

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université IBN KHALDOUN – Tiaret

Faculté des sciences, et de la technologie et des sciences de la matière

Département Informatique

Ecole Doctorale Sciences et Technologies de l'Information
et de la Communication (STIC)

Mémoire

Présenté pour l'obtention du diplôme de

Magister en Informatique

Option :
Systèmes d'Informations et de Connaissances (SIC)

Par:

MOKHTARI Ahmed

Thème

**Etude de l'évolution des systèmes sociaux de tagging
collaboratif et la relation à l'émergence
et l'évolution du langage**

Directrice de Mémoire: Mme HASSAS Salima (Prof. Université Claude Bernard-Lyon1)

Composition du jury

Président du jury:	Mr. BALLA Amar	(Prof. Ecole nationale Supérieure d'Informatique)
Examineurs:	Mr. CHALAL Rachid	(MCA Ecole nationale Supérieure d'Informatique)
	Mr. GHOMARI Reda A.	(MCA Ecole nationale Supérieure d'Informatique)
Directrice de Mémoire:	Mme HASSAS Salima	(Prof. Université Claude Bernard-Lyon1)

Année Universitaire 2010/2011

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Remerciements

*Je tiens à remercier ALLAH le tout puissant pour sa grâce et sa bonté
envers nous.*

*Je tiens ensuite à remercier mon encadreur la Professeur Salima HASSAS
pour l'honneur de travailler sous sa direction par cette occasion qu'elle m'a
offerte.*

*Je tiens à remercier les membres du jury, Mr. BALLA Amar,
Mr. CHALAL Rachid et Mr. GHOMARI Reda A.
pour avoir accepté d'évaluer ce modeste travail.*

Je remercie également tous mes enseignants.

Dédicace

J'ai le plaisir de dédier ce modeste travail à mes chers parents et mes frères.

Également à toute la promotion de l'école doctorale STIC en souvenir de tous les moments que nous avons passés ensemble et à tous mes amis.

Que le tout puissant les garde tous.

Sommaire

Table des matières	i
Liste des figures	iv

Partie I Etat de l'art

Introduction générale	1
1 Motivations	1
2 Travaux liés	1
3. La problématique	2
3.1 Les systèmes de tagging collaboratif	2
3.2 L'émergence et l'évolution du langage	3
3.3 Les systèmes de tagging et l'émergence du langage	4

Chapitre 1 Les systèmes de tagging

1.1 Introduction	6
1.2 Exemples de systèmes de tagging	6
1.3 La structure tripartite des systèmes de tagging	7
1.4. La folksonomy et la personomy	8
1.4.1 La folksonomy	8
1.4.2 La personomy	9
1.5 Les types de tags	9
1.6. L'évolution et la dynamique d'un système de tagging	10
1.6.1 Distributions de la loi de puissance	11
1.7 Les taxonomies et le tagging	11
1.8 Les ontologies et le tagging	11
1.9 Les systèmes de tagging pour la recherche de l'information	12
1.10 Conclusion	13

Chapitre 2 L'émergence du langage

2.1 Introduction	14
2.2. Le modèle de la cohérence de vocabulaire	14
2.3 Le modèle des réseaux de neurones	15
2.4 Le modèle de l'apprentissage réitéré (ILM)	16
2.5 Le modèle de mémoire associative	17

2.6 Les jeux du langage	19
2.6.1 Le jeu de dénomination	19
2.6.2 Le jeu de discrimination	21
2.7 Le langage comme un système complexe adaptatif	21
2.8 Conclusion	22

Chapitre 3 La modélisation à base d'agents

3.1 Introduction	23
3.2. Agent	23
3.3 Système multi-agents	25
3.4 La modélisation à base d'agent	25
3.5 Conclusion	26

Partie II Les objectifs, les hypothèses et la méthodologie de travail

4.1 Objectifs	27
4.2 Hypothèses de travail	27
4.3 Méthodologie de travail	28
4.4 La mise en pratique et réalisation	29
4.4.1 La plate forme	30
4.4.2 L'implémentation	32

Partie III La modélisation et l'expérimentation

Chapitre 5 La modélisation d'un système de tagging à base des SMAs

5.1 Introduction	33
5.1.1 L'axe Environnement	33
5.1.2 L'axe Organisation	34
5.1.3 L'axe Agent	34
5.1.4 L'axe Interaction	34
5.2 La modélisation de la dynamique de tagging	35
5.2.1 Environnement	35
5.2.2 Organisation	36
5.2.3 Agent	37

5.2.4 Interaction	37
5.2.5 Algorithme d'interaction	40
5.2.6 Adaptation	42
5.2.7 Résultats et discussion	42
5.2.8 Conclusion	49
Chapitre 6 La modélisation linguistique mono-tagging	
6.1 Introduction	50
6.2 Environnement	51
6.3 Organisation	52
6.4 Agent	52
6.5 Interaction	53
6.6 Adaptation	56
6.7 Algorithme d'interaction et d'adaptation	57
6.8 Résultats est discussion	58
6.9 Conclusion	73
Chapitre 6 La modélisation linguistique multi-tagging	
7.1 Introduction	74
7.2 Environnement	75
7.3 Organisation	76
7.4 Agent	76
7.5 Interaction	77
7.6 Adaptation	80
7.7 Algorithme d'interaction multi tagging et d'adaptation	81
7.8 Résultats est discussion	82
7.9 Conclusion	95
Conclusion générale	97
Bibliographie	98

Table des figures

FIG. 1.1 - La structure tripartite des systèmes de tagging	7
FIG. 1.2 - a) la croissance de nombre de tags dans un système de tagging réel	10
FIG. 1.2- b) la croissance de nombre de tags par une loi de puissance	10
FIG. 2.1- a) la structure du réseau de neurones	16
FIG. 2.1- b) le scénario d'apprentissage	16
FIG. 2.2 - schéma de l'apprentissage réitéré dans les générations successives	17
FIG. 2.3 - l'expérience de « talking heads »	18
FIG. 2.4 -Deux agents jouant le jeu de dénomination	20
FIG. 4.1 - Le modèle de référence pour une plate-forme multi-agents FIPA	31
FIG. 5.1 -Les différents axes d'un SMA (approche voyelle)	33
FIG. 5.2 - l'environnement d'un système de tagging et ses trois parties	36
FIG. 5.3 - la communauté des utilisateurs d'un système de tagging	36
FIG. 5.4 -l'architecture d'un agent taggeur avec les comportements	37
FIG. 5.5 - Evolution de nombre de tags	44
FIG. 5.6 - Evolution de l'entropie	44
FIG. 5.7 - l'évolution de nombre de ressources	45
FIG. 5.8 - Evolution de nombre de tags	46
FIG. 5.9 - Evolution de l'entropie	46
FIG. 5.10 - l'évolution de nombre de ressources	47
FIG. 5.11 - Evolution de nombre de tags	47
FIG. 5.12 - Evolution de l'entropie	48
FIG. 5.13 - l'évolution de nombre de ressources	48
FIG. 6.1 - l'environnement et les trois parties d'un système de tagging	51
FIG. 6.2 – les sous communautés dans un système de tagging	52
FIG. 6.3 – l'architecture d'un agent taggeur avec les comportements	53
FIG. 6.4 – exemple de matrice de codage	59
FIG. 6.5 – les courbes d'évolution du nombre de tags et l'entropie	60
FIG. 6.6 – l'évolution du nombre de ressources	61
FIG. 6.7 – un extrait de la matrice de codage	61
FIG. 6.8 – les catégories émergentes	62
FIG. 6.9 – un extrait d'une mémoire associative	63

FIG. 6.10 – un extrait de la folksommie (les premiers tags)	63
FIG. 6.11 – les ressources assignées par le tag « gofa »	63
FIG. 6.12 – les courbes d'évolution du nombre de tags et l'entropie	64
FIG. 6.13 – l'évolution du nombre de ressources	64
FIG. 6.14 – un extrait de la matrice de codage	65
FIG. 6.15 – les catégories émergentes	65
FIG. 6.16 – un extrait d'une mémoire associative	66
FIG. 6.17 – un extrait de la folksommie (les premiers tags)	66
FIG. 6.18 – les ressources assignées par le tag « hoya »	67
FIG. 6.19 – les courbes d'évolution du nombre de tags et l'entropie	67
FIG. 6.20 – l'évolution du nombre de ressources	68
FIG. 6.21 – un extrait de la matrice de codage	68
FIG. 6.22 – les clusters émergents	69
FIG. 6.23 – les ressources assignées par le tag « joka »	69
FIG. 6.24 – les courbes d'évolution du nombre de tags et l'entropie	70
FIG. 6.25 – l'évolution du nombre de ressources	70
FIG. 6.26 – un extrait de la matrice de codage	71
FIG. 6.27 – les clusters émergents	71
FIG. 6.28 – un les ressources assignées par le tag « pope »	72
FIG. 7.1 – l'environnement (du système de tagging) et ses trois parties	75
FIG. 7.2 – la communauté des utilisateurs dans un système de tagging	76
FIG. 7.3 – l'architecture d'un agent taggeur avec les comportements	77
FIG. 7.4 – un exemple de la matrice de codage	83
FIG. 7.5 – la dynamique de système de tagging	84
FIG. 7.6 – un extrait de la matrice de codage	85
FIG. 7.7 –un extrait de clusters émergents	85
FIG. 7.8 – un extrait d'une mémoire associative	86
FIG. 7.9 – un extrait de la folksommie (les premiers tags)	86
FIG. 7.10– les ressources assignées par les tags « jado », «wiwo », « jesi »	87
FIG. 7.11 – la dynamique de système de tagging	88
FIG. 7.12 – un extrait de la matrice de codage	88
FIG. 7.13 – un extrait de clusters émergents.	89

FIG. 7.14 – un extrait d’une mémoire associative	89
FIG. 7.15 – un extrait de la folksommie (les premiers tags)	90
FIG. 7.16 – les ressources assignées par les tags « dise », «roha», « qaro»	90
FIG. 7.17 – la dynamique de système de tagging	91
FIG. 7.18 – un extrait de la matrice de codage	91
FIG. 7.19 – un extrait de clusters émergents	92
FIG. 7.20 – les ressources assignées par les tags «joho », «cemi», « gofa»	92
FIG. 7.21 – la dynamique de système de tagging	93
FIG. 7. 22 – un extrait de la matrice de codage	93
FIG. 7. 23 – un extrait de clusters émergents	94
FIG. 7. 24 – les ressources assignées par les tags «hago», «bava »	94

INTRODUCTION

GENERALE

Introduction générale

1 Motivation

Dans les systèmes de tagging collaboratif, tout le monde annote les ressources par les tags pour créer un système de catégorisation, mais est ce qu'il y a une possibilité de l'émergence d'un langage de tagging partagé par une communauté d'utilisateurs de ces systèmes ?

Notre motivation à ce travail réside dans la tentative d'améliorer les systèmes de tagging collaboratif existant en étudiant cette possibilité c-à-d l'émergence d'un langage de tagging.

Ce langage, si son émergence est possible, a le caractère d'être conventuel par l'ensemble d'utilisateurs qui le rend comme une sorte de consensus de la communauté pour catégoriser une réalité sur le web.

Et il sera utilisé par la suite par les utilisateurs de ces systèmes pour décrire les ressources dans leurs requêtes afin de les découvrir lors de la phase de recherche des ressources.

La compréhension de la dynamique complexe qui gouverne l'évolution des systèmes de tagging collaboratif en permettant l'émergence d'un langage de tagging est elle-même une motivation pour traiter ce sujet.

2 Travaux liés

Beaucoup de travaux de recherche sont réalisés ayant comme objet de recherche l'étude des systèmes de tagging. Parmi les travaux qui sont proches de notre travail qui est l'évolution des systèmes de tagging collaboratif et l'émergence des folksonomies, nous citons à titre d'exemples quelques uns :

Le travail présenté dans (Golder et Huberman 2006) s'intéresse à la structure des systèmes de tagging collaboratif et à la dynamique observée dans ces systèmes.

(Cattuto et Loreto 2007) présentent une étude de leur évolution dans une perspective sémiotique, dans (Dellschaft et Staab 2008) un modèle épistémique est présenté pour les systèmes de tagging collaboratif.

Une étude de la dynamique complexe des systèmes de tagging est présentée dans (Halpin et Robu 2007) et dans le travail de (Santos-Neto et Condon 2009) le comportement individuel et social dans les systèmes de tagging sont analysés.

Il y a d'autres travaux de recherche qui s'intéressent aux systèmes de tagging selon différents angles comme (Benz et Grobelnik 2008) qui s'intéressent à la mesure de similarité inter-tags et inter-folksonomies. Les auteurs de (Schenkel et Crecelius 2008) étudient les systèmes de tagging d'un angle de la recherche de l'information (IR) pour améliorer les méthodes de la recherche pour construire des moteurs de recherche à base des tags. Dans (Yeung et Gibbins 2008), les systèmes de tagging sont utilisés pour la génération des profils des utilisateurs pour des systèmes de recommandation à base de tags.

3. La problématique

Les travaux réalisés dans la modélisation de l'émergence et l'évolution du langage ont montré qu'un ensemble d'agents peut établir un langage commun. Ces modèles permettent aux chercheurs d'établir un certain nombre de principes et de conditions qui aident à l'émergence et l'évolution du langage comme par exemple l'auto-organisation, l'imitation et l'apprentissage par renforcement.

Ce travail a pour but l'étude de l'évolution des systèmes de tagging collaboratif et les conditions qui permettent l'émergence des folksonomies, et d'étudier ces systèmes d'une perspective linguistique en considérant les folksonomies comme étant des langages à base des catégories, en appliquant et testant les principes utilisés dans l'émergence du langage sur les systèmes de tagging pour répondre à la question suivante: Est ce qu'il existe une relation entre l'émergence du langage et l'évolution des systèmes de tagging ? Autrement dit quels sont les fondements linguistiques qui peuvent expliquer leur évolution vers l'émergence des folksonomies?

En utilisant une modélisation à base de systèmes multi-agents.

3.1 Les systèmes de tagging collaboratif

L'apparition du web 2 (le web social) a changé la façon dont les utilisateurs utilisent et interagissent avec le web. À l'aide d'un ensemble de technologies, l'utilisateur peut participer à la création, l'organisation et même la rédaction des ressources (par exemple des pages web), les systèmes de tagging collaboratif tombent dans cette catégorie des technologies de web social.

L'idée de base derrière le principe de fonctionnement des systèmes de tagging est de donner à l'utilisateur la possibilité d'annoter les ressources à travers des mots clés (les tags) afin de les décrire ou les catégoriser pour faciliter l'organisation des ressources

personnelles ainsi la recherche et la découverte de ces ressources déjà annotées ou des ressources annotées par les autres utilisateurs du système.

Dans les systèmes de tagging où le processus du tagging se fait d'une façon collaborative, on observe l'émergence d'un système de catégorisation utilisé par la communauté des utilisateurs, ce système est dénommé « Folksonomie » pour la taxonomie des folks (Smith 2004).

Le processus de tagging collaboratif est un processus interactionnel complexe qui permet de considérer les systèmes de tagging « des systèmes complexes adaptatifs », qui s'adaptent à travers les changements faites par les utilisateurs lorsqu'ils annotent les ressources du système, ces interactions simples permettent l'émergence des structures macroscopiques (les folksonomies).

La modélisation à base des systèmes multi-agents peut bien servir dans la modélisation des systèmes de tagging, chaque agent représente un utilisateur et le système avec ses ressources représente l'environnement avec lequel ils interagissent par processus de tagging.

3.2 L'émergence et l'évolution du langage

Parmi les caractéristiques qui nous distinguent comme des êtres humains est sans doute notre faculté d'utiliser le langage dans la communication. Ce langage est utilisé pour différents buts comme l'expression des idées, la demande des informations et même pour penser symboliquement.

Une question naturelle s'est posée, quelle est l'origine du langage et comment elle évolue ? Cette question a été posée par les philosophes grecs et a été abordée scientifiquement dans le domaine des linguistiques dans le siècle passé où l'essai de répondre à cette question a engendré deux écoles de pensée principales.

La première école, qui s'appelle «cognitivist» et représentée par Chomsky (Chomsky 1995), suppose que le langage est une capacité innée et qu'elle est réalisée par un système cognitif qui stocke la description du langage (la syntaxe), ce système est donné comme nom l'organe du langage. La deuxième école, appelée « fonctionnaliste » (Dik 1997), fait l'hypothèse de l'invention du langage pour répondre au besoin de la communication, autrement dit le langage est inventé pour réaliser la fonction de la communication et la recherche sociale de l'information.

Un domaine de recherche très actif dans l'émergence du langage est la modélisation informatique « computational modeling », c'est une approche bio-inspirée et qui se base sur les modèles de vie artificielle (Steels 1997).

Dans cette approche, le langage est comparé aux systèmes biologiques et considéré comme étant « un système complexe adaptatif » qui s'adapte à son environnement par l'apprentissage où l'auto-organisation fait l'émergence du langage par les interactions des individus les uns avec les autres et avec l'environnement.

La plus part des modèles proposés dans cette approche se basent sur des scénarios de communication qui s'appelle « les jeux de langage » dans lesquels les individus essaient d'imiter des mots produits par les autres ou de donner des noms (mots) à des objets (sens) pour établir un lexique commun.

La modélisation à base de systèmes multi-agents (agent based modeling) est généralement utilisée pour modéliser les individus possédant et utilisant un langage dans ce type de modèles.

3.3 Les systèmes de tagging et l'émergence du langage

L'un des points communs entre les systèmes de tagging collaboratif et les langages est de considérer les deux « des systèmes complexes adaptatifs » où l'auto-organisation et les interactions au niveau du système jouent un rôle très important pour expliquer l'émergence des langages et les folksonomies. Par les systèmes de tagging, on s'intéresse au processus et la dynamique qui produit les folksonomies qui sont une sorte de vocabulaire ou un langage de catégorisation utilisé par la communauté des utilisateurs.

En plus, l'opération d'annoter une ressource par un tag pour la nommer ou la catégoriser par le processus « de tagging collaboratif » ressemble aux principes utilisés dans les jeux du langage (jeu de dénomination et jeu d'imitation par exemple) qui sont utilisés dans les modèles informatiques réalisés pour l'étude de l'émergence du langage.

La différence entre les jeux du langage et le tagging collaboratif réside dans le mode d'interaction entre les agents, bien qu'elle se fait d'une manière directe ou explicite « agent à agent » dans le premier alors que dans le deuxième, elle se fait d'une manière indirecte ou implicite « agent à système » ensuite « système à agent » autrement dit l'environnement d'interaction (le système de tagging) joue le rôle d'un milieu de communication.

Ainsi, la modélisation à base des systèmes multi-agents peut être utilisée pour modéliser et simuler un ensemble d'individus échangeant les mots et leurs sens ou une communauté d'utilisateurs annotent des ressources sur un système de tagging collaboratif.

L'organisation de travail

Ce travail est divisé en trois parties :

- La première partie, l'état de l'art, est composée de trois chapitres comme suit
 - Chapitre 1 : Les systèmes de tagging.
 - Chapitre 2 : L'émergence du langage.
 - Chapitre 3 : La modélisation à base d'agents.

- La deuxième partie est composée d'un seul chapitre
 - Chapitre 4 : Les objectifs, les hypothèses et la méthodologie de travail.

- La troisième partie : La modélisation et l'expérimentation, est composée de trois chapitres comme suit
 - Chapitre 5 : La modélisation d'un système de tagging à base des SMAs.
 - Chapitre 6 : La modélisation linguistique mono-tagging.
 - Chapitre 7 : La modélisation linguistique multi-tagging.

ETAT

DE

L'ART

1.1 Introduction

Une nouvelle famille d'applications sémantiques et sociales forme le WEB 2.0, la deuxième génération du Web. Une des forces conduisant à cette évolution est un paradigme de classification distribué connu sous le nom de « *le tagging collaboratif* », qui a été avec succès déployé dans des applications Web conçues pour organiser et partager en ligne diverses ressources telles que des pages Web, des articles scientifiques, des photos et la musique. Les utilisateurs du Web interagissent avec un système de tagging collaboratif en envoyant un contenu (ressource) dans le système, et librement associer des étiquettes (tags) à ce contenu, l'unité de base d'information dans un système de tagging collaboratif est une **entrée**, celle-ci est un triple (**utilisateur, ressource, {tag}**), dans de tels systèmes la communauté est composée de centaines de milliers d'utilisateurs et avec des millions de ressources. Les utilisateurs sont exposés donc aux ressources et aux tags déjà existant dans le système, et associent librement des tags avec les ressources nouvellement entrées.

Au niveau macroscopique, l'ensemble des tags évolue dans le temps et mène vers des patrons de l'utilisation des termes qui sont partagés par la communauté entière. Par conséquent on observe l'émergence d'un système de catégorisation généralement désigné par le nom « *folksonomie* », qui peut être utilisé efficacement pour la navigation dans un ensemble des ressources volumineux et hétérogène.

1.2. Exemples de systèmes de tagging

Parmi les systèmes de tagging les plus connus, on trouve les deux sites « Delicious » et « Flickr », le premier est concerné par les pages web (URL) et le deuxième est un site des photos personnelles.

1.2.1 Delicious, les données sont rassemblées en novembre 2006 (Benz et Grobelnik 2008). Au total 667.128 utilisateurs de la communauté Delicious ont été rassemblés, comportant 2.454.546 tags, 18.782.132 ressources, et 140.333.714 attributions de tag.

1.2.2 Flickr les données sont rassemblées pendant 2006 et 2007, l'ensemble de données se compose de 3.074.947 utilisateurs, 5.556.568 tags, 41.278.715 ressources, et 187.168.654 attributions de tag.

1.3 La structure tripartite des systèmes de tagging

Pour étudier un système, nous avons besoin d'un modèle conceptuel. Ceci s'applique sur les systèmes de tagging collaboratifs qui est capable de formaliser afin de faire des prévisions basées sur des données empiriques, le modèle tripartite a été déjà théorisé (Halpin et Robu 2007).

Pour ce modèle, il y a trois entités principales qui composent n'importe quel système de tagging collaboratif :

- les utilisateurs du système (les gens qui font réellement le tagging).
- les tags elles-mêmes.
- les ressources étant étiquetées.

Chacun de ces dernières peut être vue comme des espaces séparés qui constituent des ensembles de nœuds, liés par des arrêtes (voir FIG. 1.1).

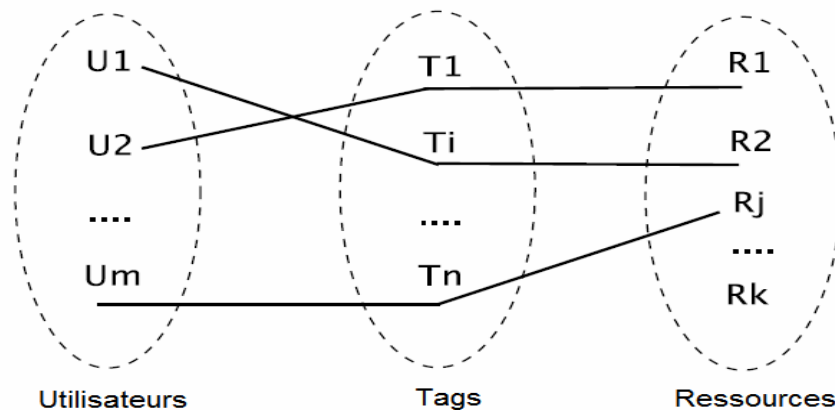


FIG. 1.1 -La structure tripartite des systèmes de tagging (De Halpin et Robu 2007)

Le premier espace, l'*espace d'utilisateur*, comprend l'ensemble de tous les utilisateurs du système de tagging, où chaque nœud est un utilisateur. Le deuxième espace est l'*espace de tag*, l'ensemble de toutes les tags, où un tag correspond à un terme (music, java, à lire). Le troisième espace est l'*espace de ressource*, l'ensemble de toutes les ressources. Chaque instance de tagging peut être vue comme deux arrêtes qui lient un utilisateur à un tag et ce tag à une ressource donnée.

1.4. Les folksonomies et les personomies

Le concept de folksonomie est le plus utilisé et étudié dans les systèmes de tagging collaboratif, ce terme (Folksonomy en anglais) se compose de ‘Folks’ et ‘taxonomy’ et l’autre terme (Personomy en anglais) se compose de ‘Personel’ et ‘taxonomy’.

1.4.1 La folksonomie

Le terme folksonomie a été inventé par T.Vander Wal (Smith 2004) pour décrire les structures taxonomiques qui émergent quand de grandes communautés d’utilisateurs étiquettent collectivement des ressources.

La collection de tous les utilisateurs, ressources et tags aussi bien que les attributions des tags aux ressources s'appellent folksonomie.

Une folksonomie est définie formellement comme suit (Dellschaft et Staab 2008) :

Définition 1. Une folksonomie est un tuple $F := (U, T, R, Y, pt)$ où

- U, T, et R sont des ensembles finis, dont les éléments s'appellent utilisateurs, tags et ressources respectivement.
- Y est une relation ternaire entre eux, c-à-d, $Y \subseteq U \times T \times R$, attribution de tag.
- pt est une fonction $pt : Y \rightarrow \mathbb{n}$ qui assigne à chaque attribution de tag de Y un marqueur temporel n. Il correspond au temps où un utilisateur a assigné un tag à la ressource.

Les attributions de tag peuvent être groupées dans plusieurs expéditions, une expédition contient toutes les attributions de tag faites par le même utilisateur à la même ressource en même temps.

Le marqueur temporel est associé aux attributions de tag et permet leur classement temporel dans un flux des événements de tagging. L'analyse des flux permet de faire des observations intéressantes de la dynamique continue dans un système de tagging ainsi que les mécanismes fondamentaux. Elle aide à une meilleure compréhension des systèmes de tagging social.

1.4.2 La personomie

Si nous voulons comprendre les intérêts d'un utilisateur simple, son profile, nous devons seulement se concentrer sur les tags et les documents qui sont associés à cet utilisateur particulier. Un tel ensemble de données est nommé *une personomie* (Yeung et Gibbins 2008).

Définition 2. Une personomie P_u d'un utilisateur u est une restriction d'une folksonomie F à u : càd. $P_u = (Tu; Du; Au)$, où

Au est l'ensemble d'annotations de l'utilisateur : $Au = \{(t; d) / (u; t; d) \in A\}$,

Tu est l'ensemble de tags de l'utilisateur: $Tu = \{t / (t; d) \in A\}$, et

Du est l'ensemble de documents de l'utilisateur: $Du = \{d / (t; d) \in A\}$.

1.5 Les types de tags

Dans un système de tagging, les utilisateurs utilisent différents types de tag pour organiser ou identifier ou qualifier une ressource. Les tags, selon Golder et Huberman (Golder et Huberman 2006), sont divisés en sept catégories :

1. **Identifiant de quoi s'agit-il.** En principe, les tags identifient les thèmes des éléments étiquetés. Ces tags incluent les noms communs de niveaux de spécificités différentes, aussi bien autant de noms propres.
2. **Identifiant ce qu'est-il.** Les tags peuvent identifier quel est le genre d'un élément, en plus de quoi s'agit-il. Pour exemple, photo, logiciel et livre.
3. **Identifiant qui le possède.** Quelques éléments sont étiquetés selon qui possède ou a créé la ressource.
4. **Catégories de raffinage.** Leur rôle est de raffiner ou qualifier les catégories existantes, les nombres peuvent accomplir cette fonction.
5. **Identification des qualités ou des caractéristiques.** Adjectifs comme effrayante, drôle, stupide étiquètent la ressource selon l'opinion de tagger.

6. **Autoréférence.** Les tags commençant par « mon » comme monobjet et mescommentaires identifient le contenu en termes de sa relation au tagger.

7. **Organisation de tâche.** Quand le rassemblement de l'information s'est relié à l'accomplissement d'une tâche, cette information pourrait être étiquetée selon cette tâche comme àlire.

1.6. L'évolution et la dynamique d'un système de tagging

L'étude des systèmes de tagging collaboratif montre qu'ils évoluent dans le temps par une dynamique complexe. Les interactions des utilisateurs avec le système en annotant les ressources permettent cette évolution.

L'observation de l'utilisation et la réutilisation des tags montre que la croissance de la taille de l'ensemble des tags du système est décrite par une loi de puissance avec un exposant plus petit qu'un et que cette distribution donne lieu à l'émergence des folksonomies, il faut noter que les vocabulaires ordinaires comportent une propriété similaire, cette propriété est liée à leur croissance. Si on balaye un texte écrit dans un langage naturel et compte le nombre de différents mots qui sont apparus en fonction du nombre total de mots lus, on trouve que leur croissance suit une loi en puissance comme celle observée dans les systèmes de tagging collaboratif (voir FIG. 1.2).

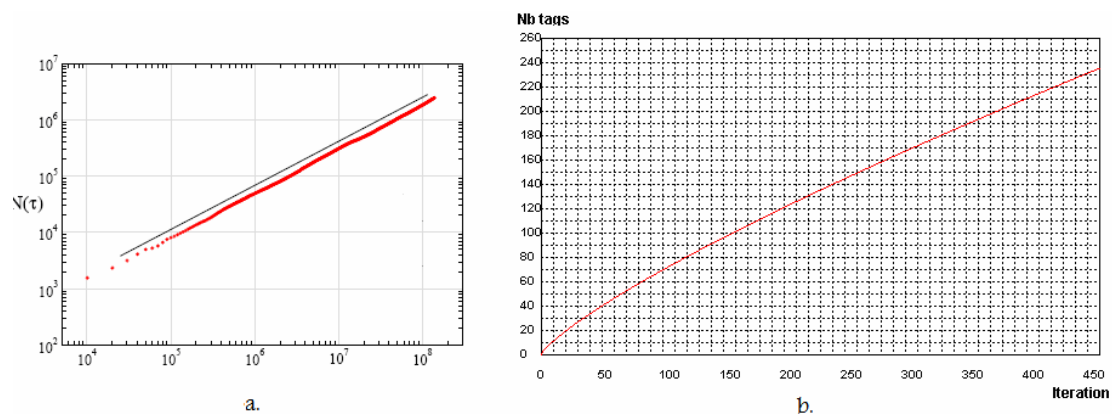


FIG. 1.2 - a) la croissance de nombre de tags dans un système de tagging réel

(De Cattuto et Baldassarri 2007)

- b) la croissance de nombre de tags par une loi de puissance

(un exposant =0.8)

1.6.1 Distributions de la loi de puissance

Certains phénomènes sont décrits par une distribution qui s'appelle la loi de puissance qui est définie mathématiquement comme suit (Halpin et Robu 2007).

Définition 3 Une loi de puissance est un rapport entre deux quantités scalaires x et y de la forme :

$$y = cx^\alpha$$

Où α et c sont des constantes caractérisant la loi de puissance.

1.7 Les taxonomies et le tagging

Dans la communauté des utilisateurs des systèmes de tagging, une différenciation est faite entre les systèmes à base de tagging et les taxonomies (Golder et Huberman 2006). Tandis que les derniers sont hiérarchiques et exclusifs, les premiers sont non-hiérarchiques et inclusifs. Des exemples des taxonomies incluent le système de Linnaean de classification des êtres vivants, la classification décimale de Dewey pour des bibliothèques, et le système des fichiers dans un système d'exploitation. Dans de tels systèmes, chaque animal, livre, fichier, est dans une catégorie non ambiguë qui est dans une catégorie plus générale.

Par contre, le tagging n'est ni exclusif ni hiérarchique, ça veut dire qu'un objet peut appartenir à plusieurs catégories et cela peut donc dans quelques circonstances avoir un avantage plus que les taxonomies hiérarchiques.

1.8 Les ontologies et le tagging

Le problème de génération des métadonnées pour des ressources du Web avec la plus grande efficacité continue à être le souci central comme la quantité de l'information sur le Web se développe (Halpin et Robu 2007). Les systèmes de tagging collaboratifs comme « Del.icio.us » et « Flickr » permettent aux utilisateurs d'affecter aux objets des mots clés pour faciliter la récupération pour lui et pour d'autres utilisateurs. L'ensemble des catégories qui sont dérivées à base des tags (les folksonomies) sont utilisées pour organiser l'information en ligne. Cette approche est différente des

ontologies formelles qui sont imposées par des experts, pas par les utilisateurs (Shirky 2005). Le tagging est considéré comme un procédé de catégorisation, contrairement au processus de classification pré optimisé comme les ontologies de Web sémantiques créées par expert.

Il y a des avantages et des inconvénients de l'approche de tagging, les systèmes de tagging permettent une malléabilité et une adaptabilité beaucoup plus grande dans l'organisation de l'information que les systèmes formels de classification, Cependant, un certain nombre de problèmes proviennent de l'organisation de l'information dans de tels systèmes comprenant l'ambiguïté dans la signification des tags et l'utilisation des synonymes qui crée la redondance informationnelle.

1.9 Les systèmes de tagging pour la recherche de l'information

Les moteurs de recherche utilisent le contenu d'une page web ou les métadonnées liées à cette page pour trouver les résultats pertinents et un moteur de recherche populaire comme « Google » utilise l'algorithme PageRank (Page et Brin 1998) basé sur l'analyse des liens pour classer les résultats. Une extension de ces méthodes et algorithmes de recherche est l'utilisation des tags attribués à une ressource (page web, photo) pour juger la pertinence de cette ressource vis-à-vis d'une requête d'un utilisateur, l'ensemble des tags dans un système de tagging comme « Del.icio.us » ou « Flickr » par exemple représente une connaissance collective construite par la communauté des utilisateurs de ces systèmes, ce que on appelle la sagesse sociale (Schenkel et Crecelius 2008). Par la sagesse sociale, nous désignons les folksonomies émergentes qui servent comme étant un système de catégorisation et des métadonnées pouvant être intégrés dans la recherche.

La reprise des ressources précédemment annotées est simple et intuitive, en tant que la plupart des applications de tagging collaboratif souvent présentent les tags de l'utilisateur dans l'interface. Le choix d'un tag montre toutes les ressources annotées par l'utilisateur avec ce tag. L'utilisateur recherchant les ressources particulières qu'il n'a pas encore annotées peut choisir un tag approprié et regarder des ressources annotées par d'autres utilisateurs.

Cependant, le procédé de découverte peut être beaucoup plus complexe. Un utilisateur peut naviguer dans la folksonomie, dirigeant par des tags, des ressources, ou même d'autres utilisateurs. Cette capacité de naviguer dirigé par la folksonomie à travers plusieurs dimensions (les tags, les ressources, les utilisateurs, le temps) est une raison de la popularité des systèmes de tagging collaboratif.

1.10 Conclusion

Dans cette partie, nous avons présenté un aperçu sur les systèmes de tagging collaboratif et quelques notions liées à eux comme les tags, la folksonomie et la dynamique.

Ce type de systèmes est devenu de plus en plus populaire grâce à leur philosophie de conception qui se fait autour de la participation de l'utilisateur et du partage des ressources en ligne dans la communauté et en particulier la possibilité d'annoter en utilisant des tags les ressources personnelles afin de les organiser et les décrire.

Beaucoup de travaux de recherche s'intéressent aux folksonomies pour étudier leur émergence, leur évolution et leur dynamique.

2.1 Introduction

Les langages possèdent un certain nombre de propriétés qui sont spécifiques au langage humain : elles le distinguent à la fois des autres systèmes de communication animale, et des autres modes d'expression et de communication existant dans les sociétés humaines. Ces propriétés relèvent de tous les niveaux linguistiques : phonologie, morphologie, syntaxe, lexique, etc.

Comment ont-elles pu émerger, et pourquoi ? Comment ont-elles pu évoluer ? Quelle est la nature des contraintes générales dans lesquelles s'est produite l'émergence du langage et son évolution ?

Pour aborder ce problème, un certain nombre de chercheurs ont réalisé des modélisations de l'émergence du langage, la modélisation informatique permet de réaliser des systèmes simples d'interactions entre agents, dans lesquels on introduit un certain nombre de contraintes et l'on observe les propriétés du système de communication qui se stabilisent au cours des échanges.

Ces modèles permettent de représenter de manière très simplifiée les interactions langagières entre agents et les processus d'acquisition d'un langage par ces agents. L'objectif essentiel de ces travaux n'est pas de modéliser de façon « réaliste » ces phénomènes hautement complexes, mais de rechercher les conditions minimales dans lesquelles peuvent émerger les propriétés linguistiques auxquelles on s'intéresse. Par la suite, on cite quelques modèles de l'émergence du langage lexical.

2.2 Le modèle de la cohérence de vocabulaire

Dans ce modèle, il est assumé que la langue humaine commence à partir d'un ensemble cohérent de correspondances entre des sens et des expressions «meanings-utterances mappings » (**M-U mappings**) (Ke et al. 2002).

Ces correspondances, considérées comme étant les premiers vocabulaires, sont le résultat des conventions établies parmi une population des agents. Chaque agent a sa propre manière d'appeler un ensemble des objets et sans toute conception explicite ou implicite, un vocabulaire commun cohérent peut être conventionné comme propriété émergente de la population.

Dans ce modèle, chaque agent a deux matrices, une pour ses correspondances M-U qui peut être parlées et autre matrice pour ses correspondances M-U écoutées, dont les éléments lui indique la probabilité de la corrélation entre les différentes correspondances M-U. Dans la production, selon sa matrice des parlants, le locuteur code un sens dans une expression parmi les correspondances M-U, en prenant la probabilité la plus haute dans la ligne de ce sens. Dans la compréhension, selon sa matrice des écoutes, l'interlocuteur décode le signal (l'expression écoutée) dans un sens selon les correspondances M-U, en prenant toujours la probabilité la plus haute dans la colonne de cette expression.

Après qu'une vérification directe si le sens codé par le locuteur assortit celui décodé par l'interlocuteur, un mécanisme d'auto-organisation ajuste des probabilités des correspondances M-U dans leurs matrices parlant/écoutant. Si elles s'assortissent, la probabilité de la correspondance M-U choisie dans la matrice parlante du locuteur est augmentée et d'autres probabilités dans la même ligne sont diminuées. En plus, la probabilité de la correspondance M-U choisie dans la matrice d'écoute de l'interlocuteur est augmentée et d'autres probabilités dans la même colonne sont diminuées. Autrement, l'ajustement inverse est exécuté. Toutes les probabilités sont aléatoirement initialisées au début de la simulation.

2.3 Le modèle des réseaux de neurones

Munroe et Cangelosi (Munroe et Cangelosi 2002), en utilisant un réseau de neurones, ont implémenté un modèle de fourragement de champignon pour démontrer comment l'apprentissage et la sélection naturelle interagissent dans des conditions différentes. C'est un modèle « situé » ; tous les agents fourragent dans un monde artificiel avec des « champignons » toxiques et comestibles.

La structure du réseau de neurones, qui simule l'état interne de chaque agent, est représentée sur la figure FIG. 2.1 (a). La couche d'entrée inclut la propriété visuelle du champignon et l'expression linguistique pour décrire le champignon. La couche de sortie inclut une partie pour l'action, reliant des communications langagières avec les actions efficaces, et une sortie linguistique, qui peut être employée pour entraîner d'autres agents. L'entrée et la sortie du réseau de neurones sont des signaux de

longueur constante. La compétence linguistique est stockée comme poids de liaison dans la connexion parmi les différentes couches.

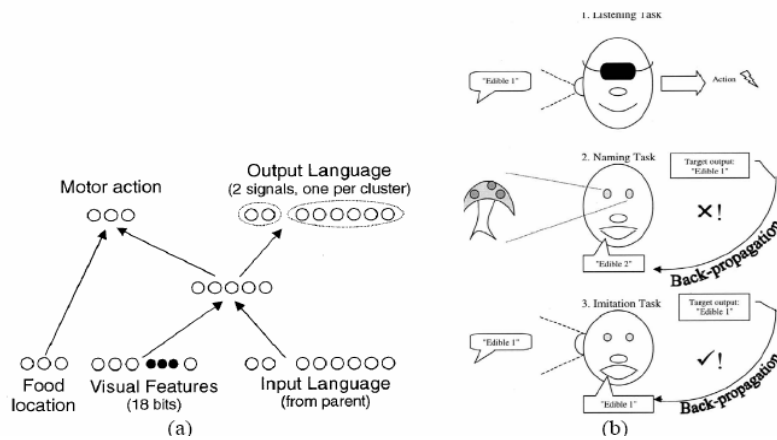


FIG. 2.1 - a) la structure du réseau de neurones ; b) le scénario d'apprentissage.

(De Cangelosi & Parisi 2002)

Après une génération de fourragement, vingt fourrageurs experts sont choisis comme parents basés sur leur fitness. Puis, par la reproduction asexuelle, chacun d'eux produit cinq enfants, qui copient les poids de liaison initiaux de leurs parents. L'opération de mutation dans les *algorithmes génétiques* (GA), un changement aléatoire du poids de liaison d'un certain agent enfant, présente la variance.

Les poids de liaison sont ajustés pendant un processus d'apprentissage. Dans l'étape 1 (procédé de fourragement), sans communication linguistique, les agents jugent la comestibilité des champignons seulement par l'information visuelle du champignon rencontré dans le monde artificiel. Après plusieurs générations, l'étape 2 commence, dans laquelle on permet la communication linguistique.

Les vingt fourrageurs experts sont reportés dans la prochaine génération en tant que « maîtres ». Chaque nouvel agent continue à fourrager comme avant mais une entrée linguistique additionnelle est toujours donnée par le parent.

2.3 Le modèle de l'apprentissage réitéré (ILM)

Le modèle de l'apprentissage réitéré (The Iterative Learning Model) est un cadre général pour modéliser la transmission culturelle de la langue verticalement (Kirby 2002).

Le modèle réitère à travers des générations en lesquelles on divise la population en des adultes et des apprenants. Chaque génération consiste en une population des agents qui apprennent la langue à partir des expressions produites par les générations précédentes. L'apprenant commence sa vie en tant qu'utilisateur débutant et acquiert la langue en observant le comportement des adultes, en supposant qu'ils maîtrisent la langue.

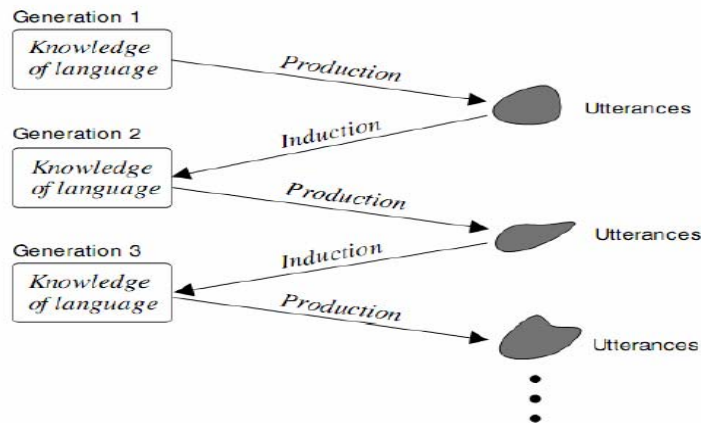


FIG. 2.2 - schéma de l'apprentissage réitéré dans les générations successives
À la fin de chaque itération, les adultes partent et sont remplacés par des nouveaux apprenants, ce cycle est répété comme le montre FIG. 2.2. ILM est un modèle simplifié de la dynamique de population et de la transmission culturelle de la langue.

2.5 Le modèle de mémoire associative

Dans les modèles développés par Steels et Kaplan (Steels 1996, Steels et Kaplan 1999, 2001, Kaplan 2000, 2001), les agents sont munis d'une mémoire associative, qui s'enrichit au cours des interactions. Autrement dit, contrairement aux modèles génétiques, les agents ne « naissent » pas avec un code langagier inscrit dans leur génome : ils « apprennent », tout au long de leur existence, à associer des sons avec des sens.

Dans les premières expériences, tout le processus était entièrement simulé sur ordinateur. Par la suite, ces chercheurs ont construit de véritables robots (*talking heads*), munis de caméras orientables et d'un système de traitement d'image qui permettaient à ces robots de percevoir et de « pointer » (par l'orientation de la caméra) différentes formes géométriques sur un tableau (voir FIG. 2.3).



FIG. 2.3 - l'expérience de « *talking heads* » (De Steels2001)

Ces conditions plus réalistes, présentent l'intérêt de faire apparaître des phénomènes nouveaux, difficilement simulables sur ordinateur.

Pour les premières expériences, Kaplan a étudié les conditions dans lesquelles un lexique stable pouvait émerger dans un tel système, en testant divers mécanismes d'apprentissage par les agents.

Dans toutes ces expériences, les agents disposent d'un certain nombre d'objets (typiquement une dizaine) dans leur environnement, et d'un certain nombre de mots qu'ils peuvent prononcer. Au départ, il n'y a aucun lien privilégié entre objets et mots dans la mémoire associative des agents : les locuteurs choisissent donc au hasard un mot qui est décodé aussi au hasard par leur interlocuteur. Au fur et à mesure des interactions, les agents modifient les liens dans leur mémoire associative.

La question qui se pose est de savoir quel type de modification est susceptible d'aboutir à la stabilisation d'un lexique commun entre les agents.

(1) Kaplan a d'abord testé un simple mécanisme d'*imitation*. Quel que soit le résultat du jeu, on augmente dans la mémoire associative de l'interlocuteur le lien entre le mot prononcé et l'objet désigné par le locuteur. Les résultats obtenus montrent que ce mécanisme est insuffisant pour obtenir un lexique commun.

(2) Dans une deuxième expérience, Kaplan utilise un apprentissage par *essai et*

erreur. Cette fois-ci, le résultat du jeu compte. S'il a été un échec, le lien entre le mot et l'objet désigné par l'interlocuteur est diminué. S'il a été un succès, ce lien est renforcé. Là encore, il n'y a pas de convergence dès que le nombre d'objets à désigner est assez grand.

(3) Enfin, Kaplan a combiné les deux mécanismes, en procédant de la manière suivante. En cas de succès du jeu, on procède comme dans l'expérience (2). Mais en cas d'échec, on ne se contente pas de diminuer le lien qui a provoqué l'erreur, on augmente aussi, comme dans l'expérience (1), le lien entre le mot prononcé par le locuteur et l'objet qu'il avait sélectionné. Autrement dit, l'interlocuteur utilise pour corriger ses erreurs l'indication fournie par le locuteur. On observe systématiquement la stabilisation d'un lexique commun en un temps très raisonnable.

Ainsi ces expériences montrent que l'établissement d'une convention lexicale peut être obtenue de manière « culturelle », c'est-à-dire par apprentissage au cours des interactions entre agents, à condition de disposer d'un mécanisme d'acquisition appropriée, qui conjugue une faculté d'imitation et l'aptitude à corriger ses propres erreurs.

2.6 Les jeux du langage

Pour étudier l'émergence du langage, les chercheurs ont inventé un ensemble de scénarios de communication permettant de modéliser un ensemble d'agents engagés dans une communication verbale, ces scénarios sont appelés les jeux du langage « langage game » ; on va citer deux de ces jeux qui sont les jeux de dénomination et les jeux de discrimination.

2.6.1 Le jeu de dénomination

Dans le jeu de dénomination (the Naming Game NG), une population des individus convient sur l'utilisation d'une convention simple (par exemple un nom donné à un objet) sans utiliser une coordination centrale, mais au contraire exploitant seulement les interactions locales (Steels 1995, Steels 1996). C'est un modèle simple dans lequel l'idée que la langue est un système complexe adaptatif a été exploitée,

Le modèle

Le jeu de dénomination (NG) est joué par une population de N agents qui jouent par paires des interactions afin de négocier les conventions, c.-à-d., associations entre les formes et les sens, et il peut décrire l'apparition d'un consensus global parmi elles. Un exemple d'un tel jeu est celui d'une population à la laquelle on doit atteindre le consensus sur le nom (c.-à-d., la forme) assigné à un objet (c.-à-d., la signification) exploitant seulement des interactions locales.

Chaque agent a un répertoire interne, dans lequel un nombre illimité de mots peut être stocké. Comme condition initiale, tous les répertoires doivent être vides. Les interactions se déroulent de la manière suivante :

- Un premier agent, le « locuteur », sélectionne un objet de son environnement et le nomme en prononçant un « mot ».
- Le deuxième agent, « l'interlocuteur », entend le mot émis, et doit désigner l'objet correspondant. Si l'objet désigné est bien celui qu'avait sélectionné le locuteur, le jeu est un succès, sinon c'est un échec (voir FIG. 2.4).

L'interlocuteur modifie alors sa mémoire associative en fonction du résultat du jeu en renforçant les scores en cas de succès et diminuant les scores en cas d'échec.

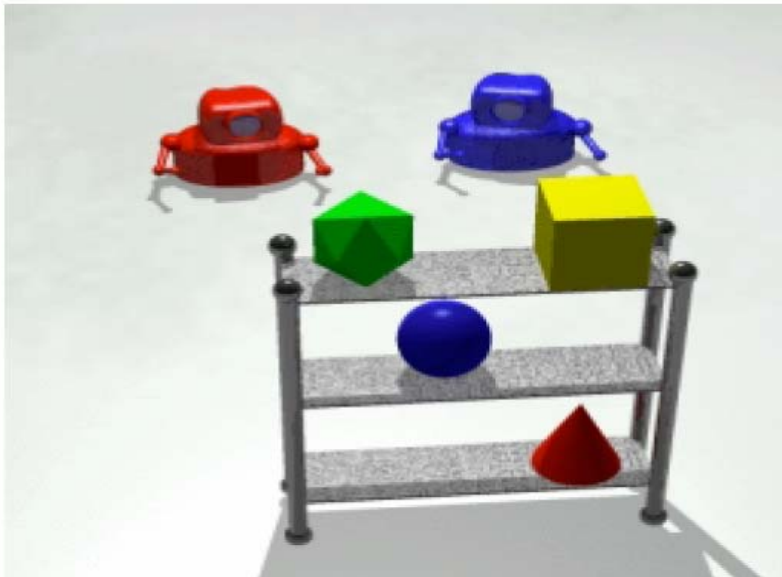


FIG. 2.4 -Deux agents jouant le jeu de dénomination (Du Kaplan 2000)

2.6.2 Le jeu de discrimination

Dans ce type de jeu, un ensemble d'agents avec des règles de communication de base et sans supervision externe peut évoluer un ensemble de catégories par la discrimination d'une réalité, en réalisant un système de communication caractérisé par quelques catégories linguistiques (Puglisi et autres 2007).

Le modèle

Le modèle fait participer une population de N individus, engagés dans la catégorisation d'un canal perceptuel analogique simple, chaque stimulus étant représenté comme un nombre à valeurs réelles dans l'intervalle $[0, 1]$, ici la catégorisation est identifiée comme le partitionnement de l'intervalle $[0, 1]$ en sous intervalles discrets.

Il devrait maintenir dans l'esprit que le but est d'étudier pourquoi le continuum de significations perceptibles dans le monde est organisé dans un nombre fini et petit de sous-ensembles avec différents noms.

Chaque individu a un répertoire dynamique des associations de forme-sens liant des catégories perceptuelles (sens) aux mots (formes). Les catégories et les mots perceptuels associés à elles co-évoluent dynamiquement à travers des interactions de communication élémentaires. Tous les joueurs sont initialisés avec seulement la catégorie perceptuelle triviale $[0, 1]$, sans nom associé à elle. À chaque itération un paire d'agents (en jouant comme locuteur et l'autre comme interlocuteur) sont choisis et présentés avec une nouvelle « scène », c.-à-d. un ensemble de M objets ($M \geq 2$), dénoté comme $o_i \in [0, 1]$ avec $i \in [1, M]$. Le locuteur distingue la scène, ajoutant de nouvelles frontières des catégories pour isoler le thème, ensuite il appelle un objet et l'interlocuteur essaye de le deviner. Le joueur réussi s'il devine un mot correct. Basé sur les résultats du jeu les individus peuvent mettre à jour leurs frontières de catégorie et le répertoire des mots associés.

2.7 Le langage comme un système complexe adaptatif

Le langage est un système complexe adaptatif. Il comporte les interactions écologiques de beaucoup de joueurs : les gens qui veulent communiquer et un monde à parler. Il fonctionne à travers plusieurs niveaux différents (neurones, cerveaux, et

corps ; phonèmes, morphèmes, lexèmes, constructions, interactions, et discours), différents niveaux sociaux (individus, groupes sociaux, réseaux, et cultures). Comme système complexe, ses systémativités émergent en suivant des principes adaptatifs et darwiniens.

Des changements du système sont engendrés par l'adaptation des agents à leur environnement (van Lier 2004), souvent comprenant la rétroaction réciproque qu'ils reçoivent en raison de leurs activités communes. Ainsi, le caractère principal du système linguistique peut être défini comme l'adaptation dynamique à un contexte spécifique (Tucker et Hirsch-Pasek 1993). En raison de son propriété à auto-organisation, la nouvelle organisation du système de langage qui émerge est qualitativement différente et nouvelle des organisations précédentes.

L'apprentissage des langages et leur utilisation sont des processus dynamiques dans lesquels les régularités et le système résultent de l'interaction des personnes, cerveaux, individus, sociétés et cultures en utilisant des langages dans le monde.

2.8 Conclusion

Dans cette partie, nous avons introduit l'émergence des langages, une question qui a été toujours posée par les chercheurs dans les différents domaines.

La problématique est étudiée en informatique dans beaucoup de travaux dans lesquels des modèles sont proposés pour modéliser l'émergence du langage ou autres propriétés. Nous avons présenté quatre modèles de l'émergence du langage: le modèle de la cohérence de vocabulaire, le modèle des réseaux de neurones, le modèle de l'apprentissage réitéré et le modèle de mémoire associative.

Dans ces modèles, le langage est considéré comme étant un système complexe adaptatif.

Les jeux de langage sont un modèle général de simulation pour étudier les phénomènes linguistiques comme l'émergence, l'évolution et la stabilité.

3.1 Introduction

Les phénomènes observés dans les systèmes complexes ne se produisent pas comme résultat des comportements isolés faits par des individus, mais plutôt comme résultat des interactions répétées entre les individus à travers le temps, si on prend comme exemple les systèmes de tagging collaboratif, les structures émergentes (les folksonomies) ne sont pas le résultat d'une action ou un comportement fait par un seul utilisateur, mais le résultat d'un processus collaboratif « tagging collaboratif ».

On commence cette section par présenter les concepts de base de cette approche de modélisation comme agent et système multi-agents. Ensuite on présente l'approche *la modélisation à base d'agents (Agent Based Modeling ABM)*, qui comporte la simulation d'un grand nombre d'agents autonomes qui agissent l'un sur l'autre et avec un environnement, afin de modéliser et simuler ce types de systèmes et l'observation des structures émergentes de leur interactions nous aide à mieux comprendre et expliquer leur comportement.

3.2 Agent

Le concept d'agent a été l'objet des études dans différentes disciplines. Il a été non seulement utilisé dans les domaines de l'intelligence artificielle, mais aussi dans des disciplines comme la philosophie et la psychologie. Aujourd'hui, avec l'avènement de nouvelles technologies et l'expansion de l'Internet, ce concept est encore associé à plusieurs nouvelles applications comme *agent ressource*, *agent courtier*, *assistant personnel*, *agent interface*, *agent ontologique*, etc. Dans la littérature, on trouve une multitude de définitions d'agents. Elles se ressemblent toutes, mais diffèrent selon le type d'application pour laquelle est conçu l'agent. Parmi les premières définitions de l'agent on cite celle de Ferber (Ferber 95) :

« Un agent est une entité autonome, réelle ou abstraite, qui est capable d'agir sur elle-même et sur son environnement, qui, dans un univers multi-agents, peut communiquer avec d'autres agents, et dont le comportement est la conséquence de ses observations, de ses connaissances et des interactions avec les autres agents. »

Il ressort de cette définition des propriétés clés comme l'*autonomie*, l'*action*, la *perception* et la *communication*. D'autres propriétés peuvent être attribuées aux agents. Nous citons en particulier la *réactivité*, la *rationalité*, l'*engagement* et l'*intention*.

Dans (Jennings et Wooldridge 98), la définition suivante pour un agent a été proposée:

« Un agent est un système informatique, *situé* dans un environnement, et qui agit d'une façon *autonome* et *flexible* pour atteindre les objectifs pour lesquels il a été conçu. »

Les notions «situé», «autonomie» et «flexible» sont définies comme suit :

- *situé* : l'agent est capable d'agir sur son environnement à partir des entrées sensorielles qu'il reçoit de ce même environnement. Exemples : systèmes de contrôle de processus, systèmes embarqués, etc.
- *autonome* : l'agent est capable d'agir sans l'intervention d'un tiers (humain ou agent) et contrôle ses propres actions ainsi que son état interne ;
- *flexible* : l'agent dans ce cas est :
 - *capable de répondre à temps* : i.e., l'agent doit être capable de percevoir son environnement et élaborer une réponse dans les temps requis ;
 - *proactif* : l'agent doit exhiber un comportement proactif et opportuniste, tout en étant capable de prendre l'initiative au «bon» moment ;
 - *social* : l'agent doit être capable d'interagir avec les autres agents (logiciels et humains) quand la situation l'exige afin de compléter ses tâches ou aider ces agents à accomplir les leurs.

Un exemple simple est un agent simulé qui se déplace autour un environnement simulé en cherchant de la nourriture, et consommant de la nourriture quand elle la trouve.

Dans la plupart des modèles de la psychologie social, un agent est une version simplifiée et abstraite d'un être humain. Cependant, d'autres niveaux des agents sont

également possibles; un agent peut représenter un neurone dans un réseau de neurones simulé, ou a acteur économique à grande échelle telle qu'une société.

3.3 Système multi-agents

Un système multi-agents, alors, est un système qui contient un ensemble d'agents agissant l'un sur l'autre et/ou avec leurs environnements. Les SMA sont conçus et implantés idéalement comme un ensemble d'agents interagissants, Le plus souvent, selon des modes de *coopération*, de *concurrence* ou de *coexistence* (Chaib-draa 94). Un système de multi-agents (SMA) est généralement caractérisé par :

1. chaque agent a des informations ou des capacités de résolution de problèmes limités; ainsi chaque agent a un point de vue partiel ;
2. il n'y a aucun contrôle global du système multi-agents ;
3. les données sont décentralisées ;

Les SMA sont des systèmes idéaux pour représenter des problèmes possédants de multiples méthodes de résolution, de multiples perspectives. Ces systèmes possèdent les avantages traditionnels de la résolution distribuée et concurrente de problèmes comme la modularité, la vitesse (avec le parallélisme), et la fiabilité (due à la redondance). Ils héritent aussi des bénéfices envisageables de l'Intelligence Artificielle comme le traitement symbolique (au niveau des connaissances), la facilité de maintenance, la réutilisation et la portabilité mais surtout, ils ont l'avantage de faire intervenir des schémas d'interaction sophistiqués. Les types courants d'interaction incluent la coopération (travailler ensemble à la résolution d'un but commun) ; la coordination (organiser la résolution d'un problème de telle sorte que les interactions nuisibles soient évitées ou que les interactions bénéfiques soient exploitées) ; et la négociation (parvenir à un accord acceptable pour toutes les parties concernées).

3.4 La modélisation à base d'agent

Un modèle à base d'agent ou à base des systèmes multi-agents (Agent based modeling ABM) est un système multi-agents simulé construit pour un but particulier : capturer les éléments théoriques principaux d'un certain processus biologique ou psychologique ou social.

Les agents dans ces modèles interagissent les uns avec les autres et avec un environnement à travers le temps. Cette approche tient compte de l'observation des conséquences à grande échelle des hypothèses théoriques quand les comportements sont effectués dans le contexte de beaucoup d'autres agents, et réitéré dynamiquement sur une période prolongée.

Dans un ABM, les agents sont autonomes, ils poursuivent indépendamment ces différents buts basés sur leur propre information locale. Il n'y a aucune autorité centrale. Cette propriété signifie que les structures émergent d'un multi-agent sont dues aux processus de l'*auto-organisation* plutôt que la conception et la planification centralisées

Ainsi, les agents suivent des règles extrêmement simples, l'un des buts de ce type de modélisation est d'identifier les règles les plus simples qui génèrent les phénomènes complexes en étude (la notion de l'émergence).

3.5 Conclusion

La modélisation à base des systèmes multi-agents représente une approche alternative des approches analytiques basées sur des modèles mathématiques. Cette approche sert mieux à modéliser les systèmes complexes où les structures émergentes des interactions et la dynamique qui gouverne leurs comportements peuvent être générés et expliqués par le processus de l'auto-organisation.

HYPTHESES
ET
METHODOLOGIE

4.1 Objectifs

L'objectif de ce travail est d'étudier l'évolution des systèmes de tagging collaboratif et de répondre à la question suivante:

Quelles sont les conditions qui permettent leur dynamique et leur évolution qui mènent à l'émergence des folksonomies au sein de ces systèmes ?

Le deuxième objectif est d'étudier l'évolution des systèmes de tagging d'un perspective linguistique et cognitif et dans cette phase de travail, nous essayons de répondre à la question suivante:

Est-ce que un langage de tagging partagé par une communauté d'utilisateurs peut émerger dans tels systèmes ? Ou bien Est-ce que une modélisation à base linguistique, où nous considérons la folksonomie émergente un langage émergent et les tags comme les mots de ce langage et les ressources comme les sens de ces mots, peut mener à l'émergence d'un langage de tagging, et peut générer la même dynamique observées dans les systèmes de tagging collaboratif?

4.2 Hypothèses de travail

Pour atteindre les objectifs fixés pour ce travail, nous posons un certain nombre d'hypothèses.

La première hypothèse concernant l'évolution des systèmes de tagging, on suppose que les systèmes de tagging sont des systèmes complexes adaptatifs où :

- L'auto-organisation (la capacité de système de s'évoluer vers des structures organisationnelles bien définies ou vers un ensemble d'états stables sans supervision extérieure) qui se fait par les interactions des utilisateurs avec le système.
- L'adaptation du système se fait par l'apprentissage lorsque les utilisateurs apprennent quels sont les tags appropriés à une ressource et la même chose lorsque le système utilise des tags choisis à partir des connaissances individuelles d'un utilisateur.
- Le mécanisme du sélectionisme (choisir une configuration, la plus utilisée ou la plus sélectionnée) qui se fait au niveau locale par l'imitation individuelle (la tendance de l'utilisateur d'utiliser des tags qu'il a déjà utilisé) ainsi qu'au niveau globale par

l'imitation sociale (la tendance de l'utilisateur d'utiliser des tags que des autres utilisateurs ont déjà utilisé).

Qui sont responsables de la dynamique observée au sein des systèmes de tagging et de l'émergence des folksonomies.

La deuxième hypothèse pour l'évolution des systèmes de tagging d'un perspective linguistique, on suppose, en plus les principes cités dans la première hypothèse, que:

-Doter les agents par des mémoires associatives pour stocker des connaissances linguistiques (des correspondances « tag-ressource»).

- L'adaptation de ces connaissances linguistiques qui se fait par l'apprentissage par renforcement c-à-d ils renforcent les scores des tags les plus appropriés à une ressource donnée et ils diminuent les scores des tags mal attribués à une ressource donnée.

-L'utilisation des jeux du langage (jeu d'imitation et jeu de dénomination) qui peuvent servir comme modèle avec une adaptation aux systèmes de tagging, car dans ce type de système il n'y a pas une communication directe entre les utilisateurs pour jouer le rôle d'un locuteur ou un interlocuteur comme celle trouvée dans ces jeux,

Sont responsables de l'émergence d'un langage de tagging partagé par les agents au niveau de leurs mémoires associatives, ce langage sert comme un système de catégorisation appelée une folksonomie.

4.3 Méthodologie de travail

Notre Méthodologie de travail se base principalement sur des simulations et des expérimentations informatiques pour étudier l'évolution des systèmes de tagging collaboratif.

Pour cela nous utilisons le paradigme des systèmes multi-agents pour modéliser ces systèmes qui sont considérés comme des systèmes complexes adaptatifs et pour l'organisation de travail, nous inspirons des travaux de Kaplan (Kaplan 2000) en adoptant une approche incrémentale des modélisations au niveau de l'architecture de l'agent et au niveau des interactions des agents avec le système de tagging.

4.3.1 Interactions

Les interactions entre les agents et le système commencent par des schémas simples constitués principalement par les opérations de tagging faites par les agents, dans cette étape l'agent fait des assignations des tags à des ressources et il les envoie au système qui va les sauvegarder.

Dans la deuxième phase de modélisation, les interactions deviennent plus complexes et se base sur les schémas de communication utilisés dans les scénarios des jeux du langage, autrement dit, il y aura une conversation entre un agent et le système pour négocier l'assignation d'un tag à une ressource donnée, l'agent et le système peuvent avoir le rôle du locuteur ou de l'interlocuteur.

4.3.2 Agents

L'architecture des agents dans les premières expériences est constituée d'un ensemble de comportement pour assurer les opérations de tagging. Dans cette phase, l'architecture ne contient pas une mémoire, la seule mémoire se trouve au niveau de l'environnement (système de tagging collaboratif).

Par la suite et comme les modèles deviennent complexe, nous dotons l'architecture de l'agent par une mémoire associative qui contient des associations entre les tags et les ressources c-a-d les assignations effectuées « tag-ressource».

4.3.3 Environnements

L'environnement dans lequel se trouvent les agents et avec lequel ils interagissent est le système de tagging collaboratif qui constitue selon le modèle tripartite (Halpin et Robu 2007) d'un ensemble d'utilisateurs (agents) et un ensemble de ressources à tagger par un ensemble de tags.

Par la suite, lorsque nous ajoutons les mémoires associatives aux agents nous faisons la même chose pour le système de tagging pour étudier le système de catégorisation qui peut émerger dans la mémoire associative du système.

4.4 La mise en pratique et réalisation

Dans cette partie nous présentons la plate-forme JADE utilisée pour supporter les systèmes multi-agents modélisés et le langage Java utilisé dans l'implémentation.

4.4.1 La plate-forme

Le meilleur moyen pour construire un système multi-agents (SMA) est d'utiliser une plate-forme multi-agents. Cette dernière est un ensemble d'outils nécessaire à la construction et à la mise en service d'agents au sein d'un environnement spécifique. Ces outils peuvent servir également à l'analyse et au test du SMA ainsi créé. Ces outils peuvent être sous la forme d'environnement de programmation (API) et d'applications permettant d'aider le développeur.

Actuellement, il existe plusieurs plates-formes de développement des agents et des SMA parmi ceux on trouve la plate forme JADE :

JADE est une plate-forme multi-agents développée en Java par **CSELT** (Groupe de recherche de Gruppo Telecom, Italie) qui a comme but la construction des systèmes multi-agents et la réalisation d'applications conformes à la norme **FIPA** (Foundation for Intelligent Physical Agents) (FIPA, 1997).

JADE comprend deux composantes de base : une plate-forme agents compatible FIPA et un paquet logiciel pour le développement des agents Java.

La plate forme multi-agents JADE offre les avantages suivants :

- Plate forme assez facile à mettre en place.
- Disponibilité de packages sur lesquels nous sommes appuyés pour développer le démonstrateur.
- Documentation claire et complète.
- Licence gratuite.

La norme FIPA pour les systèmes multi-agents

Les premiers documents de spécification de la norme FIPA, appelés spécifications **FIPA97**, établissent les règles normatives qui permettent une société d'agents d'interopérer. Tout d'abord, les documents FIPA décrivent le modèle de référence d'une plate-forme multi-agents (voir FIG. 4.1) où ils identifient les rôles de quelques agents clés nécessaires pour la gestion de la plate-forme, et spécifient le contenu du langage de gestion des agents et l'ontologie du langage.

Le standard spécifie aussi le **Langage de Communication d'Agents** (Agent Communication Language - **ACL**), la communication des agents est basée sur l'envoi

de messages. Le langage **FIPA ACL** est le langage standard des messages et impose le codage, la sémantique et la pragmatique des messages. La norme n'impose pas de mécanisme spécifique pour le transport interne de messages. Plutôt, puisque les agents différents pourraient s'exécuter sur des plates-formes différentes et utiliser des technologies différentes d'interconnexion, **FIPA** spécifie que les messages transportés entre les plates-formes devraient être codés sous forme textuelle. On suppose que l'agent est en mesure de transmettre cette forme textuelle.

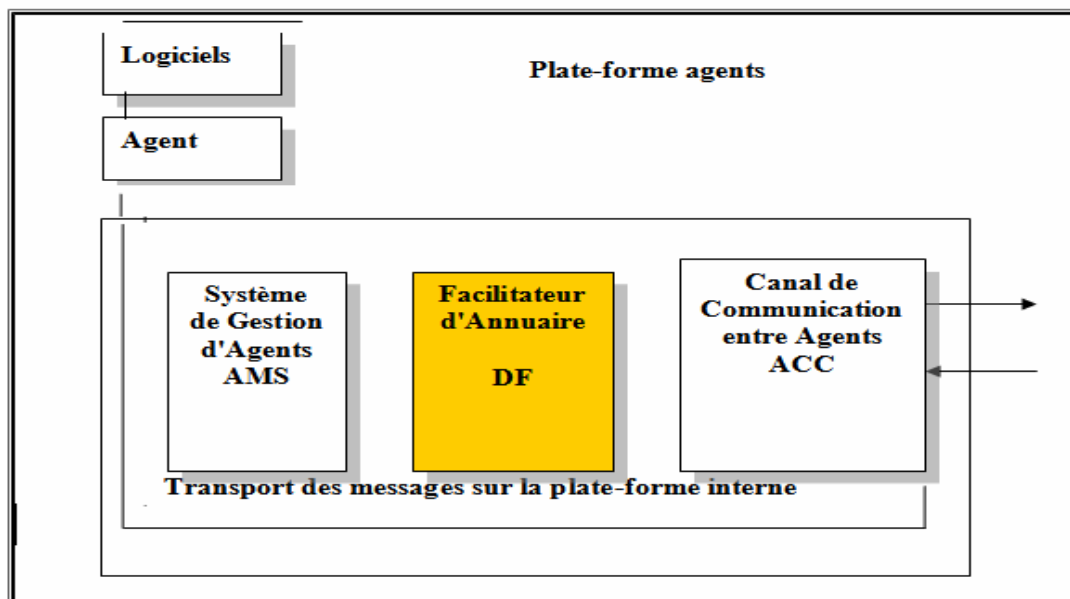


FIG. 4.1 - Le modèle de référence pour une plate-forme multi-agents FIPA

Dans la Fig. 4.1, on voit qu'il existe trois rôles principaux dans une plate-forme multi-agents FIPA :

- Le **Système de Gestion d'Agents** (Agent Management System - **AMS**) est l'agent qui exerce le contrôle de supervision sur l'accès à l'usage de la plate-forme; il est responsable de l'authentification des agents résidents et du contrôle d'enregistrements.
- Le **Canal de Communication entre Agents** (Agent Communication Channel - **ACC**) est l'agent qui fournit la route pour les interactions de base entre les agents dans et hors de la plate-forme; c'est la méthode de communication implicite qui offre

un service fiable et précis pour le routage des messages; il doit aussi être compatible avec le protocole **IIOP** (Protocole de communication permettant d'intégrer et de mettre en réseau des applications de provenance diverse) pour l'interopérabilité entre les différentes plates-formes multi-agents.

- Le **Facilitateur d'Annuaire (Directory Facilitator - DF)** est l'agent qui fournit un service de pages jaunes à la plate-forme multi-agents.

4.4.2 L'implémentation

La plate forme JADE est développée en Java pour cela tous les agents gérés par cette plate forme doivent être implémentés en Java, en bénéficiant ainsi de tous les avantages de la programmation en Java.

Le langage Java

Le langage Java est un langage de programmation orienté objet créé par James Gosling et Patrick Naughton, présenté officiellement le 23 mai 1995.

Le langage Java a la particularité principale que les logiciels écrits avec ce dernier sont très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux avec peu ou pas de modifications. C'est la plate-forme qui garantit la portabilité des applications développées en Java.

Le langage reprend en grande partie la syntaxe du langage C++, très utilisé par les informaticiens. Néanmoins, Java a été épuré des concepts les plus subtils du C++ et à la fois les plus déroutants, tels que les pointeurs et les références, et l'héritage multiple remplacé par l'implémentation des interfaces. Les concepteurs ont privilégié l'approche orientée objet de sorte qu'en Java, tout est objet à l'exception des types primitifs (nombres entiers, nombres à virgule flottante, etc.)

Philosophie

Lors de la création du langage Java, il avait été décidé que ce langage devait répondre à 5 objectifs :

- simple, orienté objet et familier ;
- robuste et sûr ;
- indépendant de la machine employée pour l'exécution ;
- très performant ;
- interprété, multi-tâches et dynamique

**MODELISATION
ET
RESULTATS**

5.1 Introduction

L'étude de l'évolution des systèmes de tagging collaboratif et les phénomènes qui accompagnent cette évolution exige une modélisation adéquate. Nous utilisons la modélisation des systèmes complexes à base des systèmes multi agents que nous jugeons appropriée à ce type de systèmes où le système consiste en un ensemble d'entités (les utilisateurs) modélisées par un système multi-agents et les agents sont en interaction complexe et forte entre eux et avec l'environnement.

Un système multi-agents est modélisé et décrit par rapport à des axes définis par l'approche voyelle: Agent, Environnement, Interaction, Organisation.

Dans chaque axe de cette approche, nous nous intéressons à un aspect du système multi-agents (voir FIG. 5.1).

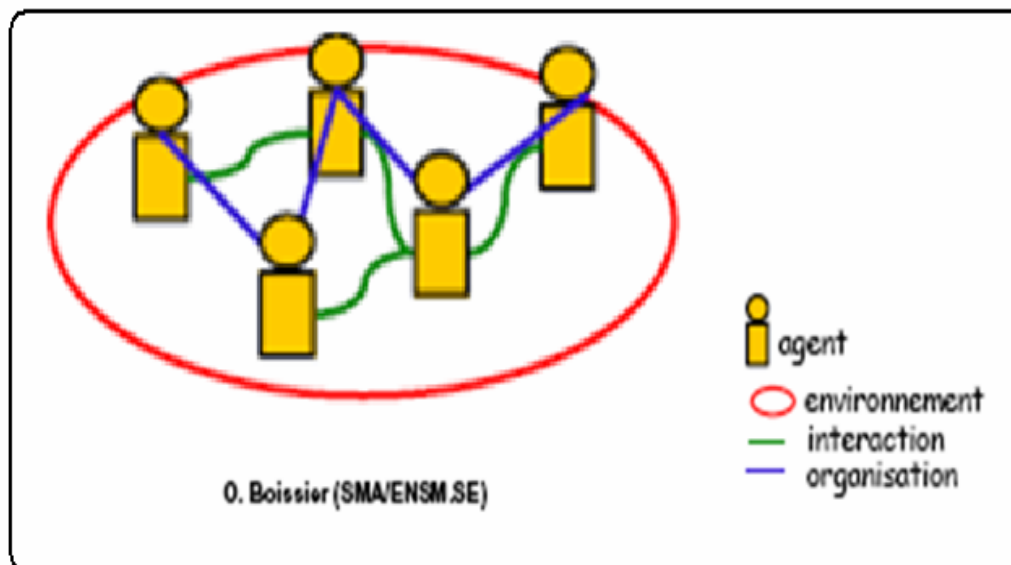


FIG. 5.1 -Les différents axes d'un SMA (approche voyelle)

5.1.1 L'axe Environnement

Nous modélisons le monde physique ou virtuel qui contient les agents et les autres entités comme les ressources. Il peut être à titre d'exemple un réseau informatique ou un espace réel où les agents peuvent se déplacer pour collecter de la nourriture ou autre. Dans notre cas de système de tagging collaboratif, il représente l'espace des ressources à tagger, leurs tags et leurs taggeurs par exemple.

5.1.2 L'axe Organisation

Dans cet axe nous précisons les organisations sociales prédéfinies comme une équipe de joueurs ou des organisations émergentes comme dans les réseaux sociaux, au fur et à mesure que le réseau évolue, il aura des groupes et des communautés émergentes.

5.1.3 L'axe Agent

Nous décrivons l'architecture des différents agents qui constituent le système, en précisant les perceptions/actions et les règles de comportement suivies par les agents.

5.1.4 L'axe Interaction

Nous définissons la dynamique et les interactions qui gèrent le système en y permettant l'émergence des propriétés distinguables, ces interactions sont entre les agents et entre les agents et l'environnement.

Nous gardons cette approche de modélisation dans tous les modèles que nous allons établir dans ce travail, en enrichissant le modèle à chaque étape de modélisation pour atteindre l'objectif final de ce travail, l'émergence d'un langage de tagging partagé par une communauté des utilisateurs d'un système de tagging collaboratif.

Nous commençons par un modèle qui a comme but l'étude de la dynamique et l'évolution des systèmes de tagging collaboratif.

Ensuite nous présentons le deuxième modèle en enrichissant le modèle précédent au niveau de l'architecture de l'agent taggeur par une mémoire associative et au niveau des interactions qui vont suivre les schémas des jeux de langage, dans ce modèle nous utilisons un processus mono tagging pour étudier l'émergence d'un langage de tagging lexicale.

Enfin nous présentons le troisième modèle en enrichissant le modèle précédent au niveau du processus de tagging. Dans ce modèle nous utilisons un processus multi tagging pour étudier l'émergence d'un langage de tagging à structure grammaticale partagé par une communauté des utilisateurs d'un système de tagging collaboratif.

5.2 La modélisation de la dynamique de tagging

Dans cette partie, nous proposons notre premier modèle pour répondre à notre premier objectif qui a pour but l'étude de la dynamique de tagging qui permet l'émergence des folksonomies.

Par ce premier modèle proposé, nous voulons régénérer les dynamiques observées au sein d'un système de tagging collaboratif et en particulier l'évolution de nombre de tags qui suit une distribution appelée « une loi en puissance » dans tels systèmes.

En testant certaines hypothèses que nous mettons concernant la dynamique générale que nous supposons fortement interactionnelle et auto-organisationnel entre les utilisateurs et le système (la dynamique des systèmes complexes) et concernant la dynamique locale qui gouverne le comportement d'un utilisateur du système que nous le supposons comme étant influencés par la communauté (ce que nous appelons l'imitation sociale) et ainsi que par la culture et les connaissances individuelles de cet utilisateur (ce que nous appelons l'imitation individuelle).

Ce modèle se base sur le travail de Santos-Neto et Condon (Santos-Neto & Condon 2009) où le comportement individuel et social dans les systèmes de tagging sont analysés.

Un système de tagging collaboratif consiste selon le modèle tripartite (Halpin et Robu 2007) en trois parties, un ensemble d'utilisateurs, un ensemble de tags et un ensemble de ressources.

Ces trois ensembles sont reliés par des liens réalisés par l'opération de tagging qui permet d'associer pour un utilisateur donné un tag à une ressource.

Dans notre modèle les utilisateurs du système sont simulés par de agents logiciels, les tags sont des mots choisis aléatoirement et les ressources sont réduites à un mot clé généré aléatoirement et on considère ce mot clé comme étant un représentant abrégé de son contenu .

Après cette description générale du système à modéliser, nous détaillons les différents axes de la modélisation d'un système multi agents.

5.2.1 Environnement

L'environnement est le système de tagging qui consiste d'un ensemble de ressources et un ensemble de tags à attribuer à ces ressources, ces assignations sont effectuées par les utilisateurs du système (agents) comme le montre FIG. 5.2.

Il faut rappeler que les tags sont des mots libres générés aléatoirement à partir d'un ensemble de consonnes et de voyelles pour qu'ils ressemblent aux tags réels et que les ressources sont réduites à un mot clé généré aléatoirement qui représente une méta donnée sur le contenu de la ressource comme celles qui se trouvent dans l'entête (header) dans une page web (des méta données sur son contenu).

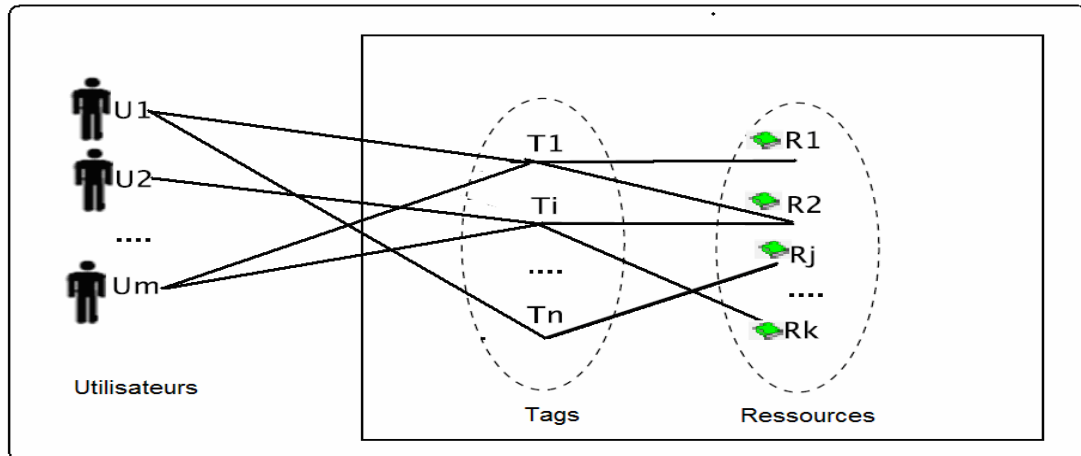


FIG. 5.2 - l'environnement d'un système de tagging et ses trois parties

5.2.2 Organisation

Les systèmes de tagging font partie des logiciels sociaux où les organisations sociales sous forme des communautés d'intérêts communs peuvent émerger dans tels systèmes si nous observons le comportement des utilisateurs qui ont la tendance de se rapprocher à un intérêt particulier qui répond à ses besoins pour former des groupes virtuels.

Pour les systèmes de tagging, les organisations émergentes correspondent aux groupes formés autour d'un ensemble des ressources et de tags comme dans la figure (voir FIG. 5.3) où le système contient quatre sous communautés.

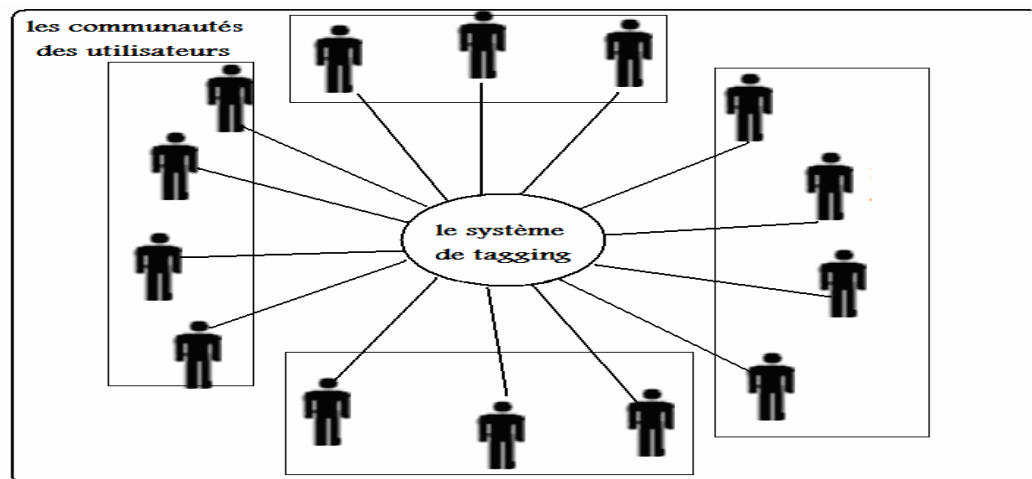


FIG. 5.3 - la communauté des utilisateurs d'un système de tagging

5.2.3 Agent

Chaque agent est capable d'exécuter des comportements simples pour faire une assignation d'un tag à une ressource (voir FIG. 5.4).

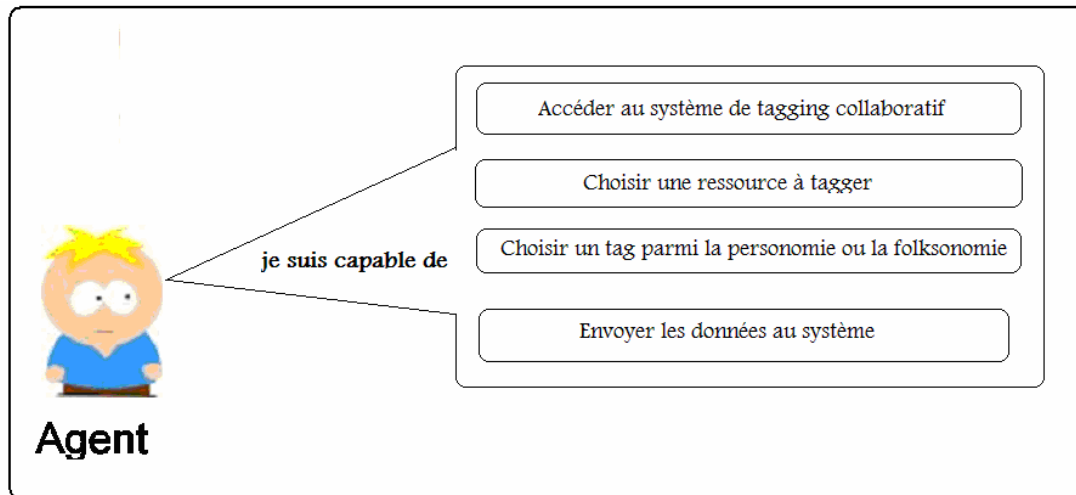


FIG. 5.4 -l'architecture d'un agent tagueur avec les comportements

À ce niveau de modélisation les agents ne possèdent pas des mémoires associatives mais toutes les actions de tagging se font dans l'environnement (où elles sont mémorisées dans la mémoire collective de tous les utilisateurs) qui fait office de mémoire partagée pour l'ensemble des utilisateurs.

5.2.4 Interaction

L'interaction des agents avec le système se fait de la manière suivante:

1. l'agent accède au système de tagging par son compte déjà créé.
2. le système donne accès pour cet utilisateur à l'ensemble de tags qu'il a utilisé dans les session précédente (la personomie), le système donne aussi accès à un ensemble de tags les plus utilisés par les autres utilisateurs.
3. l'agent choisit une nouvelle ressource ou une ressource déjà existante dans le système, si il est en train de faire l'opération de retagging.
4. l'agent a trois possibilités pour choisir un tag pour l'attribuer à la ressource choisie:
 - a. il choisit un tag parmi ses tags personnels dans ce cas nous avons une imitation individuelle.

- b. Il choisit un tag parmi les tags les plus fréquemment utilisés par la communauté du système de tagging ou parmi tous les tags du système dans ce cas nous avons une imitation sociale s'il ne trouve pas parmi les tags les plus utilisés.
 - c. il crée un nouveau tag s'il juge que la ressource appartient à une nouvelle catégorie qui n'existe pas dans sa personomie ou dans le système.
5. il envoie l'identificateur de la ressource et le tag qu'il a choisi au système sous la forme d'une entrée « U_i, T_i, R_i, t_i »
- Où U_i, T_i, R_i, t_i représentent respectivement l'utilisateur, le tag, la ressource, un marquage de temps.

Parmi les étapes précédentes, l'étape 4 est la plus importante c'est à dire le choix du tag approprié à une ressource donnée, pour cela nous allons expliquer en détail comment se fait le choix.

L'agent décide de manière autonome de faire une imitation individuelle ou une imitation sociale et il procède de la manière suivante:

1. Dans le premier cas (l'imitation individuelle), l'agent cherche parmi les « m » ressources personnelles la ressource la plus proche à la ressource en question « r_i », en mesurant sa mesure de similarité sémantique par rapport toutes les ressources personnelles.

En prenant la ressource « r_{max} » qui a la similarité maximale et s'elle est supérieur à une tendance calculée proportionnellement à un seuil prédéfini « S ». L'agent cherche le tag « t_{max} » attribué à la ressource « r_{max} » et attribuer ce même tag à la ressource « r_i » c-à-d les deux ressources « r_i » et « r_{max} » sont assignées par le même tag « t_{max} ». Formellement ceci est exprimé par

$$\text{Assigner}(r_i) = t_{max} \Leftrightarrow \exists r_{max} / \text{Assigner}(r_{max}) = t_{max} \text{ et}$$

$$\text{Sim}(r_i, r_{max}) = \text{Max}_{j=1,m}(\text{Sim}(r_i, r_j)) \text{ et}$$

$$\text{Sim}(r_i, r_{max}) > S$$

Où m est le nombre des ressources personnelles,

$\text{Assigner}(r_i) = t_j$ est une fonction de l'ensemble des ressources Res vers l'ensemble des tags T qui permet d'assigner une ressources « r_i » par un tag « t_j ».

Et $\text{Sim}(r_i, r_j) = s$ est une fonction de mesure de similarité sémantique entre les ressources qui part de l'ensemble $(\text{Res} \times \text{Res})$ vers l'intervalle $[0,1]$ et qui est définie comme suit :

Soit $r_i = \{\text{mot_clé}_i\} \in \text{Res}$ et $r_j = \{\text{mot_clé}_j\} \in \text{Res}$ deux ressources représentées par leurs mots clés, nous donnons la similarité entre elles par

$$\text{Sim}(r_i, r_j) = \frac{2 * \text{Communs_préfix}(\text{mot_clé}_i, \text{mot_clé}_j)}{\text{Long}(\text{mot_clé}_i) + \text{Long}(\text{mot_clé}_j)}$$

La fonction $\text{Communs_préfix}(\text{mot_clé}_i, \text{mot_clé}_j)$ fait retourner le nombre de caractères communs en préfixe entre les deux chaînes de caractères mot_clé_i et mot_clé_j . (Par exemple $\text{Communs_préfix}(\text{produit}, \text{production})=5$).

La fonction $\text{Long}(\text{mot_clé}_i)$ retourne la longueur ou la taille de la chaîne de caractères mot_clé_i (resp. la longueur de mot_clé_j).

Si toutes les ressources personnelles ayant des mesures de similarités inférieures à la tendance calculée par rapport au seuil « S », dans ce cas là, l'agent crée un nouveau tag « t_j » à la ressource « r_i ».

Assigner(r_i) = $t_j / t_j = \text{Nouveau_tag}()$;

Où $\text{Nouveau_tag}()$ est une fonction qui permet de créer un nouveau tag.

2. Dans le cas de l'imitation sociale, il cherche parmi les ressources du système en sélectionnant que les « n » ressources qui sont assignées par un tag parmi l'ensemble de tags les plus utilisés qui sont affichés pour lui, la ressource la plus proche à la ressource en question « r_i », en mesurant sa mesure de similarité sémantique par rapport toutes les ressources sélectionnées.

En prenant la ressource « r_{max} » qui a la similarité maximale et s'elle est supérieur à une tendance calculée proportionnellement à un seuil prédéfini « S ».

L'agent cherche le tag « t_{max} » attribué à la ressource « r_{max} » et attribuer ce même tag à la ressource « r_i » c à d les deux ressources « r_i » et « r_{max} » sont assignées le même tag « t_{max} ».

Formellement ceci est exprimé par

Assigner (r_i)= t_{max} $\Leftrightarrow \exists r_{max}$ / Assigner (r_{max})= t_{max} et
 $Sim(r_i, r_{max}) = \text{Max}_{j=1,n}(Sim(r_i, r_j))$ et
 $Sim(r_i, r_{max}) > S$

Où n est le nombre des ressources assignées par l'ensemble des tags les plus utilisés

Assigner (r_i)= t_j est une fonction de l'ensemble des ressources Res vers l'ensemble des tags T qui permet d'assigner une ressources « r_i » par un tag « t_j ».

Et $Sim(r_i, r_j) = s$ une fonction de mesure de similarité sémantique entre les ressources déjà définie dans le cas précédent de l'imitation individuelle

S'il ne trouve pas une ressource parmi les « n » ressources du système, l'agent fait élargir l'ensemble de ressources limité aux tags les plus utilisés, à toutes les tags de système (toutes les ressources de système) en procédant exactement comme la procédure citée ci-dessus.

Et si même toutes les ressources du système ont des distances inférieures à la tendance calculée par rapport au seuil « S », dans ce cas là, l'agent crée un nouveau tag « t_j » à la ressource « r_i ».

Assigner (r_i)= t_j / $t_j = \text{Nouveau_tag}()$;

Où $\text{Nouveau_tag}()$ est une fonction qui déjà mentionnée pour créer un nouveau tag.

La décision de la proximité d'une ressource par rapport à une autre se fait par l'utilisateur du système, sur la base du contenu des ressources résumé en ses mots clé.

Ci après, nous présentons l'algorithme des interaction expliquées ci-dessus

5.2.5 Algorithme d'interaction

Mode_tagging = générer aléatoirement un mode de tagging (); // tagging ou retagging.

Si (Mode_imitation == opération_tagging)

Alors

Ressource = générer aléatoirement une ressource ();

Sinon

Ressource = choisir une ressource du système ();

Fin si

Mode_imitation = générer aléatoirement un mode d'imitation (); // sociale ou indiv.

Si (Mode_imitation == Imitation_individuelle)

```
Alors
Pour i=1 à m // m est la taille de la personomie
  Faire
    Dis_sem= calculer la distance de similarité sémantique (Ressource, Ressourcei) ;

Si (Dis_sem > Dis_sem_max)
  Alors
    Ressources_proche= Ressourcei ;
    Dis_sem_max= Dis_sem ;
  Fin si
Fin pour

Sinon // imitation sociale
// n est la taille de la collection des ressources assignées par les tags les plus utilisés
Pour i=1 à n
  Faire
    Dis_sem= calculer la distance de similarité sémantique (Ressource, Ressourcei) ;
    Si (Dis_sem > Dis_sem_max)
      Alors
        Ressources_proche= Ressourcei ;
        Dis_sem_max= Dis_sem ;
      Fin si
    Fin pour
  Fin si
Tendance = calculer une probabilité proportionnelle à (seuil);
Si (Dis_sem_max < Tendance)
  Alors Tag= générer un nouveau tag () ;
  Sinon Tag= trouver le tag assigné à (Ressources_proche) ;
Fin si
Assigner (Tag, Ressource) ;
```

Remarque

Pour le traitement exécuté lorsque l'agent fait une imitation sociale étendue sur toutes les ressources du système, il exécute le bloc de l'imitation sociale mais en initialisant n par le nombre total des ressources du système ($n=Nb_total_ressources_système$), au lieu de sélectionner que les ressources assignées par les tags les plus fréquemment utilisés.

5.2.6 Adaptation

Dans cette phase de modélisation, l'adaptation se fait au niveau du système du tagging qui enregistre l'entrée « U, T, R, t » dans sa base de données et en effectuant une mise à jour des fréquences d'utilisation du tag T en l'incrémentant de un. Ces fréquences vont servir dans le tri des tags les plus utilisés dans les prochaines sessions du même utilisateur ou d'autres utilisateurs du système.

5.2.7 Résultats et discussion

Dans cette partie, nous allons présenter quelques résultats obtenus par un démonstrateur conçu sur la base de modèle décrit ci-dessus. Notre but dans cette phase est de régénérer la dynamique complexe des systèmes de tagging collaboratif. Par la suite, des exemples des scénarios simulés sont présentés en modifiant les paramètres de notre modèle comme par exemple :

- Le nombre d'agents qui accèdent le système et participent dans l'opération de tagging.
- Le seuil de mesure de la similarité entre les ressources qui permet d'assigner les ressources qui sont proches par le même tag.
- Le nombre de tags les plus utilisés présentés à un agent pour permettre une imitation sociale.
- les probabilités de faire une imitation individuelle ou sociale.
- les probabilités de faire les opérations de tagging ou retagging.

Pour les probabilités de faire un imitation individuelle ou sociale ainsi pour le choix pour l'opération de tagging ou retagging, nous utilisons dans les exemple cités ci-dessous des probabilités inspirées du travail de Santos-Neto et Condon (Santos-Neto & Condon 2009) qui sont:

50% de temps, les utilisateurs font une imitation individuelle

50% de temps, les utilisateurs font une imitation sociale et que

80% de temps, les utilisateurs introduisent des nouvelles ressources

20% de temps, les utilisateurs choisissent des ressources existant dans le système pour faire un retagging.

Ce choix est fait pour avoir une dynamique de système de tagging la plus proche à celle observée dans les systèmes réels. Mais, nous pouvons utiliser des probabilités quelconques dans les simulations.

Pour analyser la dynamique du système, nous analysons la courbe de l'évolution du nombre de tags qui donne une image globale de sa dynamique et caractérise les systèmes de tagging c à d une loi de puissance avec un exposant plus petit qu'un.

Nous analysons aussi la courbe de l'évolution de nombre de ressources dans le système.

Nous utilisons une mesure connue dans la théorie de l'information, l'entropie, pour tracer une courbe qui nous permet de voir la quantité de l'information dans le système.

Nous définissons l'entropie E d'un système de tagging collaboratif S constitué de T tag comme suit

$$E(S) = - \sum_{i=1}^T p(t_i) \cdot \log_2 p(t_i)$$

Où T est le nombre des tags qui constitue la folksonomie

Et $p(t_i)$ est la probabilité d'un tag t_i donnée par

$$p(t_i) = \frac{\text{nombre_utilisation}(t_i)}{\sum_{j=1}^T \text{nombre_utilisation}(t_j)}$$

Nous utilisons un logarithme à la base 2 pour avoir la quantité d'information d'un système de tagging en bit et qui représente le nombre de bits nécessaires à la codification de T tags de la folksonomie.

5.2.7.1 Exemple 1

Dans les premières simulations, nous utilisons un nombre d'agents très limité comme par exemple $N=10$ agents et un seuil de mesure de similarité $S=0.4$.

Les figures ci-dessous représentent l'évolution du nombre de tags dans le temps et l'évolution du nombre de ressources ainsi que l'évolution de la quantité d'information (entropie) dans le système.

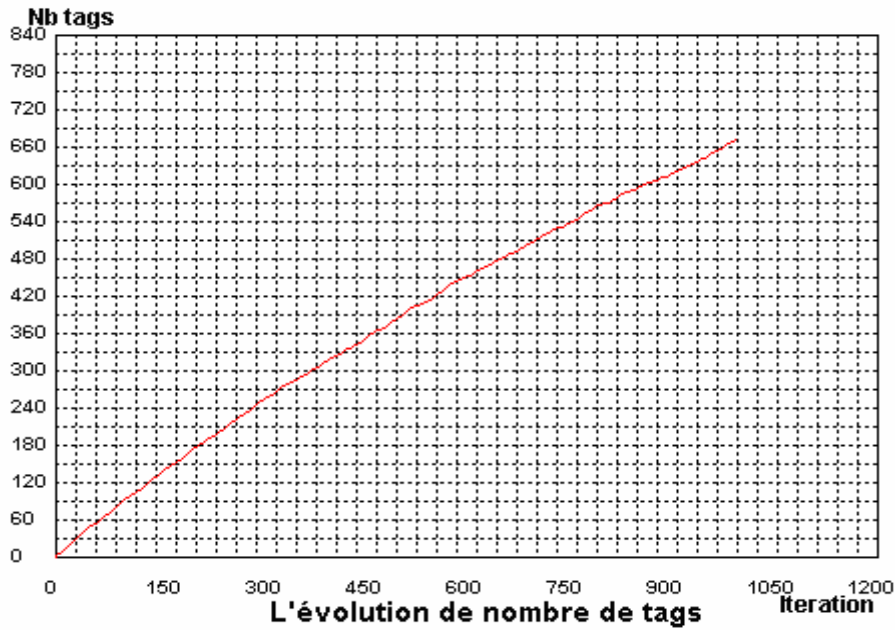


FIG. 5.5 - Evolution de nombre de tags

D'après les figures de cet exemple, on observe la convergence du nombre de tags vers 671 tags après 1019 itérations et c'est la même dynamique observée dans un système de tagging (une loi en puissance).

Ceci est dû au principe de l'imitation sociale et individuelle utilisé dans notre modèle qui se base sur la réutilisation des tags du système qui permet la convergence et l'émergence de la folksonomie constituée des 671 tags pour catégoriser 844 ressources dans cet exemple.

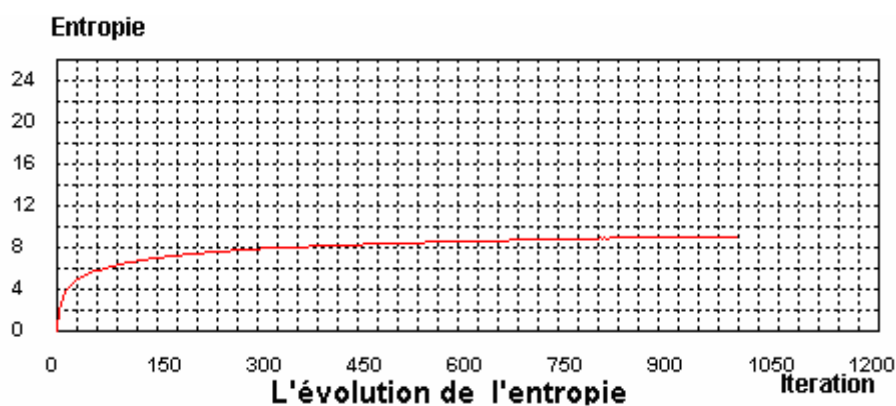


FIG. 5.6 - Evolution de l'entropie

Dans la courbe de l'entropie, on remarque une stabilisation ou une convergence de la quantité de l'information vers la valeur de 9.5 bits qui représente le nombre de bits nécessaire pour coder les 671 tags.

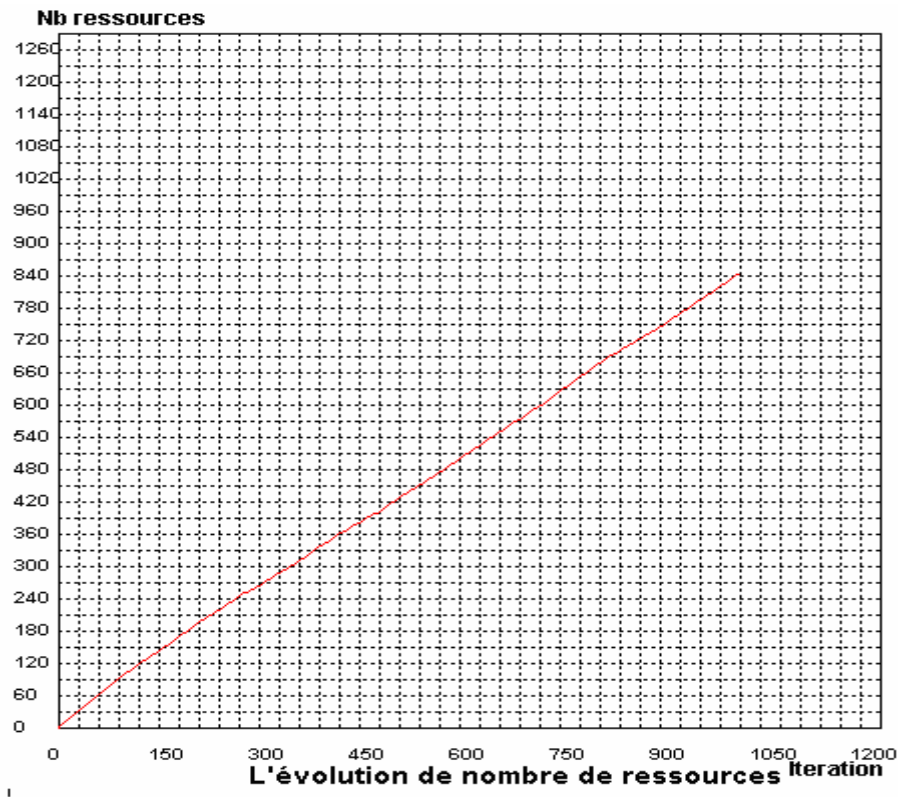


FIG. 5.7 - l'évolution de nombre de ressources

Pour le nombre de ressource, on observe une croissance rapide dans l'évolution du nombre de ressources dans le système par rapport au nombre de tags qui se converge. Ceci est expliqué par le fait que le plus souvent les agents introduisent des nouvelles ressources au système avec une petite possibilité de choisir une ressource existant dans le système, dans les 20% de situations les agents choisissent une ressource parmi les ressource du système pour faire ce qu'on appelle un retagging.

5.2.7.2 Exemple 2

Dans cet exemple, nous utilisons un nombre d'agents $N=500$ agents et un seuil de mesure de similarité $S=0.4$.

Les figures ci-dessous représentent pour cet exemple l'évolution du nombre de tags dans le temps et l'évolution du nombre de ressources ainsi l'évolution de l'entropie dans le système.

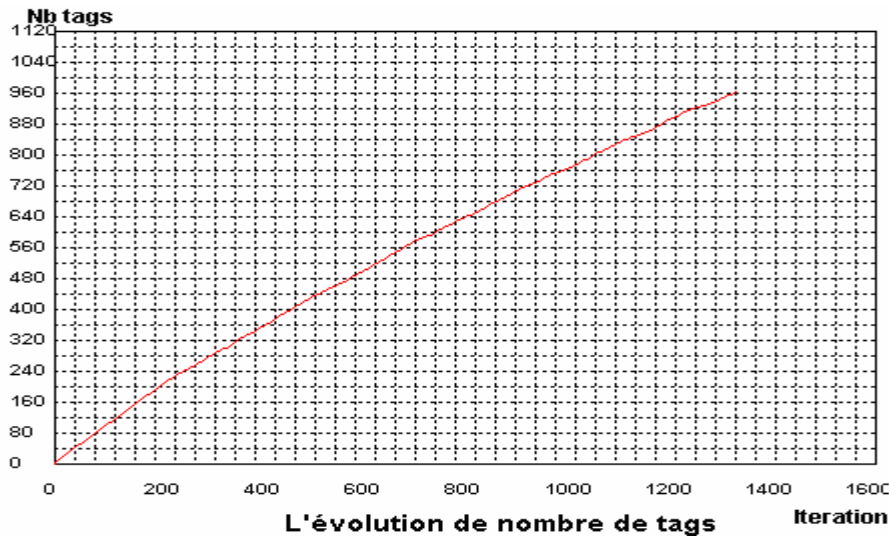


FIG. 5.8 - Evolution de nombre de tags

Dans la figure ci-dessus, il y a une convergence du nombre de tags vers 962 tags après 1363 itérations et c'est la même dynamique observée dans l'exemple précédent. La folksonomie émergente dans cet exemple est constituée des 962 tags pour catégoriser 1104 ressources.

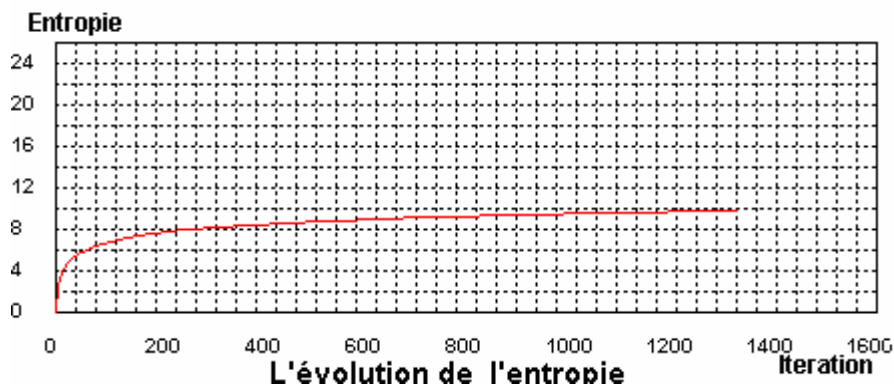


FIG. 5.9 - Evolution de l'entropie

Dans la courbe de l'entropie, on remarque aussi une stabilisation de la quantité de l'information. La convergence est vers la valeur de 10 bits qui permet de coder les 962 tags.

Pour le nombre de ressource, on observe toujours une croissance rapide dans l'évolution du nombre de ressources dans le système.

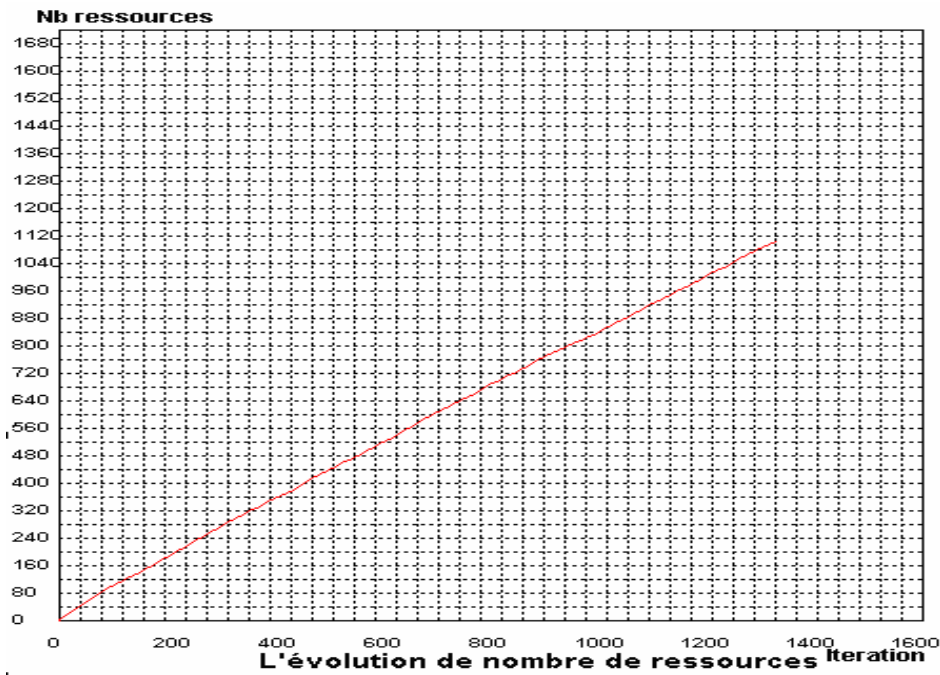


FIG. 5.10 - l'évolution de nombre de ressources

5.2.7.3 Exemple 3

Dans cet exemple, nous utilisons un nombre d'agents $N=10$ agents.

Nous avons fixé le seuil de la similarité dans les deux exemples précédents par $S=0.4$

Pour permettre une flexibilité dans la réutilisation des tags nous fixons sa valeur dans cet exemple par $S=0.1$ pour observer son effet sur la dynamique du système.

Les figures ci-dessous représentent pour cet exemple l'évolution du nombre de tags, l'évolution du nombre de ressources et l'évolution de l'entropie dans le système.

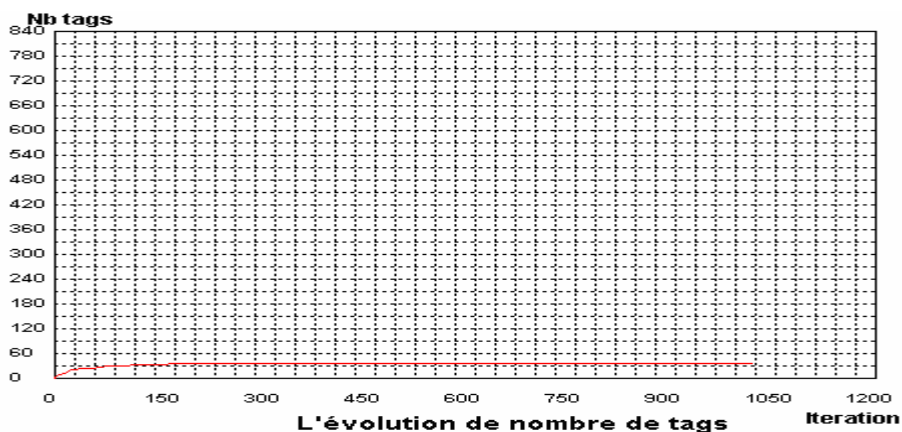


FIG. 5.11 - Evolution de nombre de tags

Avec cette nouvelle valeur du seuil, on remarque une convergence très rapide du système vers l'utilisation d'un nombre de tags très limité (34) par rapport aux autres exemples (exemple 1 et 2) après 1046 itérations.

La folksonomie émergente est constituée des 34 tags pour catégoriser toutes les 853 ressources.

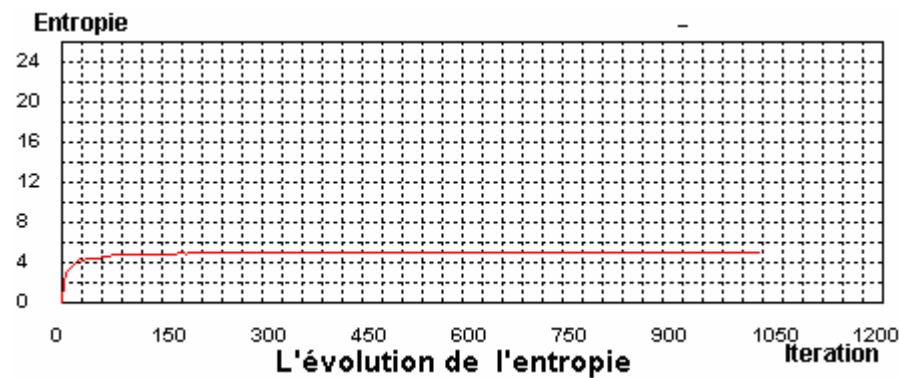


FIG. 5.12 - Evolution de l'entropie

Pour l'entropie, on remarque dans la courbe aussi une stabilisation de la quantité de l'information vers 5.1 pour coder les 34 tags.

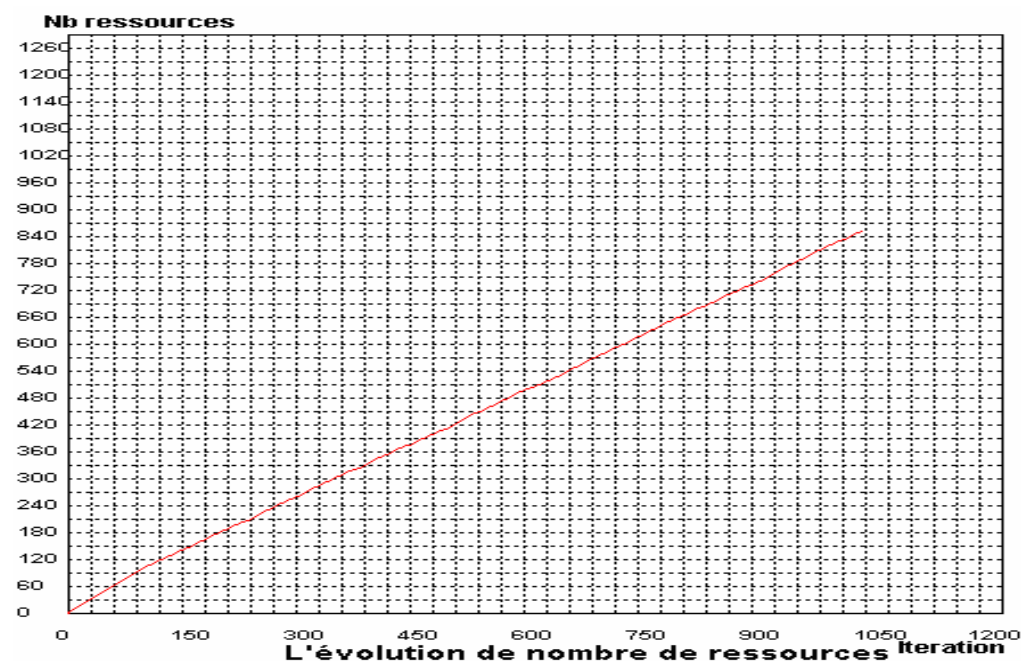


FIG. 5.13 - l'évolution de nombre de ressources

Pour les ressources, nous avons obtenu une courbe qui a la même allure que celle de l'exemple 2 (voir FIG. 5.10).

5.2.7.4 Résumé

D'après les exemples cités ci-dessus, nous remarquons l'émergence des folksonomies dans le système de tagging collaboratif et la génération de la dynamique de tels systèmes.

La qualité de la folksonomie en terme de nombre de tags qui la constitue dépend du seuil de la mesure de la similarité inter-ressources qui définit en quelque sorte le degré de liberté pour l'agent de considérer deux ressources comme ayant le même contenu.

5.2.8 Conclusion

D'après les résultats présentés ci-dessus, une modélisation des systèmes complexes à base des systèmes multi-agents adaptée pour les systèmes de tagging collaboratif où l'auto-organisation joue un rôle dans la dynamique du système et une approche d'imitation individuelle et sociale peuvent générer la dynamique de tels systèmes que nous observons sur l'évolution de nombre des tags dans le système qui suit une loi en puissance avec un exposant plus petit que un.

Ce qui nous permet de dire que la dynamique observée dans les systèmes de tagging et générée par les interactions fortes entre les utilisateurs et le système par le processus de tagging et retagging collaboratif et le mécanisme sélectionniste exprimé par le principe d'imitation sociale et individuelle.

6.1 Introduction

Après avoir modélisé la dynamique et l'évolution d'un système de tagging collaboratif à l'aide d'une modélisation à base des systèmes multi-agents et voir l'effet du mécanisme de l'imitation individuelle et sociale que les agents utilisent lorsque ils font une assignation d'un tag à une ressource donnée qui peut expliquer l'émergence des folksonomies, nous nous intéressons dans ce modèle à l'aspect cognitif et en particulier l'aspect linguistique du processus de tagging collaboratif au niveau des agents.

Le but de ce modèle est de réaliser une étape intermédiaire vers notre objectif final de ce travail qui concerne la possibilité de l'émergence d'un langage de tagging partagé par la communauté des utilisateurs du système de tagging collaboratif.

Pour aborder la problématique générale et atteindre notre objectif final (l'émergence d'un langage de tagging), nous commençons dans ce modèle par un sous objectif pour répondre à une sous problématique que nous décrivons comme suit:

Est-ce que dans un système de tagging collaboratif, la communauté des utilisateurs de ce système (les agents) peut partager un langage de tagging à structure lexicale ? Autrement dit, la possibilité de l'émergence d'un langage de tagging proche une lexique naturelle par un processus de mono-tagging, ce langage de tagging lexicale se caractérise par des correspondances de type « tag, ressources », où chaque tag représente une catégorie d'objets (ressources).

Dans cette étape de modélisation, nous faisons en plus des hypothèses que nous citons ci-après une autre hypothèse concernant le processus de tagging collaboratif que nous supposons qu'il est mono-tagging, c-à-d chaque agent de notre système multi-agents assigne une ressource par un seule tag pour la mettre dans une catégorie représentée par ce tag, ce principe va permettre l'émergence d'un ensemble des catégories partagées par les agents.

Pour étudier l'aspect cognitif (linguistique) du processus de tagging, nous enrichissons le modèle précédent comme suit :

-Nous dotons les agents par des systèmes linguistiques sous forme des mémoires associatives (Kaplan 2000), ces mémoires ont pour objectif de stocker des connaissances linguistiques concernant les assignations des tags aux ressources (les couples tag / ressource).

- Un mécanisme d'apprentissage par renforcement (Kaplan 2000) pour que les agents puissent mettre à jour leurs mémoires associatives afin d'insérer de nouvelles connaissances linguistiques ou corriger une connaissance déjà existante dans la mémoire qu'il est jugé qu'elle est fautive dans un scénario d'apprentissage.
- Les agents interagissent avec le système de tagging par des scénarios des jeux de langage (Kaplan 2000) .

Ce modèle se base comme le modèle précédent sur le travail de Santos-Neto et Condon (Santos-Neto & Condon 2009) et dans son coté linguistique sur les modèles de l'émergence du langage (voir chap.2) et principalement sur les modèles proposés dans les travaux de Kaplan (Kaplan 2000).

En utilisant toujours l'approche voyelle pour modéliser notre système multi-agents, nous détaillons ses différents axes (environnement, agent, organisation, interaction)

6.2 Environnement

L'environnement reste le même comme celui du modèle précédent c-à-d il consiste d'un ensemble de ressources et un ensemble de tags à attribuer à ces ressources (voir FIG. 6.1).

On garde la même représentation pour les tags qui sont des mots libres générés aléatoirement à partir d'un ensemble de consonnes et de voyelles ainsi que pour les ressources qui sont représentées par un mot clé généré aléatoirement (c'est une méta-donnée sur le contenu de la ressource).

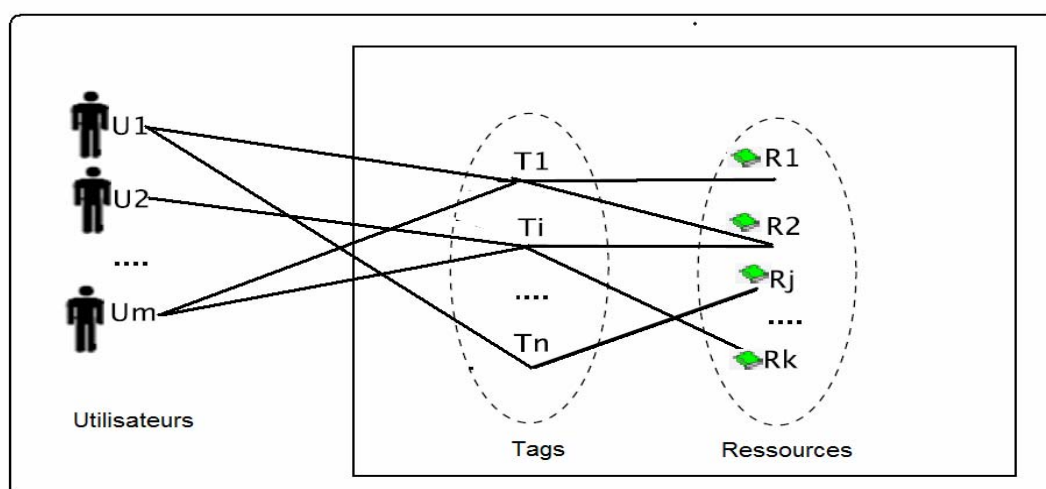


FIG. 6.1 - l'environnement et les trois parties d'un système de tagging

6.3 Organisation

Dans les systèmes de tagging, les organisations émergentes correspondent aux groupes formés autour d'un ensemble de ressources et de tags comme dans FIG. 6.2.

Ces communautés émergentes correspondent, si on s'intéresse à l'aspect linguistique, à des parties du vocabulaire du langage partagé par la communauté, chaque partie du vocabulaire globale représente un intérêt commun entre les membres de ce groupe émergent.

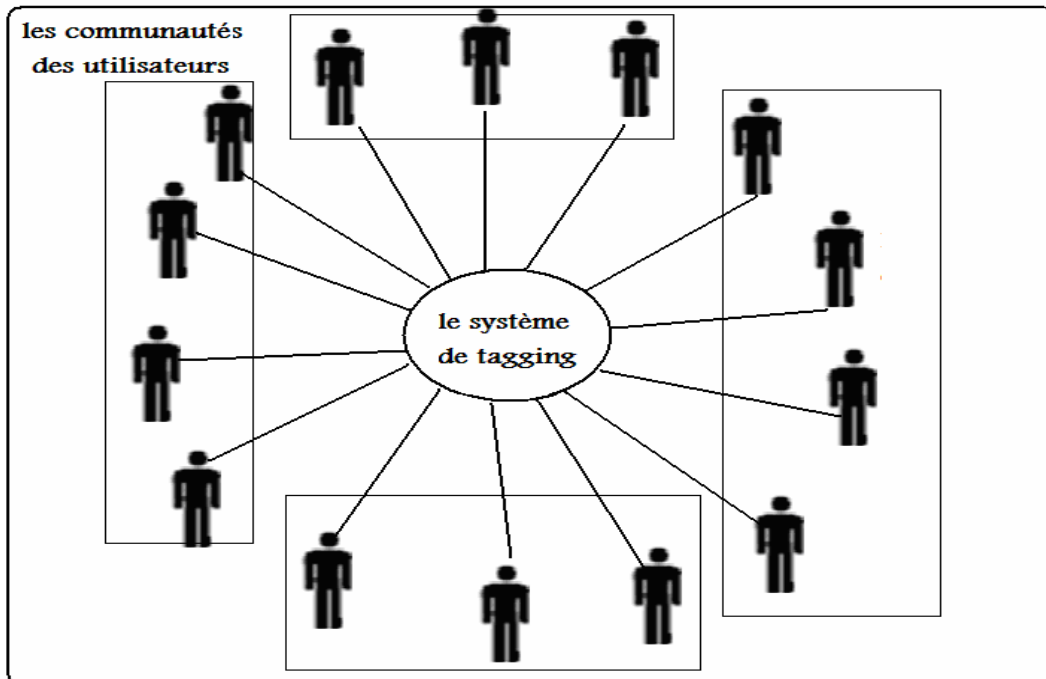


FIG. 6.2 – les sous communautés dans un système de tagging

6.4 Agent

À ce niveau de modélisation, nous enrichissons l'architecture des agents en dotant chaque agent par une mémoire associative.

Pour stocker les assignations qu'il réalise et qui permettent l'émergence d'un langage de tagging au niveau de cette mémoire comme le montre la figure FIG. 6.3.

Les agents interagissent avec le système de tagging en laissant des traces sur l'environnement sous forme des assignations stockées dans l'environnement (où elles sont mémorisées dans la mémoire collective de tous les utilisateurs).

Comme dans le modèle précédent, les agents sont capables d'exécuter des comportements simples pour faire une assignation d'un tag à une ressource et négocier ce choix avec le système et garder cette assignation en adaptant sa mémoire associative.

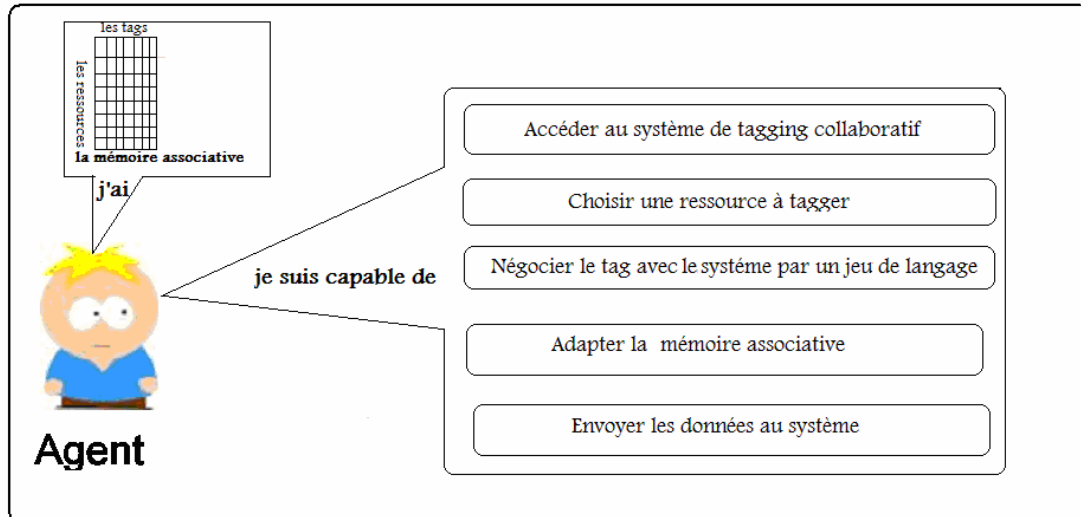


FIG. 6.3 – l’architecture d’un agent tagueur avec les comportements

6.5 Interaction

La nouveauté dans les interactions par rapport à celles du modèle précédent est dans le fait qu’elles suivent des schémas d’interaction inspirés des jeux de langage adapté aux systèmes de tagging collaboratif.

Les interactions se font entre un agent et le système (pas entre deux agents comme dans les jeux de langage standards) où l’agent (resp. le système) joue le rôle de locuteur ou interlocuteur pour négocier l’assignation d’un mot (un tag) à un sens (une ressource).

L’interaction des agents avec le système se fait de la manière suivante:

- 1.l’agent accède au système de tagging collaboratif.
- 2.le système donne accès pour cet utilisateur à l’ensemble de tags qu’il a utilisé dans les session précédente (la personomie), le système donne aussi accès à un ensemble de tags les plus utilisés par les autres utilisateurs.
- 3.l’agent choisit une nouvelle ressource ou une ressource déjà existante dans le système, s’ il est en train de faire l’opération de retagging.
- 4.l’agent a deux possibilités pour choisir un tag pour l’attribuer à la ressource choisie:

- a- il joue le rôle d’un locuteur et dans ce cas le système joue le rôle de l’interlocuteur et nous sommes dans une situation d’imitation individuelle,

autrement dit, le système apprend de l'agent en imitant les connaissances linguistiques individuelles de l'agent.

b- il joue le rôle d'un interlocuteur et dans ce cas le système joue le rôle du locuteur et nous sommes dans une situation d'imitation sociale, autrement dit l'agent apprend du système en imitant les connaissances linguistiques globales du système qui se trouvent dans la mémoire collective au niveau du système.

Dans les deux situations, il aura une négociation entre le locuteur ou l'interlocuteur selon un schéma de jeux de langage pour décider l'assignation appropriée à la ressource en question.

5. il adapte sa mémoire associative en insérant des nouvelles assignations ou en cortégeant une assignation par le renforcement.

6. il envoie l'identificateur de la ressource et le tag qu'il a choisi au système sous la forme d'une entrée « U_i, T_i, R_i, t_i »

Où U_i, T_i, R_i, t_i représentent respectivement l'utilisateur, le tag, la ressource, un marquage de temps.

L'étape 4 est toujours la plus importante (choix du tag approprié à une ressource donnée), pour cela nous allons expliquer comment se fait l'interaction.

L'agent décide de manière autonome de faire une imitation individuelle en jouant le rôle du locuteur ou une imitation sociale en jouant le rôle de l'interlocuteur et il procède de la manière suivante:

1. Dans le premier cas (le rôle du locuteur), l'étape est presque comme l'étape de l'imitation individuel du modèle précédent. L'agent cherche parmi les « m » ressources personnelles la ressource la plus proche à la ressource en question « r_i », en mesurant sa mesure de similarité sémantique par rapport à toutes les ressources personnelles. En prenant la ressource « r_{max} » qui a la similarité maximale et s'elle est supérieur à une tendance calculée proportionnellement à un seuil prédéfini « S ».

L'agent cherche le tag « t_{max} » attribué à la ressource « r_{max} » et attribuer ce même tag à la ressource « r_i », c-à-d les deux ressources « r_i » et « r_{max} » sont assignées par le même tag « t_{max} ». Sinon l'agent crée un nouveau tag « t_j » à la ressource « r_i ».

Formellement ceci est exprimé par

Assigner (r_i) = t_{max} \Leftrightarrow $\exists r_{max} /$ Assigner (r_{max}) = t_{max} et

Sim (r_i , r_{max}) = $\text{Max}_{j=1,m}(\text{Sim}(r_i, r_j))$ et

Sim (r_i , r_{max}) > S

Où m est le nombre des ressources personnelles,

Assigner (r_i) = t_j est une fonction de l'ensemble des ressources Res vers l'ensemble des tags T qui permet d'assigner une ressources « r_i » par un tag « t_j ».

Et Sim (r_i , r_j) = s une fonction de mesure de similarité sémantique entre les ressources qui part de l'ensemble (Res x Res) vers l'intervalle [0,1] et qui est définie comme suit :

Soient $r_i = \{\text{mot_clé}_i\} \in \text{Res}$ et $r_j = \{\text{mot_clé}_j\} \in \text{Res}$ deux ressources représentées par leurs mots clés, nous donnons la similarité entre elles par

$$\text{Sim}(r_i, r_j) = \frac{2 * \text{Communs_préfix}(\text{mot_clé}_i, \text{mot_clé}_j)}{\text{Long}(\text{mot_clé}_i) + \text{Long}(\text{mot_clé}_j)}$$

2. Dans le deuxième cas (le rôle d'interlocuteur), le processus se fait par une négociation entre le locuteur (le système) et l'interlocuteur (l'agent) en passant par deux étapes intermédiaires en commençant par le locuteur :

a- le système (le locuteur) cherche parmi ses ressources pour attribuer un tag à la ressource en question en se basant sur les ressources du système et les tags du système. En procédant comme suit :

Le système cherche parmi ses ressources en ne sélectionnant que les « n » ressources qui sont assignées par un tag parmi l'ensemble de tags les plus utilisés qui sont affichés pour lui, la ressource la plus proche à la ressource en question « r_i », en mesurant sa mesure de similarité sémantique par rapport à toutes les ressources sélectionnées.

En prenant la ressource « r_{max} » qui a la similarité maximale et s'elle est supérieur à une tendance calculée proportionnellement à un seuil prédéfini « S ».

L'agent cherche le tag « t_{max} » attribué à la ressource « r_{max} » et attribue ce même tag à la ressource « r_i » c-à-d les deux ressources « r_i » et « r_{max} » sont assignées le même tag « t_{max} ».

Formellement ceci est exprimé par

$$\text{Assigner}(r_i) = t_{\max} \Leftrightarrow \exists r_{\max} / \text{Assigner}(r_{\max}) = t_{\max} \text{ et}$$
$$\text{Sim}(r_i, r_{\max}) = \text{Max}_{j=1,n}(\text{Sim}(r_i, r_j)) \text{ et}$$
$$\text{Sim}(r_i, r_{\max}) > S$$

S' il ne trouve pas une ressource parmi les « n » ressources du système, l'agent élargit l'ensemble de ressources limité aux tags les plus utilisés à tous les tags de système (toutes les ressources de système) en procédant exactement comme la procédure citée ci-dessus.

Et si même avec toutes les ressources du système, l'agent n'arrive pas à trouver une ressource la plus proche, l'agent crée un nouveau tag « t_j » à la ressource « r_i ».

Le système décide qu'il assigne le tag « t_{sys} » à la ressource « r_i » ($t_{\text{sys}} = t_{\max}$ ou $t_{\text{sys}} = t_j$).

b- L'agent (l'interlocuteur) fait une imitation individuelle pour attribuer un tag à la ressource en question en se basant sur ses ressources et les tags personnels en procédant exactement comme l'étape précédente, il décide qu'il assigne le tag « t_{indiv} » à la ressource « r_i ».

A la fin de ces deux étapes l'agent négocie avec le système le tag approprié à la ressource :

Si « t_{indiv} » et « t_{sys} » sont les mêmes, nous avons un succès et il aura une adaptation.

Sinon, nous avons un échec et il aura une autre adaptation.

6.6 Adaptation

L'adaptation se fait au niveau des mémoires associatives des agents en renforçant ou diminuant les scores des assignations « tag, ressources » dans les deux scénarios d'interaction (imitation individuelle ou sociale)

a. pour le premier cas où l'agent joue le rôle du locuteur, l'adaptation se fait de la manière suivante : L'agent renforce le score du couple « t_{indiv}, r_i » dans la mémoire par une unité.

b. pour le deuxième cas où l'agent joue le rôle de l'interlocuteur l'adaptation dépend du résultat de jeux de langage (succès ou échec)

1- dans le cas de succès (« t_{indiv} » et « t_{sys} » sont les mêmes), l'agent renforce le score du couple « t_{sys}, r_i » dans la mémoire par une unité.

2- dans le cas de échec (« t_{indiv} » et « t_{sys} » sont différents), l'agent renforce le score du couple « t_{sys}, r_i » dans la mémoire par une unité et il diminue le score du couple « t_{indiv}, r_i » par une unité.

Au niveau du système du tagging, l'adaptation se fait par l'enregistrement de l'entrée « U, T, R, t » dans sa base de données et en effectuant une mise à jour des fréquences d'utilisation du tag T en l'incrémentant de un. Ces fréquences vont servir dans le tri des tags les plus utilisés dans les prochaines sessions du même utilisateur ou d'autres utilisateurs du système.

6.7 Algorithme d'interaction mono-tagging et d'adaptation (jeu de langage)

Mode_tagging=générer aléatoirement un mode de tagging (); // tagging ou retagging.

Si (Mode_tagging == opération _tagging)

Alors

Ressource = générer aléatoirement une ressource ();

Sinon

Ressource = choisir une ressource du système (); //opération de retagging

Fin si

Mode_imitation = générer aléatoirement un mode d'imitation (); // sociale ou indiv.

Si (Mode_imitation == imitation_individuelle) // le rôle de locuteur

Alors

Tag_indiv= faire une imitation individuelle (); // voir l'algorithme du modèle 1

Adapter la mémoire associative (Tag_indiv, Ressource, Renforcement);

Sinon // le rôle de interlocuteur (imitation sociale)

Tag_sys = faire une imitation sociale (); // voir l'algorithme du modèle 1

Tag_indiv= faire une imitation individuelle (); // voir l'algorithme du modèle 1

Si (Tag_indiv == Tag_sys) // c'est un succès en jeu

Alors

Adapter la mémoire associative (Tag_ sys, Ressource, Renforcement) ;

Sinon // c'est un échec en jeu

Adapter la mémoire associative (Tag_ indiv, Ressource, Diminution) ;

Adapter la mémoire associative (Tag_ sys, Ressource, Renforcement) ;

Fin si

Fin si

6.8 Résultats et discussion

Nous allons présenter quelques résultats obtenus par un démonstrateur conçu sur la base du modèle décrit ci-dessus. Notre but dans cette phase est d'étudier la possibilité de l'émergence d'un langage mono-tagging partagé par les utilisateurs du système.

Des exemples des scénarios simulés sont présentés en modifiant les paramètres de notre modèle comme :

- Le nombre d'agents qui accèdent le système et participent dans l'opération de tagging.
- Le seuil de mesure de la similarité entre les ressources qui permet d'assigner les ressources qui sont proches par le même tag.
- Le nombre de tags les plus utilisés présentés à un agent pour permettre une imitation sociale.
- les probabilités de faire une imitation individuelle (locuteur) ou sociale (interlocuteur).
- les probabilités de faire les opérations de tagging ou retagging.

Pour les probabilités de faire un imitation individuelle (avoir le rôle du locuteur) ou sociale (avoir le rôle du interlocuteur) ainsi pour le choix de l'opération de tagging ou retagging, nous utilisons aussi dans les exemple cités ci-dessous des probabilités inspirées du travail de Santos-Neto et Condon (Santos-Neto & Condon 2009) qui sont: 50% pour l'imitation individuelle, 50% pour l'imitation sociale.

80% pour le tagging et 20% pour le retagging.

Nous pouvons utiliser des probabilités quelconques dans les simulations.

Pour analyser le langage émergent, nous utilisons la matrice de codage inspiré de (Kaplan 2000) et qui nous définissons comme étant une matrice dont les colonnes sont « les mots » du langage émergent (les tags du système) et les lignes sont « les sens » codés par ces tags (les ressources du système) et dans l'intersection d'un tag « t » avec une ressource « r » le nombre d'agents qui assigne « r » par « t » devisé par

le nombre d'agents qui ont assigné la ressource « r », c-à-d la probabilité de mettre la ressource « r » dans la catégorie représentée par le tag « t ».

Le nombre d'agents qui assigne « r » par « t » est calculé à partir des mémoires associatives des tous les agents du système.

La figure FIG. 6.4 ci-dessous montre une matrice de codage prise de (Kaplan 2000) dont les W_i en colonne ,qui sont pour notre travail les tags T_i , représentent les catégories partagées par un ensemble d'agents et en lignes se trouvent les sens (les ressources pour un système de tagging) .

Matrice de codage après 10000 jeux

Referent	W1	W 2	W3	W4	W5
sens1	1.00				
sens2	1.00				
sens3	1.00				
sens4	1.00				
sens5	1.00				
sens6	1.00				
sens7	1.00				
sens8	1.00				
sens9	1.00				
sens10		0.46			0.54
sens11		0.17			0.83
sens12		0.96			0.04
sens13					1.00
sens14		0.27			0.73
sens15		0.17			0.83
sens16		0.98			0.02
sens17		1.00			
sens18				1.00	
sens19				1.00	
sens20				1.00	
sens21			0.22	0.78	
sens22				1.00	
sens23				1.00	
sens24				1.00	
sens25				1.00	
sens26			1.00		
sens27			1.00		
sens28			0.33	0.67	
sens29			1.00		
sens30			1.00		

FIG. 6.4 – Exemple de matrice de codage (Kaplan 2000)

Pour analyser la dynamique du système, nous utilisons toujours

-La courbe de l'évolution du nombre de tags qui donne une image globale de sa dynamique et caractérise les systèmes de tagging.

-La courbe de l'évolution de nombre de ressources dans le système.

-La courbe de l'évolution de l'entropie (la quantité de l'information) dans le système.

6.8.1 Exemple 1

Nous utilisons un nombre d'agents $N=2$ agents et un seuil de mesure de similarité $S=0.3$.

Les courbes suivantes représentent l'évolution du nombre de tags dans le temps et l'évolution du nombre de ressources ainsi que l'évolution de l'entropie dans le système.

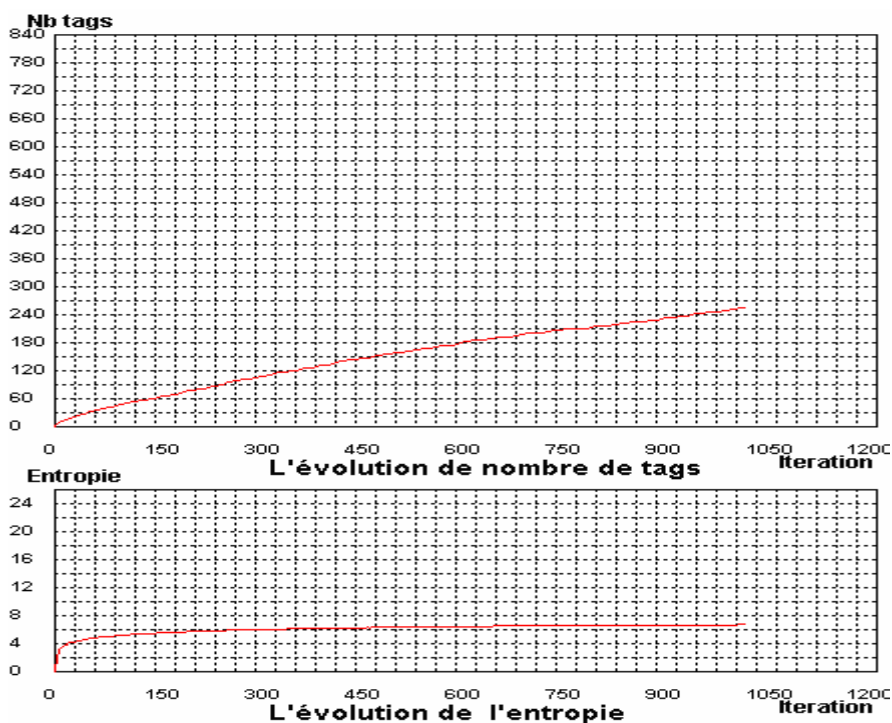


FIG. 6.5 – les courbes d'évolution du nombre de tags et l'entropie

Dans les courbes ci-dessus, nous remarquons la convergence d'une folksonomie à 255 tags pour une collection de 835 ressources pour cette simulation (1033 itérations).

Dans la courbe de l'entropie, nous avons une stabilisation de la quantité de l'information nécessaire pour ce système de codage.

Pour les ressources, nous remarquons une croissance rapide du nombre de ressources ceci est justifié comme le modèle précédent par le processus de tagging qui fait introduire des nouvelles ressources au systèmes (avec une probabilité de 80% pour cet exemple).

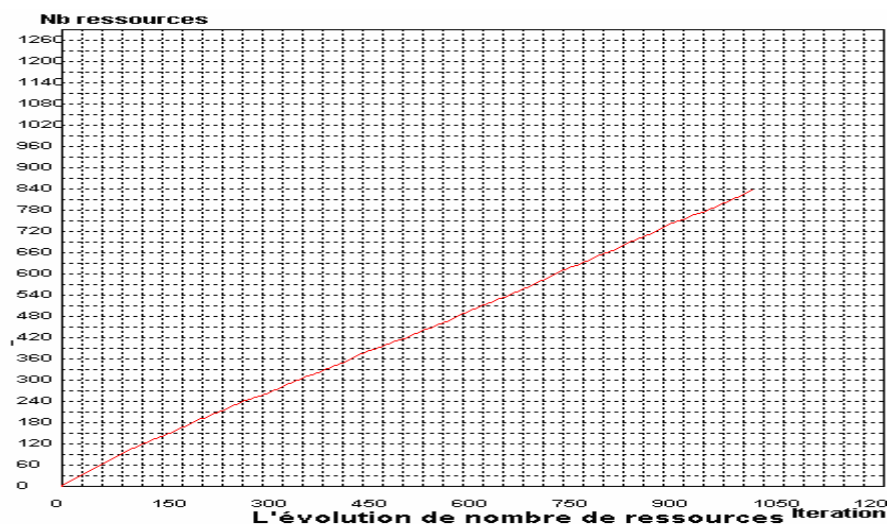


FIG. 6.6 – l'évolution du nombre de ressources

Pour voir le langage émergent, nous devons calculer la matrice de codage pour visualiser et voir les catégories émergentes dans le système de tagging collaboratif pour cet exemple.

Vu le grand nombre de ressources et de tags, nous ne pouvons pas afficher toute la matrice de codage, la figure suivante est un extrait de la matrice de codage globale en se limitant aux 10 tags les plus utilisés et un sous ensemble de ressources.

RES/TAG	gofa	bofo	yfo	jaso	hojo	rove	wose	kozo	womi	rtje
lea			1.0							
joi			1.0							
idf			1.0							
jth			1.0							
lta			1.0							
kam		1.0								
kar		1.0								
kai		1.0								
kwb		1.0								
kop		1.0								
kuj		1.0								
kso		1.0								
kms		1.0								
kps		1.0								
xom	1.0									
wom	1.0									
ku	1.0									
kw	1.0									
kw	1.0									
xp	1.0									
kl	1.0									
rb	0.5									
rv						1.0				
sp						1.0				
sp						0.5				
rs					1.0					
lca					1.0					
lu					1.0					
lga					0.5					
sd					0.5					
oc				1.0						
ocm				1.0						
opt				1.0						
omf				1.0						
oal				0.5						
nsa									1.0	
nch									1.0	
nea									1.0	
nqb									1.0	
noq									1.0	
ndv									0.5	
uik								1.0		
uim								1.0		
uie								1.0		
qsa							1.0			
qk							1.0			
grw							1.0			
qbs							1.0			
qkr							1.0			
qak							1.0			
qfj							1.0			
hkr										1.0
hes										0.5

FIG. 6.7 – un extrait de la matrice de codage

La visualisation graphique est toujours mieux qu'un tableau ou un texte pour cela nous visualisons les résultats sous forme des clusters, où chaque cluster est une catégorie émergente représentée par un tag donné et nous ne montrons que quelques clusters et que 10 ressources par cluster si le cluster contient plus que 10 ressources à cause de la contrainte de l'espace.

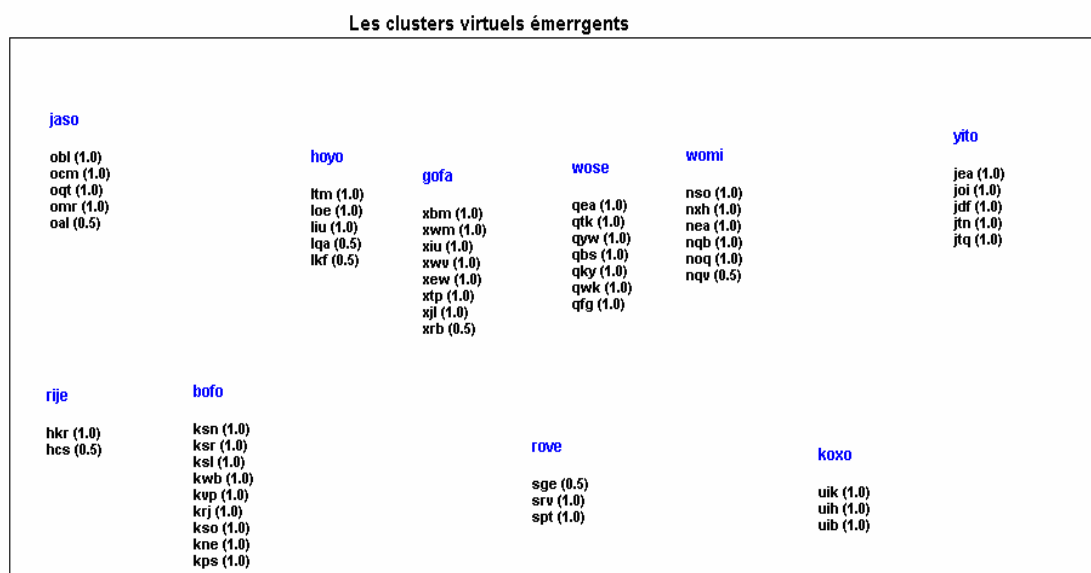


FIG. 6.8 – les catégories émergentes

Dans la matrice de codage et dans les clusters, nous voyons bien le langage émergent sous forme des catégories émergentes et les ressources qui appartiennent à chaque catégorie.

Par exemple, si nous prenons la catégorie représentée par le tag « gofa » nous remarquons que ces éléments sont des ressources qui commencent par le caractère « x », selon la fonction de mesure de similarité que nous définissons, ces ressources sont proches. Ce qui fait que pour les agents, ces ressources appartiennent à un domaine commun. La même chose pour le tag « jaso » et les ressources commençant par « o » ainsi que pour le tag « womi » et les ressources commençant par « n ».

La figure suivante représente un extrait d'une mémoire associative d'un agent de la communauté dont les valeurs positives sont des scores renforcés et les valeurs négatives sont des scores diminués par un échec dans un jeu de langage avec le système.

RES/TAG	gofa	yefe	hihi	yito	wici	diqe	deno	hoyo	soga	miti
xbm	1									
ixc		-1	2							
jea				1						
nqv					1					
wba						-1	1			
ltm								2		
dwr									1	
xwm	1									
liu								3		
vsq										1
joi				2						
edg										
dsh									-1	

FIG. 6.9 – un extrait d’une mémoire associative

La figure suivante montre les tags les plus utilisés avec leurs fréquences d’utilisation.

TAG	FRQ UTILIS
gofa	44
bofo	44
yito	42
jaso	40
hoyo	38
rove	36
wose	35
koxo	31
womi	30
rije	25
foka	24
jeya	23
faqo	22
lawo	22
hise	20
more	18
koha	18
gejo	18
riti	17
hihi	17
deno	17

FIG. 6.10 – un extrait de la folksommie (les premiers tags)

Au niveau de l’environnement (le système de tagging), nous pouvons voir, en analysant les traces des assignation dans le système, les catégories émergentes comme le montre la figure suivante où les ressources commençant par le caractère « x » sont assignées par le tag « gofa ».

Autrement dit, ce qui se trouve dans les mémoires associatives des agent est une image de ce qui se trouve dans l’environnement et vice versa.

Nous pouvons expliquer ça par la puissance de l’imitation sociale qui force l’agent à suivre la communauté dans ces conventions pour mettre une ressource dans une catégorie.

Iteration	USER	TAG	Ressource
2	AGENT_1	gofa	xbm
17	AGENT_1	gofa	xwm
38	AGENT_1	gofa	xiu
45	AGENT_1	gofa	xwv
67	AGENT_1	gofa	xew
68	AGENT_1	gofa	xtp
77	AGENT_2	gofa	xjl
84	AGENT_1	gofa	xrb
104	AGENT_1	gofa	xwk
134	AGENT_2	gofa	xwm
189	AGENT_2	gofa	xud
191	AGENT_1	gofa	xrb
212	AGENT_1	gofa	xwv
217	AGENT_2	gofa	xow
249	AGENT_2	gofa	xfm
295	AGENT_1	gofa	xtr
323	AGENT_2	gofa	xfa
345	AGENT_2	gofa	xya
357	AGENT_2	gofa	xkc
370	AGENT_1	gofa	xwv
377	AGENT_1	gofa	xjl
383	AGENT_2	gofa	xwm
425	AGENT_2	gofa	xbn
444	AGENT_2	gofa	xvp
490	AGENT_2	gofa	xhs

FIG. 6.11 – les ressources assignées par le tag « gofa »

6.8.2 Exemple 2

Nous utilisons pour cet exemple un nombre d'agents $N=10$ agents, et un seuil de mesure de similarité $S= 0.3$.

En commençant toujours par l'analyse de la dynamique du système, les figures ci-dessous représentent les courbes de l'évolution du nombre de tags dans le temps et l'évolution du nombre de ressources ainsi que l'évolution de la quantité d'information dans le système.

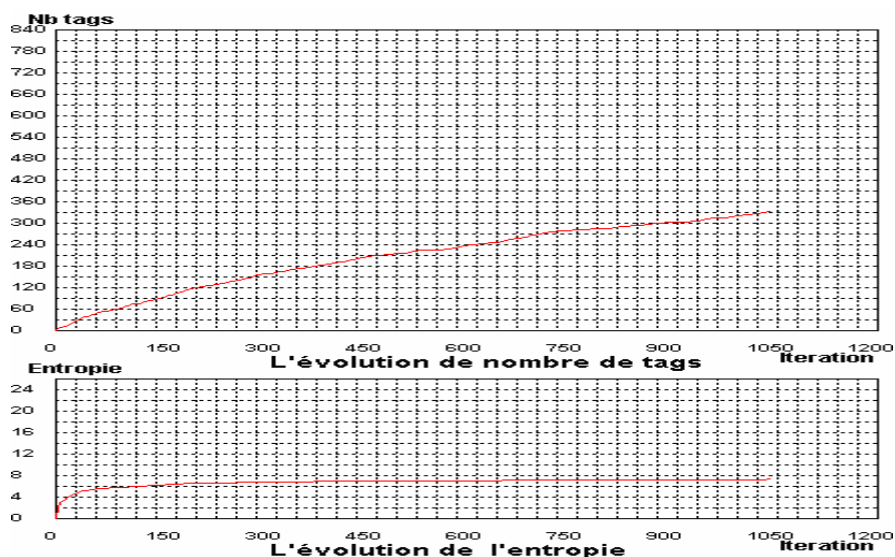


FIG. 6.12 – les courbes d'évolution du nombre de tags et l'entropie

Dans les courbes ci-dessus de l'évolution du nombre de tags et de l'entropie, nous voyons bien une convergence d'une folksonomie à 332 tags pour une collection de 884 ressources pour la simulation de cet exemple (1070 itérations).

Dans la courbe de l'entropie, nous avons une stabilisation de la quantité de l'information nécessaire pour ce système de codage.

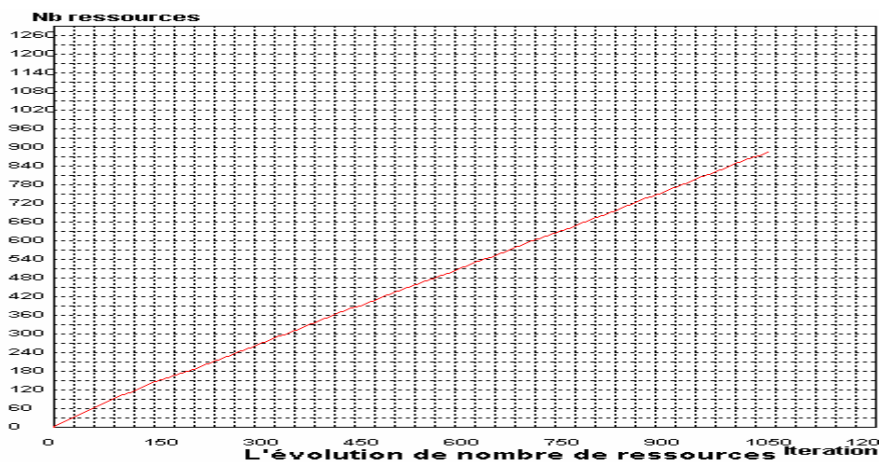


FIG. 6.13 –l'évolution du nombre de ressources

Nous calculons la matrice de codage pour analyser le langage émergent constitué des catégories émergentes dans le système de tagging collaboratif pour cet exemple, la figure suivante est un extrait de la matrice de codage globale.

RESSTAG	hoya	xeja	dige	toqo	joca	xage	koyi	losa	fame	jaki
pom			1.0							
pck			1.0							
psh			1.0							
pcr			0.66							
rni		1.0								
rnu		0.66								
rnk		0.66								
raq		0.66								
rnh		0.5								
rhj	1.0									
hbp	0.75									
hac	0.5									
hfm	0.5									
qhi	0.4									
hkq	0.4									
qbi	0.25									
fum						1.0				
fgx						0.66				
fgx						0.33				
uye					1.0					
utc					0.66					
ucd					0.5					
ykh				0.83						
yai				0.66						
ybg				0.33						
iqh									0.5	
tia								1.0		
tka								1.0		
tyn								1.0		
ths								0.66		
tqm								0.5		
tjh								0.5		
tbk								0.5		
acr							1.0			
acw							1.0			
awi							0.5			
acw							0.5			
awi							0.33			
njd										1.0
nca										0.5
nux										0.33
nti										0.33
iqh									0.5	
tia								1.0		
tka								1.0		
tyn								1.0		
ths								0.66		
tqm								0.5		
tjh								0.5		
tbk								0.5		

FIG. 6.14 – un extrait de la matrice de codage

Après avoir calculer la matrice du codage, on montre les catégories émergentes sous forme des clusters, la figure suivante représente quelques clusters émergents dans le système de tagging.

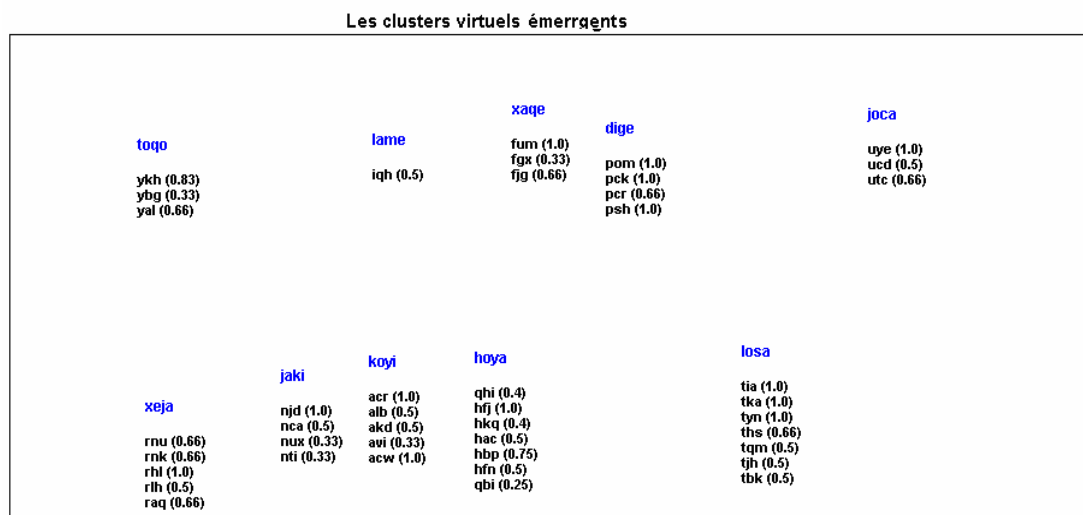


FIG. 6.15 – les catégories émergentes

Dans la matrice de codage et dans les clusters visualisés, nous avons un langage émergent sous forme des catégories émergentes et les ressources qui appartiennent à chaque catégorie.

On peut citer comme dans l'exemple précédent quelques catégories, pour la catégorie représentée par le tag « xaqe » nous remarquons que ses éléments sont des ressources qui commencent par le caractère « f », pour les agents, ces ressources appartiennent à un domaine commun. La même chose pour catégorie du tag « joca » et les ressources commençant par « u ».

La figure suivante représente un extrait d'une mémoire associative d'un agent de la communauté pour montrer l'effet de l'apprentissage par renforcement.

RES/TAG	miva	vibi	dege	navo	jeka	jeva	jisi	fita	kapa	qemi
rsl	1									
rnu	-1	1								
gdh			2							
bnl				1						
wkn					-1	1				
nca							-1	1		
vso									1	
wik						1				
dja										1
sbh										
gix										
riy	-1									

FIG. 6.16 – un extrait d'une mémoire associative

Dans cette mémoire, la ressource « gdh » par exemple est assignée par le tag « dege » deux fois. La même chose pour les autres ressources.

La figure suivante montre les tags les plus utilisés avec leurs fréquences d'utilisation pour cet exemple (10 agents).

TAG	FRQ UTILIS
hoya	57
xeja	40
dige	35
toqo	34
joca	30
xaqe	30
koyi	28
losa	27
lame	24
jaki	23
fido	21
woha	18
jeva	17
xehe	17
liro	16
baye	16
qoki	14
yexa	14
dege	14
xolo	14
waxi	14

FIG. 6.17 – un extrait de la folksommie (les premiers tags)

En analysant l'environnement, nous constatons l'émergence des catégories. On cite par exemple la catégorie des ressources commençant par le caractère « h » qui sont assignées par le tag « hoya » comme le montre la figure suivante.

Iteration	USER	TAG	Ressource
24	AGENT_8	hoya	hfj
64	AGENT_9	hoya	hac
89	AGENT_7	hoya	hbp
91	AGENT_5	hoya	hfn
127	AGENT_7	hoya	hep
142	AGENT_5	hoya	qic
205	AGENT_1	hoya	hpo
228	AGENT_9	hoya	hoc
249	AGENT_9	hoya	hxc
294	AGENT_4	hoya	hri
305	AGENT_3	hoya	qvi
313	AGENT_6	hoya	qbi
332	AGENT_3	hoya	qhi
339	AGENT_9	hoya	hbp
355	AGENT_7	hoya	hvx
389	AGENT_8	hoya	qoy
403	AGENT_7	hoya	qjv
405	AGENT_10	hoya	qum
436	AGENT_4	hoya	hbp
486	AGENT_4	hoya	qtn
501	AGENT_8	hoya	hdp
506	AGENT_10	hoya	hvr
520	AGENT_7	hoya	hgv
523	AGENT_4	hoya	hbf
551	AGENT_2	hoya	qfj

FIG. 6.18 – les ressources assignées par le tag « hoya »

6.8.3 Exemple 3

Nous utilisons pour cet exemple un nombre d'agents $N=100$ agents et un seuil de mesure de similarité $S= 0.3$.

Pour chaque simulation, nous vérifions que la dynamique d'un système de tagging est régénérée, les figures ci-dessous représentent l'évolution du nombre de tags dans le temps, l'évolution du nombre de ressources et l'évolution de l'entropie dans le système.

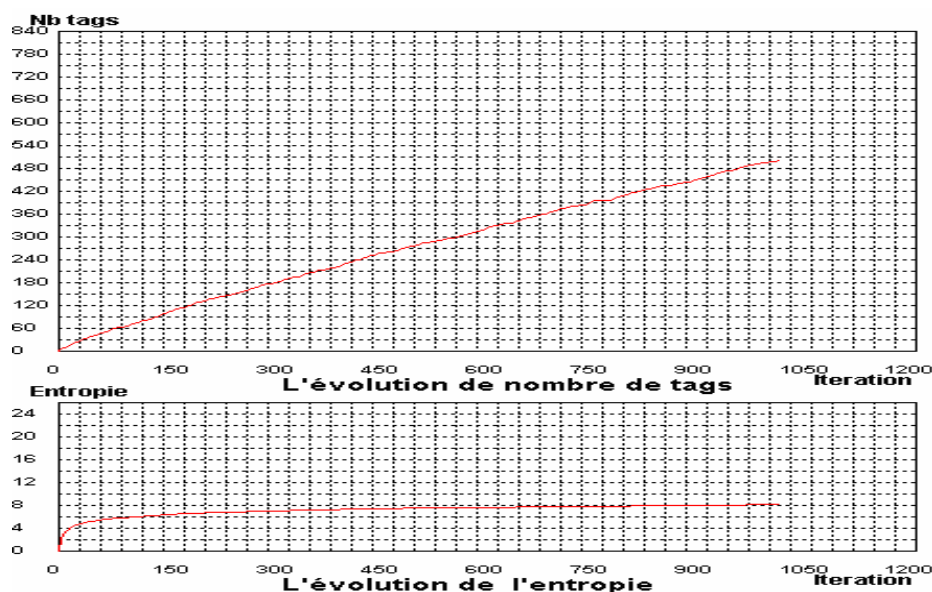


FIG. 6.19 – les courbes d'évolution du nombre de tags et l'entropie

Dans les courbes ci-dessous, nous remarquons l'émergence d'une folksonomie à 500 tags pour une collection de 875 ressources pour cet exemple après 1035 itérations.

Dans la courbe de l'entropie, nous avons aussi une stabilisation de l'entropie nécessaire pour ce système de codage.

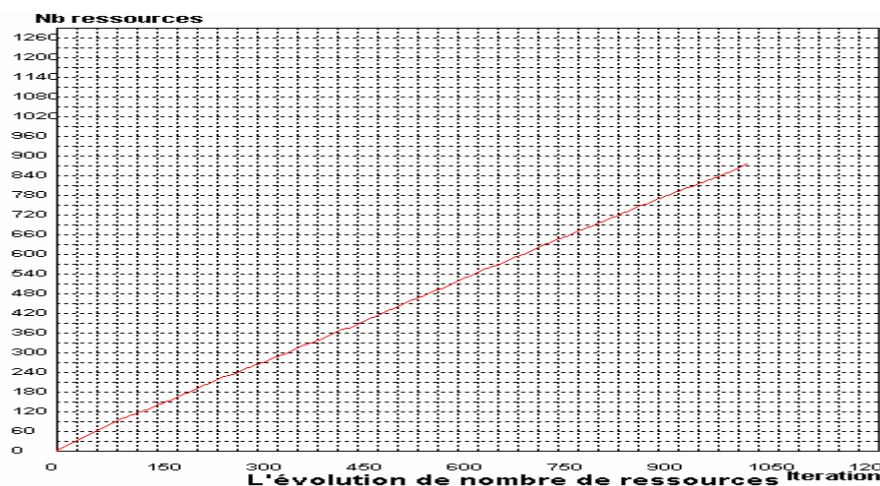


FIG. 6.20 – l'évolution du nombre de ressources

En passant à l'analyse linguistique de cet exemple, nous calculons la matrice de codage illustrée dans la figure suivante où on a montré un extrait de la matrice de codage de cet exemple.

Par exemple, la catégorie représentée par le tag « howe » contient des ressources qui commencent par le caractère « n ». La même chose pour le tag « joka » et les ressources commençant par « m ».

RES/TAG	howe	joka	tiji	jeso	nebe	hoke	nahi	fiio	cidi	topa
flu			1.0							
mt			1.0							
foz			1.0							
fluq			0.4							
mfc		1.0								
mre		1.0								
mrb		0.8								
mrah		0.75								
hpa	0.75									
jsa	0.5									
jyf	0.5									
jt	0.5									
jwo	0.33									
jesw	0.25									
ngm	0.22									
ah						1.0				
apf						0.75				
ah						0.5				
ahc						0.33				
rod					1.0					
igd					1.0					
ity					1.0					
inc					1.0					
ih					0.33					
hnp			1.0							
hny			0.66							
hnn			0.66							
hup			0.5							
hsp			0.5							
chs										0.5
yrm									1.0	
yth									1.0	
yiv									1.0	
kwih								1.0		
kcr								0.66		
kbrn								0.4		
ipl						1.0				
icf						1.0				
im						0.66				
imf						0.6				
ipk						0.5				
iqn						0.25				
imf						0.6				
ipk						0.5				
iqn						0.25				

FIG. 6.21 – un extrait de la matrice de codage

La figure suivante représente quelques clusters émergents dans le système de tagging.

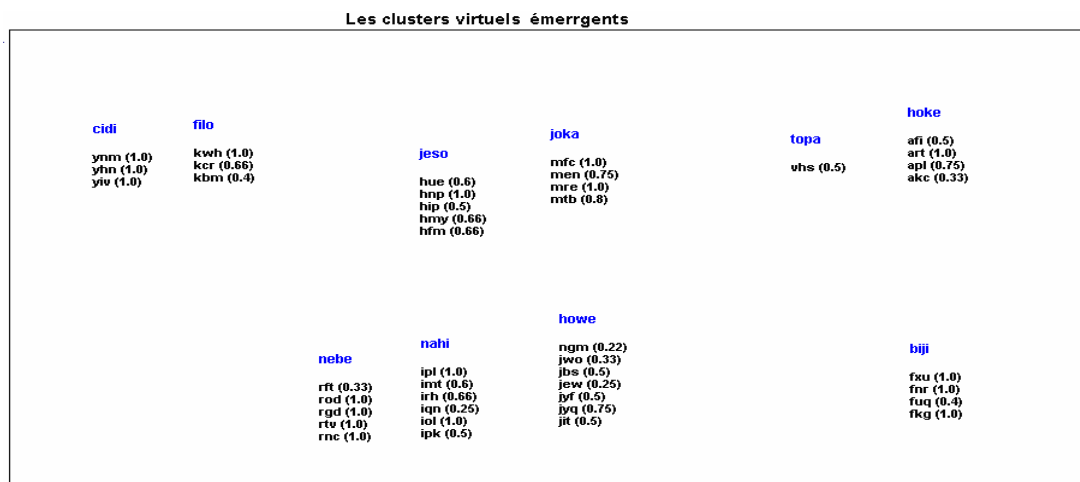


FIG. 6.22 – les clusters émergents

Au niveau du système de tagging, nous pouvons voir les catégories émergentes comme par exemple les ressources commençant par le caractère « m » sont assignées par le tag « joki » montrées dans la figure suivante .

Iteration	USER	TAG	Ressource
3	AGENT_50	joki	mfc
40	AGENT_91	joki	men
54	AGENT_77	joki	mre
65	AGENT_62	joki	mtb
153	AGENT_15	joki	mku
155	AGENT_25	joki	mre
203	AGENT_36	joki	mur
252	AGENT_82	joki	mtb
310	AGENT_77	joki	men
312	AGENT_2	joki	mxy
399	AGENT_63	joki	mci
416	AGENT_74	joki	mdk
423	AGENT_2	joki	mxx
436	AGENT_13	joki	men
462	AGENT_41	joki	mit
547	AGENT_100	joki	mds
558	AGENT_75	joki	mfi
563	AGENT_39	joki	mbv
564	AGENT_81	joki	mfj
621	AGENT_12	joki	mdt
663	AGENT_34	joki	mnw
669	AGENT_55	joki	mwo
720	AGENT_48	joki	mtb
724	AGENT_51	joki	mfc
779	AGENT_26	joki	mbq

FIG. 6.23 – les ressources assignées par le tag « joki »

6.8.4 Exemple 4

Nous utilisons pour cet exemple un nombre d'agents $N=10$ agents.

Dans les exemples précédents, nous avons gardé une valeur du seuil de mesure de similarité $S= 0.3$, pour voir l'effet de ce seuil sur la qualité du langage émergent, nous utilisons pour cet exemple un seuil de mesure de similarité $S= 0.5$.

En s'intéressant à la dynamique du système, nous observons la convergence d'une folksonomie 661 tags pour une collection de 839 ressources pour cette simulation après 1016 itérations.

La même chose pour la courbe de l'entropie de cet exemple, nous avons une stabilisation de la quantité de l'information nécessaire pour ce système de codage.

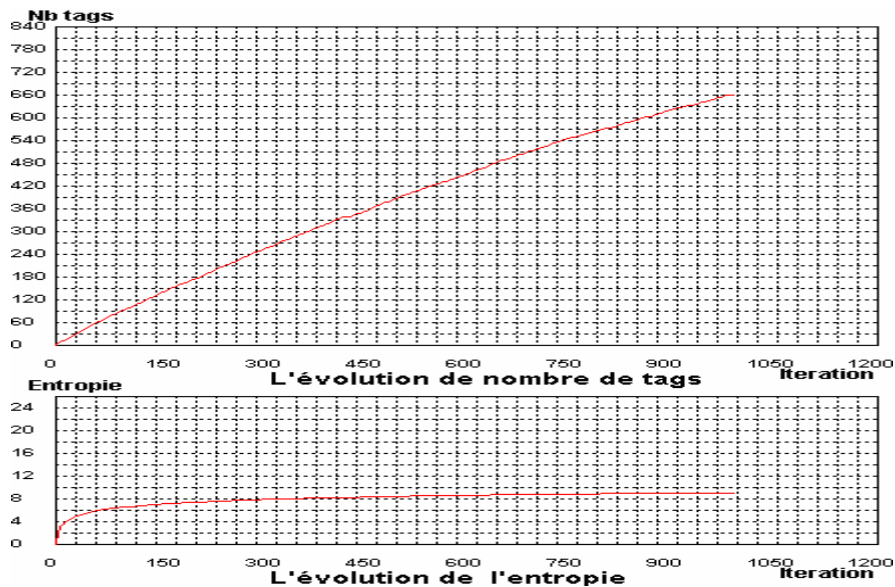


FIG. 6.24 – les courbes d'évolution du nombre de tags et l'entropie

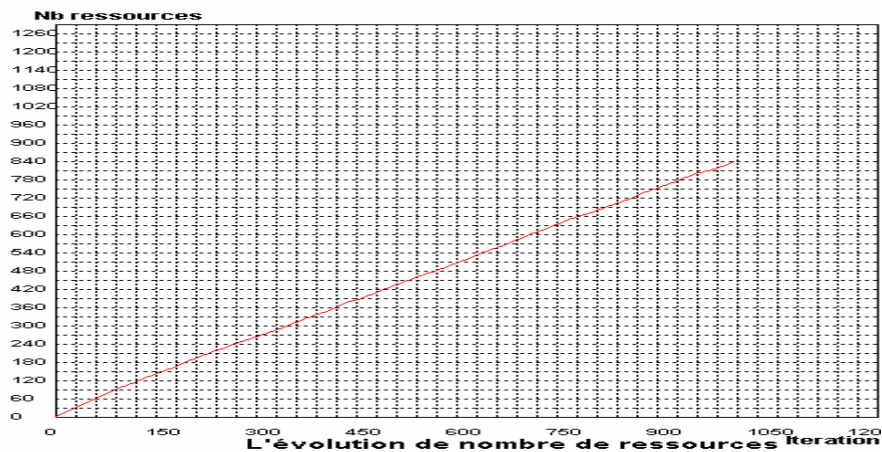


FIG. 6.25 – l'évolution du nombre de ressources

Nous calculons la matrice de codage pour cet exemple pour montrer le langage émergent, nous ne montrons toujours qu'une partie de la matrice de codage vue l'espace limité.

RESITAG	pope	dehe	yero	yipe	rawe	rafi	wone	vera	qoca	jeda
byf			0.33							
iuy			0.33							
bvh			0.25							
hyd		0.5								
hak		0.5								
hyv		0.4								
xoh	1.0									
xod	0.66									
xoq	0.33									
xif	0.2									
uth									1.0	
usi									0.5	
utr									0.5	
lvt								0.5		
dpk							0.33			
ayu							0.33			
ntf										0.66
nca										0.5
nwb										0.5
nab										0.33
uth									1.0	
usi									0.5	
utr									0.5	
kir						0.5				
sqi					0.5					
lvi					0.33					
jgl				1.0						

FIG. 6.26 – un extrait de la matrice de codage

La figure suivante représente quelques clusters émergents dans le système de tagging pris de la matrice du codage.

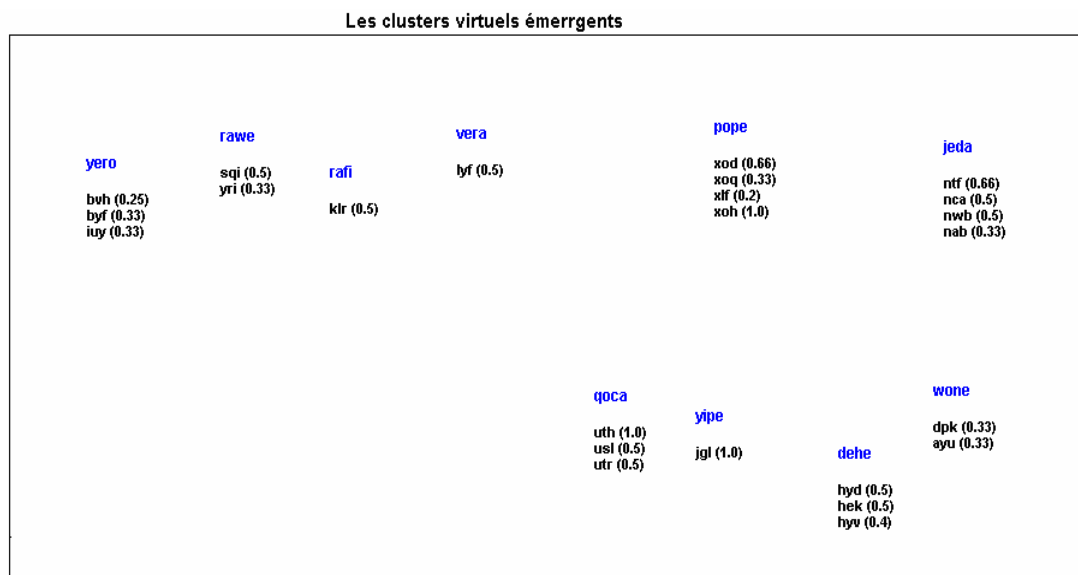


FIG. 6.27 – les clusters émergents

Dans la matrice de codage et les clusters émergents, la catégorie représentée par les tag « pope » contient des ressources qui commencent par les caractères « xo » et une ressource commence par « xl », la même chose pour le tag « qoca » et les ressources commençant par « ut » et une ressource commence par « us ».

Dans la figure suivante, nous montrons une catégorie émergente représentée par le tag « pope » qui contient les ressources commençant par le caractère « x ».

Iteration	USER	TAG	Ressource
44	AGENT_2	pope	xod
67	AGENT_1	pope	xoq
96	AGENT_6	pope	xoh
105	AGENT_7	pope	xai
111	AGENT_3	pope	xod
176	AGENT_6	pope	xbj
192	AGENT_6	pope	xag
294	AGENT_2	pope	xat
428	AGENT_3	pope	xmg
577	AGENT_2	pope	xhf
580	AGENT_2	pope	xlf
597	AGENT_6	pope	xdw
699	AGENT_8	pope	xcu
739	AGENT_1	pope	xny
960	AGENT_3	pope	xoh
983	AGENT_6	pope	xcq
992	AGENT_5	pope	xmv

FIG. 6.28 – un les ressources assignées par le tag « pope »

6.8.5 Résumé

En comparons la matrice de codage et les catégories émergentes de l'exemple 4 avec un seuil de similarité $S=0.5$ par les catégories dans les autres exemples (exp.1, exp.2, exp.3) où nous avons utilisé un seuil $S=0.3$, nous remarquons que dans les premiers exemples (1,2 et 3), les catégories émergentes ont des cardinalités (nombre de ressources pour cette catégorie) grandes mais dans l' exp.4, les catégories émergentes ont des cardinalité très petites (entre 1 et 3).

Nous expliquons ça dans les deux situations comme suit :

-dans les exemples 1, 2 et 3, un seuil de similarité $S=0.3$ permet une tolérance en terme de la possibilité de mettre deux ressources ou plus dans la même catégorie représenté par le tag T, c-à-d le degré de généralité entre les ressources est très élevée, ce qui permet aux agents taggeurs d'avoir un système de catégorisation (folksonomie) très générale (une catégorie à plusieurs ressources).

-mais dans l'exemple 4, un seuil de similarité $S=0.5$ permet une restriction en terme de la possibilité de mettre les ressources dans la une catégorie représenté par le tag T, c-à-d le degré de spécificité entre les ressources est très élevée, ce qui permet aux agents taggeurs d'avoir un système de catégorisation (folksonomie) très restreint (une catégorie à nombre de ressources limité).

6.9 Conclusion

Dans ce modèle et d'après les résultats des simulations cités ci-dessus, un langage de tagging partagé par les utilisateurs du système de tagging collaboratif modélisés par des agents taggeurs est émergé dans le système et au niveau des mémoires associatives des agents puisque nous sommes intéressés à l'aspect linguistique de ce processus.

Nous pouvons dire que :

Des utilisateurs (agents) du système de tagging collaboratif doté par des mémoires associatives et,

Suivant un mécanisme d'apprentissage approprié (l'apprentissage par renforcement)

Avec le principe de l'auto-organisation au niveau des interactions à travers le processus de tagging (mono-tagging pour ce modèle) et,

Un mécanisme sélectionniste exprimé par le principe d'imitation sociale et individuelle.

En utilisant les scénarios des jeux de langage adaptés à notre contexte,

Permettent l'émergence d'un langage de tagging partagé par la communauté des utilisateurs du système.

Pour notre modèle, la qualité du langage de tagging partagé qui est au même temps un système de catégorisation partagé par la communauté des utilisateurs dépend du niveau de généralité et de spécificité que les agents utilisent dans le processus de tagging que nous contrôlons par le changement de la valeur du seuil de mesure de similarité (plus la valeur du seuil est petite, le niveau de généralité augmente et plus la valeur du seuil est grande, le niveau de spécificité augmente).

Introduction

Dans le modèle précédent, nous avons étudié l'émergence d'un langage de tagging à structure lexicale (une ressource est associée à un seul tag) à travers une modélisation linguistique adaptée aux systèmes de tagging collaboratif en utilisant les jeux de langage et des mémoires associatives dans l'architecture des agents.

L'étape suivante et naturelle de ce travail est d'étudier la possibilité de l'émergence d'un langage de tagging à structure grammaticale partagé par les utilisateurs d'un système de tagging collaboratif qui représente l'objectif final de notre travail.

Ce langage de tagging grammatical se caractérise par des correspondances de type « expression/ressource » (expression= {tag₁, tag₂, ..., tag_n}), où chaque expression formée d'un ensemble de tags représente une catégorie de ressources.

L'ensemble de tags assignés à une ressource donnée est considéré comme étant une phrase nominale (les agents n'utilisent pas les verbes) d'un langage naturel comme par exemple « une rose blanche » pour caractériser une ressource.

Par exemple pour une ressource «r » assignée par {tag₁, tag₂, tag₃}, nous considérons le « tag₁ » comme étant ce que la ressource «r » représente (une page web, un logiciel, une photo...) et le « tag₂ » comme étant un domaine auquel la ressource «r » appartient (l'informatique, la médecine, la littérature ...) et le « tag₃ » comme étant un sous-domaine auquel la ressource «r » appartient (par exemple pour le domaine d'informatique, on trouve la génie logiciel, réseaux, systèmes d'exploitation...).

Nous utilisons les hypothèses des modèles précédents avec des modifications appropriés à notre objectif pour étudier l'émergence de langage de tagging partagé par les utilisateurs d'un système de tagging en enrichissant le modèle précédent comme suit :

- Au niveau des interactions qui vont suivre un processus multi-tagging autrement dit les agents assignent chaque ressource par un ensemble de tags.
- Nous modifions la structure de la mémoire associative d'un agent de sorte qu'elle permette de stocker à chaque colonne une expression (un ensemble de tags) choisie pour une ressource donnée.
- Et un mécanisme d'apprentissage par renforcement pour que les agents puissent mettre à jour leurs mémoires associatives
- Les agents interagissent avec le système de tagging par des scénarios des jeux de langage.

Ce modèle se base, comme les modèles précédents, sur le travail de Santos-Neto et Condon (Santos-Neto & Condon 2009), ainsi que sur les modèles proposés dans les travaux de Kaplan (Kaplan 2000).

Ceci, en gardant toujours la même approche voyelle pour modéliser notre système multi agents.

7.2 Environnement

L'environnement est le même que celui des modèles précédents, il contient un ensemble de ressources et un ensemble de tags à attribuer à ces ressources comme le montre FIG. 7.1.

Pour les tags, on garde la même représentation (des mots libres générés aléatoirement à partir d'un ensemble de consonnes et de voyelles)

Pour les ressources, elles sont réduites à « n » mots clés générés aléatoirement qui représentent des méta-données sur le contenu de la ressource.

L'agent fait un multi-tagging en utilisant des tags choisis uniquement parmi ces « n » mots clés. Il fait une opération de tagging de cette ressource en la comparant avec les autres ressources pour trouver la ressource la plus proche par rapport à ce mot clé (nous aurons comme résultats « n » tags pour chaque ressource où chaque tag est une dimension (par exemple un domaine, un sous domaine et un type etc.) ou une façon de catégoriser cette ressource qui est de la forme $\text{ressource}=\{\text{mot clé}_1, \text{mot clé}_2, \dots, \text{mot clé}_n\}$.)

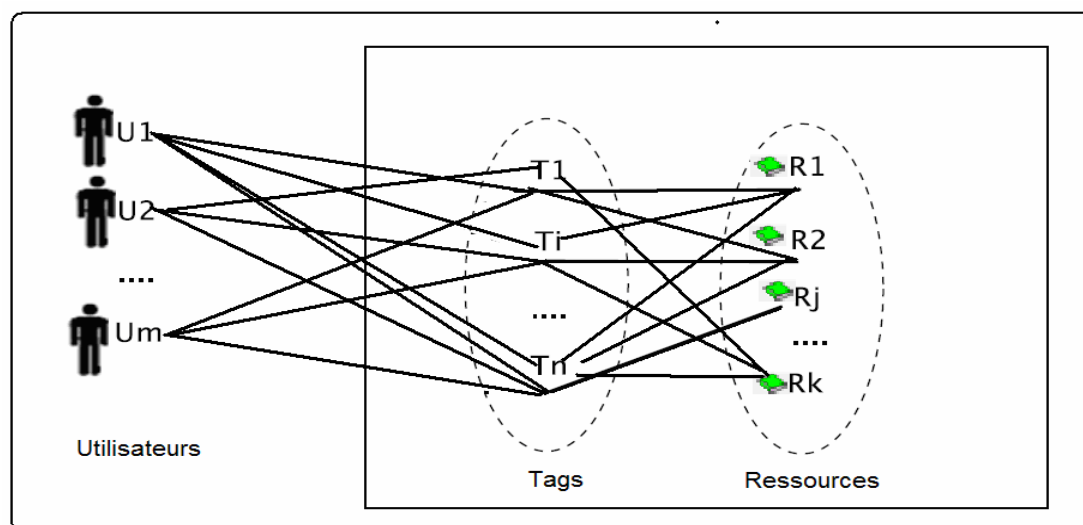


FIG. 7.1 – l'environnement (du système de tagging) et ses trois parties

7.3 Organisation

Les organisations émergentes correspondent toujours aux groupes formés autour d'un ensemble de ressources et de tags partagés par les membres de ces organisations comme dans la figure FIG. 7.2.

Mais en s'intéressant à l'aspect linguistique, ces groupes virtuels correspondent à des expressions ou des phrases pour désigner des ressources partagées par une sous communauté des utilisateurs des systèmes de tagging.

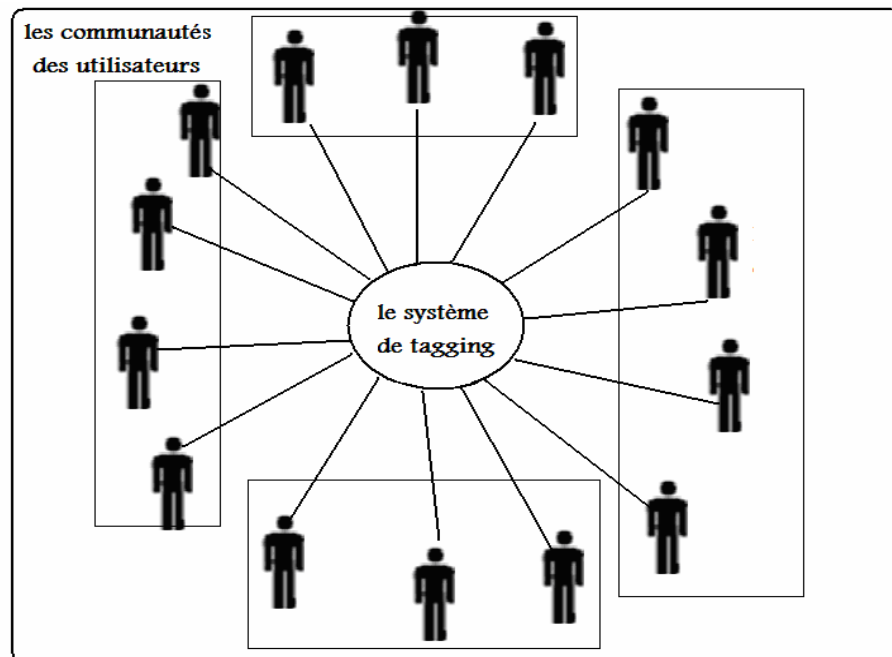


FIG. 7.2 – la communauté des utilisateurs dans un système de tagging

7.4 Agent

Pour ce modèle, l'architecture de l'agent ne change pas globalement, c-à-d l'agent est doté d'une mémoire associative et quelques règles de comportement.

Mais nous faisons des modifications et des enrichissements au niveau des règles de comportement avec le système qui vont suivre une logique de multi-tagging et au niveau de la mémoire associative.

Pour que la mémoire associative permette l'émergence d'un langage à structure grammatical (une expression de « n » tags par ressource au lieu d'un tag), il faut adapter sa architecture en sorte qu'elle peut stocker les connaissances linguistiques sous forme des expressions liées à des ressources (des couple de la forme {tag₁, tag₂, ..., tag_n}/ressource) comme le montre la figure FIG. 7.3.

Les agents interagissent avec le système de tagging en laissant des traces sur l'environnement sous forme des assignations stockées dans l'environnement (où elles sont mémorisées dans la mémoire collective de tous les utilisateurs).

Comme nous avons mentionné ci-dessus, les agents sont capables d'exécuter des comportements simples pour faire une assignation de « n » tags à une ressource et négocier ces choix avec le système et garder cette assignation en adaptant sa mémoire associative.

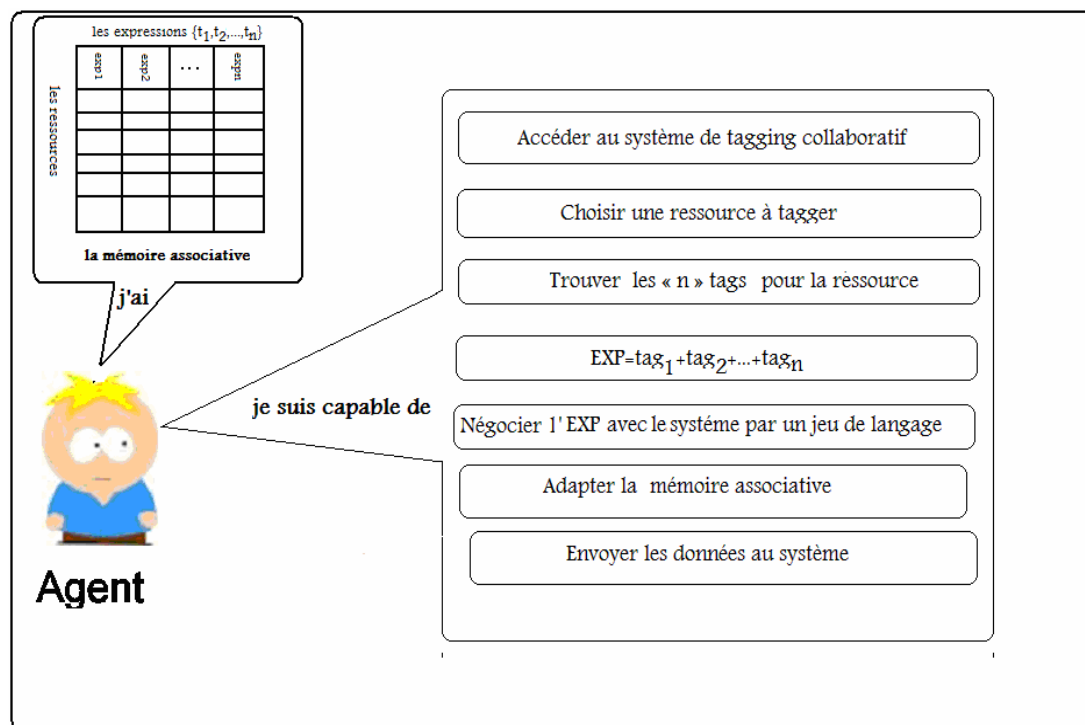


FIG. 7.3 – l'architecture d'un agent taggeur avec les comportements

7.5 Interaction

Les interactions de ce modèle suivent aussi des schémas d'interaction inspirés des jeux de langage adapté aux systèmes de tagging collaboratif.

L'enrichissement par rapport au modèle précédent se fait au niveau de mode de tagging qui n'est plus mono mais un processus de multi-tagging collaboratif.

Les interactions se font entre un agent et le système où l'agent (resp. le système) joue le rôle de locuteur ou interlocuteur pour négocier l'assignation d'une expression (un ensemble de « n » tags) à un sens (une ressource constituée de « n » mot clés).

L'interaction des agents avec le système se fait de la manière suivante:

1. l'agent accède au système de tagging collaboratif.

2. le système donne accès pour cet utilisateur à l'ensemble de tags qu'il a utilisé dans les sessions précédentes (sa personomie), le système donne aussi accès à un ensemble de tags les plus utilisés par les autres utilisateurs.

3. l'agent choisit une nouvelle ressource ou une ressource déjà existante dans le système, s'il est en train de faire l'opération de retagging.

4. l'agent fait « n » opérations de tagging par rapport aux « n » mots clés de la ressource choisie (mot clé₁, mot clé₂,..., mot clé_n) où dans chaque opération élémentaire, il a deux possibilités pour choisir un tag pour l'attribuer à la ressource choisie:

a- il joue le rôle d'un locuteur et dans ce cas le système joue le rôle de l'interlocuteur (c'est une imitation individuelle).

b- il joue le rôle d'un interlocuteur et dans ce cas le système joue le rôle du locuteur (c'est une imitation sociale).

Dans les deux situations, il y aura une négociation entre le locuteur et l'interlocuteur selon un schéma de jeux de langage pour décider l'assignation appropriée à la ressource en question.

5. il adapte sa mémoire associative en insérant des nouvelles assignation ou en corrigeant une assignation par le renforcement.

6. il envoie l'identificateur de la ressource et l'ensemble de tags qu'il a choisi au système sous la forme d'une entrée « $U_i, \{T_{i1}, T_{i2}, \dots, T_{in}\}, R_i, t_i$ »

Où $U_i, \{T_{i1}, T_{i2}, \dots, T_{in}\}, R_i, t_i$ représentent respectivement l'utilisateur, l'expression, la ressource, un marquage de temps.

En détaillant toujours l'étape 4 que nous jugeons comme étant la plus importante comme suit :

L'agent décide de manière autonome de faire une imitation individuelle en jouant le rôle du locuteur ou une imitation sociale en jouant le rôle de l'interlocuteur et il procède de la manière suivante:

1. Dans le premier cas (le rôle du locuteur), l'agent fait « n » opérations de tagging pour assigner une expression constituée de « n » tags ($Expression_{indiv} = \{T_{indiv1}, T_{indiv2}, \dots,$

T_{indivn}). Dans chaque opération il fait une imitation individuelle, cette imitation se fait par rapport à l'un des mots clés de la ressource (en commençant par mot clé₁ ensuite mot clé₂ jusqu' au mot clé_n).

L'agent cherche parmi les ressources personnelles la ressource la plus proche à la ressource par rapport au mot clé_i « $r_i = \{\text{mot clé}_i\}$ », en mesurant sa similarité sémantique par rapport à toutes les ressources personnelles en ne prenant en compte que les mots clés_i de même ordre que celle de la ressources en question (1 ou 2 ou n).

L'agent assigne le tag assigné à la ressource qui a la similarité maximale s'elle est supérieure à une tendance calculée proportionnellement à un seuil prédéfini « S ».

Sinon il crée un nouveau tag.

2. Dans le deuxième cas (le rôle d'interlocuteur), le processus se fait par une négociation entre le locuteur (le système) et l'interlocuteur (l'agent) en passant par deux étapes intermédiaires en commençant par le locuteur :

a- le système (le locuteur) fait une imitation sociale en cherchant parmi ses ressources pour attribuer une expression constitué de « n » tags ($\text{Expression}_{\text{sys}} = \{T_{\text{sys}1}, T_{\text{sys}2}, \dots, T_{\text{sys}n}\}$) à la ressource en question en se basant sur les ressources du système et les tags du système.

b- L'agent (l'interlocuteur) fait une imitation individuelle pour attribuer une expression constitué de « n » tags ($\text{Expression}_{\text{indiv}} = \{T_{\text{indiv}1}, T_{\text{indiv}2}, \dots, T_{\text{indiv}n}\}$) à la ressource en question en se basant sur ses ressources et les tags personnels en précédant exactement comme l'étape précédente.

Pour la mesure de similarité entre deux ressources, nous utilisons la même fonction utilisée dans les modèles précédent qui est définie comme suit :

Soient $r_i = \{\text{mot_clé}_i\} \in \text{Res}$ et $r_j = \{\text{mot_clé}_j\} \in \text{Res}$ deux ressources représentées par leurs mots clés, nous donnons la similarité entres elles par

$$\text{Sim}(r_i, r_j) = \frac{2 * \text{Communs_préfix}(\text{mot_clé}_i, \text{mot_clé}_j)}{\text{Long}(\text{mot_clé}_i) + \text{Long}(\text{mot_clé}_j)}$$

$\text{Sim}(r_i, r_j)$ est une fonction de mesure de similarité sémantique entres les ressources qui part de l'ensemble (Res x Res) vers l'intervalle [0,1]

A la fin de ces deux étapes intermédiaires, l'agent négocie avec le système l'expression appropriée à la ressource :

Si « Expression_{indiv} » et « Expression_{sys} » sont les mêmes, nous avons un succès et il y aura une adaptation.

Sinon, nous avons un échec et il y aura une autre adaptation.

Adaptation

Dans ce modèle, l'adaptation est comme celle du modèle précédent mais les agents utilisent des expressions de « n » tags au lieu d'un seul tag.

Elle se fait au niveau des mémoires associatives en renforçant ou diminuant les scores des assignations «expressions, ressources » dans les deux scénarios d'interaction (imitation individuelle ou sociale)

a. pour le premier cas où l'agent joue le rôle du locuteur, l'adaptation se fait de la manière suivante : L'agent renforce le score du couple « Expression_{indiv}, r_i » dans la mémoire par une unité.

b. pour le deuxième cas où l'agent joue le rôle de l'interlocuteur l'adaptation dépend du résultat de jeux de langage (succès ou échec)

1- dans le cas de succès (« expression_{indiv} » et « Expression_{sys} » sont les mêmes), l'agent renforce le score du couple « Expression_{sys}, r_i » dans la mémoire par une unité.

2- dans le cas de l'échec (« Expression_{indiv} » et « Expression_{sys} » sont différentes), l'agent renforce le score du couple « Expression_{sys}, r_i » dans la mémoire par une unité et il diminue le score du couple « Expression_{indiv}, r_i » par une unité.

Au niveau du système du tagging, l'adaptation se fait par l'enregistrement de « n » entrées de la forme « U, T_i, R, t » dans sa base de données et en effectuant une mise à jour des fréquences d'utilisation des tags T_i en les incrémentant de un. Ces fréquences vont servir dans le tri des tags les plus utilisés dans les prochaines sessions du même utilisateur ou d'autres utilisateurs du système.

Algorithme d'interaction multi-tagging et d'adaptation (jeu de langage)

Mode_tagging=générer aléatoirement un mode de tagging (); // tagging ou retagging.

Si (Mode_tagging == opération _tagging)

Alors

Ressource = générer aléatoirement une ressource () ; // composé de 3 mot clés

Sinon

Ressource = choisir une ressource du système () ; //opération de retagging

Fin si

Mode_imitation = générer aléatoirement un mode d'imitation () ; // sociale ou indiv.

Si (Mode_imitation == imitation_individuelle) // le rôle de locuteur

Alors

Pour i=1 jusque n Faire

Tag_indiv= faire une imitation individuelle par rapport au mot clé (i) ;

Expression_indiv= Expression_indiv + Tag_indiv ;

Fin faire

Adapter la mémoire associative (Expression_indiv, Ressource, Renforcement) ;

Sinon // le rôle de interlocuteur (imitation sociale)

Pour i=1 jusque n Faire

Tag_sys= faire une imitation sociale par rapport au mot clé (i) ;

Expression_sys = Expression_sys + Tag_sys ;

Fin faire

Pour i=1 jusque n Faire

Tag_indiv= faire une imitation individuelle par rapport au mot clé (i) ;

Expression_indiv= Expression_indiv + Tag_indiv ;

Fin faire

Si (Expression _indiv == Expression _sys) // c'est un succès en jeu

Alors

Adapter la mémoire associative (Expression_sys, Ressource, Renforcement) ;

Sinon // c'est un échec en jeu

Adapter la mémoire associative (Expression_indiv, Ressource, Diminution) ;

Adapter la mémoire associative (Expression_sys, Ressource, Renforcement) ;

Fin si

Fin si

7.8 Résultats et discussion

Dans cette partie, nous présentons quelques résultats obtenus par un démonstrateur conçu sur la base du modèle décrit ci-dessus. Notre but dans cette phase est d'étudier la possibilité de l'émergence d'un langage à structure grammaticale par un processus de multi-tagging, ce langage est partagé par les utilisateurs du système

Les paramètres de notre modèle que nous modifions dans les simulations sont :

- Le nombre d'agents qui accèdent le système et participent dans l'opération de multi-tagging.
- Le nombre « n » de tags qui constituent les phrases du langage émerger.
- Le seuil de mesure de la similarité entre les ressources qui permet d'assigner les ressources qui sont proches par le même tag.
- Le nombre de tags les plus utilisés présentés à un agent pour permettre une imitation sociale.
- les probabilités de faire une imitation individuelle (locuteur) ou sociale (interlocuteur).
- les probabilités de faire les opérations de tagging ou retagging.

Pour les probabilités de faire une imitation individuelle (avoir le rôle du locuteur) ou sociale (avoir le rôle du interlocuteur) ainsi pour le choix de l'opération de tagging ou retagging, nous utilisons aussi dans les exemples cités ci-dessous des probabilités inspirées du travail de Santos-Neto et Condon (Santos-Neto & Condon 2009).

Ce choix est fait pour étudier la possibilité de l'émergence dans une dynamique de système de tagging la plus proche à celle observée dans les systèmes réels.

Mais, nous pouvons utiliser des probabilités quelconques dans les simulations.

Pour le nombre « n » de tags par expression, nous pouvons utiliser d'autres valeurs de « n » qui varient entre 1 et 10 ;

Remarque

Pour $n=1$, on obtient un langage à structure lexicale que nous avons déjà étudié dans le modèle précédent de notre travail.

Pour analyser le langage émergent, nous utilisons une version adaptée de la matrice de codage que nous avons utilisé dans le modèle précédent pris de (Kaplan 2000) que nous redéfinissons comme étant une matrice dont les colonnes sont « les expressions » du langage émergent (les expressions sont formées de « n » tags du système) et les lignes

sont « les sens » codés par ces tags (les ressources du système) et dans l'intersection d'une expression « $exp = \{t_1, t_2, \dots, t_n\}$ » avec une ressource « $res = \{\text{mot clé}_1, \text{mot clé}_2, \dots, \text{mot clé}_n\}$ » le nombre d'agents qui assignent « res » par « exp » divisé par le nombre d'agents qui ont assigné la ressource « res », c-à-d la probabilité de mettre la ressource « res » dans la catégorie représentée par l'expression « exp ».

La figure FIG. 7.4 montre un exemple d'une matrice de codage où les EXPi en colonne sont les expressions de langage émergent et qui représentent des catégories partagées par un ensemble d'agents et en lignes se trouvent les ressources d'un système de tagging .

Referent	EXP1	EXP2	EXP3	EXP4	EXP5
Ress1	1.00				
Ress2	1.00				
Ress3	1.00				
Ress4	1.00				
Ress5	1.00				
Ress6	1.00				
Ress7	1.00				
Ress8	1.00				
Ress9	1.00				
Ress10		1.00			
Ress11				0.78	
Ress12				1.00	
Ress13				1.00	
Ress14				1.00	
Ress15				1.00	
Ress16			1.00		
Ress17			1.00		
Ress18			0.33		
Ress19			1.00		
Ress20			1.00		

FIG. 7.4 – un exemple de la matrice de codage

Pour analyser la dynamique du système, nous utilisons toujours

- La courbe de l'évolution du nombre de tags.
- La courbe de l'évolution du nombre de ressources dans le système.
- La courbe de l'évolution de l'entropie (la quantité de l'information) dans le système.

7.8.1 Exemple 1

Nous utilisons dans cet exemple un nombre d'agents $N=10$ agents, un nombre de tags par expression $n=3$ Et un seuil de mesure de similarité $S= 0.3$.

Pour analyser la dynamique générée au sien du système, nous traçons les courbes de l'évolution du nombre de tags dans le temps et l'évolution du nombre de ressources ainsi que l'évolution de la quantité d'information dans le système.

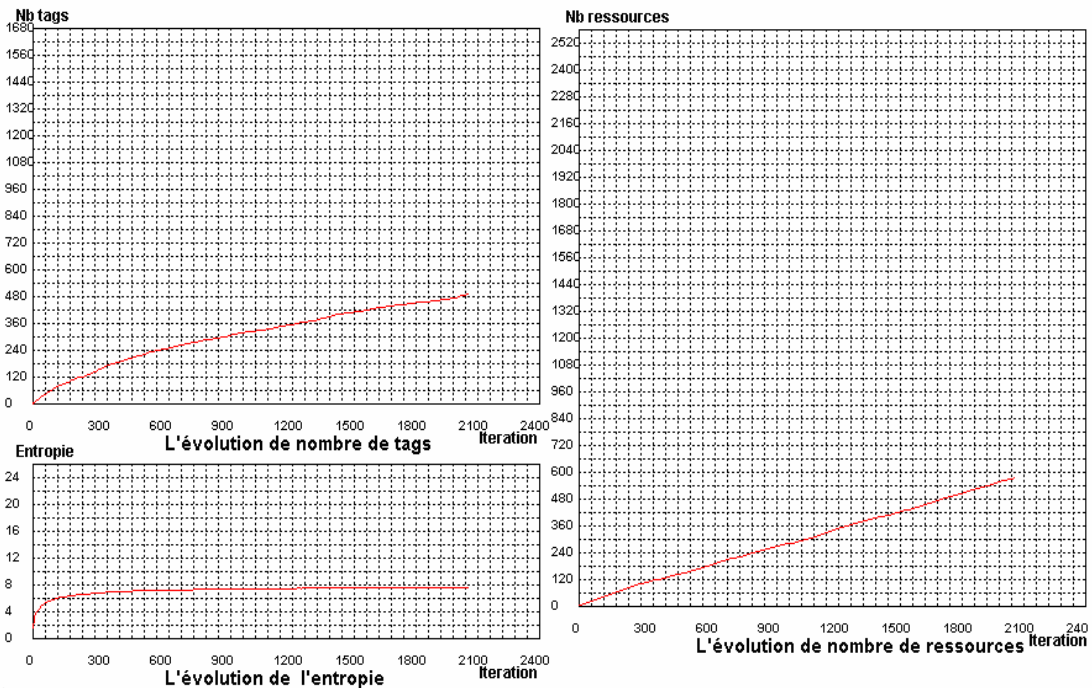


FIG. 7.5 – la dynamique de système de tagging

Pour la dynamique, nous remarquons la convergence d'une folksonomie à 489 tags pour une collection de 572 ressources pour cette simulation (2115 itérations). Dans la courbe de l'entropie, nous avons une stabilisation de la quantité de l'information nécessaire pour ce système de codage.

C'est la même dynamique comme celle des systèmes de tagging collaboratif qui suit une loi en puissance avec un exposant moins que un.

En s'intéressant à l'aspect linguistique de cette simulation, nous calculons la matrice de codage des catégories émergentes dans le système. La figure suivante est un extrait de la matrice de codage globale vu la contrainte de l'espace.

RES/TAG	jado wiwo jesi	fexe piya sawo	coro wiwo jefi	cevo hake sawo	seci piya kego	cipe vixa sawo	seci hake kego	gada sari loki
hfu,ndf,rjh			1.0					
hqd,nhq,ruf			1.0					
hxo,pwx,sga		1.0						
hma,psk,sqw		1.0						
cik,nti,vdg	1.0							
cvl,nto,vod	1.0							
gcm,kgv,sly						1.0		
bhw,pty,tix					1.0			
brm,pjw,twa					1.0			
ath,lhn,sect				1.0				
cpj,osw,qij								1.0
bvx,lep,thg							1.0	
gcm,kgv,sly						1.0		

FIG. 7.6 – un extrait de la matrice de codage

Dans la figure suivante, nous visualisons les résultats de cet exemple sous forme de clusters, où chaque cluster est une catégorie émergente représentée par une expression de 3 tags et nous montrons que quelques clusters.

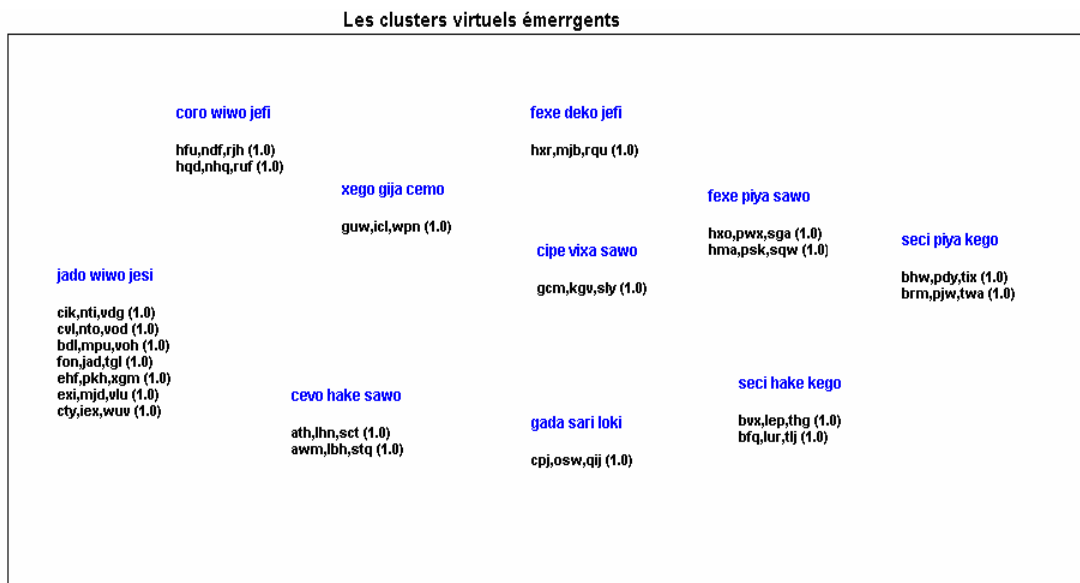


FIG. 7.7 – un extrait de clusters émergents

Dans la matrice de codage et dans les clusters, nous constatons qu'un langage à structure grammatical émerge sous forme des catégories.

On cite quelques exemples pour bien montrer le langage, si nous prenons la catégorie représentée par l'expression « jado wiwo jesi » nous remarquons que ses éléments sont des ressources dont leur mot clé 1 commence par le caractère « c » et

leur mot clé 2 commence par le caractère « n » et leur mot clé 3 commence par le caractère « v », selon la fonction de mesure de similarité que nous définissons, ces ressources son proches.

La même chose pour l'expression « seci piya kego » et les ressources commençant par « b », « p », « t » par rapport les 3 dimensions.

La figure suivante représente un extrait d'une mémoire associative d'un agent de la communauté dont les valeurs positives sont des scores renforcés et les scores négatifs sont des scores diminués.

RES/TAG	jado wiwo jesi	bahi bivi qeba	yena vabi koya	gita vabi jesi	jado woni sari	gici teqi kiti	jado wawi kiti	jado piya loki	weqo siqi kiti
cik,nti,vdg	1								
fbo,jue,uvj		1							
gnu,iso,usv			1						
auc,lok,vfc				1					
cbg,npq,xkw					1				
dji,nit,qbx						1			
crt,plg,qdc							-1	1	
aji,kar,qhl									1
fum,itb,tqg									

FIG. 7.8 – un extrait d'une mémoire associative

La figure suivante montre les tags les plus utilisés avec leurs fréquences d'utilisation.

TAG	FRQ UTILIS
piya	69
sawo	66
pewo	64
vixa	58
cemo	52
kego	49
jefi	47
wamo	46
loki	46
jesi	45
lexo	45
sari	42
seci	42
deko	41
gija	41
hake	39
nita	37
wiwo	35
yipi	33
cevo	33
cipe	28
jado	28

FIG. 7.9 – un extrait de la folksommie (les premiers tags)

En passant à une analyse des catégories émergentes au niveau de l'environnement par l'analyse des traces des assignation, les catégories émergentes représentées par des expressions de 3 tags comme le montre la figure suivante où les ressources commençant par le caractère « c » sont assignées par le tag « jado », les ressources commençant par le caractère « n » sont assignées par le tag « wiwo » et les ressources commençant par le caractère « v » sont assignées par le tag « jesi ».

Si nous prenons l'intersection de ces 3 sous-ensembles de ressources on trouve les éléments de la catégorie représentée par l'expression « jade wiwo jesi » comme la ressource « cok, njc, vmd » qui se trouve dans les 3 sous-ensembles de chaque tag.

On peut dire que ce qui se trouve dans les mémoires associatives des agents se trouve dans l'environnement et vice versa.

Iteration	USER	TAG	Ressource
25	AGENT_7	jado	cik,nti,vdg
127	AGENT_7	jado	cbg,npq,xkw
148	AGENT_7	jado	crt,plg,qdc
157	AGENT_8	jado	cok,njc,vmd
226	AGENT_2	jado	cvk,olv,vah
232	AGENT_8	jado	ctq,mtk,usw
346	AGENT_4	jado	ckh,nwl,vnc
613	AGENT_6	jado	clo,loi,wbe
655	AGENT_5	jado	chb,phm,shc
664	AGENT_5	jado	chu,kje,uty
697	AGENT_7	jado	cvi,nto,vod
700	AGENT_1	jado	cvo,jmr,toy
835	AGENT_4	jado	ckx,ike,swm
979	AGENT_3	jado	cmv,owr,wnu
1018	AGENT_5	jado	ctq,mtk,usw
1153	AGENT_4	jado	cie,mdk,qdu
1159	AGENT_4	jado	cvk,olv,vah
1222	AGENT_3	jado	csd,omp,veb
1408	AGENT_4	jado	cih,mfp,rid
1531	AGENT_3	jado	cou,pme,xsk
1702	AGENT_7	jado	cpf,kyr,xok
1705	AGENT_7	jado	cjo,jtc,tau
1747	AGENT_3	jado	cjh,jmo,tmd
1762	AGENT_4	jado	cif,mya,rhw
1789	AGENT_1	jado	ctr,ina,tlw

Iteration	USER	TAG	Ressource
26	AGENT_7	wiwo	cik,nti,vdg
158	AGENT_8	wiwo	cok,njc,vmd
242	AGENT_5	wiwo	btn,nkt,rcd
251	AGENT_4	wiwo	heu,nqy,tuy
335	AGENT_6	wiwo	gqm,nvh,tbk
347	AGENT_4	wiwo	ckh,nwl,vnc
395	AGENT_6	wiwo	cyo,nqe,tmd
440	AGENT_7	wiwo	ftd,nxt,xil
449	AGENT_8	wiwo	acm,nlf,tdv
488	AGENT_7	wiwo	hfu,ndf,rjh
500	AGENT_5	wiwo	hqd,nhq,ruf
620	AGENT_5	wiwo	bas,nil,ujm
650	AGENT_6	wiwo	hvr,mx,ulo
698	AGENT_7	wiwo	cvi,nto,vod
749	AGENT_7	wiwo	frw,naw,ukr
842	AGENT_1	wiwo	cqu,nfj,qcy
956	AGENT_5	wiwo	aie,nhf,wrg
1010	AGENT_9	wiwo	cyq,nlq,rau
1055	AGENT_5	wiwo	att,njm,vfk
1214	AGENT_2	wiwo	are,nhc,qto
1328	AGENT_5	wiwo	axs,nwb,smn
1439	AGENT_4	wiwo	frw,naw,ukr
1514	AGENT_1	wiwo	fcv,nyd,gle
1535	AGENT_8	wiwo	hti,npm,xqp
1577	AGENT_9	wiwo	fhe,nsg,wpd

Iteration	USER	TAG	Ressource
27	AGENT_7	jesi	cik,nti,vdg
75	AGENT_1	jesi	hjp,mlo,vrv
117	AGENT_7	jesi	auc,lok,vfc
183	AGENT_6	jesi	ano,noh,vcp
204	AGENT_6	jesi	hmv,jqb,vse
228	AGENT_2	jesi	cvk,olv,vah
375	AGENT_4	jesi	hmv,kht,vgd
387	AGENT_1	jesi	cqv,pca,vju
456	AGENT_6	jesi	ebw,ivh,vxf
573	AGENT_7	jesi	fyd,mqh,vgd
582	AGENT_8	jesi	gsb,lqs,vaw
609	AGENT_5	jesi	cpf,pig,vxf
618	AGENT_4	jesi	aou,isd,vfp
684	AGENT_10	jesi	bqr,lnu,vpj
699	AGENT_7	jesi	cvi,nto,vod
723	AGENT_3	jesi	gcv,bi,vqk
792	AGENT_4	jesi	adq,lap,vlh
846	AGENT_10	jesi	hon,kfl,vrv
888	AGENT_2	jesi	fjx,njf,vaf

FIG. 7.10 – les ressources assignées par les tags « jado », « wiwo », « jesi ».

7.8.2 Exemple 2

Nous utilisons pour cet exemple un nombre d'agents $N=100$ agents, un nombre de tags par expression $n=3$ et un seuil de mesure de similarité $S= 0.3$.

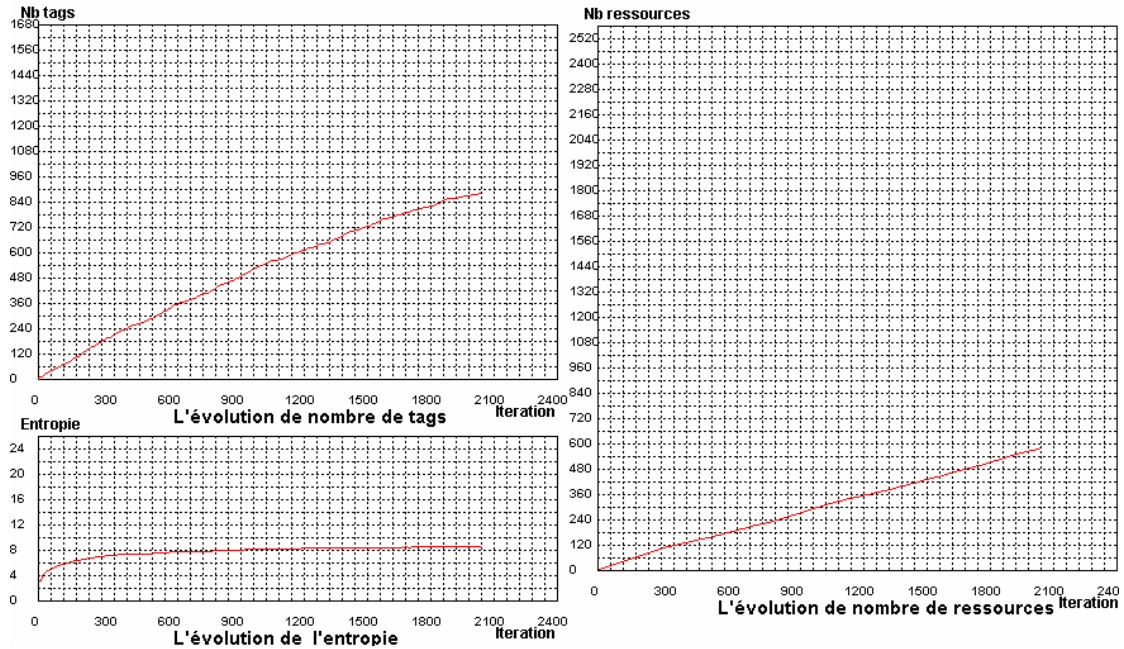


FIG. 7.11– la dynamique de système de tagging

Nous avons obtenus une convergence d'une folksonomie à 882 tags pour une collection de 579 ressources pour cette simulation après 2100 itérations.

Nous calculons toujours la matrice de codage pour savoir s'il y a un langage émergent dans le système pour cet exemple.

RES/TAG	beta tasi ceti	qowe veje yico	dise fimi qaro	miva fimi dina	qowe tasi koce	qowe hayo haxi	dise roha qaro	beta sowa ceti
etj,nir,qid			1.0					
etw,ngx,qep			1.0					
ekm,nqk,qfe			1.0					
bpy,isc,vmq		1.0						
bop,ifc,vgw		1.0				1.0		
cpy,kvf,xcs	1.0							
cti,kvl,xfo	1.0							
cug,kvf,xip	1.0							
cti,khq,xct	1.0							
bop,ifc,vgw		1.0				1.0		
bsm,kja,udm					1.0			
beq,kbu,uxh					1.0			
bin,kvn,ulx					1.0			
frh,nts,sgx				1.0				
fer,nsj,sbh				1.0				
cvi,lsf,xei								1.0
ctu,lbn,xpm								1.0
efy,oxl,qbg							1.0	
epc,oqf,qci							1.0	

FIG. 7.12 – un extrait de la matrice de codage

En utilisant une représentation graphique, nous visualisons les résultats sous forme de clusters.

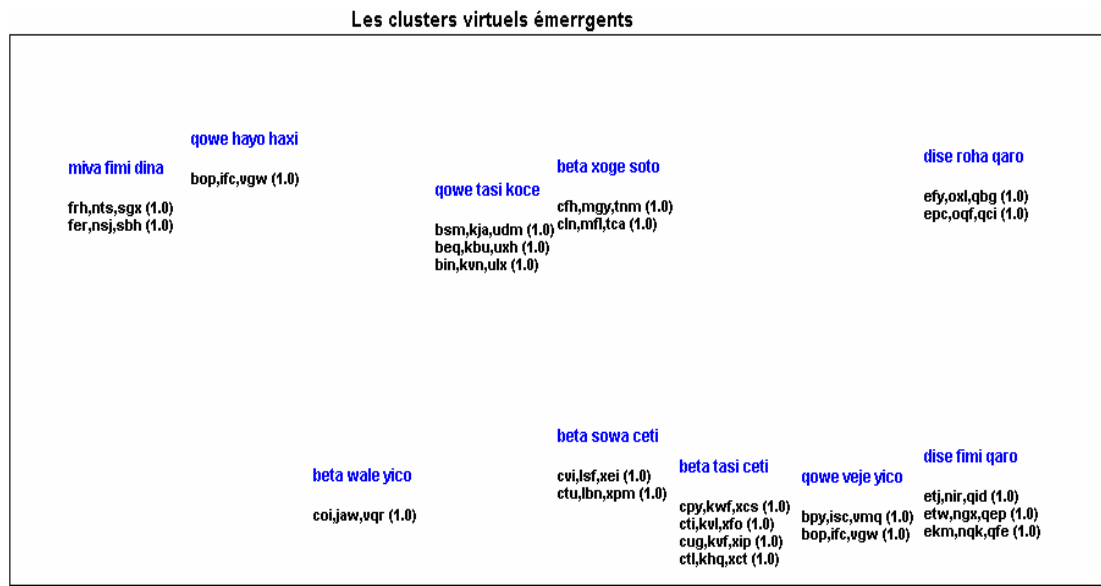


FIG. 7.13 – un extrait de clusters émergents

A partir des clusters, nous pouvons dire qu'un langage à structure grammatical émerge sous forme des catégories groupant des ressources proches entre eux.

Pour la catégorie représentée par l'expression « dise roha qaro », nous voyons que ces éléments sont des ressources dont leur mot clé 1 commence par le caractère « e » et leur mot clé 2 commence par le caractère « o » et leur mot clé 3 commence par le caractère « q », selon la fonction de mesure de similarité que nous définissons, ces ressources sont proches. On peut dire la même chose pour l'expression « miva fimi dina » et les ressources commençant par « f », « n », « s ».

La figure suivante représente un extrait d'une mémoire associative d'un agent de la communauté pour montrer l'effet de l'apprentissage par renforcement.

RES/TAG	fefa weba lini	bemo laqo ...	vema nagi j...	vema nagi l...	xafe tasi ceti	jidi weba s...	qowe hayo ...	rogi java m...	miva xera t...	foxi boho n...
gfo,ivm,xbc	1									
eia,out,rgo		1								
dnm,kfy,thc			1							
dyn,kov,xpo				-1	1					
bsy,ifa,wpq						-1	1			
frh,nts,sgx								-1	1	
exi,lpx,wig										-1

FIG. 7.14 – un extrait d'une mémoire associative

La figure suivante montre les tags les plus utilisés avec leurs fréquences d'utilisation.

TAG	FRQ UTILIS
dise	82
miva	50
tasi	48
gawe	48
beta	47
veje	45
xemo	42
koce	42
soto	41
ceti	40
roha	38
qowe	37
xoge	35
qaro	35
sowa	33
pewi	31
dina	29
fimi	25
yico	23
xafe	17
cela	17
haxi	16

FIG. 7.15 – un extrait de la folksommie (les premiers tags)

Au niveau de l'environnement, il y a aussi des catégories émergentes représentées par les expressions de 3 tags. Par exemple dans figure suivante, les ressources commençant par le caractère « e » sont assignées par le tag « dise », les ressources commençant par le caractère « o » sont assignées par le tag « roha » et les ressources commençant par le caractère « q » sont assignées par le tag « qaro ».

En prenant l'intersection de ces 3 sous ensembles de ressources on trouve les éléments de la catégorie représentée par l'expression « dise roha qaro ».

Iteration	USER	TAG	Ressource
4	AGENT_71	dise	eoh,jrc,rkv
10	AGENT_73	dise	elm,out,tqh
19	AGENT_76	dise	etj,nir,qid
37	AGENT_69	dise	ecq,mit,twy
67	AGENT_56	dise	eox,icw,wtc
91	AGENT_6	dise	edn,okd,rvs
115	AGENT_84	dise	efy,oxl,qbg
118	AGENT_85	dise	eul,lyx,wkf
193	AGENT_37	dise	egy,mse,xfb
230	AGENT_1	dise	etx,puq,vrk
247	AGENT_57	dise	eni,jmv,wup
253	AGENT_7	dise	ech,puq,sbl
254	AGENT_7	dise	ech,puq,sbl
256	AGENT_4	dise	enc,put,shb
257	AGENT_4	dise	enc,put,shb
334	AGENT_51	dise	edy,nmc,rhu
340	AGENT_49	dise	exd,mgt,wta
367	AGENT_37	dise	ewy,kyl,xij
409	AGENT_50	dise	epx,lkv,qih
449	AGENT_85	dise	cpe,puk,rjx
479	AGENT_30	dise	dim,puc,qwr
496	AGENT_51	dise	efn,nby,url
505	AGENT_48	dise	elo,ngw,bf
532	AGENT_83	dise	epc,oqf,qci
535	AGENT_42	dise	ewd,pol,tfv

Iteration	USER	TAG	Ressource
8	AGENT_72	roha	fia,oha,tsd
11	AGENT_73	roha	elm,out,tqh
14	AGENT_74	roha	aqq,odf,tce
65	AGENT_57	roha	cbj,obj,svk
92	AGENT_6	roha	edn,okd,rvs
116	AGENT_84	roha	efy,oxl,qbg
440	AGENT_83	roha	dyl,oyg,rhn
500	AGENT_50	roha	cbj,obj,svk
533	AGENT_83	roha	epc,oqf,qci
599	AGENT_49	roha	fte,ohm,xii
686	AGENT_7	roha	aue,okx,rkm
701	AGENT_5	roha	dlu,oyd,rhv
767	AGENT_88	roha	gtr,orf,vnc
797	AGENT_84	roha	ekh,oto,xbt
818	AGENT_90	roha	ftk,okd,rjq
833	AGENT_70	roha	fsa,oyr,ril
1112	AGENT_61	roha	cny,oks,qaw
1121	AGENT_46	roha	hbm,oge,wsd
1169	AGENT_40	roha	fhc,ohw,rjb
1208	AGENT_27	roha	dbt,ovd,rjg
1262	AGENT_92	roha	frb,oyp,qnk
1283	AGENT_96	roha	hed,odm,uyr
1334	AGENT_20	roha	frb,oma,xqw
1343	AGENT_10	roha	dys,oej,tvm
1355	AGENT_29	roha	fia,oha,tsd

Iteration	USER	TAG	Ressource
21	AGENT_76	qaro	etj,nir,qid
117	AGENT_84	qaro	efy,oxl,qbg
240	AGENT_53	qaro	bcb,nnd,qub
411	AGENT_50	qaro	epx,lkv,qih
447	AGENT_59	qaro	ccg,mod,gro
480	AGENT_30	qaro	dim,puc,qwr
534	AGENT_83	qaro	epc,oqf,qci
621	AGENT_45	qaro	hrb,joc,qxm
888	AGENT_97	qaro	etr,pld,qmf
960	AGENT_15	qaro	dym,nvq,gok
990	AGENT_82	qaro	gkd,jom,qub
1059	AGENT_99	qaro	ceb,jqt,qev
1074	AGENT_81	qaro	hts,kmq,qmh
1113	AGENT_61	qaro	cny,oks,qaw
1134	AGENT_39	qaro	abk,jfd,qtr
1137	AGENT_2	qaro	cdj,ins,qsw
1146	AGENT_39	qaro	fts,ncl,qnk
1182	AGENT_21	qaro	etw,ngx,qep
1242	AGENT_22	qaro	ckt,pnb,qcm

FIG. 7.16 – les ressources assignées par les tags « dise », « roha », « qaro ».

7.8.3 Exemple 3

Nous utilisons pour cet exemple un nombre d'agents $N=10$ agents, un nombre de tags par expression $n=3$ et un seuil de mesure de similarité $S= 0.6$.

Nous voulons tester l'effet de la valeur du seuil de mesure de similarité sur le langage émergent (les catégories de ce langage).

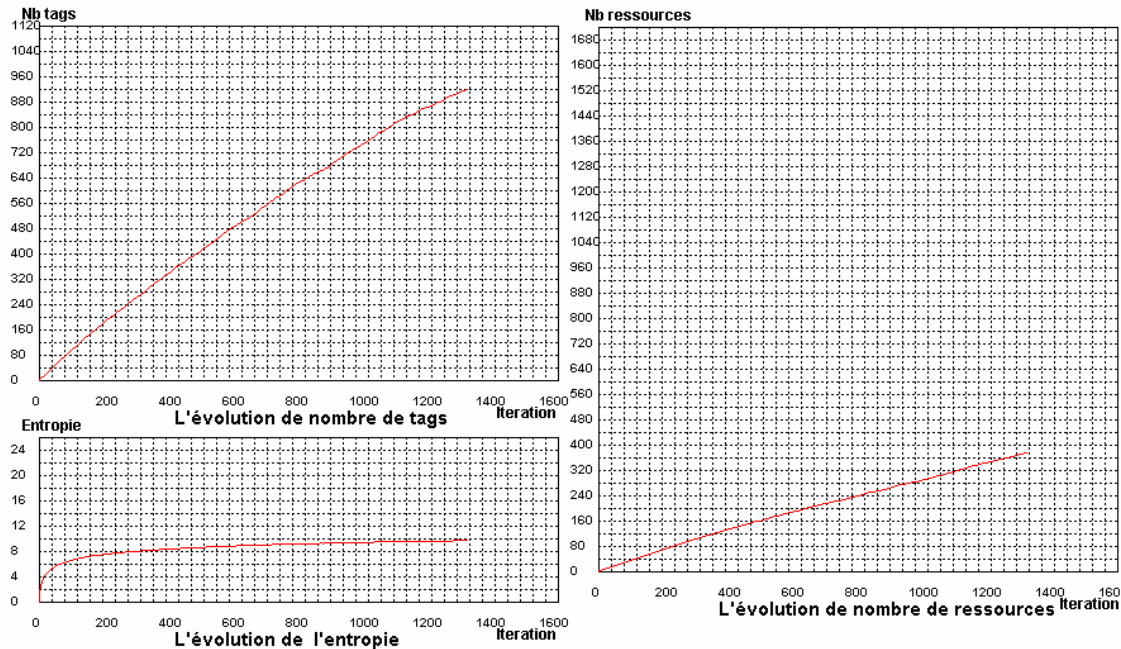


FIG. 7.17 – la dynamique de système de tagging

En analysant la dynamique dans les courbes ci-dessus, nous observons comme dans les exemples précédents la convergence d'une folksonomie à 917 tags pour une collection de 374 ressources pour cette simulation (1353 itérations).

Et pour l'entropie, nous avons une stabilisation de la quantité de l'information dans le système.

Pour étudier l'aspect linguistique de cet exemple, nous devons calculer la matrice de codage à partir de laquelle nous analysons le langage émergent et ses catégories.

RES/TAG	rowe diko gofe	hajo jele qeqi	joho cemi gofa	mice jete gofe	soqo pefi yohe	ceye haqi sapa	yido digi noma	yofi voye gami	kejo jele sapa
eko_mkp_xcb	1.0								
axb_mca_tsx		1.0							
guc_jmo_wxe			1.0						
fax_job_xgs				1.0					
fno_kst_wvb					1.0				
dch_leo_ros						1.0			
fnx_osf_sxh							1.0		
efj_okl_vpb								1.0	
hap_mcx_rj									1.0

FIG. 7. 18 – un extrait de la matrice de codage

La figure suivante représente les résultats obtenus dans la matrice » du codage sous forme des clusters.

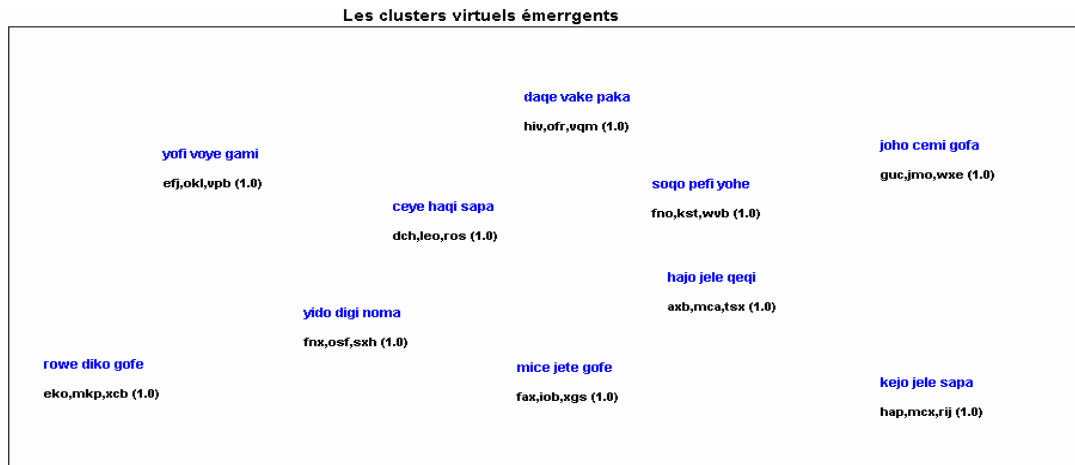


FIG. 7. 19 – un extrait de clusters émergents

Par exemple, si nous prenons la catégorie représentée par l’expression « joho cemi gofa » nous remarquons qu’elles contient un seul élément dont le mot clé 1 commence par le caractère «g » et le mot clé 2 commence par le caractère «j » et le mot clé 3 commence par le caractère «w » ,

Dans le système nous avons aussi des catégories émergentes représentées par les expressions de 3 tags. Si nous prenons l’intersection des 3 sous ensembles de ressources qui sont les ressources commençant par le caractère « g » et assignées par le tag « joho », les ressources commençant par le caractère « j » et assignées par le tag « cemi » et les ressources commençant par le caractère « w » et assignées par le tag « gofa », on trouve les éléments de la catégorie représentée par l’expression « joho cemi gofa». Comme la ressource « guc, jmo, wxe» qui se trouve dans les 3 sous ensembles de chaque tag.

Iteration	USER	TAG	Ressource
7	AGENT_4	joho	guc,jmo,wxe
391	AGENT_4	joho	gml,mfc,whm
931	AGENT_2	joho	gus,lht,tpw

Iteration	USER	TAG	Ressource
8	AGENT_4	cemi	guc,jmo,wxe

Iteration	USER	TAG	Ressource
9	AGENT_4	gofa	guc,jmo,wxe
642	AGENT_4	gofa	hgi,owp,wya

FIG. 7. 20 –les ressources assignées par les tags « joho », «cemi», « gofa».

7.8.4 Exemple 4

Nous utilisons pour cet exemple un nombre d'agents $N=100$ agents, un nombre de tags par expression $n=2$ et un seuil de mesure de similarité $S=0.3$.

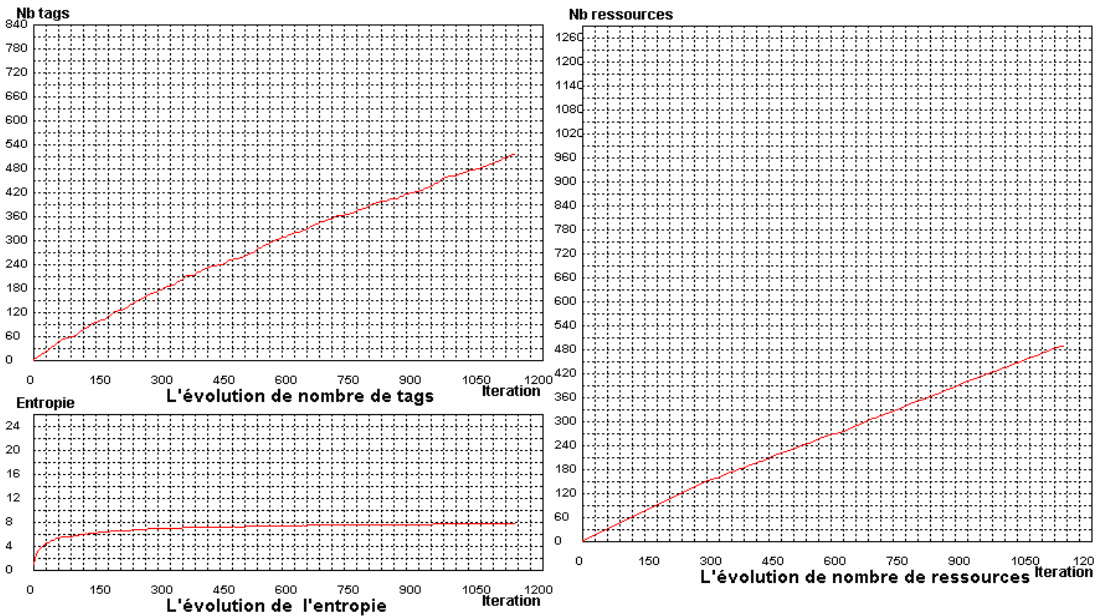


FIG. 7.21 – la dynamique de système de tagging

Nous remarquons la convergence d'une folksonomie à 515 tags pour une collection de 489 ressources pour cette simulation (1162 itérations). Nous avons une stabilisation de la quantité de l'information nécessaire pour ce système de codage.

Nous utilisons la matrice de codage pour montrer les catégories émergentes dans le système.

RES/TAG	vaxo feni	hago bava	hago qeqe	faro qoco	hago haga	hago feni	hago boyi	vaxo qoco	viye boyi
bs.phf			1.0						
bxq.pyr			1.0						
bht.pyv			1.0						
bqp.nhg		1.0							
bqv.nma		1.0							
bhs.ncj		1.0							
bct.nfe		1.0							
fgr.hwp	1.0								
fws.isv	1.0								
fla.lbc	1.0								
byg.lrx						1.0			
oka.lav						1.0			
lxc.lhg						1.0			
bed.lae					1.0				
baq.law					1.0				
byd.lxm					1.0				
cnh.jkh			1.0						
cvs.jmh			1.0						
clh.jde			1.0						
car.jlo			1.0						
hle.mux									1.0
hvj.mlg									1.0
hsg.mih									1.0
hor.mxb									1.0
fla.jml							1.0		
fem.jat							1.0		
flj.jik							1.0		
fes.jei							1.0		
bcc.mwx							1.0		
baf.mds							1.0		
buy.mqd							1.0		
bcs.mdb							1.0		
bpf.mjd							1.0		

FIG. 7.22 – un extrait de la matrice de codage

Nous utilisons toujours une visualisation sous forme de clusters pour montrer chaque catégorie émergente et ses ressources.

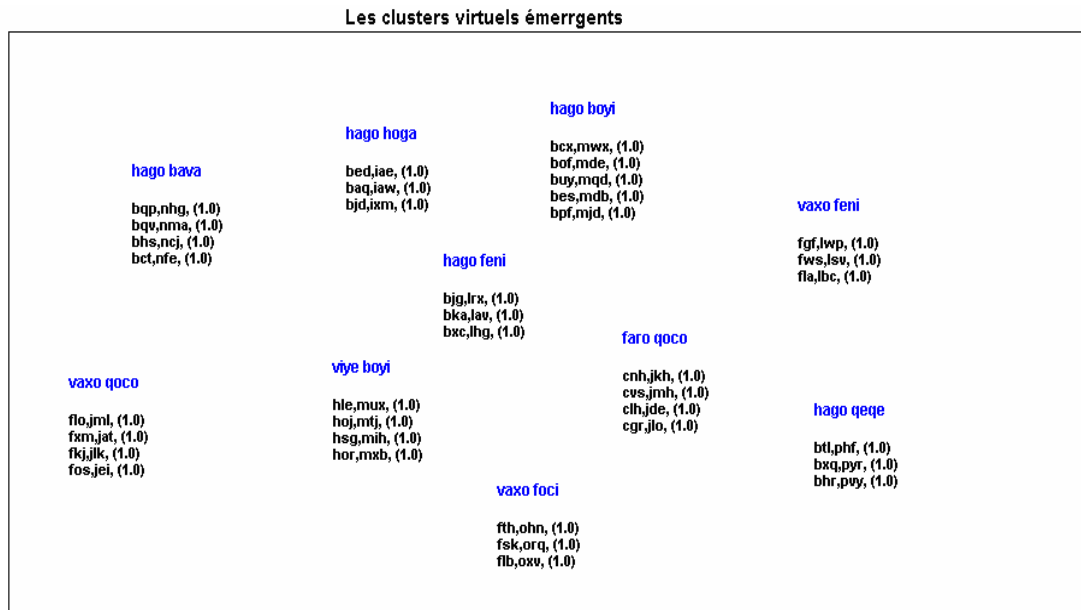


FIG. 7.23 – un extrait de clusters émergents

Pour $n=2$, nous avons obtenu un langage à structure grammaticale émergent sous forme des catégories des 2 tags.

En citant quelques exemples, la catégorie représentée par l'expression « hago bava » contient des ressources dont leur mot clé 1 commence par le caractère « b » et leur mot clé 2 commence par le caractère « n ».

Iteration	USER	TAG	Ressource
7	AGENT_52	hago	bjg,lrx,
17	AGENT_17	hago	btl,phf,
33	AGENT_40	hago	bed,iae,
49	AGENT_33	hago	bcx,mwx,
73	AGENT_23	hago	bof,mde,
81	AGENT_1	hago	bqp,nhg,
83	AGENT_3	hago	baq,iaw,
165	AGENT_34	hago	bqv,nma,
169	AGENT_32	hago	bhs,ncj,
199	AGENT_2	hago	bji,kox,
209	AGENT_6	hago	bql,oum,
219	AGENT_55	hago	bxq,pyr,
251	AGENT_17	hago	bcx,lhg,
379	AGENT_21	hago	bjg,lrx,
437	AGENT_23	hago	bes,mdb,
441	AGENT_10	hago	bct,nfe,
449	AGENT_63	hago	bep,ory,
457	AGENT_6	hago	bxw,ljp,
461	AGENT_42	hago	brt,kwa,
475	AGENT_57	hago	buy,mqd,
491	AGENT_12	hago	bhr,pvy,
495	AGENT_8	hago	bhj,ose,
519	AGENT_68	hago	bgq,kra,
559	AGENT_76	hago	bon,oau,
569	AGENT_89	hago	bid,ovg,
36	AGENT_38	bava	agv,nfj,
82	AGENT_1	bava	bqp,nhg,
96	AGENT_53	bava	bie,nft,
150	AGENT_40	bava	fbv,ngj,
166	AGENT_34	bava	bqv,nma,
170	AGENT_32	bava	bhs,ncj,
294	AGENT_31	bava	cki,ndh,
350	AGENT_57	bava	dje,nhq,
370	AGENT_8	bava	eif,nfi,
374	AGENT_18	bava	emw,nhl,
442	AGENT_10	bava	bct,nfe,
490	AGENT_52	bava	heo,nkh,
546	AGENT_68	bava	fqv,nmd,
582	AGENT_69	bava	dhq,nxp,
584	AGENT_75	bava	dvk,nsq,
592	AGENT_18	bava	esi,nym,
700	AGENT_98	bava	bdl,ndj,
746	AGENT_34	bava	bje,njh,
802	AGENT_10	bava	adv,nkt,
832	AGENT_50	bava	bfn,ncu,
850	AGENT_56	bava	gtn,njp,
878	AGENT_11	bava	eyv,nhw,
920	AGENT_79	bava	cpd,ner,
924	AGENT_67	bava	dky,nrd,
930	AGENT_90	bava	fbp,njs,

FIG. 7.24 – les ressources assignées par les tags «hago», «bava ».

Au niveau de l'environnement, le système de tagging collaboratif, les catégories émergentes représentées par les expressions de 2 tags comme le montre la figure précédente où les ressources commençant par le caractère « b » sont assignées par le tag « hago » et les ressources commençant par le caractère « n » sont assignées par le tag « bava ». En prenant l'intersection de ces 2 sous ensembles de ressources on trouve les éléments de la catégorie représentée par l'expression « hago bava ».

7.8.5 Résumé

En comparant les catégories émergentes de ce modèle avec celles du modèle précédent, nous remarquons qu'elles contiennent un petit nombre de ressources. Ceci est dû au fait que lorsqu'un agent catégorise les ressources selon trois dimensions (par exemple le domaine et le sous domaine et le type), il doit satisfaire trois contraintes pour juger qu'une ressource appartient à une catégorie.

Cette restriction ne permet pas l'émergence des catégories ayant un grand nombre de ressources proches.

Et si nous observons les catégories émergentes de l'exemple 3 où nous utilisons un seuil de similarité $S=0.6$, nous remarquons que les catégories émergentes ont une seule ressource.

Nous expliquons cela comme suit : Un seuil de similarité $S=0.6$ génère un degré de spécificité entre les ressources très élevée pour mettre les ressources dans une catégorie représentée par l'expression Exp.

Nous pouvons dire qu'avec un seuil de mesure de similarité sémantique élevée, nous aurons un langage où les correspondances entre les mots et les sens sont de type « expression/ressource » au lieu d'avoir un langage où les correspondances sont de type « expression/{ensemble de ressources} ».

7.9 Conclusion

Nous avons étudié dans ce modèle l'émergence d'un langage de tagging à structure grammaticale partagé par les utilisateurs du système de tagging collaboratif.

D'après les résultats des simulations cités ci-dessus, nous pouvons dire qu'un langage de tagging à structure grammaticale a émergé dans le système et au niveau des mémoires associatives des agents. Ce langage se caractérise par des correspondances de type (expressions/catégories de ressources) où nous considérons que chaque expression est une phrase si nous les comparons à un langage naturel.

Nous pouvons dire aussi que :

Des utilisateurs (agents) du système de tagging collaboratif doté par des mémoires associatives adaptées pour être capables de stocker des expressions grammaticales et,

Un processus de multi-tagging pour générer des expressions de n tags et,

Avec le principe de l'auto-organisation au niveau des interaction et,

En utilisant les scénarios des jeux de langage adapté au système de tagging collaboratif.

Peuvent partager un langage de tagging à structure grammaticale pour désigner l'ensemble des ressources du système.

CONCLUSION

GENERALE

Conclusion générale

Dans ce modeste travail, nous avons étudié la dynamique qui gouverne les systèmes de tagging collaboratif à travers le premier modèle en utilisant une modélisation des systèmes complexes à base des systèmes multi-agents afin de comprendre les principes qui génèrent leur évolution et qui mènent vers l'émergence des folksonomies, des systèmes de catégorisation créés et partagés par la communauté des utilisateurs de ces systèmes.

Cette approche de modélisation permet de la régénération de la dynamique observée au sein des systèmes de tagging collaboratif par l'auto-organisation et un mécanisme de sélectionnisme exprimé par le principe de l'imitation sociale et individuelle.

Notre objectif principal de ce travail est l'étude de l'émergence d'un langage de tagging partagé par les utilisateurs du système, nous avons traité cet objectif à travers deux modèles

Dans le deuxième modèle, en enrichissant le modèle précédent par des structures cognitif sous forme des mémoires associatives au niveau des agents taggeurs et des scénarios d'interactions à base de jeux de langage, nous avons trouvé que les utilisateurs du système de tagging collaboratif peuvent partager un langage de tagging émergent à structure lexicale.

Dans le troisième modèle, par la modification des règles d'interactions entre les agents et le système en utilisant un processus de multi-tagging et l'adaptation des mémoires associatives avec la nouvelle structure du langage, nous avons trouvé qu'un langage à structure grammaticale peut émerger au sein des systèmes de tagging collaboratif.

Perspectives

L'application de ces résultats sur des systèmes de tagging collaboratif existant ou la conception des nouveaux systèmes de tagging en prenant en compte l'aspect linguistique de ces systèmes sont des objectifs futurs et une continuation de ce travail.

L'utilisation d'autres modèles de l'émergence de langage est aussi une autre approche de traiter cette problématique pour mieux comprendre l'aspect cognitif du processus de tagging collaboratif.

Bibliographie

Bibliographie

- 1- Benz, D., Grobelnik, M., Hotho, A., Jäschke, R., Mladenic, D., Servedio, V. D., Sizov, S., al., Analyzing Tag Semantics Across Collaborative Tagging Systems. Dagstuhl Seminar 08391–Working Group Summary, 2008.
- 2- Cangelosi, A., Parisi, D., Simulating the Evolution of Language. London: Springer Verlag (eds.) 2001.
- 3- Cattuto, C., Baldassarri, A., Servedio, V. D. P., & Loreto, V., Vocabulary growth in collaborative tagging systems. Arxiv e-print: <http://arxiv.org/abs/0704.3316>, 2007.
- 4- Chaib-draa, B., Distributed artificial intelligence : An overview. In A. Ken, J. G. Williams, C. M. Hall, and R. Kent, editors, Encyclopedia Of Computer Science And Technology, volume 31, pages 215-243. Marcel Dekker, Inc, 1994.
- 5- Chomsky, N., Language and Nature, *Mind*, 104, pp. 1-61, 1995.
- 6- Dellschaft, K., Staab, S., An epistemic dynamic model for tagging systems. Proceedings of the 19th ACM Conference on Hypertext and Hypermedia HT'08 (p. 71-80), 2008.
- 7- Dik, S., The theory of Functional Grammar, Mouton De Gruyter (2 tomes: Part 1: The structure of the clause et Part 2: Derived and complex constructions), 1997.
- 8- Ferber, J., Les systèmes multi-agents, vers une intelligence collective. InterEditions, 1995.
- 9- Golder, S., & Huberman, B. A., The structure of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208, 2006.
- 10- Halpin, H., Robu, V., & Shepherd, H., The complex dynamics of collaborative tagging. Proceedings of the 16th international conference on World Wide Web (p. 211-220), 2007.
- 11- Jennings, N.R, Wooldridge, M., Sycara, K., A roadmap of agent research and development. *Int Journal of Autonomous Agents and Multi-Agent Systems*, 7-38, 1998.
- 12- Kaplan, F. L'émergence d'un lexique dans une population d'agents autonomes, Ph.D thesis, Université Paris VI, 2000.
- 13- Kaplan, F., La naissance d'une langue chez les robots, Hermès, 2001.
- 14- Ke, Jinyun, James W. Minett, Ching-Pong Au, and William S-Y. Wang, Self-organization and selection in the emergence of vocabulary. *Complexity* 7.3:41-54, 2002.
- 15- Kirby, S., Learning, bottlenecks and the evolution of recursive syntax. *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, ed. by Ted Briscoe, 173-205. Cambridge: Cambridge University Press, 2002 .

- 16- Munroe, S., Cangelosi, A., Learning and the evolution of language: the role of culture variation and learning cost in the Baldwin Effect. *Artificial Life* 8.4:311-340,2002.
- 17- Page L., Brin S., Motwani R., and Winograd T., The PageRank citation ranking: Bringing order to the web. WWW'98, pp. 161–172, Brisbane, Australia, 1998.
- 18- Puglisi, A., Baronchelli A., and Loreto V., Cultural route to the emergence of linguistic categories. arxiv: physics/0703164, March 2007.
- 19- Santos-Neto, E., Condon, D., Andrade, N., Iamnitchi, A., et Ripeanu, M., “Individual and social behavior in tagging systems,” Proceedings of the 20th ACM conference on Hypertext and hypermedia, p. 183–192, 2009.
- 20- Schenkel, R., Crecelius, T., Kacimi, M., Neumann, T., Parreira, J., Spaniol, M., Weikum, G., et al., Social wisdom for search and recommendation. *IEEE Data Engineering Bulletin*, 31(2), 40–49, 2008.
- 21- Shirky, C., *Ontology is Overrated: Categories, Links and Tags*.
[http://www.shirky.com/writings/ontology_overnated.html], 2005
- 22- Smith, G., “Atomic: Folksonomy: social classification”.
[http://atomiq.org/archives/2004/08/folksonomy_social_classification.html], 2004.
- 23- Steels, L., A self-organizing spatial vocabulary. *Artif. Life*, 2(3):319–332, 1995.
- 24- Steels, L., Self-organizing vocabularies. In C. Langton and T. Shimohara, editors, *Artificial Life V: Proceeding of the Fifth International Workshop on the Synthesis and Simulation of Living Systems*, pages 179–184, Cambridge,MA, USA,The MIT Press, 1996.
- 25-Steels, L., The synthetic modeling of language origins. *Evolution of Communication* 1.1-35, 1997.
- 26- Steels, L., *Language Games for Autonomous Robots*. IEEE Intelligent Systems, 2001.
- 27- Steels, L., Kaplan, F., Collective learning and semiotics dynamics, in D. Floreano, J-D. Nicoud, F. Mondada F. (éds.), *Advances in Artificial Life, Lecture Notes in Artificial Intelligence*, p. 704-708, Berlin, Springer- Verlag, 1999.
- 28- Steels, L., Kaplan, F., Bootstrapping Grounded Word Semantics, in T. Briscoe (éd.) *Linguistic evolution through language acquisition: Formal and Computational Models*, Cambridge University Press, 2001.
- 29- Tucker, M., Hirsch-Pasek, K., ‘Systems and language: Implications for acquisition’ in B. Smith and E. Thelen (eds): *A Dynamical Systems Approach to Development*. Cambridge MA: MIT Press, 1993.

- 30- Van Lier, L., *The Ecology and Semiotics of Language Learning*. Dordrecht: Kluwer Academic Publishers, 2004.
- 31- Yeung, C.A., Gibbins, N., Shadbolt, N. ,“A study of user profile generation from folksonomies,” *Proceedings of the Workshop on Social Web and Knowledge Management at WWW2008*, 2008.

Résumé

L'évolution de web vers les web 2.0 ou le web social offre plusieurs nouvelles opportunités aux utilisateurs à l'aide d'un ensemble de technologies logicielles comme les réseaux sociaux, les blogs et les systèmes de tagging collaboratif.

Dans les systèmes de tagging, les utilisateurs annotent les ressources par des mots clés (tags) pour catégoriser et organiser les ressources personnelles.

Dans ce travail, nous étudions l'émergence d'un langage de tagging formé par des couples(tag/ressource) partagés par la communauté des utilisateurs.

A travers des modèles basés sur l'approche des systèmes complexes et des systèmes multi-agents, nous montrons que ce langage de tagging peut émerger dans les systèmes de tagging collaboratif.

Mots clés : Systèmes de tagging, langage, émergence, le tagging collaboratif, tag, ressource, SMA.

Abstract

The evolution of the web to the web 2.0 or the social web offers a lot of new opportunities to the users with the help of a set of software technologies like the social networks, the blogs and the collaborative tagging systems.

In the tagging systems, the users annotate the resources with key words (tags) in order to categorize and organize the personnel resources.

In this work, we study the emergence of a tagging language formed by the (tag/resource) couples, shared by the community of the users.

Across models based on complex systems approach and multi-agents systems, we show that this tagging language can emerge in the collaborative tagging systems.

Key words: tagging system, language, emergence, collaborative tagging, tag, resource, MAS