Connexionnisme
et
représentation de haut niveau

Antoine CORNUEJOLS

Equipe Inférence et Apprentissage
Laboratoire de Recherche en Informatique (LRI)
UA410 du CNRS, bât. 490
Université de Paris-Sud, Orsay
91405 ORSAY Cedex
email : (uucp) antoine@lri.lri.fr
et
Institut d'Informatique d'Entreprise
18, allée Jean Rostand 91 EVRY

Résumé :

L'approche connexionniste repose sur l'émergence de concepts et de comportements adaptés à un domaine donné à partir d'assemblée d'agents primitifs en interactions locales simples. La description de ces agents et de leurs interactions est généralement sans rapport direct avec la description donnée par un expert humain du domaine dans lequel est situé le système. Ceci contraste avec l'approche classique de l'Intelligence Artificielle fondée sur une manipulation des concepts de haut niveau à partir de règles de raisonnement directement comparable à celle de l'expert. De cette opposition entre bas niveau et haut niveau de représentation et de processus, résulte d'un côté une capacité d'adaptation naturelle au milieu et une grande robustesse alliée à une opacité complète du mode de raisonnement, et de l'autre, une transparence appréciable mais aussi une rigidité et une fragilité difficile à pallier.

Nous présentons ici un système appelé INFLUENCE qui réalise l'auto-organisation et la ré-organisation spontanée (à partir de mécanismes locaux et simples) d'une base de connaissances exprimée à l'aide de concepts de haut niveau. La tâche est d'adapter continûment cette base de connaissances au fur et à mesure que des informations nouvelles parviennent au système, et en particulier d'effectuer des ré-interprétations et des révisions de modèle quand cela s'avère nécessaire. Il s'agit donc d'un système hybride alliant haut niveau de représentation et mode de fonctionnement connexionniste. Nous montrons que les performances obtenues sont supérieures à celles des systèmes classiques, et nous discutons l'opposition entre représentations localisées généralement associées à l'Intelligence Artificielle traditionnelle, et les représentations distribuées caractéristiques des modèles connexionnistes les plus courants.

1. Introduction

La plupart pour ne pas dire la totalité des modèles connexionnistes existants sont fondés sur la notion d'assemblée de "neurones" très primitifs (réalisant des opérations mathématiques simples) reliés par des connexions affectées d'un poids (servant de coefficient multiplicatif dans le transfert des signaux d'un neurone à l'autre). Le calcul ou le

raisonnement dans de tels réseaux de neurones est le fruit de processus et d'interactions locaux dont la description n'a aucun rapport immédiat avec la signification du comportement produit. Cette caractéristique est à mettre en opposition avec les modèles de l'Intelligence Artificielle classique[1] où la signification produite est directement issue, et descriptible en termes, d'une manipulation des concepts pertinents suivant des règles de déduction propres au domaine concerné.

Dans ce contraste réside l'essentiel des différences de méthodes et de domaines d'application de l'Intelligence Artificielle "symbolique" et du connexionnisme. L'une s'occupe de raisonnement conceptuel et de représentation de connaissances directement transcriptible en termes de discours humain, l'autre d'émergence de sens à partir de données brutes de bas niveau. L'une essaie de dialoguer avec les experts mais s'enlise dès que la connaissance explicite du domaine est défaillante, l'autre se délecte de données bruitées mais est incapable d'expliquer ses raisonnements et d'utiliser une théorie du domaine existante. L'opposition entre modes de représentation employés se retrouve dans les mécanismes de fonctionnement. L'Intelligence Artificielle fait généralement appel à des modèles centralisés obéissants à un module superviseur omniscient et omnipotent opérant sur la base de connaissances. Les modèles connexionnistes au contraire reposent sur l'interaction d'automates et de processus largement indépendants et qui forment en soi la base de connaissances. Pour résumer, le tableau suivant peint un portrait rapide comparant les deux approches.

	Intelligence Artificielle symbolique	Connexionnisme
Représentation et règles de transformation :	de haut niveau	de bas niveau
- granularité	élevée	faible
- nature des informations circulant dans le système	complexe	surtout numérique
Transparence :	élevée	quasi nulle
- pouvoir explicatif	potentiellement bon	très limité
- utilisation de la théorie du domaine	naturelle	très difficile
Souplesse :	faible	bonne
- exploitation de données "corrompues"	difficile	excellente
- adaptation au contexte	limitée	naturelle
- résistance aux "fautes"	mauvaise	bonne

Les deux propriétés les plus remarquables et les plus désirables des modèles connexionnistes : l'*émergence de concepts* appropriés à un domaine à partir de données brutes, et l'*adaptation naturelle au contexte*, sont dues aux phénomènes d'auto-organisation qui prennent place dans une collection nombreuse d'agents autonomes en interactions

[1] Nous utiliserons fréquemment ce vocable dans la suite de ce texte. Il peut paraître curieux d'utiliser le qualificatif de classique pour une science aussi jeune et en évolution aussi rapide que l'Intelligence Artificielle, nous voulons en fait souligner que la plupart des chercheurs dans ce domaine acceptent le dogme du "knowledge level" stipulant qu'il est possible de produire une intelligence à partir uniquement de la manipulation de concepts et de règles de transformations qui soient directement adaptés à la description du monde selon des termes de notre culture.

locales. Ces phénomènes opèrent à la fois au niveau de la représentation des connaissances lorsque celle-ci est de bas niveau au départ et au niveau des processus de calcul donnant naissance au "raisonnement" suivi par le système. Au niveau de la représentation, le phénomène d'auto-organisation conduit généralement à une représentation distribuée des concepts adaptés à la description d'un environnement. Ainsi, la reconnaissance ou l'évocation d'un concept dans le système met en jeu plusieurs "neurones" et connexions, de même que chaque neurone participe à la représentation de plusieurs concepts. Mais ce schéma que l'on retrouve dans la majorité des modèles connexionnistes actuels n'épuise cependant pas l'essence du connexionnisme. Il est en effet également possible de concevoir des modèles dans lesquels chaque cellule ou neurone représente spécifiquement un concept du discours tel que "maison", "liberté", "grand". On parle dans ce cas de représentation locale ou localisée, et le qualificatif de connexionniste appliqué à ces systèmes vient alors de leur mode de fonctionnement au niveau des processus de calcul mettant en jeu essentiellement des interactions simples (à base de calculs numériques surtout) et locales à la représentation.

Divers modèles implémentent cette approche. Ils sont généralement fondés sur l'utilisation d'un réseau sémantique pré-établi où les noeuds du réseau représentent les concepts utiles à la description du domaine et où les liens entre noeuds représentent plus ou moins le degré de parenté entre concepts et la probabilité que si l'un est évoqué, l'autre le soit aussi, ou soit, au contraire, rejeté. Avec ces systèmes on cherche surtout à identifier une situation ou un contexte à partir de données ambiguës par le jeu des activations et inhibitions croisées entre concepts. L'un des exemples les plus célèbre et les plus cité d'une telle approche est celui décrit par [Waltz & Pollack, 1985] appliqué au domaine de l'interprétation de membres de phrases ambiguës.

Nous avons voulu pour notre part explorer plus particulièrement l'adaptation d'un processus de raisonnement et de compréhension de contexte en fonction d'un apport continu et non pas unique d'informations. Afin d'étudier ce phénomène, nous avons retenu dans un premier temps un modèle où la représentation des connaissances se fait à l'aide de concepts de haut niveau, mais où le réseau sémantique se crée dynamiquement et où les mécanismes de calcul sont simples et basés sur des interactions locales autonomes.

Plus spécifiquement, le problème étudié est celui de la compréhension non monotone d'informations (en l'occurrence des textes en langage naturel) qui intervient quand des hypothèses ou un modèle explicatif élaborés pour comprendre les informations perçues se trouvent contredits par les informations parvenant ultérieurement et doivent donc être révisés. Un exemple de ce type de situation est fourni par la lecture du texte suivant :

> ***"Lorsque Paul apprit par les journaux qu'il avait gagné le gros lot à la loterie, il prit immédiatement un billet d'avion pour l'Indonésie. Il était très excité, et fit ses bagages à toute vitesse. Dans sa précipitation il partit sans fermer le gaz. Trois jours après, à Djakarta, il apprit que les deux mafiosi qui le cherchaient étaient morts dans l'explosion de son appartement. Il respira enfin."***

On note que les deux dernières phrases conduisent à ré-interpréter le contexte global et à interpréter le départ précipité de Paul non comme le résultat heureux d'un gain important à la loterie, mais comme une fuite de la mafia qui aurait retrouvé sa trace.

Ce type de tâche de ré-interprétation se révèle très difficile dans le cadre des méthodes de l'Intelligence Artificielle classique fondées sur l'emploi de règles de déduction de haut niveau conceptuel. L'approche présentée ici et implémentée sous la forme du programme INFLUENCE consiste à considérer la base de connaissances courante comme un système dynamique composé de concepts en interaction constante entre eux, à l'image d'une molécule chimique, et à la laisser s'auto-organiser et se ré-organiser spontanément (sans usage d'un module superviseur extérieur) quand des informations nouvelles parviennent au système. Le modèle résultant est connexionniste dans son mode de fonctionnement mais "symbolique" dans le mode de représentation des connaissances utilisé.

Dans un première partie nous présentons le système INFLUENCE et les mécanismes qui régissent son fonctionnement. Puis nous montrons son comportement sur un exemple de tâche de compréhension de texte simple mais représentatif. Finalement nous discutons les propriétés du système étudié, ainsi que l'opposition

traditionnellement évoquée entre représentation localisée et représentation distribuée, nous montrons à cette occasion que cette opposition peut être dépassée.

2. Le système INFLUENCE

Deux principes fondamentaux sous-tendent le concept du système INFLUENCE. D'une part, *la représentation de la connaissance doit être active*, c'est-à-dire que les inférences qui construisent le modèle de l'univers doivent être intégrées aux éléments qui composent ce modèle et non être le fait d'un agent extérieur opérant "sur" la base de connaissance. C'est un principe d'auto-organisation. D'autre part, le modèle courant doit être soumis à des "*perturbations*" afin de générer des occasions de modifications sans lesquelles il ne saurait y avoir de possibilités de non-monotonie. Dans le cas d'INFLUENCE, ces perturbations ne résultent pas seulement de la perception de données nouvelles, mais aussi de processus de remise en cause permanente de la base de connaissances et de ré-examen d'hypothèses.

Le fonctionnement du système INFLUENCE est fondé sur l'utilisation d'une représentation de connaissances par *frames*. Ce type de représentations présente en effet les particularités intéressantes d'être largement employé dans les systèmes existant de compréhension et d'interprétation de contextes naturels (textes ou scènes visuelles) et donc de fournir des outils appropriés d'interaction modèle/données, et, par ailleurs, de pourvoir une source d'inférences et de raisonnement intégrés à la représentation des connaissances sous la forme des fonctions d'attachement procéduraux liées aux slots des frames. La première condition requise, c'est-à-dire de disposer d'une représentation des connaissances active est donc remplie. La seconde, ayant trait à la source de perturbations, est réalisée dans le système INFLUENCE par un mécanisme contrôlant une certaine *instabilité des liens* entre slots et frames. Elle fait l'objet d'une description plus détaillée ci-dessous.

Finalement, dans le cadre du projet INFLUENCE, on a imposé en plus que tous les *processus actifs soient locaux*, c'est-à-dire dont les causes et les effets soient locaux à la représentation du modèle interne et non fondés sur des informations ou des actions globales.

Le système INFLUENCE est apparenté aux systèmes de compréhension de textes issus de l'Ecole de Yale, tel BORIS [Dyer,1983]. Dans cette approche les concepts jugés utiles pour la compréhension d'un "épisode" sont représentés par des frames de différent niveaux d'abstraction. La *mémoire épisodique* d'un tel système est constituée du réseau de frames instantiés correspondant à l'épisode interprété. Les frames sont liés entre eux par les liens entre les slots et leurs cibles, elles-mêmes d'autres frames. Ces liens représentent des relations sémantiques de niveau élevé telles qu'intention, causalité, location etc. Lorsque le système est interrogé, celui-ci sélectionne les concepts/frames pertinents et parcourt le ou les liens correspondants à la question posée. L'interprétation de l'univers perçu est donc fondée sur les concepts instantiés présents en mémoire épisodique et sur les liens les joignant entre eux. Si l'on change la cible d'un lien ou si l'on introduit ou supprime un frame, on modifie du même coup l'interprétation correspondante puisque les réponses aux mêmes questions seraient différentes. Cette observation conduit au shéma suivant pour permettre des perturbations au sein du système.

Les liens sont considérés comme des agents autonomes. Ils sont affectés d'un coefficient κ ($0<\kappa<1$) caractérisant *leur stabilité*. Ce coefficient correspond à la certitude que l'on met dans l'information inscrite dans le lien en question. Par exemple, si l'on est certain que Pierre va en Californie pour profiter du soleil, le lien `sur-opérateur` de l'**OPERATEUR OP-ALLER**(en Californie) pointant vers l'**OP-PRENDRE-SOLEIL** sera affecté d'un coefficient κ élevé (proche de ou égal à 1) signifiant que ce lien sera très stable. Au contraire, si la certitude associée était faible (peut-être Pierre va en Californie pour signer un contrat) κ serait plus proche de 0 et ce lien serait instable, c'est-à-dire qu'il lui arriverait souvent de remettre en cause sa cible et de chercher une nouvelle cible (potentiellement la même).

Le mécanisme de base est qu'à chaque instant, avec une probabilité fonction de κ (élevée si κ est faible et quasi nulle quand κ approche 1), **un lien peut décider de se "désengager"** de sa cible actuelle et chercher une cible quelconque correspondant à des spécifications similaires (voir ci-dessous).

Il est à noter qu'un lien même désengagé pointe encore sur sa cible initiale, il ne changera de destination que lorsque le choix d'un nouveau frame, par un procédé précisé dans la suite, sera effectué. Cela assure que le système est capable

de fournir une réponse en permanence aux questions qu'on peut lui poser, ce qui ne serait pas le cas si un certain laps de temps s'écoulait entre l'instant du désengagement et l'instant de choix d'une nouvelle cible et de re-connection.

Le problème maintenant est d'assurer que les **choix de nouvelles cibles** et donc les ré-interprétations résultantes ne soient pas anarchiques et dénuées de sens (dans le domaine considéré). Trois niveaux de contraintes s'exercent sur ces choix.

- Le niveau 0 correspond aux "contraintes de type": un lien `sur-opérateur` par exemple ne peut pointer que vers un frame de type **OPERATEUR** et pas sur un frame de type **EVENEMENT** ou **ETRE-HUMAIN**.

- Le niveau 1 correspond à des contraintes sémantiques plus sophistiquées attachées au domaine d'application considéré. Par exemple supposons que le domaine du discours étudié concerne, entre autre, les relations familiales et que l'on sache que dans le contexte décrit le mari et la femme sont très probablement de même religion si l'un d'eux est musulman ou juif. Alors, si la mémoire épisodique contient à un certain moment un frame **M-MARIAGE** dont la femme est musulmanne et dont le lien mari est désengagé celui-ci devrait pointer préférentiellement vers un **ETRE-HUMAIN** (contrainte de niveau 0) musulman. Ces contraintes sont inscrites dans les procédures d'attachement procédural de chaque slot/lien sous la facette `contraintes`.

- Le niveau 2 est original. Il implémente un principe général de "qualité" et d'esthétique d'un modèle ou d'une interprétation. Un bon modèle est bien sûr un modèle qui fournit des réponses si possibles justes aux questions qu'on lui pose, *pouvoir d'explicativité et de prédiction* [Popper,1963]. C'est aussi un modèle qui ne multiplie pas les hypothèses auxiliaires non nécessaires. C'est le *principe de parcimonie* de Bacon, général en science : une théorie est d'autant meilleure qu'elle est économe en principes, règles et hypothèses. Une première traduction de ce principe sous forme locale serait que chaque lien lorsqu'il choisit une cible se connecte préférentiellement sur les frames les plus connectés, c'est-à-dire vers les concepts qui interviennent déjà le plus dans l'interprétation courante. Si en plus une élimination des frames les moins connectés intervenait, on obtiendrait un élagage progressif des concepts inutiles. En fait l'implémentation retenue pour INFLUENCE raffine davantage ce shéma et attribue à chaque frame un coefficient appelé "influence" qui est fonction non seulement du nombre de liens qui pointent vers ce frame mais aussi de l'influence des frames qui lui sont connectés. On propage donc ici des influences un peu à l'image du "spreading activation" utilisé par exemple par [Anderson,1983] dans le système ACT*.

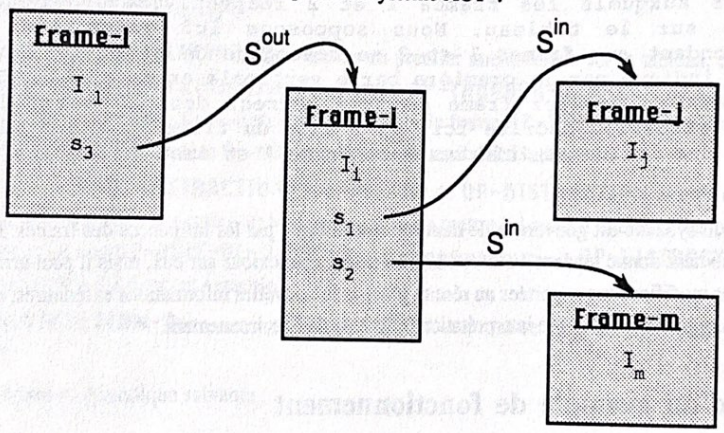


Figure 1 : Dans la situation présentée ici, l'influence du frame-i serait:

$$I_i = I_{i_0} + (S^{in}_{s1} \cdot I_j + S^{in}_{s2} \cdot I_m) + S^{out}_{s3} \cdot I_l$$

L'influence du **FRAME-I** est la somme de son influence initiale I_0 déterminée à priori par l'expert du domaine (qui, par exemple, souhaite favoriser le concept de **OP-ETRE-RICHE**), et des influences qu'il gagne en étant connecté à d'autres frames. Cette dernière partie est elle même décomposable en deux parts. La première correspond aux gains fournis par les liens originaires du **FRAME-I** grâce aux S^{in}, la seconde aux gains réalisés par les liens pointant sur le **FRAME-I**, grâce aux S^{out}. La sémantique exacte attachée aux coefficients S^{in} et S^{out} est en partie dépendante du

domaine d'application. Le coefficient α permet de régler quelle importance relative on accorde à l'influence définie a priori I_0 par rapport à l'influence gagnée grâce au contexte.

L'importance de ces influences dans le fonctionnement de l'algorithme réside dans le fait qu'elles déterminent les fréquences avec lesquelles les frames émettent une "annonce" sur un tableau visible par tous les agents (les liens). Durant chaque intervalle de temps *dt* (à chaque boucle de l'algorithme) un frame de chaque type est tiré aléatoirement avec une probabilité dépendant directement de son influence, et les liens désengagés à cet instant et cherchant une cible du type correspondant se connectent alors sur le frame "émetteur". Cela signifie que les frames déjà bien connectés vont émettre plus fréquemment que les frames moins influents et donc avoir plus de chance d'accroître encore leur connectivité et leur influence.

La figure suivante montre ce qui se passe lorsque deux frames du même type sont en compétition pour un lien donné.

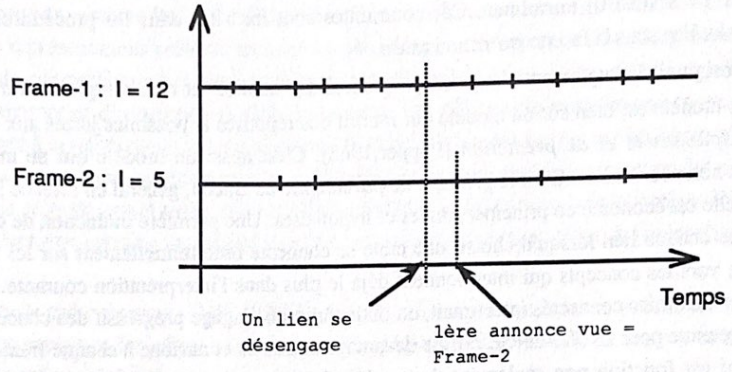


Figure 2 : Sur les deux lignes horizontales supérieures sont indiqués les instants auxquels les frames 1 et 2 respectivement "émettent" une annonce sur le tableau. Nous supposons ici qu'un lien du type correspondant aux frames 1 et 2 se désengage de sa cible à un certain instant indiqué par la première barre verticale en pointillé. Ce lien se connectera au premier frame de type correct dont il verra l'annonce. Dans la situation décrite ici, il s'agit du frame 2, alors même que ce frame a une influence inférieure au frame 1 et émet donc moins souvent.

Ainsi l'évolution du système est gouvernée de manière stochastique par les influences des frames. Les frames les plus influents à un moment donné tendent à attirer encore d'autres connexions sur eux, mais il peut arriver, par le jeu des hasards ou par des modifications apportées au réseau grâce à de nouvelles informations extérieures, que des frames peu influents deviennent centraux dans une interprétation différente de l'environnement.

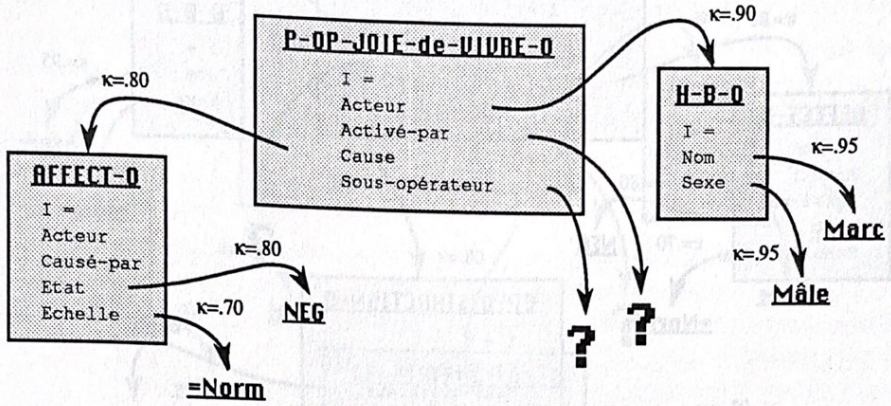
4. Esquisse d'un exemple de fonctionnement

Avant de détailler plus précisément les caractéristiques du modèle INFLUENCE, il peut-être utile d'en appréhender l'esprit sur un exemple de fonctionnement simple. Il s'agit d'une tâche de compréhension de texte en langage naturel exigeant une modeste ré-interprétation en cours d'assimilation des informations fournies. Le texte est le suivant :

"Marc s'ennuyait. Il s'empara du journal sur le fauteuil, et saisit en dessous la télécommande de la télé."

On s'attend ici à ce que le système après avoir pensé que Marc voulait se distraire en lisant le journal, ré-interprète l'épisode et conclut qu'en fait Marc veut regarder la télévision.

On fournit les éléments du texte séquentiellement au système, en les traduisant directement, ici à la main, en portions de réseaux de frames. Ainsi, par exemple, le fragment ***"Marc s'ennuyait"*** est traduit sous la forme:



Remarque : **H-B** (de Human-Being) représente le frame **ETRE-HUMAIN** dont il est question dans le texte.

C'est ainsi que ce fragment de texte pourrait être représenté par un système de compréhension issu des travaux de l'école de Yale tel que BORIS.

A ce point c'est tout ce qui se passerait dans un tel système. Avec INFLUENCE, la mémoire devient dynamique. C'est-à-dire que, d'une part les liens établis peuvent se remettre en cause à tout instant si leur force κ associée est inférieure à 1, et d'autre part les liens non engagés cherchent activement une cible.

Si donc le réseau précédemment construit est laissé en mémoire épisodique pour un certain temps avant que d'autres éléments d'information soient ajoutés, une évolution possible de ce réseau est la suivante.

Le lien `sous-opérateur` de **G-OP-JOIE-DE-VIVRE-0** va chercher une cible. Pour cela, dans cet exemple, il a le choix entre les candidats suivants affectés de leur influence initiale I_0 : **OP-DISTRACTION** (I_0=2) et **OP-TROUVER-JOB-INTERESSANT** (I_0=1)[2].

Supposons que le frame **OP-DISTRACTION** émette en premier une annonce sur le tableau, ce qui est plausible puisque son influence est supérieure à celle de son concurrent **OP-TROUVER-JOB-INTERESSANT** . Le lien **OP-JOIE-DE-VIVRE-0**:`sous-opérateur` va alors se fixer sur le frame **OP-DISTRACTION**. Ce faisant il modifie l'influence I de **OP-DISTRACTION** et la fait passer au dessus du seuil d'activation.

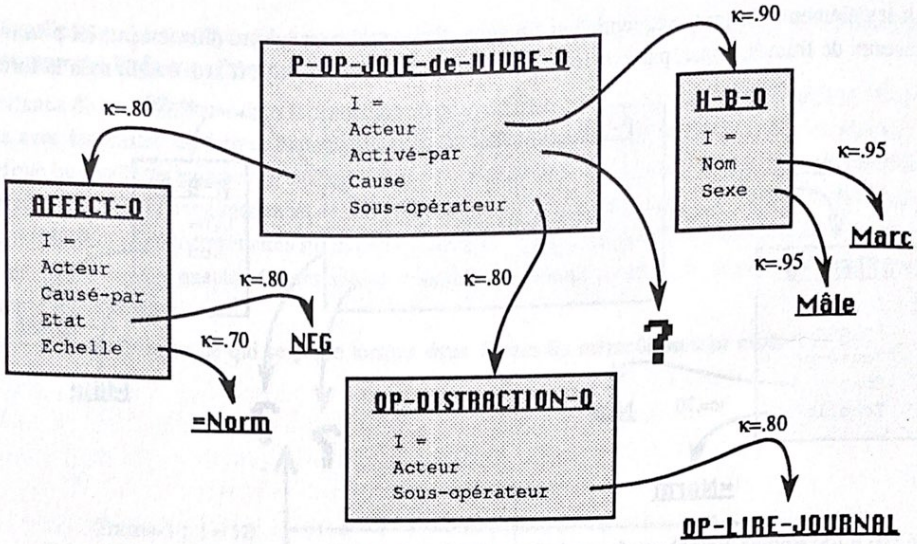
Un nouveau frame actif **OP-DISTRACTION-0**, copie active de **OP-DISTRACTION** se crée donc. Il est la cible du lien **OP-JOIE-DE-VIVRE-0**:`sous-opérateur`, et va se mettre lui aussi à chercher des cibles pour son slot `sous-opérateur`. Le lien **OP-JOIE-DE-VIVRE-0**:`sous-opérateur` --> **OP-DISTRACTION-0** est affecté du coefficient κ = reset-κ = 0.8 dans cet exemple.

Le lien **OP-DISTRACTION-0**:`sous-opérateur` peut à son tour se fixer sur une cible: par exemple **OP-LIRE-JOURNAL**.

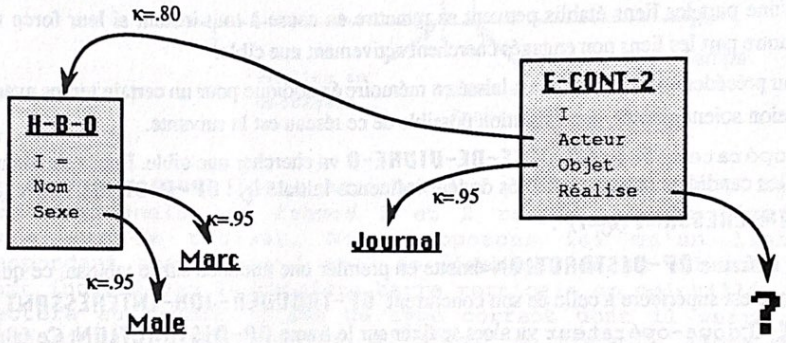
On obtient la mémoire épisodique suivante :

---

[2] Le rôle des influences initiales est d'exprimer le contexte général dans lequel se situe l'action. Ici la différence d'influences initiales signifie que l'on se trouve dans un contexte, une société, qui favorise la "distraction" par rapport au "travail" pour assurer le bonheur des individus



Supposons alors que le deuxième élément d'information ***"Il s'empara du journal sur le fauteuil ..."*** soit introduit dans la mémoire sous la forme suivante[3] :



On suppose que le pronom ***"il"*** a été reconnu comme une référence au seul acteur identifié jusqu'à présent **ETRE-HUMAIN-0.**

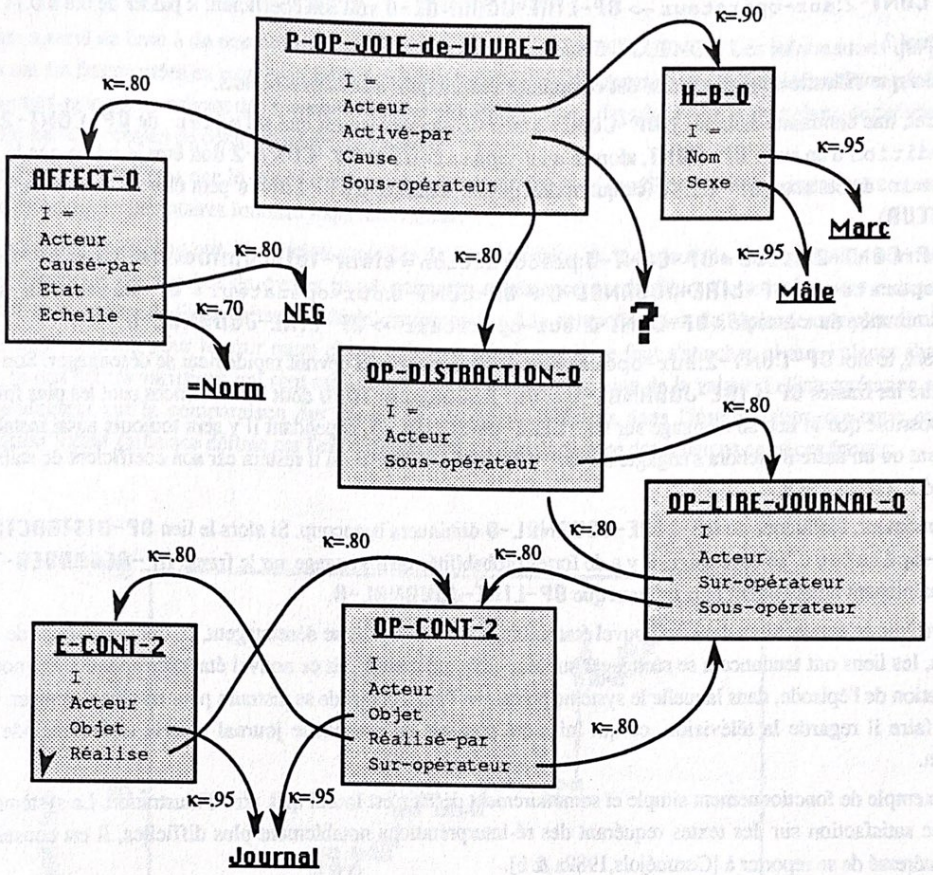
Le slot `opérateur` de l'évènement **E-CONT-2** cherche une cible. Cette recherche est en fait contrainte aux seules cibles de type **OP-CONT.** Une nouvelle entité **OP-CONT-2** est donc rapidement activée, et à son tour le slot `sur-opérateur` de **OP-CONT-2** cherche à se remplir.

Les choix possibles sont: **OP-LIRE-JOURNAL** (I_0=4), **OP-TUER-MOUSTIQUE** (I_0=1), **OP-REGARDER-TV** (I_0=4), et **OP-JOUER-MINITEL** (I_0=2). Les contraintes associées au slot `sur-opérateur` dans le frame **OP-CONT-2** signifient que seuls les plans **OP-LIRE-JOURNAL** et **OP-TUER-MOUSTIQUE** conviennent bien si l'`objet` de **OP-CONT-2** est un journal (κ = 0.2 dans les autres cas, ce qui veut dire que le candidat est alors fortement mis en cause).

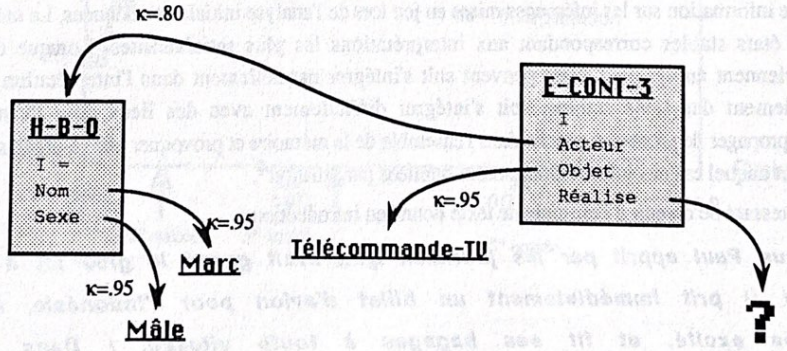
Supposons que le frame **OP-LIRE-JOURNAL** l'emporte grâce à son influence notablement plus élevée que celle de **OP-TUER-MOUSTIQUE** (4 au lieu de 2), alors le lien **OP-CONT-2**:`sur-opérateur` va se fixer sur **OP-LIRE-JOURNAL** et l'apport d'influence à **OP-LIRE-JOURNAL** va faire passer son influence au dessus du seuil

[3] Remarque : la référence au fauteuil n'est pas prise en compte ici.

d'activation, ce qui va donner naissance au frame actif **OP-LIRE-JOURNAL-0**. L'interprétation du système de ce qui lui a été dit jusque là est donc que Paul prend le journal afin de se distraire, ce qui lui restaurera sa joie de vivre.



Soumettons alors au système l'information finale : ***"... et saisit en dessous la télécommande de la télévision".***



Le lien **OP-CONT-3**:`sur-opérateur` cherche une cible qui d'après ses contraintes associées, ne peut être que **OP-REGARDER-TV** ou **OP-VERIFIER-COURANT**.

Dès que le lien **OP-CONT-2**:sur-opérateur --> **OP-REGARDER-TV-0** est établi (et cela devrait se produire car l'influence initiale de **OP-REGARDER-TV** est supérieure à l'influence initiale de **OP-VERIFIER-COURANT**), le lien **OP-CONT-2**:sur-opérateur --> **OP-LIRE-JOURNAL-0** voit son coefficient κ passer de 0.8 à 0.14.

Pourquoi ?

C'est ici que s'illustre en partie le rôle des contraintes sémantiques associées aux slots.

En effet, une contrainte associée à **OP-CONT-3**:sur-opérateur est que si l'effet de **OP-CONT-2** est la précondition d'un autre **OP-CONT**, alors le sur-opérateur de **OP-CONT-2** doit être le même que le sur-opérateur de cet autre **OP-CONT** (ce qui traduit qu'une séquence d'**OP-CONT**s peut être au service d'un même **OPERATEUR**).

Ici, **OP-CONT-2**:effet = **OP-CONT-3**:précondition = **clear-télécommande-TV**, et **OP-CONT-2**:sur-opérateur = **OP-LIRE-JOURNAL-0** ≠ **OP-CONT-3**:sur-opérateur = **OP-REGARDER-TV-0**, d'où la diminution du κ associé à **OP-CONT-2**:sur-opérateur --> **OP-LIRE-JOURNAL-0**.

Dès lors, le slot **OP-CONT-2**:sur-opérateur devient instable et devrait rapidement se désengager. Son choix réside entre les frames **OP-LIRE-JOURNAL-0** et **OP-REGARDER-TV-0** dont les influences sont les plus fortes. Il est fort possible que le lien se réengage sur **OP-LIRE-JOURNAL-0**, cependant il y sera toujours aussi instable. A un moment ou un autre il viendra s'engager sur **OP-REGARDER-TV-0** où il restera car son coefficient de stabilité k sera élévé κ = reset-κ =.8.

A ce moment, l'influence de **OP-LIRE-JOURNAL-0** diminuera beaucoup. Si alors le lien **OP-DISTRACTION-0**:sous-opérateur se désengage, il y a de fortes probabilités qu'il s'engage sur le frame **OP-REGARDER-TV-0** qui est maintenant notablement plus influent que **OP-LIRE-JOURNAL-0**.

Le système se trouve alors dans un nouvel état stable (même lorsqu'ils se désengagent, ce qui doit arriver de temps en temps, les liens ont tendance à se réengager sur leur cible antérieure). Et ce nouvel état correspond à une nouvelle interprétation de l'épisode, dans laquelle le système pense que Paul a choisi de se distraire pour ne plus s'ennuyer, et que pour ce faire il regarde la télévision, ce qui lui a été possible en prenant le journal puis la télécommande de la télévision.

Cet exemple de fonctionnement simple et sommairement décrit n'est fourni qu'à titre d'illustration. Le système a été testé avec satisfaction sur des textes requérant des ré-interprétations notablement plus difficiles, il est conseillé au lecteur intéressé de se reporter à [Cornuéjols,1989a & b].

Il faut en retenir que le processus de ré-interprétation s'opère spontanément au sein de la mémoire uniquement sur la base d'informations locales de cohérence et de parcimonie, en utilisant les connaissances sémantiques du système à l'exclusion de toute information sur les inférences mises en jeu lors de l'analyse initiale des données. La mémoire tend à évoluer vers des états stables correspondant aux interprétations les plus satisfaisantes. Lorsque de nouvelles informations parviennent au système, elles peuvent soit s'intégrer naturellement dans l'inteprétation courante en construisant facilement des liens stables, soit s'intégrer difficilement avec des liens dont l'instabilité peut éventuellement se propager de proche en proche dans l'ensemble de la mémoire et provoquer une ré-organisation, soit ne pas s'intéger du tout auquel cas ces informations seront oubliées par attrition[4].

Il peut être intéressant de revenir à l'exemple de texte donné en introduction :

> ***"Lorsque Paul apprit par les journaux qu'il avait gagné le gros lot à la loterie, / il prit immédiatement un billet d'avion pour l'Indonésie. / Il était très excité, et fit ses bagages à toute vitesse. / Dans sa précipitation il partit sans fermer le gaz. / Trois jours après, à Djakarta,***

---

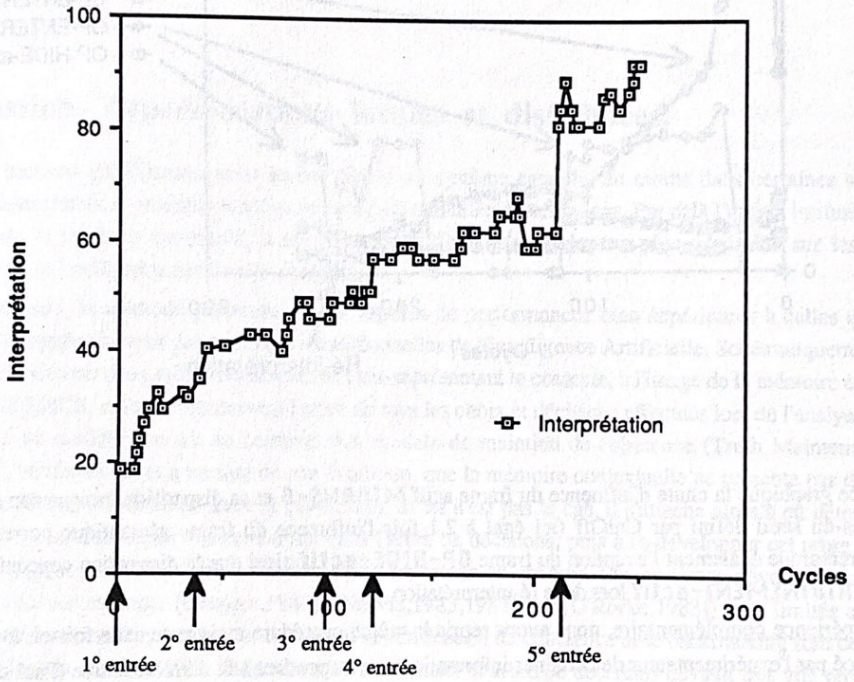
[4] Les frames dont l'influence est inférieure à un certain seuil (donc sont mal connectés) disparaissent au bout d'un certain temps.

Il apprit que les deux maflosi qui le cherchalent étalent morts dans l'explosion de son appartement. Il respira enfin."

Ce texte a servi de base à de nombreuses expériences avec le système INFLUENCE. Les informations qui y sont contenues ont été fragmentées en morceaux ici indiqués par les membres de phrases entre '/'. Ces informations partielles ont été fournies au système suivant des séquences variées afin d'étudier les effets de l'ordre et du rythme de présentation des données sur le processus de compréhension du système.

Ces expériences montrent que le système est capable de construire et de modifier dynamiquement ses croyances à partir d'informations fragmentaires fournies séquentiellement.

La figure suivante se veut une illustration grossière de la progression de la 'valeur' de l'interprétation courante du système en fonction du nombre de cycles passé pour une expérience particulière. Cette valeur est mesurée en pourcentage de la valeur de l'interprétation 'idéale' correspondant à la compréhension de l'épisode complet. Le calcul utilisé est trop simplifié pour fournir autre chose qu'une indication et il ne faut s'attacher ni aux valeurs absolues mesurées, ni aux petites variations qui sont sans significations. La quantification de la valeur de l'interprétation repose en effet seulement sur la comparaison des ensembles de frames présents dans l'interprétation courante et dans l'interprétation 'idéale' (telle que définie par l'expérimentateur) sans tenir compte des relations entre ces frames.

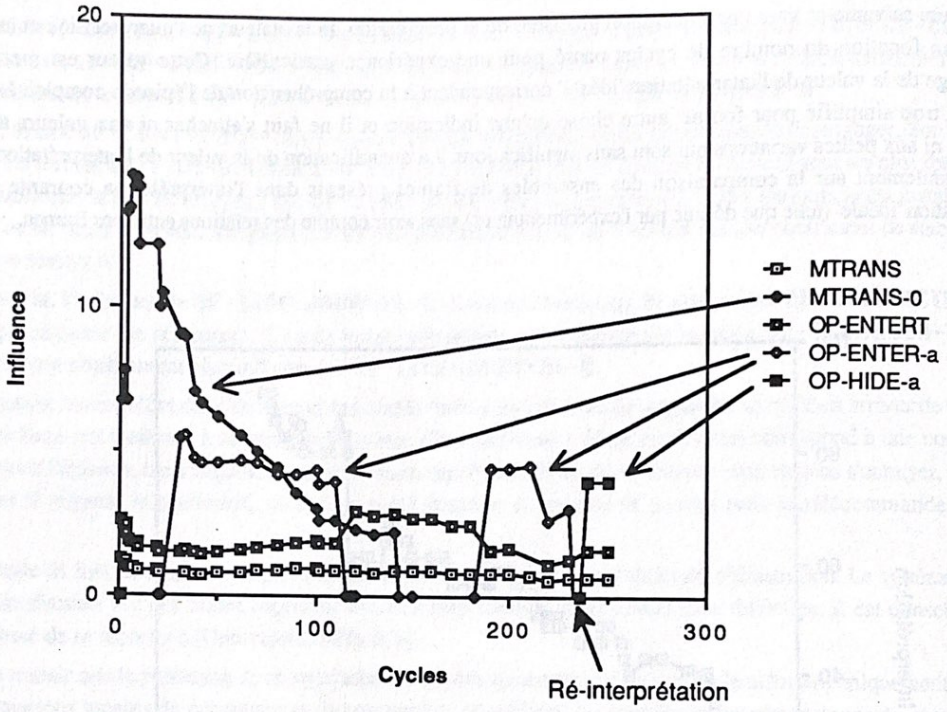


Le graphique montre clairement l'effet déterminant de la dernière entrée sur la valeur de l'interprétation finale. Cet effet correspond à la ré-interprétation rendue possible et nécessaire par le système.

Il faut noter que les cycles ne correspondent pas à des unités de temps constantes. En effet l'écriture de l'algorithme est telle que les cycles correspondent à des temps d'autant plus courts qu'il y plus de frames en mémoire (donc en particulier quand le nombre de frames actifs augmente). Ainsi donc l'équivalent durée des cycles 200-300 serait plus court que l'équivalent durée des cycles 0-100, puisqu'il y a plus de frames en mémoire à la fin du processus de

compréhension. Nous ne cherchons pas, par ailleurs, à établir une signification ou une équivalence de psychologie cognitive à ces durées.

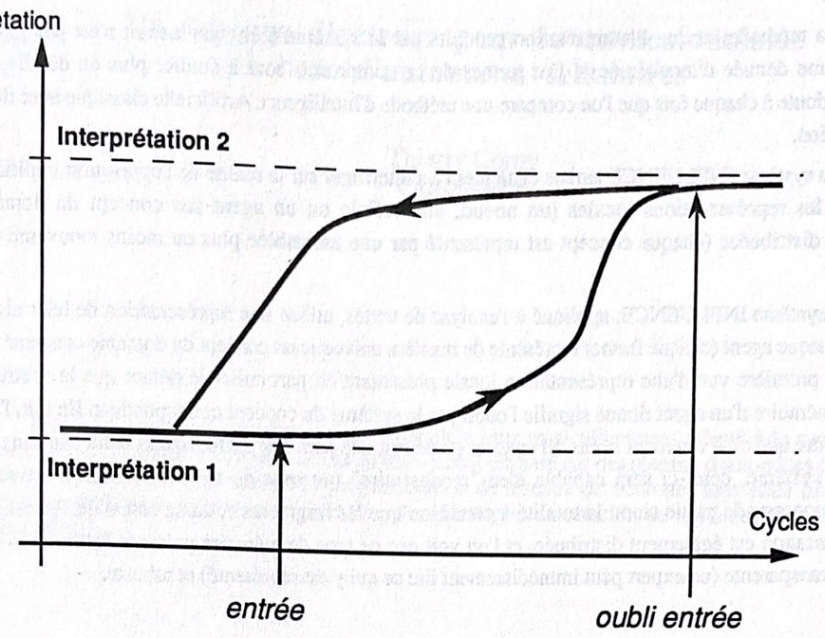
Dans la figure suivante sont reportées les évolutions des influences de quelques frames au cours du temps. Nous avons choisi les frames OP-HIDE et OP-ENTERTAINMENT qui sont révélateurs du phénomène de ré-interprétation, ainsi que le frame MTRANS-0 pour montrer comment il finit par être oublié par le système en perdant progressivement son influence.



On note sur ce graphique la chute d'influence du frame actif **MTRANS-0** et sa disparition lorsque son influence passe en dessous-du seuil défini par CutOff (ici égal à 2.1 fois l'influence du frame sémantique correspondant **MTRANS**). On remarque également l'irruption du frame **OP-HIDE-actif** ainsi que la disparition concomitante du frame **OP-ENTERTAINEMENT-actif** lors de la ré-interprétation.

En tant qu'expérience complémentaire, nous avons repris la même procédure mais avec cette fois-ci, lors du pas 270, un oubli forcé par l'expérimentateur de certaines informations correspondant à la dernière entrée (Paul savait que deux mafiosi le cherchaient pour le tuer). Le système revient alors à l'interprétation du voyage de Paul pour des raisons d'agrément au pas 305. Une dizaine de pas plus tard il reconstruit les **AFFECTs** positifs correspondant au gain à la loterie, puis oublie un peu plus tard l'existence des deux mafiosi.

Au delà des circonstances particulières de cette nouvelle expérience, il est intéressant de noter le phénomène d'hystéresis exhibé par le système.



4. Discussion. Représentations locales et distribuées.

Dans les sections précédentes, nous avons exposé un système capable, au moins dans certaines situations, de corriger une interprétation préalable inadéquate face à de nouvelles informations. Par delà l'intérêt intrinsèque de cette tentative et de la méthode présentée, il est intéressant d'en tirer des leçons plus générales sur les liens entre connexionnisme et Intelligence Artificielle classique.

En premier lieu, la méthode présentée ici est capable de performances bien supérieures à celles des systèmes analogues développées à partir des techniques traditionnelles de l'Intelligence Artificielle. Schématiquement, celles-ci impliquent de maintenir deux mémoires distinctes, l'une représentant le contexte, à l'image de la mémoire épisodique du système INFLUENCE, et l'autre conservant l'arbre de tous les choix et décisions effectués lors de l'analyse des entrées qui ont abouti au modèle courant du contexte. Un module de maintien de cohérence (Truth Maintenance System [Doyle,1979]) vérifie au fur et à mesure de son évolution que la mémoire contextuelle ne présente pas d'incohérence logique (un problème NP-complet dans sa généralité). Si tel n'est pas le cas, il s'attache alors à en détecter l'origine probable sous forme d'un choix malencontreux dans l'arbre de décisions, puis à re-développer cet arbre à partir d'un nouveau choix (processus classique du backtracking ou retour en arrière). Les systèmes développés suivant cette philosophie (voir par exemple [Granger,1980], [Norvig,1983,1987], ou [O'Rorke,1983]) sont limités au traitement d'exemples extrêmement simples dans la mesure où la vérification de cohérence et le backtracking sont des opérations délicates et difficiles, et où la taille de la mémoire à maintenir (l'arbre de décision) devient très vite prohibitive. Par ailleurs, l'arbre de décisions fait coexister au sein d'une même structure des informations de types très différents, tels des raisons syntaxiques de choix d'une interprétation (par exemple la règle de la récence pour l'interprétation d'un pronom), avec des raisons de sémantiques ou de contexte, qui sont difficiles à comparer. D'où le relatif échec de ces tentatives.

Le système INFLUENCE en revanche ne s'attache à maintenir que les informations sémantiques et contextuelles correspondant à la situation à représenter. C'est à la fois beaucoup plus naturel et suffisant dans la plupart des cas. Les conséquences positives de ce choix sont que cette méthode est facile à mettre en oeuvre (il est aisé d'augmenter un système d'interprétation existant, s'il est à base de frames, avec la méthodologie d'INFLUENCE), et qu'elle permet le fonctionnement du système même dans le cas où les informations fournies sont contradictoires. Les conséquences négatives sont que ce qui est gagné en taille de la mémoire à maintenir est compensé en partie par les nombreux calculs

impliqués par la méthode, et que l'interprétation produite par le système à chaque instant n'est pas garantie d'être optimale ni même dénuée d'incohérences. Les termes de ce compromis sont à étudier plus en détails, mais ils se retrouvent sans doute à chaque fois que l'on compare une méthode d'Intelligence Artificielle classique avec des méthodes d'auto-organisation.

L'examen du système INFLUENCE amène également à s'interroger sur la réalité de l'opposition traditionnellement évoquée entre les représentations locales (un noeud, une cellule ou un agent par concept du domaine) et les représentations distribuées (chaque concept est représenté par une assemblée plus ou moins mouvante d'entités du système).

En effet, le système INFLUENCE, appliqué à l'analyse de textes, utilise une représentation de haut niveau, à base de frames, où chaque agent (chaque frame) représente de manière univoque un concept du domaine concerné par le texte. Il s'agit donc à première vue d'une représentation locale présentant en particulier le défaut que la destruction ou la défaillance en mémoire d'un agent donné signifie l'oubli par le système du concept correspondant. En fait, l'observation du système montre que c'est rarement le cas. Si en effet on détruit une partie de l'information contenue dans la mémoire épisodique du système, celui-ci sera capable d'en "reconstruire" (au sens de mémoire reconstructive chère aux psychologues) une grande partie sinon la totalité à condition que les fragments restants soient suffisants. En un sens donc la représentation est également distribuée, et l'on voit que ce type de mémoire présente l'attrait non négligeable d'être à la fois transparente (un expert peut immédiatement lire ce qui y est représenté) et robuste.

Références

- John ANDERSON (1983) : **The Architecture of Cognition**, Harvard University Press, 1983.
- Antoine CORNUEJOLS (1989a) : **De l'Apprentissage Incrémental par Adaptation Dynamique : le système INFLUENCE**. *Thèse de doctorat* soutenue le 6 janvier 1989 à l'Université de Paris-Sud Orsay.
- Antoine CORNUEJOLS (1989b) : **"An Exploration into Incremental Learning : the INFLUENCE system"**, Proc. of *The 6th International Workshop on Machine Learning*, Ithaca, New-York, june 29th - july 1st, 1989.
- Jon DOYLE (1979) : **"A Truth Maintenance System"**, *Artificial Intelligence Journal*, 12, 1979, pp.231-272.
- Michael DYER (1983) : **In-Depth Understanding**, MIT Press, 1983.
- Richard GRANGER (1980) : **"When expectation fails: Toward a self-correcting inference system"** in *Proc. of the First National Conference on Artificial Intelligence*, Stanford, California, 1980.
- Peter NORVIG (1983) : **"Frame Activated Inferences in a Story Understanding Program"**, in Proc. *IJCAI-83*, Karlsruhe, 1983, pp.624-626.
- Peter NORVIG (1987) : **"Inference in Text Understanding"**, in Proc. of the *AAAI-87*, Seattle, Washington, july 13-17, 1987, pp.561-565.
- Paul O'RORKE (1983) : **"Reasons for beliefs in understanding: Applications of non-monotonic dependencies to story processing"**. In the Proc. of the *AAAI-83*, Washington, D.C., 22-26 août 1983, pp.306-309.
- Karl POPPER (1963) : **CONJECTURES AND REFUTATIONS. The growth of Scientific Knowledge.** Routledge and Kegan Paul, London and Henley, 1985.
- David WALTZ & Jordan POLLACK (1985) : **"Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation"**, in *Cognitive Science* 9, pp.51-74, 1985.