme nous allons le voir, ces deux composants sont homogènes à une grammaire (arbres) et à un vocabulaire (ontologie) et forment un véritable langage.

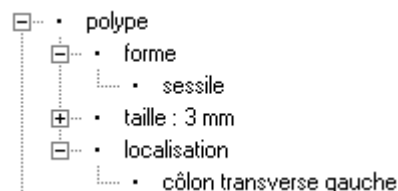
On peut trouver dans le Bescherel l'exemple d'un même « discours » exprimé sous forme d'une grammaire de dépendance et sous forme d'une représentation en langage naturel :



La première représentation est homogène à la représentation interne des informations d'Episodus (en remplaçant les mots par leur code – ce qui, par ailleurs élimine les ambiguïtés liées aux homonymies), la seconde est lisible par un être humain ; tout l'enjeu est d'organiser des transformations entre une forme et l'autre : synthèse de langage naturel dans un sens, analyse de langage naturel dans l'autre.

## Génération automatique de langage naturel

Lorsqu'un utilisateur du module de coloscopie de Nautilus souhaite lire ou imprimer un compte rendu, le système transforme « au vol » la représentation interne en arbre, du type :



en une phrase comme : « polype sessile de 3 mm situé dans le côlon transverse gauche ».

Nous allons voir les divers moyens d'arriver à un tel résultat : génération automatique, guidée ou texte « à compléter ».

## Génération automatique

Episodus possède un générateur automatique de langage naturel qui permet de transformer des

arbres simples en langage naturel.

Par exemple à partir d'un geste, réalisé sur un organe à un moment précis, il peut concevoir une phrase du type « <à tel moment>, un <geste> a été réalisé sur l'<organe> ». En organisant les données sous une forme du type :

Phrase principale :

- Sujet : <geste>
- Verbe : réaliser
- Complément circonstanciel de lieu : <organe>
- Complément circonstanciel de temps : <moment>

À partir de cette recombinaison grammaticale, le générateur automatique arrange automatiquement une structure de phrase en fonction de la langue cible.

## Génération guidée

La construction entièrement automatique n'est aujourd'hui disponible que pour des « motifs descriptifs » simples. Dans des cas plus complexes, il est nécessaire de guider la transformation en attribuant aux différents éléments de l'arbre une position dans la structure grammaticale.

Cette fonctionnalité sera mise en oeuvre au sein des Archétypes.

## Texte à compléter

Une forme de génération moins élaborée consiste à générer un texte « à trous » dont les éléments manquants seront complétés par des éléments de l'arbre ou du dossier patient.

Un exemple de texte pourrait être :

« Cher ami, <nom du patient> <né/née> le <date de naissance> présente <diagnostic>. Peux tu... »

Au sein d'Episodus, il est possible de réaliser de tels textes au format RTF (Rich Text Format), en utilisant des balises dont le contenu est interprété et remplacé par les divers éléments du dossier patient. Les balises se distinguent des éléments « fixes » du texte par le fait qu'elles commencent par les caractères [\$ et s'achèvent par les caractères \$].

L'intérieur du bloc peut contenir des libellés comme « poids » des valeurs comme « 75 kg » ou des expressions conditionnelles comme « si le patient est de sexe féminin inscrire 'née' sinon inscrire 'né' ». Le contenu des balises est constitué d'éléments du Lexique ou de chemins d'éléments du Lexique.

Par exemple, puisque le Poids est représenté au sein du Lexique par le code VPOID1, il est possible de créer les balises suivantes :

- [\$[LABEL VPOID1]\$] qui inscrira « poids »
- [\$[VALUE VPOID1]\$] qui inscrira la valeur la plus récente du poids du patient, par exemple « 75 kg »

Pour décrire un élément conditionnel en fonction du sexe du patient, et sachant que l'arbre qui contient les données administratives a pour racine ZADMI1 (données administratives), avec pour sous chapitre LIDET1 (identité) et enfin LSEX1 (sexe), donc le chemin ZADMI1/LIDET1/LSEX1 qui peut avoir pour sous-niveau soit HFEMI2 (féminin), soit HMAS2 (masculin).

La balise sera alors du type :

- [\${IF ZADMI1/LIDET1/LSEX1/HFEMI2} née [ELSE] né]\$

Soit : « si on trouve le chemin 'données administratives'/identité/'sexe'/féminin' alors inscrire 'née' sinon inscrire 'né'.

Ces approches sont complémentaires, et il est possible d'utiliser les générations automatiques et guidées au sein des balises de textes à compléter.

## **Conclusion**

Au sein d'Episodus, la logique est de décrire les observations cliniques en utilisant le langage structuré, ce qui permet de bénéficier des processus à valeur ajoutée du logiciel – dont font partie les processus de génération de textes.

L'information princeps, utilisable localement et transmissible aux divers correspondants, est donc sous forme d'arbre, et peut donner lieu à de multiples transcriptions en langage naturel.

Contrairement aux systèmes usuels, où les informations servent uniquement à combler les vides de « textes types », il est nécessaire, au sein d'Episodus, de structurer les données de façon à former un discours cohérent. Ce qui suppose généralement d'y inclure l'ensemble des informations et non simplement les données conditionnelles.