



HAL
open science

Modéliser les interactions entre une expression polysémique et son co-texte

Bernard Victorri

► **To cite this version:**

Bernard Victorri. Modéliser les interactions entre une expression polysémique et son co-texte. Claude Guimier, Salah Mejri, Bernard Victorri. Co-texte et Calcul du sens, Presses universitaires de Caen, 1997. hal-04534152

HAL Id: hal-04534152

<https://hal.science/hal-04534152>

Submitted on 16 Apr 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Modéliser les interactions entre une expression polysémique et son co-texte

Bernard VICTORRI
ELSAP - CNRS

ABSTRACT : *When analysing the behavior of a polysemous term in a sentence, it is commonly taken for granted that the other terms of the sentence (i.e. the co-text) behave in a non-polysemous way. In fact, this is not generally true, because of the ubiquity of the polysemous phenomenon. It seems therefore necessary to consider interactions between one term and its co-text as a Gestalt process : each term contributes to the global meaning and conversely the global meaning determines the meaning of each term. This paper presents a mathematical model of such a process using the theory of dynamical systems as well as a computer implementation based on an original architecture of recurrent neural networks.*

Quand on étudie le comportement sémantique d'une expression polysémique dans un ensemble d'énoncés, on est amené à considérer que le sens de l'expression est déterminé par son co-texte dans chacun des énoncés étudiés. En effet, par définition même de la polysémie, si cette expression peut prendre des sens différents suivant les énoncés dans lesquelles elle est insérée, il faut bien que ce sens dépende de quelque chose de plus que ce qu'apporte l'unité elle-même. C'est sur cette base que nous avons proposé, il y a quelques années¹, un modèle opératoire de la polysémie, dans lequel le calcul du sens d'une expression polysémique dans un énoncé était déterminé par les éléments co-textuels présents dans cet énoncé. Plus précisément, nous définissions deux espaces, l'espace sémantique associé à l'expression, et l'espace des éléments co-textuels pertinents, et le processus de détermination du sens de l'expression était modélisé par une fonction qui faisait correspondre à chaque point de l'espace co-textuel une dynamique sur l'espace sémantique. L'implémentation informatique de ce modèle, à l'aide de réseaux connexionnistes récurrents, nous a permis de faire

1. On trouvera un exposé complet et un bilan critique de ce travail dans B. Victorri et C. Fuchs, 1996, *Le Phénomène de la polysémie*, Paris, Hermès.

un certain nombre d'expérimentations concrètes sur des marqueurs grammaticaux, en particulier *encore*.

Les difficultés que nous avons rencontrées lors de ces expérimentations nous ont convaincus que notre modèle était insuffisant pour rendre compte de la complexité du phénomène de construction du sens dans un énoncé. En effet, si le sens d'une expression polysémique est bien déterminé par son co-texte, ce processus n'est pas « unilatéral » : il faut prendre en compte l'influence qu'exerce en retour l'expression sur le sens dudit co-texte. En effet, il ne faut pas oublier que ce co-texte contient lui aussi, généralement, des éléments polysémiques dont le sens est déterminé par leur co-texte, dont fait partie l'expression initiale sur laquelle on s'était focalisé. Cela nous a conduit, en collaboration avec Yves-Marie Visetti², à concevoir un nouveau modèle capable de rendre compte de la réciprocité de ces interactions, dont l'implémentation informatique est en cours. C'est ce modèle que nous présentons ici, en mettant l'accent sur une de ses caractéristiques essentielles : son architecture « adaptable » à la taille et à la structure de l'énoncé. Nous verrons en quoi cela constitue une avancée importante par rapport à d'autres modèles connexionnistes proposés dans ce domaine, mais aussi quelles sont les limites actuelles de cette adaptabilité.

1. Le problème à résoudre

Quelques exemples permettront de mieux comprendre la nature des difficultés que nous avons rencontrées avec notre premier modèle, centré sur une seule unité polysémique. Ces exemples portent sur l'unité polysémique que nous avons le plus étudiée, à savoir l'adverbe *encore*. Quand *encore* porte sur le groupe verbal, la nature lexicale du verbe, notamment le type de procès qu'il exprime, joue un rôle important dans la détermination du sens de *encore*. C'est ainsi qu'avec un verbe qui marque une progression, *encore* peut prendre une valeur de renchérissement sur cette progression, et il devient paraphrasable par *davantage* :

(1) Ces mesures ont encore accéléré la crise dans le pays.

(= Ces mesures ont accéléré davantage la crise dans le pays)

En revanche, un verbe marquant un événement ponctuel entraînera le plus souvent une valeur répétitive de *encore* :

(2) Ces mesures ont encore jeté le pays dans la crise.

(= Ces mesures ont jeté de nouveau le pays dans la crise)

On peut donc en déduire que la nature lexicale du type de procès du verbe doit faire partie des éléments co-textuels qui permettent de calculer le sens de *encore*.

2. Plusieurs étudiants de troisième cycle participent ou ont participé à ces travaux sous notre co-direction : Sophie Delamarche de l'Université de Paris VI, Fabrice Fontaine et Guillaume Rolland de l'Université de Caen.

Le problème, c'est que, pour certains verbes polysémiques, le type de procès exprimé par le verbe dépend du sens spécifique que celui-ci prend dans l'énoncé considéré. Ainsi, pour le verbe *précipiter*, qui peut vouloir dire aussi bien *accélérer* que *jeter*, les deux valeurs de *encore* sont possibles :

(3) Ces mesures ont encore précipité la crise dans le pays.

(4) Ces mesures ont encore précipité le pays dans la crise.

Si *précipiter* prend des sens différents dans ces deux énoncés, c'est bien sûr à cause d'autres éléments co-textuels : c'est essentiellement ici la classe sémantique du complément direct de *précipiter* qui compte : dans l'énoncé (3), il s'agit d'un processus (*crise*), qui peut être accéléré, tandis que dans l'énoncé (4), il s'agit d'une entité sociale incompatible avec le sens d'accélération. C'est en effet le sens que prend ici *pays*, ce qui doit à son tour être expliqué, puisque ce terme est lui-même aussi polysémique³.

Ainsi obtient-on une chaîne sans fin, du moins sans autre fin prévisible que la prise en compte dans le co-texte de tous les éléments constitutifs de l'énoncé, non pas parce qu'ils seraient directement pertinents pour déterminer le sens de *encore*, mais parce qu'ils sont susceptibles d'influencer indirectement ce sens par l'intermédiaire d'autres éléments co-textuels.

Prenons un autre exemple pour montrer toute l'ampleur du problème :

(5) Il fume encore.

On sait qu'avec un verbe de type activité comme *fumer*, on se trouve généralement en présence de deux ambiguïtés classiques : l'une, qui touche *encore*, entre une valeur répétitive (*Il fume de nouveau*) et une valeur durative (*Il continue à fumer*), et l'autre, qui touche le présent, entre le « présent de situation » (*Il est encore en train de fumer*) et le « présent de propriété » (*Il est encore fumeur*). Remarquons d'ailleurs que la valeur répétitive de *encore* correspond préférentiellement à la valeur de présent de situation (*Il est de nouveau en train de fumer*), tandis que la valeur durative s'accommode aussi bien des deux valeurs du présent (*Il continue à être en train de fumer* et *Il continue à être fumeur*).

Si l'on ajoute un complément, le déterminant qui introduit celui-ci joue un rôle décisif dans la détermination des valeurs aussi bien de *encore* que du présent. Ainsi si l'on compare les deux énoncés suivants :

(6) Il fume encore un cigare.

(7) Il fume encore le cigare.

3. On se rend alors compte de la possibilité d'un cercle vicieux. Si *pays* prend cette acception ici, c'est dû en grande partie à la présence de *précipiter* : le sens de chacun de ces termes dépend donc de l'autre. Le même phénomène peut d'ailleurs se produire entre la valeur de *encore* et le type de procès exprimé par le verbe, comme nous l'avons montré dans C. Fuchs, L. Gosselin et B. Victorri, 1991, « Polysémie, glissements de sens et calcul des types de procès », *Travaux de Linguistique et de Philologie*, n° spécial « Les typologies de procès », Strasbourg. Par exemple, si l'on compare *Une minute plus tard la bombe explosait*, et *Une minute plus tard la bombe explosait encore*, on se rend compte que la présence de *encore* modifie considérablement le sens des éléments qui sont censés constituer son co-texte.

on privilégiera une lecture répétitive de *encore*, avec un présent de situation pour (6) (*Il est de nouveau en train de fumer un cigare*, ou, cela revient – presque – au même : *Il est en train de fumer un cigare de plus*), alors que ce sera exactement le contraire pour (7) : un *encore* duratif et un présent de propriété (*Il continue à être fumeur de cigares*). Remarquons aussi que dans (6), *un* prend la valeur d'un déterminant spécifique (*un cigare bien particulier*), tandis que dans (7), *le* prend la valeur d'un déterminant générique (*le cigare en général*). Mais que l'on ajoute une relative à ce complément, et à nouveau les trois possibilités décrites pour l'énoncé (5) *Il fume encore* réapparaissent, avec chacun des deux déterminants, qui prennent d'ailleurs chacun une valeur tantôt spécifique tantôt générique :

(8) Il fume encore un cigare qu'il a allumé il y a une heure.

(9) Il fume encore le cigare qu'il a allumé il y a une heure.

(10) Il fume encore un cigare qu'on ne trouve plus en France depuis longtemps.

(11) Il fume encore le cigare qui a obtenu la médaille d'or au concours de La Havane.

(12) Il fume encore un cigare qu'il a dû prendre en cachette dans ton bureau.

(13) Il fume encore le cigare qu'il a dû éteindre quand le contrôleur est passé.

Il existe bien sûr une certaine « solidarité » entre les valeurs que prennent les trois unités polysémiques sur lesquelles nous nous sommes focalisés : le temps verbal, l'adverbe aspectuel et le déterminant. C'est pour cette raison que l'on n'obtient facilement⁴ que trois combinaisons :

– pour (8) et (9) : adverbe duratif + présent de situation + déterminant spécifique,

– pour (10) et (11) : adverbe duratif + présent de propriété + déterminant générique,

– pour (12) et (13) : adverbe répétitif + présent de situation + déterminant spécifique,

alors que si la combinatoire était libre, on aurait $2^3 = 8$ combinaisons en tout.

Cet exemple montre bien, d'une part, que l'on a affaire à des systèmes d'interaction entre marqueurs polysémiques en relation syntagmatique, plutôt qu'à l'influence unidirectionnelle du co-texte sur une unité polysémique particulière, et d'autre part, que l'on ne peut effectivement pas prédire, comme nous l'avait déjà montré l'exemple précédent, la taille du co-texte à prendre en compte pour déterminer le sens d'un élément polysémique. Le rôle que jouent les relatives dans cet exemple démontre à l'évidence ce dernier point : puisque l'on doit aller chercher dans une autre proposition des éléments co-textuels décisifs, rien n'empêche, théoriquement, que ces éléments ne nécessitent à leur tour, de manière

4. On peut sans doute, mais plus difficilement, obtenir au moins une autre combinaison, à savoir adverbe répétitif + présent de propriété + déterminant générique, dans laquelle *Il fume encore le cigare* voudrait dire *Il refume le cigare une fois de plus*.

réursive, la prise en compte d'une nouvelle subordonnée pour la détermination de leur sens.

Ces exemples, et d'autres⁵, nous ont convaincus que les limites que nous rencontrions dans les expérimentations informatiques avec notre premier modèle n'étaient pas dues à de simples problèmes techniques, mais bien à un problème de fond. Nous cherchions à calculer directement le sens d'une expression polysémique à partir des *formes* présentes dans le co-texte « proche », et c'est le *sens* de ces éléments co-textuels qui comptait. Bien sûr, en augmentant la taille du co-texte pris en compte, on pouvait espérer traiter de plus en plus d'énoncés, mais cette complexification non seulement rendait le travail informatique très pénible, mais surtout n'avait plus beaucoup d'intérêt : il était absurde qu'un système spécialisé dans le traitement d'une unité polysémique soit amené à traiter « par la bande » les problèmes de détermination du sens de toutes les autres unités polysémiques qui pouvaient être présentes dans les énoncés.

2. Un modèle de « compositionnalité gestaltiste »

Pour résoudre ce problème, il fallait donc se résoudre à abandonner le point de vue local que nous avions privilégié, pour passer à une modélisation du processus global d'interaction entre tous les éléments de l'énoncé. Autrement dit, comme l'illustre le schéma de la *figure 1*, le modèle à concevoir devait permettre de rendre compte des interactions entre les *sens* des expressions polysémiques de l'énoncé (représentées sur la figure par les flèches en gras). Bien entendu, un tel modèle est incompatible avec la conception classique de la compositionnalité⁶, puisqu'il s'oppose à tout ordre prédéterminé de calcul du sens des différentes composantes de l'énoncé. Le sens d'un énoncé est le résultat d'un double mouvement, puisque ce sens est fonction du sens des expressions qui le composent, et qu'inversement le sens de ces expressions dans cet énoncé est fonction du sens global de l'énoncé lui-même. On a donc affaire à un système qui obéit aux principes de base de la *Gestalttheorie* : le tout est plus que la somme de ses parties et « une partie dans un tout est autre chose que cette partie isolée ou dans un autre tout »⁷.

Il faut noter deux intérêts complémentaires du modèle illustré par ce schéma. D'abord, il est, malgré les apparences, plus simple que notre premier modèle, du moins si on l'examine localement. En effet, si l'on s'intéresse à une unité donnée, alors qu'il nous fallait prendre en compte dans le premier modèle toutes les unités qui, de près ou de loin, pouvaient influencer le calcul du sens de cette unité, on peut ici se limiter aux éléments qui interviennent directement dans ce

5. Cf. B. Victorri et C. Fuchs, 1996, *Le Phénomène de la polysémie*.

6. Cf. R. Montague, 1970, « English as a formal language », in R. Thomason (ed.), *Formal philosophy, selected papers of Richard Montague*, New-Haven, Yale University Press ; G. Gazdar et alii, 1985, *Generalized phrase structure grammar*, Oxford, Blackwell.

7. P. Guillaume, 1979, *La Psychologie de la forme*, Paris, Flammarion, p. 23.

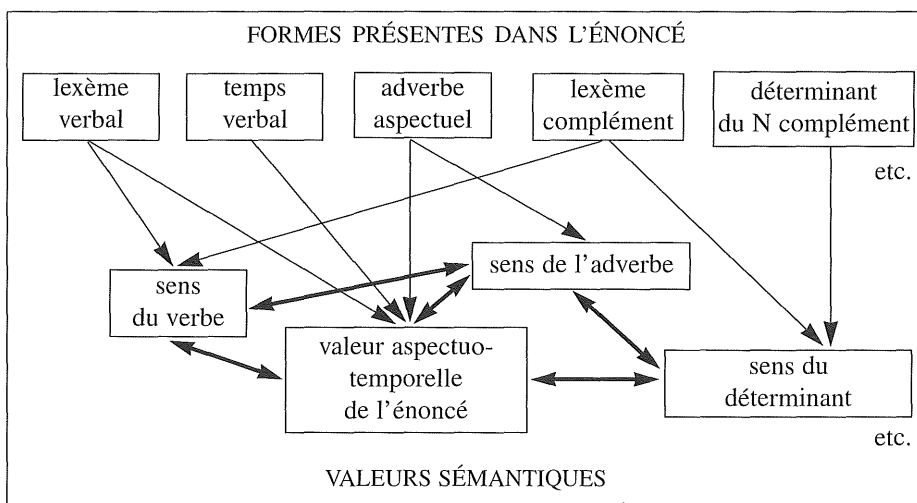


Fig. 1 : Dynamique d'interaction entre éléments polysémiques

calcul⁸: c'est ainsi qu'on peut voir sur la figure que le sens de l'adverbe aspectuel ne dépend plus que de quatre éléments : la *forme* de cet adverbe bien sûr (*encore, déjà, toujours, alors, etc.*), le *sens* du verbe, celui du déterminant, et surtout la valeur aspectuo-temporelle de l'énoncé. Bien sûr, ce schéma est loin d'être complet, mais il s'agit de montrer simplement ici le principe d'économie que cela représente, et de souligner que cela correspond assez bien à la manière dont travaille le linguiste quand il cherche à décrire le fonctionnement sémantique d'une unité polysémique en termes de règles : il répertorie en effet les éléments (locaux et globaux⁹) du co-texte qui modifie le sens de l'unité étudiée, et il utilise très spontanément comme prémisses de ses règles la donnée du sens de ces éléments.

La deuxième remarque concerne justement cet élément « global » que nous avons intitulé « valeur aspectuo-temporelle de l'énoncé ». Il s'agit d'un élément de sens qui concerne l'énoncé dans son ensemble. Il dépend à la fois du temps du verbe, de son type de procès, d'adverbes et autres compléments temporels, sans qu'il soit possible ni même souhaitable d'attribuer cet élément de sens à une seule de ces unités¹⁰. Il est important que le modèle permette l'existence de ces éléments de sens global, et qu'il respecte donc leur particularité, qui correspond, là aussi, à la manière de voir habituelle du linguiste.

8. On peut, de ce point de vue, classer ce modèle dans ce que Kleiber appelle « l'approche cognitive » de la notion de contexte, au sens où chaque unité « convoque » le contexte qui sert à déterminer son sens. Cf. G. Kleiber, 1994, « Contexte, interprétation et mémoire : approche standard vs approche cognitive », *Langue française*, 103, p. 9-22.

9. Cf. ci-après.

10. Voir par exemple L. Gosselin, 1996, *Sémantique de la temporalité en français*, Louvain-la-Neuve, Duculot, qui développe un modèle de représentation et de calcul des valeurs aspectuo-temporelles du français, tout à fait compatible avec le point de vue défendu ici.

Ce modèle peut-il prendre une forme réellement « calculatoire » ? On pourrait penser *a priori* que les « boucles » que dessinent les flèches en gras dans le schéma de la *figure 1* sont autant de « cercles vicieux » qui interdisent à tout calcul d'être mené à son terme. Cela serait vrai si ces flèches représentaient des contraintes statiques conduisant à des calculs logico-algébriques. Mais l'outillage mathématique que nous avons déjà mis en œuvre dans notre premier modèle, à savoir la théorie des systèmes dynamiques, permet de briser facilement ce cercle vicieux. En effet, nous considérons que les flèches ne contraignent pas directement les valeurs de l'élément, mais contraignent la forme d'une dynamique sur l'espace de tous les sens potentiels de l'élément, que nous appelons l'espace sémantique associé à cet élément. Il est alors possible de montrer que le schéma de la *figure 1* correspond à un système dynamique sur le produit cartésien des espaces sémantiques associés aux différents éléments, ce système dynamique étant déterminé par la donnée des formes présentes dans l'énoncé.

Nous n'allons pas entrer ici dans les détails mathématiques du modèle, mais simplement en donner une description intuitive, suffisante pour la suite de cet exposé. On peut se représenter un système dynamique (du moins ceux que nous utilisons ici¹¹) comme la donnée d'un « relief » sur un espace : à chaque point de l'espace est associé un vecteur qui indique la ligne de plus grande pente passant par ce point. La dynamique revient à se déplacer le long de ces pentes pour atteindre les points les plus bas des vallées, que l'on appelle les attracteurs de la dynamique. Ainsi, si l'on interprète les points de l'espace comme les états d'un système, se donner une dynamique sur un espace, c'est se donner une loi d'évolution qui indique comment les états se modifient à partir d'un état initial quelconque pour se stabiliser dans un état final attracteur. L'ensemble de tous les états initiaux conduisant à un même état final s'appelle le bassin de cet attracteur.

Dans notre modèle, nous définissons d'abord des dynamiques locales pour chaque élément de l'énoncé : dans l'espace sémantique associé à cet élément, pour chaque co-texte (c'est-à-dire la donnée de la *forme* et du *sens* des éléments co-textuels pertinents), on définit une dynamique locale dont le ou les attracteurs représentent le ou les sens que peut prendre l'élément considéré dans le co-texte considéré. Autrement dit, ces dynamiques locales traduisent les « règles » de fonctionnement sémantique de l'élément en question. Ce sont des règles du type : « Dans tel co-texte, *encore* prend une valeur durative, dans tel autre une valeur répétitive », avec bien sûr la possibilité de cas de figure plus complexes : « Dans tel co-texte, *encore* est ambigu entre une valeur durative et répétitive, dans tel autre, il y a indétermination entre une valeur durative et une valeur de

11. Les dynamiques décrites ici sont les dynamiques dérivant d'un potentiel, telles que nous les avons décrites dans C. Fuchs et B. Victorri, 1996, *Le Phénomène de la polysémie*. Pour une introduction aux systèmes dynamiques, voir par exemple A. Chenciner, 1992, « Systèmes dynamiques différentiables », *Encyclopædia universalis* 21, p. 1044-1080. Sur l'intérêt général de l'utilisation des systèmes dynamiques en linguistique, voir C. Fuchs et B. Victorri, (eds.), 1994, *Continuity in linguistic semantics*, Amsterdam, John Benjamins.

renchérissement», etc. En cela, il n'y a rien de nouveau par rapport à notre ancien modèle, à ceci près que, dans ce nouveau modèle, le co-texte contient le *sens* des éléments co-textuels, et non plus seulement leur *forme*, ce qui, rappelons-le, évite toutes les difficultés que nous avons rencontrées avec notre premier modèle¹².

Ensuite, nous définissons, pour chaque énoncé, un espace sémantique global associé. Celui-ci est constitué par le produit cartésien des espaces sémantiques de chaque élément de cet énoncé, c'est-à-dire qu'un point de cet espace représente un sens possible pour chacun de ces éléments. On démontre que l'on peut alors toujours définir, par simple composition, une dynamique sur cet espace global qui soit compatible avec les dynamiques locales définies ci-dessus, au sens où un attracteur de la dynamique globale est un attracteur de chacune des dynamiques locales : pour chaque élément, le sens que l'on obtient est un attracteur de la dynamique locale définie par la forme et le sens des autres éléments présents dans l'énoncé.

Ainsi, ce nouveau modèle correspond bien, au moins en théorie, à ce que nous voulions. Il suffit de donner les règles de fonctionnement sémantique de chaque unité polysémique en supposant connu le sens des autres unités polysémiques susceptibles de l'entourer, pour que, dans chaque énoncé, la dynamique globale construite à partir de ces règles prenne en compte l'interaction réciproque des diverses unités. Reprenons l'exemple de *Il fume encore*, en simplifiant au maximum le raisonnement (l'idée étant d'illustrer le plus simplement possible le mécanisme de base du modèle, et non de traiter complètement et fidèlement cet exemple). Supposons que nous ayons défini une dynamique locale pour l'adverbe *encore*, qui corresponde aux deux règles suivantes :

R₁ : *encore* + verbe de type activité + présent de propriété => adverbe duratif,

R₂ : *encore* + verbe de type activité + présent de situation => ambiguïté
entre adverbe duratif et adverbe répétitif,

et que d'autre part, nous ayons défini de même une dynamique locale pour le présent, qui corresponde aux règles :

R₃ : temps présent + verbe de type activité + adverbe répétitif => présent
de situation,

R₄ : temps présent + verbe de type activité + adverbe duratif => ambiguïté
entre présent de situation et présent de propriété.

Alors la dynamique globale, construite à partir de ces deux dynamiques locales, possèdera trois attracteurs, correspondant aux trois combinaisons « admissibles » (adverbe répétitif + présent de situation, adverbe duratif + présent de situation, et adverbe duratif + présent de propriété), sans que l'on ait à spécifier davantage la procédure de calcul. De même, si pour une raison ou une autre (par exemple en présence d'un déterminant à valeur générique), la dynamique locale associée au présent ne possédait plus qu'un attracteur, correspondant au

12. Cf. 1. *Le problème à résoudre*, supra, p. 234.

présent de propriété, alors la dynamique globale n'aurait plus elle aussi qu'un attracteur, correspondant à la seule combinaison admissible dans ce co-texte : adverbe duratif + présent de propriété.

Ainsi que le montre cet exemple (une fois de plus simplifié très grossièrement pour les besoins de l'exposé), le modèle permet donc d'utiliser les contraintes locales pour obtenir la ou les solutions qui les satisfont toutes, ou du moins, qui les satisfont le mieux possible : c'est en fait un principe d'optimalité qui est mis en œuvre dans la construction de la dynamique globale, et qui correspond bien à ce que nous entendions par « compositionnalité gestaltiste ».

3. Implémentation informatique du modèle

Reste bien sûr à savoir si l'on peut rendre ce modèle opérationnel, c'est-à-dire si l'on peut implémenter ces systèmes dynamiques sous une forme qui permette de calculer effectivement le sens d'un énoncé. Nous utilisons pour cela, comme pour notre premier modèle, des outils informatiques adaptés au calcul de systèmes dynamiques : les réseaux connexionnistes récurrents.

Pour implémenter les dynamiques locales associées à chaque unité, nous avons conçu une nouvelle architecture de réseaux récurrents, schématisée à la *figure 2*.

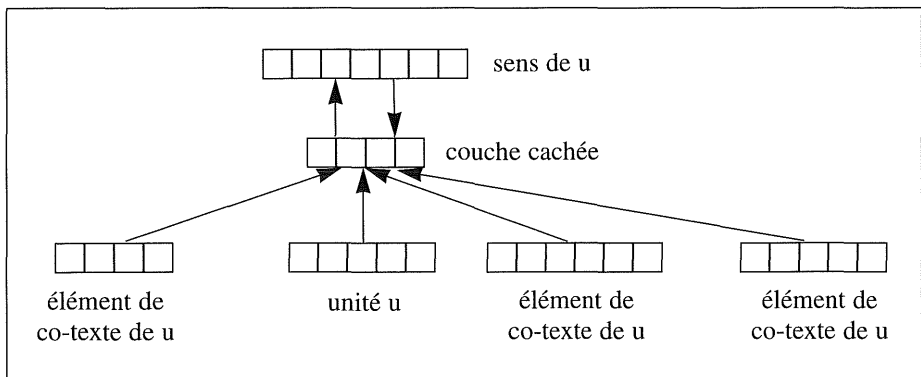


Fig. 2 : Architecture d'un réseau local

Sans entrer là non plus dans les détails techniques, cette architecture présente les caractéristiques suivantes :

- elle permet de définir une dynamique sur la couche de sortie du réseau (couche intitulée « sens de l'unité u » sur la *figure 2*), grâce à des liens récurrents entre cette couche et la couche cachée ;

- la forme de la dynamique sur la couche de sortie dépend de la couche d'entrée, c'est-à-dire d'une part de l'unité u considérée, et d'autre part des éléments pertinents du co-texte de u, qui peuvent représenter aussi bien la forme que le sens des unités dont la présence influence le sens de l'unité u ;

– l'apprentissage de cette dépendance (c'est-à-dire la relation entre la valeur de la couche d'entrée et la forme de la dynamique sur la couche de sortie) peut se réaliser de manière rapide et efficace en utilisant l'algorithme le plus classique (rétropropagation du gradient) sur un réseau unidirectionnel dérivé de ce réseau récurrent (en dupliquant la couche de sortie, de façon à ce que les liens rétroactifs entre la couche de sortie et la couche cachée deviennent des liens analogues à ceux qui relient la couche d'entrée et la couche cachée).

Ainsi un même réseau local peut-il être utilisé pour représenter une famille paradigmatique d'unités linguistiques, telle que la famille des adverbes aspectuels (*encore, déjà, toujours, alors*, etc.) ou la famille des temps verbaux (présent, imparfait, passé simple, etc.). Concrètement le travail à effectuer peut se résumer de la façon suivante. On « code »¹³ les différentes unités de la famille sur la partie de la couche d'entrée notée « unité u » sur la *figure 2*, les éléments pertinents du co-texte (forme et/ou sens) sur les parties notées « élément du co-texte de u » (le nombre de ces éléments doit être spécifié à l'avance), et on code les sens possibles des diverses unités de la famille sur la couche de sortie. On donne alors un échantillon d'apprentissage constitué d'un ensemble d'exemples qui donnent chacun la forme de la dynamique locale associée à une unité dans un co-texte donné¹⁴. Si l'échantillon d'apprentissage est suffisamment représentatif, et que l'apprentissage réussit, on obtient un réseau local capable de « calculer » le sens de n'importe quelle unité de la famille dans n'importe quel co-texte (le sens des unités co-textuelles polysémiques étant pré-déterminé : il s'agit toujours ici, ne l'oublions pas, de dynamiques locales). La réussite de l'apprentissage peut d'ailleurs être évaluée : classiquement, on constitue un échantillon de test, constitué d'exemples qui n'ont pas été utilisés dans l'échantillon d'apprentissage, et l'on évalue ce que l'on appelle les performances en généralisation du réseau sur cet échantillon de test.

Une fois réalisés les différents réseaux locaux nécessaires, on peut implémenter les dynamiques globales associées à des énoncés, en « assemblant » ces réseaux locaux. On appellera réseau global le résultat de cet assemblage, qui est illustré à la *figure 3*. Ce réseau global présente les caractéristiques suivantes :

– c'est véritablement un assemblage de réseaux locaux, au sens précis où chaque flèche représentée sur le dessin provient d'une flèche d'un des réseaux locaux associé à chaque unité de l'énoncé (plus exactement, les poids des liens schématisés par chaque flèche sont donnés par les poids des liens correspondants sur un réseau local) ;

13. Derrière ce terme de codage se cache en fait un travail linguistique considérable, notamment quand il s'agit de coder le sens ; il s'agit en fait de construire un espace sémantique approprié : cf. par exemple B. Victorri et C. Fuchs, 1992, « Construction de l'espace sémantique associé à un marqueur grammatical polysémique », *Linguisticae Investigationes* 16, 1, p. 125-153.

14. Chaque exemple correspond donc à une règle du type des règles R_1, R_2 , etc. présentées ci-dessus en 2. *Un modèle de « compositionnalité gestaltiste »*, *supra*, p. 237.

– c’est une architecture adaptable, au sens où la taille du réseau dépend de la taille de l’énoncé. Le schéma de la *figure 3* présente un réseau associé à un énoncé de trois unités, mais on généralisera sans peine au cas général : chaque unité ne sera liée qu’aux éléments pertinents de son co-texte, à savoir les éléments qui ont été utilisés lors de la confection de son réseau local ;

– enfin cette architecture permet de prendre en compte des «éléments de sens global non redistribué», telle que la valeur aspectuo-temporelle de l’énoncé dont nous avons parlé plus haut¹⁵. En fait, ils se comportent exactement comme les éléments de sens associés aux unités (en particulier, ils réclament un apprentissage sur un réseau local adapté) : la seule différence, c’est qu’aucune partie de la couche d’entrée ne leur est spécifiquement associée.

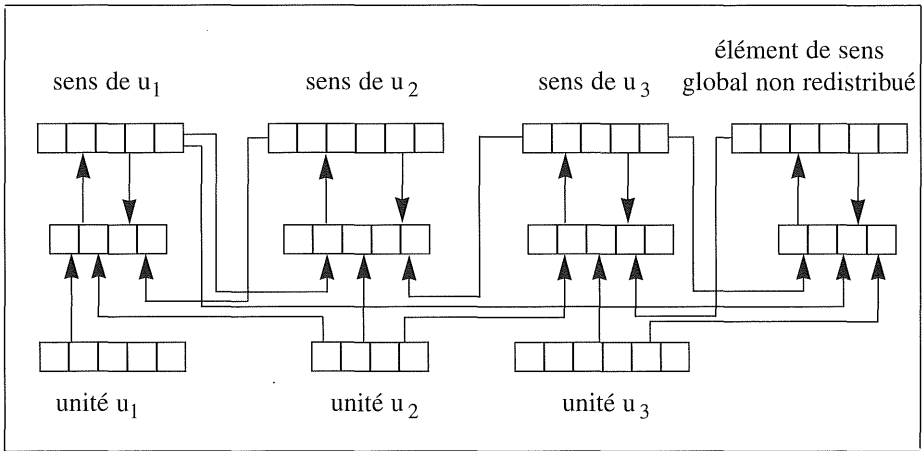


Fig. 3 : Architecture d’un réseau global

La dynamique implémentée par ce réseau correspond exactement à la dynamique globale définie sur le produit cartésien des espaces sémantiques associés à chaque unité que nous avons décrite plus haut¹⁶. Aussi un tel réseau devrait calculer automatiquement le sens d’un énoncé en appliquant le principe de compositionnalité gestaltiste que nous voulions mettre en œuvre. Un système informatique est à l’heure actuelle en cours de réalisation qui nous permettra de tester concrètement ce modèle¹⁷.

15. Cf. 2. *Un modèle de « compositionnalité gestaltiste », supra*, p. 237.

16. *Ibid.*

17. La technique informatique a été validée par Fabrice Fontaine sur des données non-linguistiques très simples (problème de perception de cubes dessinés en perspective, du type cube de Necker, cf. F. Fontaine, 1995, *Un modèle connexionniste de la compositionnalité gestaltiste*, DEA d’Intelligence artificielle, Université de Caen). Une expérimentation linguistique est actuellement menée par Sophie Delamarche en collaboration avec Pierre Cadiot qui fournit l’expertise et les données linguistiques (composés nominaux de la forme N_1 à $/de N_2$).

Conclusion

Un des principaux intérêts de ce modèle, c'est qu'il est, à notre connaissance tout au moins, l'un des rares modèles de traitement linguistique utilisant des réseaux connexionnistes à « architecture variable », dont la taille et le détail des liens peuvent s'adapter à la taille et au contenu des énoncés¹⁸. Généralement, la taille des réseaux est fixe, ce qui représente une contrainte très importante, qui force à limiter les phrases traitées à des phrases obéissant à un schéma syntaxique fixe¹⁹, ou en tout cas, même quand un mécanisme récurrent permet de traiter des phrases de longueur variable²⁰, à s'en tenir à un format fixe de représentation du contenu de la phrase.

Grâce à cette architecture, on devrait pouvoir s'adapter à la structure de l'énoncé sans perte d'information, et sans que cela nécessite un apprentissage spécifique. Notons d'ailleurs qu'il est tout à fait possible, si les résultats obtenus par un réseau global ne sont pas satisfaisants, de « prolonger » l'apprentissage effectué sur les réseaux locaux, en utilisant la configuration nouvelle présente dans l'énoncé en cours de traitement : grâce à la technique dite « des poids partagés », ce nouvel apprentissage sera en fait cumulatif, au sens où il modifiera en même temps le réseau global et les réseaux locaux qui ont servi à le construire.

Bien entendu, cette architecture présente aussi des limites : son application devra sans doute se limiter à des énoncés assez simples, dont la structure syntaxique est relativement « plate ». Pour des énoncés plus complexes, du type des exemples (8) à (13) présentés ci-dessus²¹, qui nécessitent pour leur compréhension la prise en compte d'interactions sémantiques entre différents niveaux syntaxiques (ce sont des éléments de sens contenus dans la relative qui influent sur le sens du déterminant de l'antécédent dans la principale), il faudra sans doute définir des architectures plus complexes, capable de représenter d'une manière ou d'une autre le caractère hiérarchique des structures syntaxiques. Mais nous n'en sommes pas encore là : il nous faut d'abord valider l'approche sur des données linguistiques simples avant d'envisager de telles extensions.

18. Parmi les autres tentatives d'utilisation d'architecture variable, on peut citer G. Cottrell, 1989, *A Connectionist Approach to Word Sense Disambiguation*, Pitman et Morgan Kaufmann, un des premiers à utiliser les réseaux pour des tâches de désambiguïsation sémantique, ou encore Wermter et Lehnert, 1989, « A Hybrid Symbolic/Connectionist Model for Noun Phrase Understanding », *Connection Science*, 1 (3), qui ont construit, eux, des réseaux spécialisés dans le traitement des problèmes de rattachement de groupes prépositionnels. Notre modèle emprunte d'ailleurs certaines de ses caractéristiques à ces deux modèles.

19. Comme, par exemple, pour les modèles de J.L. McClelland et A.H. Kawamoto, 1986, « Mechanisms of Sentence Processing : Assigning Roles to Constituents of Sentences », in J.L. McClelland, D.E. Rumelhart and the PDP Research Group (eds.), *Parallel Distributed Processing*, vol. 2, MIT Press ; C. Harris, 1990, « Connectionism and Cognitive Linguistics », *Connection Science*, 2 (1/2).

20. Comme, par exemple, dans les modèles de J.L. Elman, 1991, « Distributed Representation, Simple Recurrent Neural Networks, and Grammatical Structure », *Machine Learning* 7 ; M.I. Jordan, 1986, « Attractor Dynamics and Parallelism in a Connectionist Sequential Machine », *Proceedings of the 8th annual Conference on Cognitive Science*, Erlbaum.

21. Cf. 1. *Le problème à résoudre, supra*, p. 234.

RÉFÉRENCES BIBLIOGRAPHIQUES

- CHENCINER A., 1992, « Systèmes dynamiques différentiables », *Encyclopædia universalis* 21, p. 1044-1080.
- COTTRELL G., 1989, *A Connectionist Approach to Word Sense Disambiguation*, Pitman et Morgan Kaufmann.
- ELMAN J.L., 1991, « Distributed Representation, Simple Recurrent Neural Networks, and Grammatical Structure », *Machine Learning* 7.
- FONTAINE F., 1995, *Un modèle connexionniste de la compositionnalité gestaltiste*, DEA d'Intelligence artificielle, Université de Caen.
- FUCHS C. et VICTORRI B. (eds.), 1994, *Continuity in Linguistic Semantics*, Amsterdam, John Benjamins.
- FUCHS C., GOSSELIN L. et VICTORRI B., 1991, « Polysémie, glissements de sens et calcul des types de procès », *Travaux de Linguistique et de Philologie*, n° spécial, « Les typologies de procès », Strasbourg.
- GAZDAR G. et alii, 1985, *Generalized phrase structure grammar*, Oxford, Blackwell.
- GOSSELIN L., 1996, *Sémantique de la temporalité en français*, Louvain-la-Neuve, Duculot.
- GUILLAUME P., 1979, *La Psychologie de la forme*, Paris, Flammarion.
- HARRIS C., 1990, « Connectionism and Cognitive Linguistics », *Connection Science*, 2 (1/2).
- JORDAN M.I., 1986, « Attractor Dynamics and Parallelism in a Connectionist Sequential Machine », *Proceedings of the 8th annual Conference on Cognitive Science*, Erlbaum.
- KLEIBER G., 1994, « Contexte, interprétation et mémoire : approche standard (vs.) approche cognitive », *Langue française* 103, p. 9-22.
- MCCLELLAND J.L. et KAWOMOTO A.H., 1986, « Mechanisms of Sentence Processing : Assigning Roles to Constituents of Sentences », in J.L. McClelland, D.E. Rumelhart and the PDP Research Group (eds.), *Parallel Distributed Processing*, vol. 2, MIT Press.
- MONTAGUE R., 1970, « English as a formal language », in R. Thomason (ed.), *Formal philosophy, selected papers of Richard Montague*, New-Haven, Yale University Press.
- VICTORRI B. et FUCHS C., 1992, « Construction de l'espace sémantique associé à un marqueur grammatical polysémique », *Linguisticae Investigationes* 16, 1, p. 125-153.
- VICTORRI B. et FUCHS C., 1996, *Le Phénomène de la polysémie*, Paris, Hermès.
- WERMTER et LEHNERT, 1989, « A Hybrid Symbolic/Connectionist Model for Noun Phrase Understanding », *Connection Science* 1 (3).