

Revue Internationale de

ISSN 0980-1472

systemique

CONNAISSANCES IMPLICITES
ET
CONNAISSANCES EXPLICITES

Vol. 6, N° 1-2, 1992

afcet

DUNOD

AFSCET

Revue Internationale de
systemique

Revue
Internationale
de Sytémique

volume 06, numéro 1-2, pages 99 - 103, 1992

Apprentissage et connaissances implicites

Alexandre Andreewsky

Numérisation Afscet, août 2017.



Creative Commons

J.-L. LE MOIGNE, *Intelligence des mécanismes, mécanisme de l'intelligence*, Fayard, Fondation Diderot, Paris, 1986.

J.-L. LE MOIGNE, *Sur la capacité de la raison à discerner. Rationalité substantive et rationalité procédurale*. NR du GRASCE n° 91-24.

J. PITRAT, *Métacognition : futur de l'intelligence artificielle*. Hermès, Paris, 1990.

J.-C. PLANÈS, G. VEDRENNE, Un système d'analyse en langage naturel et de génération de textes de synthèse pour les enquêtes de conjoncture de la Banque De France. Avignon, EC2, mai 1991.

H. SIMON, *Models of discovery*, Pallas paperbacks D. Reidel publishing company.

T. WINOGRAD, F. FLORES, *L'intelligence artificielle en question*, PUF, Paris, 1989.

APPRENTISSAGE ET CONNAISSANCES IMPLICITES

A. ANDREEWSKY

LIMSI-CNRS ¹

Résumé

L'utilisation de méthodes d'apprentissage dans la construction des systèmes d'intelligence artificielle a pour but, en principe, d'éviter tout un travail de description exhaustive des données, objets et relations diverses indispensables au fonctionnement de ces systèmes. En d'autres termes, à l'aide d'une métastructure explicitement définie avant la phase d'apprentissage, on cherche une intégration automatique implicite de tous les éléments qui complètent le système et le rendent opérationnel.

Abstract

For Artificial Intelligence systems, learning may avoid the full description of the data, objects and relations required for these systems. We present on the concrete example of syntax learning, how such a learning process should be built upon an implicit syntactical knowledge.

1. Rappels

Dans ce qui suit on désigne par *système*, un ensemble d'éléments (qui peuvent être des objets quelconques) assujettis à des relations ou dépendances de n'importe quelle nature. Les notions de système et d'éléments en relation sont difficiles à définir indépendamment l'une de l'autre et sont acceptées comme étant des notions premières ou primitives. En mathématiques, les systèmes sont étudiés d'une manière axiomatique et formelle et ne sont pas rattachés au concret, tandis qu'en systémique on s'attache davantage au passage de la réalité à sa description formelle.

1. B.P. 133, 91403 Orsay Cedex.

1.1. *Information relative entre deux systèmes*

L'information relative fournie par un système A à un système B est définie par toute correspondance entre parties de A et parties de B qui modifie l'organisation de B.

La réalisation effective de cette correspondance suppose que l'on a mis les systèmes en interaction par un moyen quelconque, afin qu'objets et relations de B soient modifiés par les règles de correspondance définies issues de A.

On peut encore dire que l'information relative entre un système émetteur A et un système récepteur B est matérialisée par tout stimulus qui venant de A modifie l'état interne de B.

1.2. *Information implicite*

Un stimulus issu d'un système émetteur A modifie en général différemment les états internes respectifs de récepteurs quelconques B1, B2, . . . , B*i*. Cette différence peut être imputée à la spécificité de chacun des systèmes récepteurs, spécificité qui est liée à leur structure interne et à leurs acquis (dans le cas de systèmes humains on insiste parfois sur l'aspect conscient ou inconscient de ces acquis). Cette aptitude du système récepteur à interpréter, compléter, modifier et donner une résonance particulière à l'information reçue, permet et suggère de formuler le concept d'*information implicite spécifique*, pour chaque couple émetteur-récepteur. On donne encore à cette information implicite des appellations diverses : le non-dit, la présupposition, le sous-entendu, l'allusion, ...

Selon les ambiguïtés et les nécessités du dialogue usuel, une certaine explication de l'information implicite s'effectue couramment, surtout lorsque le contexte pragmatique ne permet pas la reconstitution du non-dit, et donc la compréhension de l'information. Mais l'implicite étant une constante de la communication, une explication systématique complète, outre qu'elle n'est pas toujours réalisable, rendrait le dialogue impossible.

Un autre type d'informations implicites concerne les non-dits et associations diverses des apprentissages, des mécanismes inconscients de nos activités quotidiennes, ou des tâtonnements intuitifs qui précèdent le raisonnement rigoureux. Par exemple, nous sommes capables de faire des analyses grammaticales et même de produire ces analyses au moyen de systèmes automatisés. Mais actuellement, on ne sait pas imaginer quelles sont les procédures cérébrales réelles qui opèrent en permanence au cours du dialogue

2. Le module syntaxique à apprentissage du système SPIRIT

Dans les systèmes à apprentissage qui sont conçus pour modifier et compléter leurs structures à partir des entrées qui leur sont proposées, seul l'état initial est explicité, du fait même qu'il est programmé. Dans les systèmes ayant pour but l'analyse syntaxique du langage naturel, l'apprentissage s'impose, en raison de la complexité de la tâche. La qualité des relations syntaxiques ainsi obtenues est évaluée par l'aptitude de ces systèmes à reconnaître la fonction des mots, ce qui suppose notamment leur désambiguïtation grammaticale.

Afin de mieux préciser la relation entre apprentissage et explicitation, nous allons discuter le module syntaxique utilisé dans le système documentaire SPIRIT (Andreevsky *et al.*, 1973), où la syntaxe est utilisée pour séparer les fonctionnels des sémantiques et pour « normaliser » les substantifs, adjectifs, et verbes conjugués — c'est-à-dire déterminer le masculin singulier à partir du pluriel et l'infinitif à partir des formes conjuguées.

Dans ce système le module syntaxique est construit comme suit :

— Tout d'abord, on procède à l'analyse grammaticale d'une suite d'énoncés. Cette phase représente l'initialisation de l'apprentissage et elle est entièrement manuelle. A cette étape on constitue aussi un système grammatical en rétroaction avec l'algorithme d'apprentissage et d'analyse défini.

— Ensuite on construit automatiquement un dictionnaire où sont cumulées les ambiguïtés grammaticales des formes lexicales du texte d'apprentissage. Par exemple un énoncé comme « le car part car le temps presse », s'analysera, dans l'ordre, *article * substantif * verbe conjugué * conjonction * article * substantif * verbe conjugué*, et la forme « car » cumulera les étiquettes : *substantif/conjonction*.

— Dans la troisième et dernière étape, on crée un fichier — texte des ambiguïtés, à l'aide du texte initial et du dictionnaire de cumul. En comparant ce fichier — texte au texte initial, on obtient des règles de désambiguïtation binaires, ternaires ou plus. Par exemple l'énoncé ci-dessus : *le car part...* va générer des règles de la forme :

*(article/pronom * substantif/conjonction) → (article * substantif) etc.*

L'astérisque « * » signifie : « suivi de », la flèche « → » indique la transformation subie par le terme de gauche.

Les règles que l'on obtient de cette façon peuvent être ambiguës. Par exemple le couple de mots : « la » et « ferme » peut également appartenir à l'énoncé « la ferme du voisin » ou « cette porte, il la ferme ». Pour cette

raison le couple « la ferme » va générer une règle dont la partie gauche :

(*article/pronom * verbe conjugué/substantif*)

aura plusieurs résolutions : *article * substantif* ou encore : *pronom * verbe conjugué*.

Ce fait a conduit naturellement à pondérer les résolutions par des fréquences, chaque fois que le corpus d'apprentissage était suffisamment long et caractéristique de la langue étudiée.

Si le nombre de catégories grammaticales définies pour l'analyseur syntaxique est de l'ordre de la centaine, pour un texte de l'ordre de 4 000 mots, on peut obtenir jusqu'à 600 règles binaires et 3 000 règles ternaires. Lorsque le texte d'apprentissage augmente, ces chiffres ont tendance à se stabiliser à un niveau cependant assez important. Par exemple, pour 100 000 mots on peut avoir jusqu'à 10 000 règles binaires.

— L'analyse syntaxique, au moyen des règles obtenues, se fait de proche en proche en vérifiant la compatibilité de résolution d'une règle à l'autre.

Par exemple pour l'analyse des trois mots consécutifs « le car part », la jonction des deux règles suivantes :

(*article/pronom * substantif/conjonction*) → (*article * substantif*)

(*substantif/conjonction * substantif/verbe*) → (*substantif * verbe*)

donne nécessairement la suite (*article * substantif * verbe*), où, comme on le voit, la catégorie *substantif* joue un rôle de transition entre les résolutions de la première et de la seconde règle.

Enfin, il existe dans ce « parser » une option « analyse par défaut » qui permet d'attribuer à un mot inconnu du dictionnaire une ou plusieurs catégories grammaticales compatibles avec les contextes syntaxiques gauches et droits.

3. L'apprentissage comme moyen de contourner l'explicitation

Ainsi donc l'analyseur syntaxique que nous venons de décrire réalise un apprentissage assez typique des systèmes artificiels : accumulation des données, pondération par des fréquences lexicales et syntaxiques calculées sur des corpus donnés, restructuration manuelle ou automatique à partir des sorties du système, et enfin inférence, c'est-à-dire aptitude à analyser des cas différents de ceux qui ont été introduits au cours de la phase d'apprentissage. Ce parser utilise un mode d'analyse tributaire à la fois de l'apprentissage et du texte analysé, auquel il s'adapte dynamiquement. L'explicitation des

algorithmes programmés et de la liste des règles syntaxiques obtenues au cours de l'apprentissage est toujours possible, mais difficilement exploitable. En particulier elle ne donne pas d'indications sur le mode d'enchaînement de ces règles pour un énoncé donné et ce, du fait même que la compatibilité de résolution des règles consécutives utilisées dépend des propriétés locales (à grande variabilité) des énoncés traités. La connaissance explicite des états internes du système ne donne qu'une idée très vague de ses potentialités, sans permettre de prévoir ses défaillances. L'expérience a en particulier montré que les erreurs dans l'apprentissage qui donnent lieu à la création de règles fausses sont difficilement décelables par le simple examen de ces règles, et que la meilleure détection est encore celle que l'on obtient en rétroaction avec les résultats de l'analyse. En d'autres termes l'examen des sorties du système permet de repérer les erreurs et de remonter à leurs sources pour améliorer les performances de l'analyseur.

Conclusion

Nous avons décrit un système à apprentissage qui gère automatiquement les modifications de ses états internes; une telle approche permet (entre autre) d'éviter l'explicitation des connaissances syntaxiques, acquises à partir d'un ensemble de corpus de la langue. Explicitation et apprentissage dans ce système s'articulent ainsi en une dualité implicative, tout en étant mutuellement exclusifs.

Bibliographie

- A. ANDREEWSKY, C. FLUHR, *Apprentissage, analyse automatique du langage, application à la documentation*. Documents de linguistique quantitative, N° 21, Dunod, Paris, 1973.
- A. ANDREEWSKY, M. DESI, C. FLUHR, *Méthode d'apprentissage pour l'analyse automatique morphosyntaxique et lexicale sémantique de la langue espagnole*, COLING 82, Prague, Czechoslovakia, 1982.
- L. von BERTALANFFY, *General System Theory*, George Brazillier, New York, 1968.
- J.-L. LE MOIGNE, *La théorie du système général, Théorie de la modélisation*. Puf, Paris, 1990.
- R. VALLÉE, *Subjectivité et systèmes*. Perspectives systémiques I, Centre Culturel International de Cerisy-la-Salle, 1986, in *Perspectives systémiques I*, Paulré B. (Ed.), pp. 44-51, L'Interdisciplinaire, Limonest, 1986.