

Compréhension non monotone de texte : que garder en mémoire ?

Thème : Langage naturel (Détection et correction d'erreurs)

Type : Exposé long

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
(Tel.: 69-41-66-26 / 69-41-66-29)
Telex: ORSFAC 602166F
Télécopie: 64-46-19-92

Résumé :

La compréhension automatique du langage naturel nous confronte avec le problème de la construction incrémentale d'un modèle sur la base de données ambiguës fournies séquentiellement. Il est impossible dans ces conditions d'assurer à chaque instant le choix d'interprétation le plus judicieux, et il est nécessaire de donner aux systèmes la capacité de réviser leurs opinions si la nécessité s'en fait sentir. La question est alors quelles informations conserver pour corriger une interprétation erronée ?

Ce papier présente un système appelé INFLUENCE ré-organisant sa mémoire contextuelle dynamiquement en fonction des données perçues. Il repose sur des principes originaux tels que l'auto-organisation et l'absence de module superviseur, la permanence des remises en cause au sein de la mémoire, et l'expression des incertitudes sous forme d'instabilités. Ses performances de ré-interprétation sont comparées à celles de systèmes analogues, et des considérations sur la quantité d'informations à conserver et la complexité calculatoire dans chaque cas sont abordées.

Mots clé : Compréhension non monotone de langage naturel, raisonnement non monotone, auto-organisation, détection et correction d'erreurs, quantité d'informations.

1. Introduction

La compréhension automatique de textes en langage naturel nous confronte en particulier avec le problème de l'interprétation de données souvent intrinsèquement ambiguës, fournies séquentiellement de surcroît, sans qu'il soit généralement possible d'attendre de disposer de suffisamment d'informations pour effectuer cette interprétation en toute certitude. Il est donc nécessaire de prévoir des moyens de révision d'interprétation pour les cas où des données nouvelles viendraient à infirmer les croyances ou opinions courantes.

La tâche est difficile car elle demande que soient résolus le *problème de la détection* d'une inadéquation de l'interprétation courante face aux données disponibles, puis le problème de la *découverte de la modification à apporter* pour remédier à cette inadéquation. D'un côté, outre de savoir comparer les données avec le modèle courant du monde et d'être capable de déceler des inconsistences éventuelles (un problème NP-complet en logique du premier ordre), il faut également pouvoir juger de la valeur globale de ce modèle, par exemple en estimant sa cohérence ou son degré de parcimonie. De l'autre, il est indispensable d'avoir suffisamment de connaissances à sa disposition pour pouvoir corriger une interprétation défective et rétablir une image plus fidèle du monde. Dans les deux cas la représentation des connaissances et l'organisation de la mémoire jouent un rôle essentiel. Mais, par ailleurs, la quantité et la nature des informations à conserver pour être capable d'interpréter correctement une suite de données quel que soit l'ordre de leur présentation¹, condition d'une capacité de compréhension non-monotone, est également un paramètre capital. *Que peut-on se permettre d'oublier lorsque l'on fait le choix d'une interprétation pour rester assuré de pouvoir plus tard si nécessaire corriger ce choix et modifier son image du monde ?*

Cette question précise n'a pas, à notre connaissance, été étudiée jusqu'ici, ni directement, ni systématiquement. Nous n'avons pas non plus une telle ambition dans ce papier, mais nous souhaitons faire de cette interrogation l'objectif à travers lequel nous examinerons à la fois certains systèmes existants de compréhension non-monotone et le système INFLUENCE développé par nos soins.

¹ A des contraintes de correction syntaxique près (il ne s'agit pas de présenter les mots d'un texte dans un ordre totalement arbitraire).

Le reste de cette communication est organisé comme suit. La première section rappelle les travaux antérieurs dans le domaine de la compréhension non-monotone de texte. La section suivante décrit alors le système INFLUENCE en insistant sur ce qui le distingue des autres approches dans le cadre de notre investigation. Vient alors un aperçu de son mode de fonctionnement et de ses performances sur deux exemples. Finalement, nous revenons à une discussion plus approfondie, quoique encore préliminaire, de ce qu'il est nécessaire de mémoriser pour parvenir à interpréter correctement quelqu'en soit l'ordre une suite de données fournie séquentiellement.

2. Les systèmes de compréhension non-monotone.

Une des leçons chèrement mais définitivement apprise dans le domaine de l'analyse et de la compréhension automatique de texte, est qu'il est impossible d'y faire l'économie d'une théorie de la représentation des connaissances et de l'organisation de la mémoire. C'est d'ailleurs ce qui rend les travaux portant sur le langage naturel si pertinents pour l'ensemble des recherches sur l'Intelligence Artificielle. En effet, la détermination du sens correct d'un mot, pour ne pas parler d'une métaphore ou d'une analogie, dépend en général d'inférences plausibles basées sur des caractéristiques extrêmement complexes du contexte en conjonction avec des connaissances arbitraires (a priori) sur le monde. Une grande partie des inférences rendant possible l'analyse des entrées textuelles est donc fondée d'abord sur les connaissances du monde et sur la représentation du contexte contenues en mémoire et non sur des inférences dépendant des caractéristiques syntaxiques du texte analysé.

Par ailleurs, la pertinence d'une interprétation donnée se mesure surtout, sinon intégralement, grâce à certaines caractéristiques de sa représentation en mémoire (telles la cohérence ou la parcimonie), sans faire appel à une évaluation des inférences mises en jeu lors de l'analyse qui l'a produite.

Est-il alors nécessaire de conserver une trace de ces inférences d'analyse syntaxiques ?

Si de nombreux travaux ont exploré le domaine de l'analyse et de la compréhension automatique de texte, relativement peu ont été consacrés à l'élaboration de systèmes capables de remettre en cause leur interprétation courante et de lui en substituer une autre jugée mieux appropriée. Parmi ceux-ci, les plus notables sont ceux de [Granger,1980] avec le système

ARTHUR, de [O'Rorke,1983] avec RESUND, de [Norvig,1983 & 1987] avec FAUSTUS, et de [Eiselt, 1985 & 1987]: système ATLAST. Ils partagent tous la caractéristique (jugée cruciale par [Birnbaum,1985] pour ce type de systèmes) d'incorporer des mécanismes inférentiels propres à l'évaluation des interprétations produites à côté des mécanismes inférentiels destinés au processus d'interprétation lui-même. Ces mécanismes s'appuient en particulier sur divers aspects de la représentation en mémoire de l'interprétation courante pour juger de sa valeur, tels le degré de ramification du réseau sémantique produit (parcimonie) et la connexité de ce réseau (cohérence). Cependant le deuxième point remarquable de ces approches, est qu'une fois qu'une incohérence ou situation insatisfaisante a été décelée par le mécanisme d'évaluation, la correction et la modification de l'interprétation courante s'effectue, elle, par un retour en arrière dans l'arbre de décisions d'inférences produit par le mécanisme d'interprétation. Cela signifie d'une part que le rôle des mécanismes propres à la mémoire se borne à l'évaluation de celle-ci à l'exclusion de la construction d'interprétation, et, d'autre part, qu'il est nécessaire de conserver sous une forme ou une autre les inférences et alternatives rencontrées tout le long du processus d'interprétation.

Pourtant, une fois reconnu le rôle fondamental de la connaissance et du contexte inscrits en mémoire, à la fois en cours d'interprétation des données et pour l'évaluation des modèles produits, ne serait-il pas possible de l'étendre au processus de ré-interprétation lui-même en s'affranchissant de la mémorisation, de caractère artificiel et peut-être redondant, des inférences effectuées au cours de l'histoire du système ? En d'autres termes et pour prendre un exemple, est-il nécessaire de se rappeler que tel pronom a été interprété de telle façon parce que rapporté à l'entité la plus récemment citée pour être capable éventuellement de remettre en cause cette interprétation et lui en substituer une autre ?

Si tel était le cas, cela compliquerait immensément la tâche de ré-interprétation et exigerait une double mémoire peu naturelle et encombrante, l'une contenant le résultat du processus d'analyse de texte, c'est-à-dire l'interprétation proprement dite, l'autre l'histoire des inférences effectuées pour parvenir à cette interprétation. C'est, brutalement exposée, l'approche retenue dans les systèmes cités plus haut, et en général dans les systèmes à maintien de justification comme TMS [Doyle, 1979]. Il est à noter que le caractère artificiel d'une telle solution se trouve magnifié dans le domaine de la compréhension du langage

naturel dans la mesure où bien souvent les inférences qui président à l'interprétation d'un mot par exemple sont d'ordre syntaxique (règle de récence par exemple dans l'interprétation des pronoms) c'est-à-dire n'ont rien à voir avec les raisons sémantiques ou pragmatiques qui permettent de juger dans un deuxième temps de la valeur d'une interprétation. Il serait souhaitable au contraire que les mêmes raisons qui permettent de juger d'une interprétation soit aussi celles qui fondent et réalisent le processus de ré-interprétation².

C'est en tous cas cette dernière solution qui est explorée avec le système INFLUENCE que nous décrivons dans la suite de ce papier. Il présente une alternative aux systèmes à maintien de justifications en ne retenant (tout au moins explicitement) aucune des inférences permettant d'interpréter les entrées textuelles. Les ré-interprétations ne sont ainsi guidées que par des considérations directement liées au contenu de la mémoire contextuelle.

3. 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.

² Autrement on peut se demander si il ne serait pas aussi simple et économique de conserver le texte d'origine dans son intégralité. Il semble par ailleurs que tous les types de solutions soient employés par l'homme. Ainsi de fortes évidences existent que l'ensemble des interprétations possibles des données d'entrées soient conservées pendant un court laps de temps après perception, de même que, dans les cas de ré-interprétation difficile il est observé que le regard du lecteur se reporte en arrière pour, apparemment, vérifier des données. Cependant, dans la majorité des cas les ré-interprétations semblent se réaliser sans faire appel à l'un ou l'autre des ces deux mécanismes par ré-organisation directe du contenu de la mémoire sémantique.

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 schéma 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 musulmane 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 économique 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 schéma 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*.

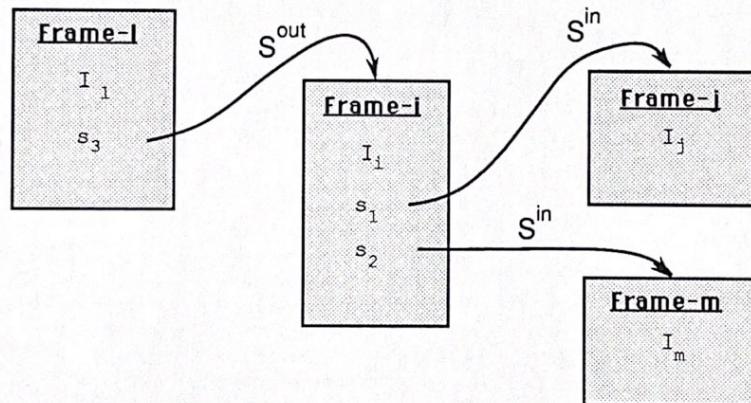


Figure 1 : Dans la situation présentée ici, l'influence du frame-i serait :

$$I_i = I_{i0} + (S_{s1}^{in} \cdot I_j + S_{s2}^{in} \cdot I_m) + S_{s3}^{out} \cdot I_l$$

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-1** grâce aux S^{in} , la seconde aux gains réalisés par les liens pointant sur le **FRAME-1**, 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.

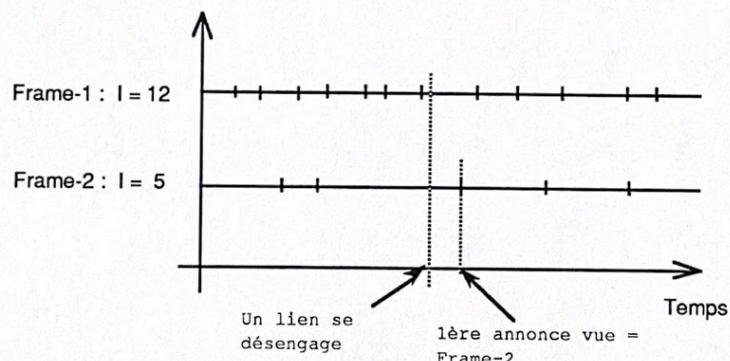


Figure 2 : Sur les deux lignes horizontales supérieures sont indiqués les instants auxquels les frames 1 et 2 respectivement "émettent" une annonce sur le tableau. Nous supposons ici qu'un lien du type correspondant aux frames 1 et 2 se désengage de sa cible à un certain instant indiqué par la première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

décrise 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.

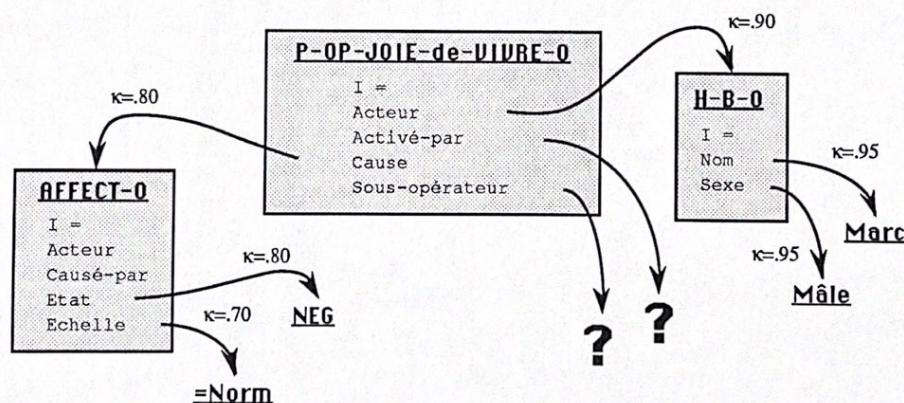
4. Esquisse d'un exemple de fonctionnement

Avant de détailler plus précisément les caractéristiques du modèle INFLUENCE, il peut-être utile d'en apprêhender 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-VIURE-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** ($I_0=1$)³.

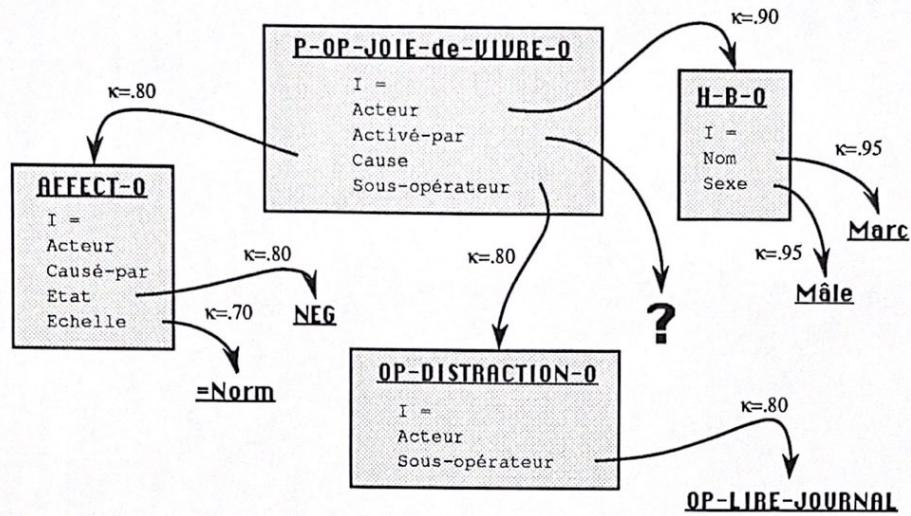
Supposons que le frame **OP-DISTRACTION** émette en premier une annonce sur le tableau, ce qui est plausible puisque son influence est supérieure à celle de son concurrent **OP-TROUVER-JOB-INTERESSANT**. Le lien **OP-JOIE-DE-VIURE-0:sous-opérateur** va alors se fixer sur le frame **OP-DISTRACTION**. 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-VIURE-0:sous-opérateur**, et va se mettre lui aussi à chercher des cibles pour son slot sous-opérateur. Le lien **OP-JOIE-DE-VIURE-0:sous-opérateur --> OP-DISTRACTION-0** est affecté du coefficient $\kappa = \text{reset-}\kappa = 0.8$ dans cet exemple.

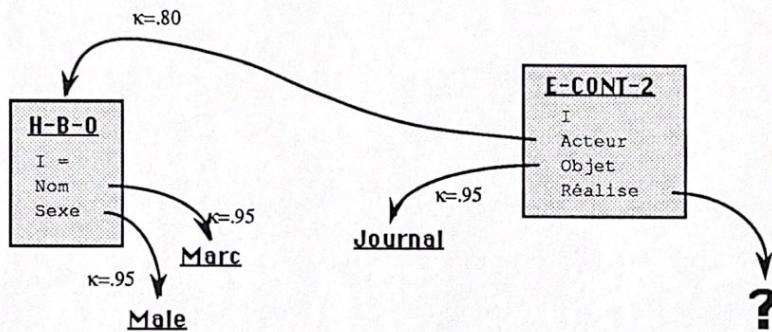
Le lien **OP-DISTRACTION-0:sous-opérateur** peut à son tour se fixer sur une cible: par exemple **OP-LIRE-JOURNAL**.

On obtient la mémoire épisodique suivante :

³ 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 suivante⁴ :



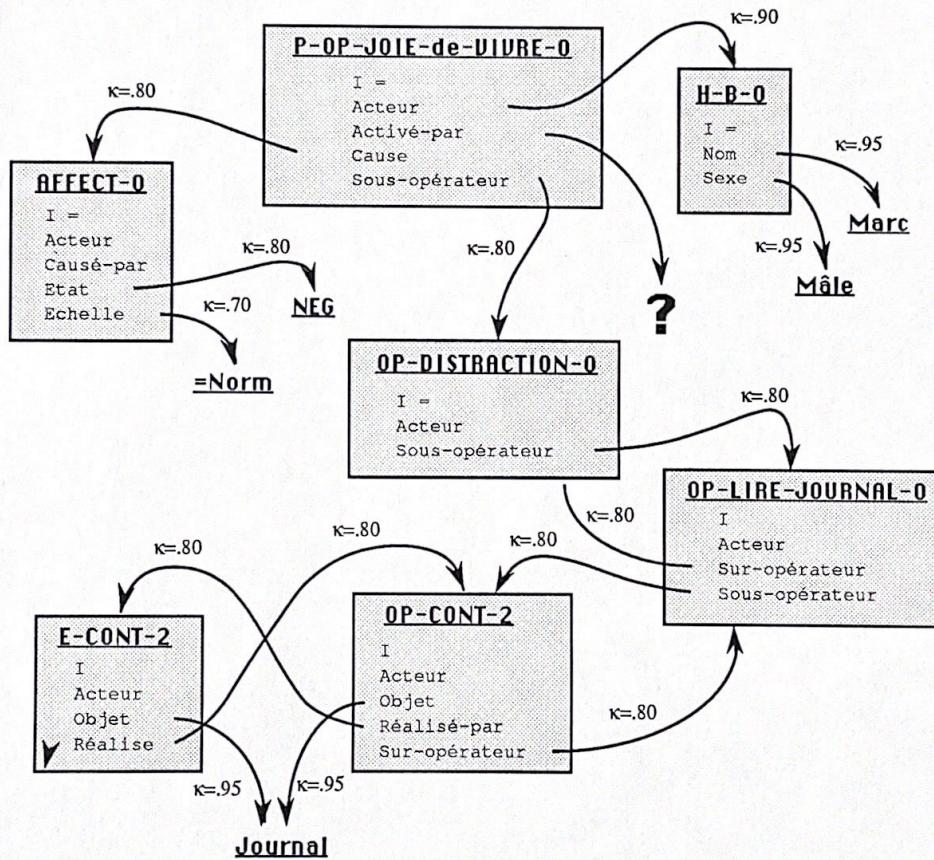
On suppose que le pronom "*Il*" a été reconnu comme une référence au seul acteur identifié jusqu'à présent **ETRE-HUMAIN-0**.

Le slot opérateur de l'événement **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.

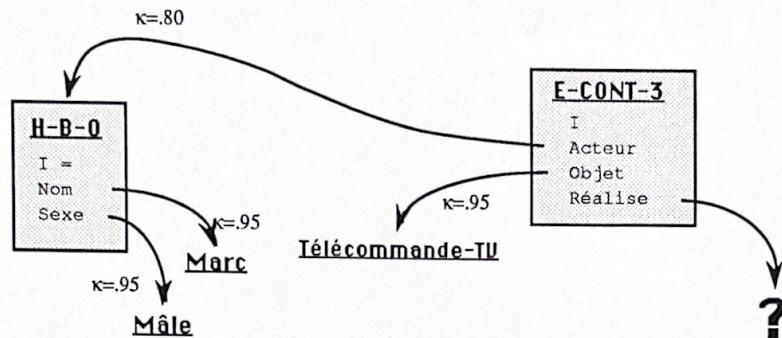
⁴ Remarque : la référence au fauteuil n'est pas prise en compte ici.

Les choix possibles sont: **OP-LIRE-JOURNAL** ($I_0=4$), **OP-TUER-MOUSTIQUE** ($I_0=1$), **OP-REGARDER-TU** ($I_0=4$), 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 ($\kappa = 0.2$ 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 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-TU** ou **OP-VERIFIER-COURANT**.

Dès que le lien **OP-CONT-2:sur-opérateur** --> **OP-REGARDER-TU-0** est établi (et cela devrait se produire car l'influence initiale de **OP-REGARDER-TU** 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-CONTs** peut être au service d'un même **OPERATEUR**).

Ici, **OP-CONT-2:effet** = **OP-CONT-3:précondition** = **clear-télécommande-TU**, et **OP-CONT-2:sur-opérateur** = **OP-LIRE-JOURNAL-0** ≠ **OP-CONT-3:sur-opérateur** = **OP-REGARDER-TU-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-TU-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 élevé $\kappa = \text{reset-}\kappa = .8$.

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 notamment 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 notamment plus difficiles⁵, il est conseillé au lecteur intéressé de se reporter à [Cornuéjols,1989a].

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'interprétation courante en construisant facilement des liens stables, soit s'intégrer difficilement avec des liens dont l'instabilité peut éventuellement se propager de proche en

⁵ Par exemple: "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."

proche dans l'ensemble de la mémoire et provoquer une ré-organisation, soit ne pas s'intégrer du tout auquel cas ces informations seront oubliées par attrition⁶.

5. Quelle mémoire pour quelle ré-interprétation ?

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, une saine réaction pourrait s'exprimer par : "Et bien voilà encore un autre système de compréhension non monotone, et après !?". Il nous semble en effet qu'il faut examiner une question d'intérêt plus général. Lorsque des données ambiguës sont fournies séquentiellement à un système, et que celui-ci ne peut avoir recours qu'à des inférences inductives donc faillibles pour les analyser, *quels types et quelle quantité d'informations doit-il conserver pour être capable de produire finalement une interprétation correcte de ces données quel que soit l'ordre dans lequel elles ont été perçues* ? C'est tout le problème du raisonnement non monotone : par quel "miracle" est-il possible de s'affranchir des circonstances historiques et fortuites de la présentation des données ? Dans un autre contexte, cette question conditionne également les possibilités d'apprentissage incrémental et conduit à la notion de stratégie d'éducation d'un système (voir [Cornuéjols,1989a & 1989b]). On retrouve aussi ce problème, sous une autre forme, en planification adaptative.

Nous distinguons dans la suite quatre types de systèmes suivant les informations qu'ils conservent afin d'être en mesure de corriger leur interprétation courante.

1. Conservation des données brutes.

C'est la solution extrême qui consiste à conserver l'ensemble des entrées. Notons que c'est encore la solution la plus employée en apprentissage incrémental (voir [Vrain & Lu,1988]) ou en planification adaptative par exemple. Cette solution d'aspect trivial n'est pas exempte de difficultés (en dehors de la mémoire nécessaire dans cette approche) car en cas d'interprétation paraissant inadéquate, il reste encore le problème

⁶ Les frames dont l'influence est inférieure à un certain seuil (donc sont mal connectés) disparaissent au bout d'un certain temps.

d'en localiser la raison puis d'effectuer la correction. En l'absence d'informations supplémentaires, cela signifie probablement l'obligation d'analyser à nouveau complètement l'ensemble des données disponibles. Il est donc utile de conserver, en plus des données, des "pointeurs" vers les points critiques. Ces pointeurs peuvent prendre en particulier la forme de traces d'inférences. D'où la solution suivante.

2. Conservation des inférences et choix effectués lors de l'analyse des données.

C'est sur ce principe que se fondent les systèmes de maintien de cohérence et de justifications à la TMS [Doyle,1979]. Elle sous-tend la quasi totalité des systèmes existants de compréhension non monotone de textes tels ARTHUR de [Granger,1980] qui conserve un tableau de pointeurs vers toutes les inférences générées durant le processus d'interprétation d'un texte, ou RESUND de [O'Rorke,1983] directement calqué sur le système TMS avec en sus des heuristiques de choix des inférences à rétracter. Le système FAUSTUS de [Norvig,1983 & 1987] conserve quant à lui l'ensemble des frames qui auraient pu être activés. Finalement le système ATLAST de [Eiselt,1987] retient les inférences potentielles sous la forme de chemins dans un graphe causal qui représente l'ensemble de toutes les interprétations possibles. Cette dernière approche appartient également à la solution 3 décrite ci-après.

Plusieurs objections viennent légitimement à l'esprit. D'abord, il est probable que la quantité d'informations à maintenir est rédhibitoire excepté pour les exemples les plus simples. Ensuite, comme nous l'avons déjà noté en introduction, le maintien de toutes les justifications implique, en particulier dans le cas de la compréhension du langage naturel, de maintenir ensemble des raisons d'ordres très différents (syntaxiques, sémantiques, pragmatiques, ...) qu'il est difficile de comparer et de combiner.

3. Maintien d'un réseau causal représentant toutes les relations possibles dans le domaine étudié.

Un notable et très intéressant exemple de cette approche est le concept de réseau d'inférence bayésien de [Pearl,1986,1987,1988]. Il est particulièrement adapté aux tâches de diagnostic (voir [Geffner & Pearl,1987]) et en général aux domaines dans lesquels le réseau causal sous-jacent est facile à exprimer et défini à l'avance. Il est

beaucoup plus discutable et difficile à mettre en oeuvre dans les domaines où les relations sont potentiellement très riches et nombreuses comme en langage naturel. De nombreux systèmes à base de propagation de marqueurs dans un réseau sémantique pré-existant ont pourtant été proposés [Anderson,1983], [Waltz & Pollack,1985], etc... Nous pensons pour notre part que cette solution n'est pas viable telle quelle en dehors de petits exemples jouets, mais qu'elle peut être fort utile en tant que mécanisme supplémentaire d'inférence pour l'analyse de texte (voir [Charniak,1983 & 1986]). Elle fait par ailleurs un pas vers un oubli des inférences de base qui produisent l'interprétation des données et vers un processus de ré-interprétation ne prenant en compte que les connaissances du domaine et du contexte inscrits dans la mémoire.

4. Oubli des inférences d'interprétation et phénomène de mémoire reconstructive.

C'est le cas du système INFLUENCE. La mémoire se ré-organise uniquement en fonction de mesures de cohérence et de parcimonie au sein de l'interprétation produite sans intervention aucune des inférences utilisées ou envisagées lors du processus d'analyse du texte. Ici, l'information nécessaire à la ré-interprétation provient entièrement de connaissances sur le domaine (inscrites dans les frames et les restrictions attachées à chaque slot) et dont les effets s'expriment en partie dans les influences associées à chaque frame et dans les coefficients de stabilité de chaque lien. Il n'y a plus de distinction entre des étapes d'évaluation d'interprétation, de localisation de la cause de l'inadéquation et de correction, mais au contraire intégration de ces trois processus. Une interprétation insatisfaisante se manifeste par une grande instabilité de la mémoire de contexte, et la ré-organisation résultante de la mémoire vers un état plus stable correspond à la correction de l'erreur initiale d'interprétation. Cette ré-organisation, guidée à la fois par les connaissances du domaine (dans les frames et les slots) et de contexte (coefficients d'influences et d'incertitude), est similaire à une mémoire reconstructive qui, lorsqu'il lui manque des éléments d'information, les reconstruit à partir des connaissances disponibles. Le résultat est que l'absence de traces explicites des données initiales ou des inférences mises en jeu au cours de leur analyse se traduit contre une charge accrue d'inférences et une tendance du système à préférer les interprétations les plus plausibles donc aussi les plus "banales". Il peut ainsi arriver

qu'il soit nécessaire de répéter au système certaines informations correspondant à des faits difficilement reconstructibles⁷.

Pour résumer, la capacité d'un système à corriger son interprétation en fonction de données nouvelles requiert la conservation d'informations redondantes. Il n'est pas facile dans un domaine comme la compréhension du langage naturel où les connaissances mises en jeu sont multiples et encore mal formalisées, de préciser quantitativement le degré de redondance nécessaire. On peut cependant dresser un panorama des solutions possibles pour obtenir cette redondance. A un bout de l'horizon toutes les données brutes sont conservées, et le système doit ré-analyser les données complètement chaque fois qu'un état insatisfaisant est constaté. A l'autre bout de l'horizon en revanche toute trace des données d'entrée et des inférences syntaxiques opérant directement sur elles sont oubliées, et la ré-interprétation résulte des processus naturels de raisonnement internes à la mémoire. La redondance provient ici de la connaissance générale du monde dont dispose le système initialement. Entre ces deux extrêmes, se trouvent les approches mixant une conservation plus ou moins complète des inférences mises en jeu lors de l'interprétation des données (celles-ci pouvant alors être oubliées dans leur forme explicite), et une connaissance du monde de toute façon importante.

Nous avons montré avec le système INFLUENCE que la solution extrême d'oubli total des données et processus initiaux d'interprétation est viable et performante. Nous pensons aussi qu'elle est la plus naturelle. Notre projet est maintenant d'analyser plus formellement les conditions à remplir en termes de mémoire et de calculs pour que soit possible la compréhension non-monotone de données.

Références

- John ANDERSON (1983) : **The Architecture of Cognition**, Harvard University Press, 1983.
- Lawrence BIRNHAUM (1985) : "Lexical Ambiguity as a Touchstone for Theories of Language Analysis", in Proc. of IJCAI-85, Los Angeles, august 18-23, 1985, pp.815-820.

⁷ Il est à noter que le système INFLUENCE est capable de poser et d'ordonner des questions en fonction de leur pertinence, ce qui permet d'aider le "professeur" du système.

- Eugene CHARNIAK (1983) : "Passing Markers: A theory of Contextual Inference in Language Comprehension", in *Cognitive Science* 7, pp.171-190, 1983.
- Eugene CHARNIAK (1986) : "A Neat Theory of Marker Passing", In Proc. of the AAAI-86, Philadelphia, Pennsylvania, August 11-15, 1986, pp.584-588.
- 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 system INFLUENCE", à paraître dans les 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.
- Hector GEFFNER & Judea PEARL (1987) : "An Improved Constraint-Propagation Algorithm for Diagnosis", dans *IJCAI-87*, Milan, 23-28 août 1987, pp.1105-1111.
- 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.
- Judea PEARL (1986) : "Fusion, Propagation, and Structuring in Belief Networks", in *Artificial Intelligence journal*, 29, pp.241-288, 1986.
- Judea PEARL (1987) : "Distributed Revision of Composite Beliefs", in *Artificial Intelligence journal*, 33, pp.173-215, 1987.
- Judea PEARL (1988) : *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufmann, 1988.
- Karl POPPER (1963) : *CONJECTURES AND REFUTATIONS. The growth of Scientific Knowledge*. Routledge and Kegan Paul, London and Henley, 1985.
- Christel VRAIN & C.R. LU (1988) : "Apprentissage Incrémental par Analogie", dans les Actes des Troisièmes Journées Françaises de l'Apprentissage (JFA-88), Cassis, France, 5-6 mai 1988.
- David WALTZ & Jordan POLLACK (1985) : "Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretation", in *Cognitive Science* 9, pp.51-74, 1985.