

EN QUOI LES ANALYSES PSYCHOLINGUISTIQUES PEUVENT-ELLES CONTRIBUER À L'ÉLABORATION DE SYSTÈMES DE RECHERCHE ET DE REPRÉSENTATIONS DES CONNAISSANCES ?

Martine CORNUÉJOLS
MoDyCo, Université Paris X

SOMMAIRE

1. Introduction
2. Normes associatives verbales et imagées
 - 2.1. Constitution des normes
 - 2.2. Proposition d'une typologie des liens associatifs
 - 2.3. Analyse des normes associatives par rapport à cette typologie
 - 2.3.1. Le rôle du contexte
 - 2.3.2. L'organisation en catégories sémantiques
 - 2.3.3. L'organisation globale tripartite
 - 2.3.4. Analyse en fonction du type de catégorie (naturel / artificiel) des entités présentées
 - 2.3.5. Les thèmes évoqués par les associations sémantiques
 - 2.3.6. Le cas particulier des associés linguistiques
3. Des normes associatives aux réseaux sémantiques
4. Comparaison des associations sémantiques avec des cooccurrences de corpus textuels
5. Perspectives pour la représentation des connaissances dans les systèmes artificiels, la recherche documentaire, le web sémantique, l'indexation multimodale, ...

Résumé : *L'étude concerne l'organisation des réseaux sémantiques verbaux et imagés. La différenciation de ceux-ci est mise en évidence par la constitution de corpus d'associations imagées et verbales et l'analyse des types de liens associatifs. Le principe organisateur en termes de relations associatives de type catégories sémantiques est remis en cause. Une structuration en termes de situation, caractéristique et catégorie sémantique est proposée.*

L'analyse plus fine selon l'appartenance aux entités naturelles ou artificielles est réalisée en rapport avec les connaissances neuropsychologiques connues en termes de déficits catégorie spécifique.

Des perspectives dans le domaine de la gestion des documents, de l'indexation multimodale et du web sémantique sont envisagées.

1. Introduction

L'étude de l'organisation des réseaux sémantiques en mémoire humaine et de l'activation des significations ou des concepts par les perceptions visuelles peut servir de base pour élaborer des systèmes de représentation sémantique pour le traitement automatique du langage, en particulier pour la constitution d'ontologies et de lexiques, et également des systèmes de recherche d'information (web sémantique, ...) en contribuant à l'élaboration d'indexation. En effet, il paraît tout à fait important d'obtenir une bonne adéquation entre la représentation mentale de l'utilisateur final qui fait la requête pour sa recherche d'information et la représentation d'informations élaborée par le concepteur du système de recherche d'information. Cependant, la majorité des études sont axées sur le lexique et le langage verbal. Or les systèmes d'information évoluent vers le multimodal. On peut donc s'interroger sur la transposition des principes de sémantique verbale dans la sémantique imagée.

La psycholinguistique apporte un nouveau savoir sur l'organisation sémantique et comporte un volet expérimental de recueil de données concernant les réseaux associatifs et une partie modélisation pour construire une représentation des connaissances qui sera utilisable pour l'élaboration de systèmes d'information.

Les recherches exposées dans cet article portent sur la constitution des corpus verbaux et imagés correspondant aux associations sémantiques présentes en mémoire (Cornuéjols 1999), à leur analyse aboutissant à l'élaboration d'une typologie des liens associatifs et à la mise en évidence de principes organisateurs (en termes de situation, caractéristiques et catégorie sémantique), à la mise

en relation des items verbaux et imagés pour modéliser les réseaux sémantiques, à la détermination des thématisations émergentes et à la comparaison avec des corpus textuels de type journaux d'actualité (Ferret & Cornuéjols 1998) et de type roman (Reza & Rastier 1999).

Ces recherches visent à contribuer à mettre au point des systèmes de recherche d'information basés sur les mots sémantiquement proches aux mots-clés (approche itérative de la requête), à la modélisation sémantique à la base du web sémantique et également à la constitution de systèmes d'indexation multimodaux qui ne soient pas nécessairement basés sur la dénomination ou la description des images, comme c'est le cas actuellement, mais également sur la proximité sémantique et l'interrelation des réseaux imagés et verbaux.

2. Normes associatives verbales et imagées

2.1. Constitution des normes

Pour déterminer si le réseau associatif imagé est différent du réseau associatif verbal, une des méthodes est de comparer le corpus d'associations verbales et d'associations imagées. La littérature ne faisant mention d'aucune norme d'associations imagées existante, nous avons entrepris d'établir une norme d'associations imagées pour 285 items. Pour pouvoir la comparer avec la norme d'associations verbales et sachant que les normes d'associations verbales existantes ne comportaient pas uniquement des items concrets imageables, nous avons établi la table d'associations verbales correspondant aux dénominations des entités représentées par les images de la norme associative imagée. Ces données sont obtenues par une tâche d'association libre. Le dépouillement des résultats se fait en listant les occurrences citées par chacun des 134 participants face à l'item inducteur (soit environ 40000 données par norme) et à compter le nombre de fois (dénombrement et classement) où chaque occurrence a été citée sur l'ensemble des participants.

L'analyse globale comparative montre que les normes verbale et imagée divergent pour 77,26% des associés sémantiques cités. Une analyse plus fine sur l'associé majoritairement cité uniquement, montre que les normes associatives diffèrent dans 60,21% des cas. Le recouvrement de l'associé majoritaire sur les deux normes est donc d'environ 40%.

Pour donner une idée sur un exemple concret, l'image du « zèbre » évoque spontanément dans la tête des personnes l'idée de « savane », alors que le mot « zèbre » lorsqu'il est lu, évoque « rayures » qui est une des caractéristiques de l'entité.

Tous ces résultats confortent l'idée de réseaux associatifs sensiblement différents pour l'image et pour le mot. Ces réseaux bien que différenciés sont intimement liés.

Ces corpus ont été validés par des expériences d'amorçage sémantique entre mots et images se basant sur les normes associatives ainsi établies. Un effet d'amorçage similaire est obtenu entre images et mots associés, sur la base de la norme associative imagée précédemment établie, et entre entités verbales sur la base de la norme associative verbale. En d'autres termes, l'image peut évoquer automatiquement le mot associé, tel qu'établi dans la norme associative imagée, mais ne peut évoquer automatiquement le mot que l'on supposerait associé, si l'on se base sur la norme associative verbale. La norme associative verbale ne permet pas de rendre compte des liens sémantiques entre images et entités verbales.

2.2. Proposition d'une typologie des liens associatifs

La grande dispersion des résultats nous a incitée à proposer une typologie des liens associatifs. Dans tous les modèles de mémoire sémantique (Dubois 1972 ; Denhière 1975 ; Ehrlich & Tulving 1976) on retrouve l'idée de catégories, dimensions ou composantes sur lesquelles les éléments stockés peuvent être classés et entretiennent entre eux un certain système de relations. La force associative et la nature des relations « associant-associés » reflètent certaines propriétés de la mémoire sémantique.

Une typologie en neuf points a été proposée : situation (spatiale, temporelle, événement), catégorie sémantique (super ordonné, classe, instance), attribut ou caractéristique, fonction (ex. : « couteau »-« couper ») ou produit (ex. : « miel » pour « abeille »), objet naturellement associé (ex. : « selle » pour « cheval »), association de type linguistique (ex. : « chapeau »-« melon »), symbole (ex. : « cœur »-« amour »), nul ou sans réponse (révèle la difficulté d'association), illisible (l'impondérable des expériences papier-crayon).

2.3. Analyse des normes associatives par rapport à cette typologie

La classification des associations sémantiques des normes verbales et imagées a été opérée par trois juges de façon indépendante. Le degré d'accord a été évalué par un coefficient de corrélation de Bravais – Pearson à .78 ($p < .001$).

Une analyse à deux niveaux peut être réalisée :

- l'une sur l'associé majoritaire uniquement ;
- l'autre sur l'ensemble des associés cités face à un item.

Les différences entre les tables d'associations verbales et imagées sont significatives pour les catégories typologiques Situation, Instance, Attributs, Fonction, Nul, Illisible.

Les associés de type Instances, Attributs et Fonctions sont plus souvent donnés en réponse à des stimuli verbaux qu'à des stimuli imagés. À l'opposé, les stimuli présentés sous forme d'images engendrent plus de réponses de type Nul ou Illisible, ce qui laisse supposer que la tâche d'évocation d'une image mentale est plus ardue que l'association simple entre mots.

Ce résultat conforte l'idée que les structures associatives correspondant aux images et aux mots ne sont pas similaires.

2.3.1. Le rôle du contexte

Le résultat le plus important de cette étude est la constatation que, quel que soit le niveau d'analyse, la présentation d'une image évoque préférentiellement la situation (ou contexte spatial ou temporel ou encore événementiel), dans laquelle l'entité représentée par l'image est généralement rencontrée. Le contexte le plus souvent évoqué est un contexte spatial, plutôt que temporel. Le sujet évoque préférentiellement une localisation plutôt qu'un événement situé dans le temps. La présentation d'une entité lexicale de type substantif évoque elle préférentiellement une caractéristique qui spécifie l'entité représentée par le mot. Ce qui voudrait dire que quand on voit un objet isolé, spontanément on le re-situe dans un contexte, un environnement, alors que lorsqu'on lit sa dénomination, on évoque spontanément un trait qui le caractérise.

Les associations majoritaires sont donc de type situation (spatiale essentiellement et événementielle) et objets associés ou contigus. Ceci corrobore les résultats de Yeh & Barsalou (2000) qui montrent que les concepts apparaissent en situation et non abstraits de celles-ci. Ceci explique aussi pourquoi les concepts peuvent apparaître sous différentes formes selon les situations, où chaque forme contient les propriétés pertinentes à la spécificité de la situation (Yeh & Barsalou 2000). Donc il semblerait que les principes organisationnels de la mémoire sémantique diffèrent en fonction du contexte. Par exemple, *un piano* est un instrument de musique, mais dans un contexte situationnel de déménagement, son poids devient un trait saillant. Le fait que chaque item présenté isolément évoque le même associé cité par environ 60% des personnes suggère que le stockage du poids ou de la saillance des traits reste le même à l'intérieur de la population, mais que le contexte peut faire varier ce facteur. Cela suggère donc que la saillance des propriétés des items isolés peut être différente de celle des items en situation. Le fait que le contexte apparaisse spontanément dans les associés majoritaires peut inciter à penser que le contexte lui-même fait partie des traits qui définissent un objet. Le rôle du contextuel dans l'organisation des associés sémantiques est analysé également en fonction du type d'item source.

2.3.2. L'organisation en catégories sémantiques

Un autre résultat important issu de cette analyse est la constatation que les liens associatifs de type catégorie sémantique et instance, qui révèlent des liens de hiérarchie taxinomique, ne représentent que 10,23% pour les associations imagées et 21,84% pour les associations verbales. Ce résultat suggère donc que l'organisation taxinomique n'est pas prédominante dans le réseau associatif verbal ou imagé, alors même que la littérature la pose comme principe organisateur prépondérant dans l'organisation conceptuelle (Collins & Quillian 1969 ; Rosch 1975 ; Sartori & Job 1988).

De plus, la majorité des liens taxinomiques sont de type instance ou classe équivalente, ce qui suppose donc que le sens privilégié est du général vers le spécifique et non de l'instance (pourtant représentée par l'item source) vers la catégorie plus globale.

2.3.3. Organisation globale tripartite

L'étude des associations montre que l'on peut regrouper celles-ci en deux grands groupes : les associations qui spécifient l'objet par opposition aux éléments qui le replacent dans un contexte.

Dans le premier groupe sont regroupées les classes instance, attributs et fonction, alors que dans le deuxième se retrouvent les classes situation, items associés ou contigus, et de catégorie (super ordonné, comme contexte hiérarchique).

En ce qui concerne les *associations verbales*, 30,58% des associations produites spécifient les entités considérées, alors que 66,74% les contextualisent.

Pour ce qui est des *associations imagées*, 11,15% des associations spécifient l'entité tandis que 88,85% des associés fournis les contextualisent. La contextualisation, bien que majoritaire dans les deux tables d'associations, est fortement accentuée dans le cas des images, puisqu'elle apparaît dans presque 90% des cas.

Pour ce qui est des associations verbales, dans plus d'un quart des cas l'associé produit vise à spécifier l'entité dénommée. Ces résultats confirment que les caractéristiques permettant de différencier les entités sont plus associées aux mots qu'aux images.

Si on dissocie les catégorisations sémantiques, qui sont un principe organisateur à part entière, on peut considérer trois modes d'organisation, le mode dominant étant la situation ou contexte (contexte essentiellement spatial ou événementiel), le deuxième étant la caractéristique (attribut ou trait, fonction, produit, ...) et le troisième, la catégorie sémantique.

2.3.4. Analyse en fonction du type de catégorie (naturel / artificiel) des entités présentées

Une analyse des associations par catégorie sémantique a été réalisée pour déterminer si la nature de la catégorie sémantique de l'amorce (ou item source) avait une incidence sur la cohésion du noyau associatif central. En effet le corpus d'items source comprend 114 entités naturelles (animaux, végétaux, insectes, ...) et 170 entités artificielles (véhicules, outils, monuments, vêtements, instruments de musique, ...). Les données neuropsychologiques proviennent de cas de patients présentant des déficits catégorie spécifique ou des déficiences sélectives dans certains domaines sémantiques (Hillis & Caramazza 1991 ; Warrington & Shallice 1984 ; Saffran & Schwartz 1992). La configuration la plus fréquemment rencontrée est l'atteinte touchant les connaissances du domaine des entités vivantes, alors que la sémantique des catégories sémantiques artificielles semble préservée. Parfois la configuration inverse est observée (Hillis & Caramazza 1991 ; Moss & Tyler 1997, 2000 ; Sacchett & Humphreys 1992 ; Warrington & McCarthy 1983, 1987).

Un de nos objectifs, par l'analyse des normes associatives et des types de liens sémantiques, est de mettre en évidence les différences entre entités naturelles et artificielles et de rendre compte de ce qui peut expliquer les déficits catégorie spécifique. Notre question était de savoir quel type de lien associatif peut être altéré qui expliquerait les déficits sélectifs observés et d'ainsi mettre en évidence le type de relation qui pourrait être dominant pour chaque catégorie d'item.

Une première remarque est que cette étude ne révèle aucun effet catégorie sémantique dépendant. Quelle que soit la catégorie de l'item source, en moyenne, environ 63% des sujets ont cité le même pool d'items associés.

	"SITUATION"				"CARACTERISTIQUE"					"CATEGORIE"			
	sit	Ass- o	evt	act	Attr	prod	fct	sym	cont	Cl eq	inst	Cl cat	CL
Entités naturelles													
moy %	21	3	4	3	21	8	3	3	2	9	10	7	6
ET	18	8	6	9	20	16	8	7	7	15	17	10	10
%	31				37					32			
Entités artificielles													
moy %	11	11	6	5	9	7	10	1	8	7	16	3	6
ET	18	19	9	11	15	17	16	6	20	14	26	7	14
%	33				35					32			

Sit : Situation (localisation) ; **Act** : Acteurs ; **Evt** : Evénement ; **Ass Obj** : Objet Associé ; **Sym** : Symbole ; **Attr** : Attributs ; **Prod** : Produit ; **Fct** : Fonction ; **Cont** : Contenu ; **Cl equ** : Classe équivalente ; **Cl Inst** : Instance ; **Cl Cat** : Nom de la Catégorie.

Tableau 1 : Pourcentage des 3 liens associatifs des 3 modes d'organisation entre entités naturelles et entités artificielles

Les résultats du tableau 1 montrent que les entités naturelles sont également réparties entre situation, caractéristique et catégorie.

Une analyse plus fine montre que les différences apparaissent entre catégories naturelles et artificielles en termes de situation (majoritaire pour les entités naturelles) et d'objets associés (prédominant pour les entités artificielles) pour ce qui concerne le mode de structuration de type situation ; en termes d'attribut (majoritaire pour les entités naturelles) et de fonction (logiquement plus associée aux entités artificielles) pour ce qui concerne les liens caractérisants ; en termes d'instance (majoritaire pour les entités artificielles) et les noms des catégories sémantiques (ou super ordonné) (majoritaires pour les entités naturelles) pour le mode de structuration de type catégorie.

Une analyse encore plus fine consiste à différencier les différents types d'entités naturelles et d'entités artificielles et de déterminer le type de lien prédominant selon le type spécifique d'items.

	SITUATION					CARACTERISTIQUE					CATEGORIE			
	Sit	Act	Evt	Ass- obj	sym	Char	prod	fct	sp	cont ents	Cl equ	Cl inst	Cl cat	Cl
WA	22.2 2	0	5.55	0	0	33.3 3	0	0	11.1 1	0	5.55	16.6 6	0	5.55
DA	8.33	0	0	0	0	16.6 6	58.3 3	0	8.33	0	8.33	0	0	0
B	45.4 5	0	0	0	0	27.2 7	0	0	0	0	0	9.09	18.1 8	0
I	33.3 3	0	0	0	0	11.1 1	22.2 2	0	0	0	22.2 2	11.1 1	0	0
SA	80	0	0	0	0	20	0	0	0	0	0	0	0	0
BP	17.6 4	0	0	23.5 2	5.88	29.4 1	0	5.88	0	0	5.88	0	5.88	5.88
UP	16.6 6	8.33	8.33	0	0	25	8.33	0	0	8.33	16.6 6	8.33	0	0
VF	28	0	0	4	0	20	12	0	0	0	12	12	4	0

En quoi les analyses psycholinguistiques peuvent-elles contribuer

WA : Wild Animals (animaux sauvages) ; **DA** : Domestic Animals (animaux domestiques) ; **B** : Birds (oiseaux) ; **I** : Insects (insectes) ; **SA** : Sea Animal (animaux marins) ; **BP** : Body Parts (éléments du corps) ; **UP** : Universe Parts (éléments de l'univers) ; **VF** : Vegetables and Fruits (végétaux et fruits).
Sit : Situation ; **Act** : Actors ; **Evt** : Event ; **Ass Obj** : Associated Object ; **Sym** : Symbol ; **Char** : Characteristic ; **Prod** : Product ; **Fct** : Function ; **Sp** : Cultural specificity ; **Cont** : Contents ; **Cl equ** : Equivalent class ; **Cl Inst** : Instance ; **Cl Cat** : Name of Category.

Tableau 2 : Répartition des associés dans les catégories pour chaque type d'entité naturelle.

	SITUATION					CARACTERISTIQUE					CATEGORIE			
	Sit	Act	Evt	Ass-obj	sym	Char	prod	fct	sp	cont ents	Cl equ	Cl inst	Cl cat	Cl
Cloth	10.5 2	5.26	0	36.8 4	0	5.26	0	0	0	10.5	10.5 2	15.7 8	0	5.26
Vehi	40	0	10	30	0	0	0	0	0	0	10	0	0	10
Furn	5.88	0	5.88	23.5 2	0	0	17.6 4	11.7 6	0	5.88	17.6 4	11.7 6	0	0
Build	25	8.33	16.6 6	8.33	8.33	8.33	0	0	0	8.33	0	16.6 6	0	0
Gam	12.5	12.5	12.5	0	0	0	0	12.5	0	0	12.5	37.5	0	0
Mus	10	0	10	10	0	0	60	0	0	0	0	10	0	0
Cont	11.7 6	0	5.88	0	0	0	0	0	0	52.9	0	23.2 5	0	5.88
O S	0	0	12.1 2	9.09	0	3.03	15.1 5	15.1 5	0	6.06	6.06	33.3 3	0	0
Weap	0	0	0	40	0	0	20	0	0	0	20	20	0	0
Tools	0	13.3 3	0	40	0	0	0	0	0	0	6.66	13.3 3	6.66	20
other	8.69	4.34	4.34	13.0 4	4.34	0	4.34	13.0 4	8.69	0	8.69	17.3 9	0	13,0 4

Cloth : Clothes (vêtements) ; **Vehi** : Vehicles (véhicules) ; **Furn** : Furnitures ; **Build** : Buildings (monuments) ; **Gam** : Games (jeux) ; **Mus** : Musical Instruments (instruments de musique) ; **Cont** : Containers (récipients) ; **O S** : Office Supplies (outils de bureau) ; **Weap** : Weapons (armes).

Tableau 3 : Répartition des associés pour chaque type d'entité artificielle

Comme le montrent les tableaux ci-dessus, l'associé majoritaire dépend de la catégorie de l'objet. Ceci est à mettre en rapport avec les données de déficits neuropsychologiques.

2.3.5. Les thèmes évoqués par les associations sémantiques

Cette étude fait émerger une organisation des connaissances en termes de situation et de thématique. Plusieurs associés peuvent être regroupés par thème évoqué.

Par exemple, dans la norme associative verbale, pour « abeille » on retrouve le thème de la production de miel, le thème de la piqûre, ...

1	ABEILLE	miel	48	34,78%	produit	s	fct
2	ABEILLE	ruche	20	14,49%	sit (lieu)	s	lieu
	ABEILLE	nid	1	0,72%	sit (lieu)	s	lieu
3	ABEILLE		15	10,87%	rien		
5	ABEILLE	guêpe	9	6,52%	cl (equ)	m	cat
	ABEILLE	frelon	1	0,72%	cl (equ)	m	cat
	ABEILLE	Maya	2	1,45%	cl (inst)	m	cat
	ABEILLE	insecte	3	2,17%	cl (cat)	t	cat
	ABEILLE	rayures	2	1,45%	car	a	struct
	ABEILLE	jaune	1	0,72%	car	a	struct
	ABEILLE	ailles	3	2,17%	car	a	struct
	ABEILLE	tache	1	0,72%	car	a	struct
	ABEILLE	bzzzzzzzz	2	1,45%	produit	a	struct
	ABEILLE	dard	1	0,72%	car	a	struct
4	ABEILLE	piqûre(s)	12	8,69%	evt	s	fct
	ABEILLE	butiner	2	1,45%	fct	s	fct
	ABEILLE	fleur	2	1,45%	obj-ass	s	lieu
	ABEILLE	prés	1	0,72%	sit (lieu)	s	lieu
	ABEILLE	chat	1	0,72%			?

Pour le même exemple dans la norme imagée :

	ABEILLE	2	miel	29	21,17%	23,36%	prod
	ABEILLE		pot de miel	3	2,19%		prod
	ABEILLE	1	piqûre	31	22,63%	28,47%	Sit (evt)
	ABEILLE		piqûre d'insecte	1	0,73%		
	ABEILLE		piquer	2	1,46%		
	ABEILLE		danger piqûre	1	0,73%		
	ABEILLE		dard	1	0,73%		
	ABEILLE		allergie au venin de guêpe	1	0,73%		
	ABEILLE		oedème	1	0,73%		
	ABEILLE		méchant	1	0,73%		
	ABEILLE	3	ruche	13	9,49%	19,71%	Sit (lieu)
	ABEILLE		nid	1	0,73%		
	ABEILLE		essaim	2	1,46%		
	ABEILLE		toile d'araignée	1	0,73%		
	ABEILLE		grenier	1	0,73%		
	ABEILLE		cave	1	0,73%		
	ABEILLE		boîte transparente	1	0,73%		
	ABEILLE		camping l'été	1	0,73%		
	ABEILLE		confiture	1	0,73%		
	ABEILLE		été	5	3,65%		Sit (t)
	ABEILLE		prairie	1	0,73%		Sit (lieu)
	ABEILLE	4	fleur(s)	10	7,30%		Sit (lieu)
	ABEILLE		pollen	1	0,73%		
	ABEILLE		bouton	3	2,19%		
	ABEILLE		arbre	1	0,73%		
	ABEILLE		insecte	5	3,65%	11,68%	Cl (cat)
	ABEILLE		Maya	1	0,73%		Cl (inst)

En quoi les analyses psycholinguistiques peuvent-elles contribuer

ABEILLE	mouche	4	2,92%	CI (equ)
ABEILLE	ABEILLE	3	2,19%	
ABEILLE	guêpe	1	0,73%	
ABEILLE	araignée	1	0,73%	
ABEILLE	moustique	1	0,73%	
ABEILLE	odeur	1	0,73%	
ABEILLE		1	0,73%	
ABEILLE	bzzzzzzzz	1	0,73%	
ABEILLE	homme	1	0,73%	
ABEILLE	horreur	1	0,73%	

2.3.6. Le cas particulier des associés linguistiques

Pour ce qui est des associations de type linguistique, elles sont très minoritaires et ne représentent que 3,67% des associations. Cette classe, bien que spécifique du verbal, pose des problèmes de classification. Les associations de type linguistique pourraient être assimilées à la catégorie « objets associés ». Doit-on la classer comme une composante de la catégorie instance (ex. « Chapeau – melon » est une instance de « chapeau ») ou comme un objet (linguistique) associé (ex. « rouge – gorge » qui n'est pas une instance de « rouge » et qui ne définit l'objet que par l'association des deux items) ?

Les associations linguistiques montrent bien que ce n'est pas l'associé qui est catégorisé, mais bien l'association.

Les données montrent que 30,58% des associations linguistiques produites spécifient ou caractérisent l'entité (instancier), alors que dans 66,74% des cas, elles la contextualisent (contexte linguistique).

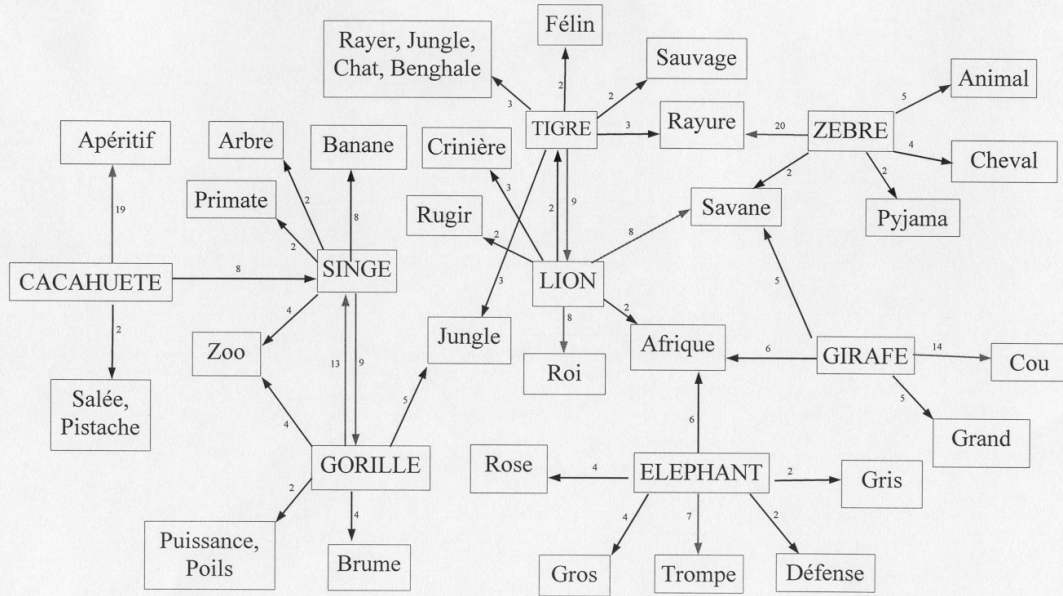
3. Des normes associatives aux réseaux sémantiques

La démarche que nous avons suivie a permis à partir de chaque item inducteur de percevoir le champ sémantique couvert par les associés cités. Une autre démarche possible est celle consistant à partir d'une notion commune à rechercher comment une notion se réalise dans différents signifiants.

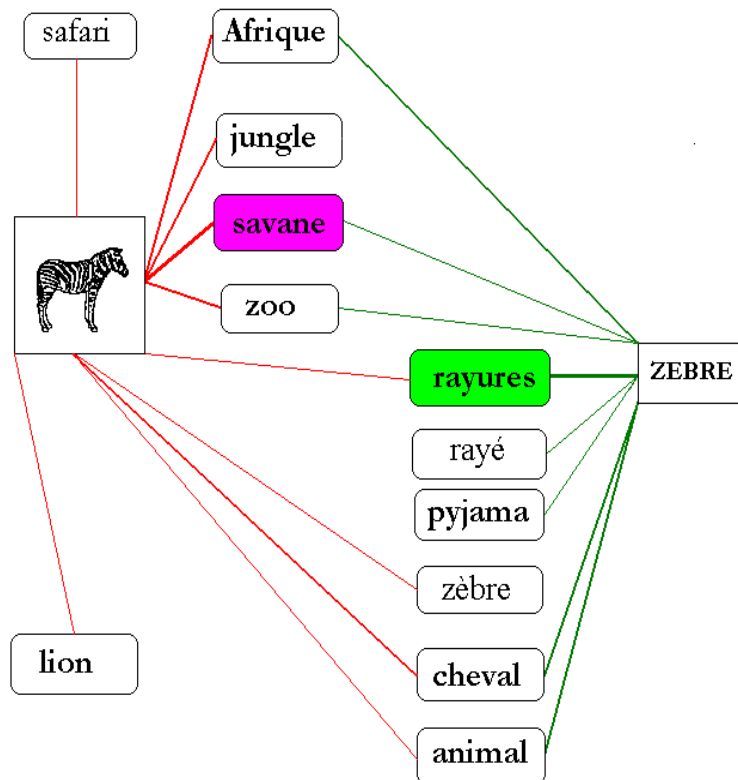
La mise en relation des éléments lexicaux est donc opérée au niveau des catégories d'objets pour tenter de représenter les relations entre items appartenant à la même catégorie sémantique et de constituer ainsi les réseaux sémantiques des entités à la manière de Collins et Loftus (1975).

Schéma de relations 'Animaux Sauvages'

Seules les relations supérieures à une occurrence sont mentionnées. Le rouge indique les relations les plus fortes.



Nous avons également mis en relation réseau sémantique verbal et imagé.



4. Comparaison des associations sémantiques avec des cooccurrences de corpus textuels

Le postulat de base est que les relations associatives reflètent l'usage verbal. Les données d'associations de mots recueillies dans la norme associative verbale que nous avons constituée sont comparées aux associations (cooccurrences de mots) trouvées dans des contextes phrastiques issus de corpus textuels de type *Le Monde* (Ferret & Cornuéjols 1998) et de type roman (Reza & Rastier 1999). Le pourcentage de recouvrement des associations verbales avec les cooccurrences textuelles est faible dans le cas du corpus issu du *Monde* car ce journal est très spécifique (politique, économique), alors qu'il est de 40% avec un corpus de type roman. Ceci accrédite la validité des normes associatives.

Les normes associatives imagées et verbales constituent des éléments de base utilisables dans les systèmes d'indexation d'images et de mots du web sémantique.

5. Perspectives pour la représentation des connaissances dans les systèmes artificiels, la recherche documentaire, le web sémantique, l'indexation multimodale,

...

Une des problématiques communes que nous avons avec le domaine du traitement du langage naturel est celle de la nature et de l'organisation des connaissances et le formalisme adéquat pour les représenter.

Les associations reflètent l'organisation en mémoire sémantique mais également l'usage verbal qui en est fait en production écrite. Ces corpus peuvent être un bon indicateur des associations spontanées entre termes que vont faire les opérateurs utilisateurs d'un système de recherche d'information. Ils permettent d'évaluer les phénomènes de fréquence d'occurrence et de co-occurrence de mots qui révèlent certains aspects de la structure sémantique sous-jacente à l'usage verbal.

L'une des perspectives est utiliser ces réseaux associatifs conjointement à des formalismes de représentation sémantique des textes pour élaborer des graphes conceptuels qui permettent une indexation des documents plus efficace, et donc une recherche d'information plus efficace. Cela permet aussi de constituer des dictionnaires électroniques à partir des associations de mots permettant de mieux cerner la requête de l'utilisateur si le mot ou le groupe de mots le plus efficace pour sa recherche ne lui vient pas à l'esprit lors de la consultation du système. Par ces associations de mots et ces dictionnaires, le système pourra fournir des résultats plus proches de la demande réelle de l'utilisateur. Ceci est encore du domaine de la recherche et des perspectives d'avenir. Constituer ainsi des réseaux de mots-clés, plutôt que des requêtes sur de simples mots isolés pourrait être intéressant.

Le problème particulier des images en indexation est de savoir quel est le type de trait de l'image qui peut jouer le rôle de descripteur. La possibilité de faire porter l'indexation, non plus seulement sur des termes isolés, mais sur des nœuds dans un réseau de relations sémantiques, pourrait aider à obtenir des systèmes de traitement de l'information plus adaptés, non seulement pour cerner plus précisément le sens de la requête, mais également pour être plus en adéquation avec les représentations sémantiques des utilisateurs finaux.

BIBLIOGRAPHIE (TRÈS SUCCINCTE)

CORNUÉJOLS, M. 2001. *Sens du mot, sens de l'image*, Paris, Editions L'Harmattan.

Les références citées ci-dessus y sont répertoriées.