



République Algérienne Démocratique Et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique  
Université de Saad Dahlab Blida  
Faculté des Sciences  
Département d'informatique



**Mémoire de fin d'études**  
**Pour l'obtention du diplôme de Master en Informatique**  
**Option: ingénierie du logiciel**

**Thème : Etude des performances des  
approches de calcul du contenu  
informationnel pour le calcul de la  
similarité sémantique**

**Promoteur:**

**-Mr Djilali nahal**

**Encadreur:**

**-Mr Cherif zakar amine**

**Réaliser par:**

**- Ferid Djamel Eddine**

**-Kheddam Redouane**

**Année Universitaire**

**2013/2014**

MA-004-216-1

# Dédicace

---

Toutes les lettres ne sauraient trouver les mots qu'il faut...  
Tous les mots ne sauraient exprimer la gratitude, l'amour, le respect, la  
Reconnaissance...  
Aussi c'est tout simplement que...

Je dédie ce mémoire ...

A Mes parents,

A mes frères,

Et à tous mes amis

***-Kheddam Redouane***

# Remerciements

---

On dit souvent que le trajet est aussi important que la destination. Les cinq années de master m'ont permis de bien comprendre la signification de cette phrase toute simple. Ce parcours, en effet, ne s'est pas réalisé sans défis et sans soulever de nombreuses questions pour lesquelles les réponses nécessitent de longues heures de travail.

Je tiens à la fin de ce travail à remercier ALLAH le tout puissant de m'avoir donné la foi et de m'avoir permis d'en arriver là.

Mes remerciements vont également à mes parents Lakhdar et aïcha, de tous les sacrifices qu'ils ont consentis pour me permettre de suivre mes études dans les meilleures conditions possibles et n'avoir jamais cessé de m'encourager tout au long de mes années d'études.

Je remercie infiniment le professeur Mr Djilali nahal , mon promoteur de mémoire dont la disponibilité, le savoir-faire et le soutien ne m'ont jamais fait défaut.

Je remercie tous les professeurs de la faculté des sciences de l'université de Blida, ainsi que tous les étudiants,

Je remercie également les responsables de la faculté des Sciences à l'université de Blida pour leur générosité et leur esprit d'ouverture qu'ils m'ont manifesté durant nos contacts.

Enfin, je remercie aussi tous ceux qui m'ont aidé de près ou de loin à réaliser ce travail.

**- Ferid Djamel Eddine**



## Table des matières

INTRODUCTION GÉNÉRALE.....	9
PROBLÉMATIQUE.....	10
OBJECTIF .....	10
<b>CHAPITRE1: LA SIMILARITE SÉMANTIQUE .....</b>	<b>11</b>
<b>Introduction .....</b>	<b>11</b>
1-Définition .....	12
2-Domains d'application .....	13
3-Propriétés Des Mesures de la similarité sémantique:.....	14
4-Terminologie et notation .....	15
5-Classification des approches de mesure de similarité .....	16
5.1-Approches basées sur les arcs .....	16
5.2-Approches basées sur le contenu informationnel.....	16
6-Calcul de la similarité sémantique.....	17
A. Approches basées sur les arcs.....	17
1-la mesure de Rada et al .....	17
2-La mesure de Hirst et St-Onge.....	17
3-La mesure de Wu & Palmer .....	18
4-La Mesure de Zargayouna.....	20
5-La Mesure de Sussna.....	21
6-La mesure de Leacock et Chodorow .....	22
B. Approches basées sur le contenu informationnel .....	22
Définition de contenu informationnel.....	22
1-la mesure de Resnik:.....	24
2-La mesure de Giang&Conrath.....	24
3-La mesure de lin.....	25
Discussion .....	25
Evaluation des mesures de similarité:.....	26

Conclusion.....	27
<b>CHAPITRE2:WORDNET .....</b>	<b>28</b>
<b>1-Introduction .....</b>	<b>28</b>
2-Histoire et membres de l'équipe .....	29
3-Wordnet Dans d'autres langues.....	29
4-WordNet comme ontologie.....	30
5-Structure de la connaissance .....	32
6-La hiérarchie est-un dans WordNet.....	33
7-Notion de synset .....	34
8-Relations entre les synsets définis dans WordNet.....	36
9-Partie de nom de WordNet.....	39
10-Notion d'instance hyponyme .....	39
11-Partie des adjectives de WordNet.....	40
12-Partie des adverbes de WordNet.....	41
13-Partie des verbes de WordNet.....	41
14-WordNet et les relations sémantiques .....	42
15-Applications .....	46
16-Limites de WordNet.....	47
17-Wordnet et la Mesure de la similarité sémantique.....	48
18-Conclusion .....	48
<b>CHAPITRE3: LES APPROCHES DE CALCUL DE CONTENU INFORMATIONNEL .....</b>	<b>49</b>
<b>1-Introduction .....</b>	<b>49</b>
2-Approche basé sur une ressource externe .....	50
2.1-L'Approche de Resnik .....	50
3-L'Approche basées sur WordNet.....	51
3.1- L'approche de Seco.....	51
3.2- L'approche de Zhou.....	53
3.3 -L'approche de Sanchez .....	54
3.4- L'approche de Sebti .....	56
Exemple de calcul et d'évaluation d'CI pour (Seco, Zhou et Sanchez, Sebti) .....	59
3.5- L'approche de Hanif Seddiqui.....	62
3.6. L'approche de Hadj Taieb.....	66
4-Comparaison de différents modèles d'IC .....	69
5-Discussion et solutions proposées.....	70
6-Conclusion .....	70

<b>CHAPITRE4:IMPLÉMENTATION .....</b>	<b>71</b>
<b>4.1Réalisation .....</b>	<b>71</b>
Introduction.....	71
Définition Java.....	71
Présentation de l'application .....	72
Explication.....	72
<b>4.2Tests et Résultats .....</b>	<b>73</b>
Introduction.....	73
4.2.1-Miller & Charles [1991] .....	74
4.2.2-Rubenstein&Goudenough [1965].....	82
Conclusion .....	93
<b>CONCLUSION GÉNÉRALE.....</b>	<b>103</b>
<b>RÉFERENCE.....</b>	<b>104</b>

## LISTE DES FIGURES

FIG1.1–Les relations conceptuelles [Wu & Palmer, 1994] .....	18
FIG1.2–Relations entre concepts.....	19
FIG1.3–Extrait de Wordnet .....	20
FIG1.4–Extrait de WordNet avec les probabilités correspondantes aux différents concepts .....	23
FIG2.1–Extrait de WordNet 2.0 .....	31
FIG2.2–Extrait de WordNet 2.0 .....	32
FIG2.3–Un schéma de la hiérarchie est-un dans WordNet.....	33
FIG2.4–Une illustration de <i>synsets</i> et <i>les relations</i> dans WordNet .....	37
FIG2.5–Structure d'adjectifs descriptifs.....	41
FIG3.1–les probabilités correspondantes aux différents concepts [Lin, 1998].	50
FIG3.2–Extraits du WordNet 2.0 .....	52
FIG3.3–Exemple de calcul d'IC pour quelques concepts avec l'approche de sebti .....	57
Fig3.4–Extrait de WordNet 2.0 .....	58
Fig3.5–Un exemple d'une taxonomie des concepts .....	59
Fig3.6–nombre de relations dans une taxonomie d'une ontologie.....	64
Fig3.7–Les sous arbre définissant CI (atropine) et CI (obidoxime chloride).....	66
Fig3.8–Extrait du WordNet 2.0 .....	67
Fig4.1–la Fenêtre principale de notre application .....	81

## LISTE DES Tableaux

TAB1.1 – Résultat de simWP avec trace d'exécution.....	19
TAB1.2 – Les paires des mots dans les expérimentations avec les jugements humains extrait de [Miller & Charles, 1991] .....	26
TAB2.1 –Relations entre les synsets définis dans WordNet.....	36
TAB2.2 –un comptage des relations de WordNet 2.1 par catégorie.....	38
TAB2.3 – les Indicateurs relationnel de WordNet .....	38
TAB2.4 – les Indicateurs réflexive de WordNet.....	38
TAB3.1 – les valeurs d'IC pour chaque concept dans la figure.....	60
TAB3.2 – contient des valeurs d'IC mesurées par Seco et Hanif Seddiqui .....	65
TAB3.3 – Comparaison des différents modèles d'IC.....	69



## Introduction générale:

Les êtres humains sont capables de juger de la parenté des mots (concepts) relativement facilement, et sont souvent en un accord général sur la façon dont les mots sont liés deux. Par exemple, peu de gens en désaccord que «crayon» est plus lié à "papier" que de "bateau". Miller et Charles [1991] attribuent cette perception humaine de parenté avec le chevauchement des représentations contextuelles de mots dans l'esprit humain, et il est à moins une certaine compréhension de la façon dont les humains sont capables d'effectuer cette tâche. Cependant, il reste un ouvert question sur la façon de créer des méthodes de calcul automatique qui attribuent les niveaux de relation ou scores pour des paires de concepts. Un certain nombre de mesures de la parenté ont été proposées, la plupart d'entre eux se fondant les informations issues de la base de données lexicale de WordNet, et peut-être augmentée avec corpus en fonction statistiques.

WordNet est structuré autour de la notion de synsets, c'est-à-dire en quelque sorte un ensemble de synonymes qui forment un concept. Un synset représente un sens de mot. Les synsets sont reliés entre eux par des relations, soit lexicales (antonymie par exemple) ou taxonomiques (hyperonymie, méronymie, etc).

L'utilisation de la similarité sémantique est une problématique qui ne cesse de prendre de l'essor d'une année à une autre dans le traitement automatique de langages naturels (TALN) et dans la recherche d'information (RI). Ses racines remontent à (Quillian, 1968) et (Collins, 1975) qui ont pensé à numériser le sens ou la sémantique. Cela permet de dépasser de nombreux problèmes liés aux comparaisons terme à terme. De nombreuses applications comme la désambiguïsation du sens des mots (Resnik, 1999) vu le problème de polysémie, extraction et la recherche d'information, détection et correction des lapsus (Budanitsky, 2001), segmentation du texte (Kozima, 1994), recherche d'images (Smeulders et al., 2000), l'indexation sémantique (Zargayouna, 2004a et 2004b) et le résumé automatique (Lin, 2003) utilisent désormais la similarité sémantique dans le but d'obtenir de meilleurs résultats.

Un problème récurrent de ces applications est la mesure de proximité entre concepts. Elle a été étudiée par de nombreux auteurs, et deux grandes approches basées sur les arcs (Hirst-St-Onge, 1998), (Leacock et Chodorow, 1998), (Wu et Palmer, 1994) en exploitant la structure de la ressource sémantique utilisée, et les approches utilisant le contenu informationnel (Resnik, 1995), (Jiang et Contrath, 1997) et (Lin, 1998).

Le contenu informationnel (CI) a été introduit pour la première fois dans la mesure de la relation sémantique entre concepts par Resnik en 1995. Le CI d'un concept traduit la pertinence de ce dernier dans un corpus en tenant compte de la fréquence d'apparition des mots auxquels il se réfère ainsi que de la fréquence d'apparition des concepts qu'il généralise. Il dénote la spécificité ou généralité d'un concept: plus un concept est spécifique plus son CI est important et plus il est général, plus son CI est petit.

Le contenu informationnel (CI) est une mesure de spécificité pour un concept. Des valeurs plus élevées sont associées à des concepts plus spécifiques (par exemple, voiture), tandis que ceux avec des valeurs plus basses être plus général (par exemple, objet).

## Problématique :

Le contenu informationnel joue un rôle important dans des mesures sémantiques de similarité.  
Comment acquérir une valeur appropriée d'CI ?

Il existe plusieurs approches pour le calcul de cette valeur. Deux grands axes sont distingués: le premier utilise une analyse statistique d'un corpus pour attribuer des probabilités à chaque concept, tandis que la deuxième méthode elle exploite la structure de WordNet sans faire recours à une ressource externe. Il inclut le modèle de CI basé sur les hyponymes, le modèle basé les feuilles basée, le modèle basé sur les relations et le modèle qui prend en considération la structure détaillée de l'ontologie Wordnet.

## Objectif :

L'objectif est d'étudier les performances des différentes approches de calcul du contenu informationnel et de dégager l'approche qui donne le meilleur résultat de calcul du CI. Les différentes approches de calcul du CI seront ensuite couplées avec les différentes méthodes de mesures de la similarité sémantique afin de donner les meilleurs couples (approche de calcul du CI et mesures de similarité sémantique). Les performances de chaque couple seront comparées aux jugements humains.

# -Chapitre 1 : La similarité sémantique

## Introduction :

Dans les dictionnaires explicatifs comme le Petit Robert, on trouve très souvent des synonymes ou des antonymes pour un mot quelconque. Par exemple, le mot « maison » et le mot « logement » sont synonymes, le mot « maison » et le mot « abri » sont aussi synonymes. Mais, comment peut-on dire que la connexion entre « maison » et « logement » est plus forte que celle entre « maison » et « abri » ? La réponse à cette question implique la notion de Similarité sémantique des mots qui peut se représenter par une valeur qui définit comment 2 mots se relie. Plus concrètement, si la similarité entre le mot  $m_1$  et le mot  $m_2$  est quantifiée par  $\text{sim}(m_1, m_2)$ , on peut dire que « maison » est plus proche de « logement » que de « abri » si on a  $\text{sim}(\text{maison}, \text{logement}) > \text{sim}(\text{maison}, \text{abri})$  et vice versa.

La formalisation et la quantification de la similarité des mots ont été introduites depuis très Longtemps. Cela remonte au moins à l'époque d'Aristote (384 – 322 B.C) (Budanisky, 1999), Mais ces préoccupations n'avaient pas, jusqu'à il y a peu, trouvé beaucoup d'applications concrètes.

Dans le domaine du Traitement Automatique du Langage Naturel (TALN), les relations Sémantiques comme la synonymie, l'antonymie, l'hyponymie, la méronymie, etc. sont des Notions particulièrement importantes. Avec le développement de l'informatique et du Web, il Y a chaque jour plus d'information disponible sur les pages du Web et sur les archives électroniques .C'est un avantage remarquable que l'on n'aurait jamais imaginé au début du siècle précédent. Cependant, pour les informaticiens, ça pose un problème pratique : comment trouver les informations utiles et gérer ces informations ? C'est la question qui a mené au développement du domaine de la Recherche d'Information (RI). Jusqu'à présent, la plupart des méthodes utilisées dans les moteurs de recherche sont basées sur l'analyse statistique des Occurrences des mots dans les documents. Cela marche bien dans beaucoup de cas, mais il y a

Encore des cas où ces méthodes ne sont pas satisfaisantes. Par exemple, quand on lance une Recherche avec le mot clé : «oiseau», on ne veut pas seulement obtenir les documents contenant le mot « oiseau». On attend aussi les documents qui contiennent les synonymes ou les mots étroitement liés avec le mot «oiseau» au niveau sémantique. Dans ce contexte, les Mesures de similarité entre les mots sont particulièrement utiles.

L'identification de la similarité dans les ontologies est un concept fondamental qui est adopté par plusieurs techniques telles que le regroupement, la fouille de données (datamining), le Web sémantique et en particulier, le domaine de la recherche de l'information.

Cette dernière repose largement sur des mesures pour l'identification de la similarité entre les documents [Bar 99] [Sal 83]. La majorité des approches de la recherche de l'information ne prennent en compte que des mots simples et/ou des fragments des mots pour la recherche des documents pertinents et ignorent l'idée essentielle qui prend en compte les rapports ontologiques des mots. Les rapports ontologiques entre les mots peuvent être détectés par un processus de calcul de similarité entre des paires d'objets contenus dans l'ontologie. Dans la littérature, plusieurs travaux sur la mesure de similarité sémantique entre les objets d'une ontologie ont été développés dans différents contextes.

Dans ce chapitre, nous étudions un certain nombre de ces mesures ,On peut distinguer deux grandes familles d'approches pour l'identification de la similarité sémantique. Les approches basées sur les nœuds [Res95] [Lin 98] [Giang97] utilisant typiquement des mesures du contenu informationnel pour déterminer la similarité conceptuelle. En plus, la similarité entre les concepts est déterminée par le degré de partage de l'information. L'autre famille d'approches repose uniquement sur la hiérarchie ou sur les distances des arcs [Rada 89] [Wup 94] Ce type de mesure s'appuie sur la structure de la ressource sémantique en proposant un comptage plus ou moins élaboré du nombre d'arcs séparant deux concepts. Ces mesures se servent de la structure hiérarchique de l'ontologie pour déterminer la similarité sémantique entre les concepts. Le principe de calcul de similarité avec cette approche est basé sur l'idée suivante : plus le chemin entre deux nœuds est court plus ils sont plus semblables. L'autre notion qui caractérise cette deuxième approche est que les arcs d'une taxonomie représentent des distances uniformes, par conséquent, cette approche présente l'inconvénient que tous les liens sémantiques possèdent le même poids ce qui impose des difficultés au niveau de la définition et du contrôle des distances des liens.

## **1-Définition :**

La similarité sémantique ou la distance sémantique est une notion utilisée pour désigner la mesure de ressemblance entre informations contenues dans des entités descriptives telle que les textes, pages web, images.

La similarité sémantiques réfère au degré auquel deux concepts sont liés (ou non). Les êtres humains sont capables de juger facilement si une paire de concepts sont liés d'une certaine façon. Par exemple, la plupart conviennent que le papier et le crayon sont plus liés que sont la voiture et brosse à dents. Cette thèse examine la question de savoir comment la similarité

sémantique peut être abordée d'un point de vue informatique, et se traduit par une nouvelle mesure de parenté qui sera être démontré à la fois efficace et adaptable.

L'objectif des mesures de similarité sémantique est d'évaluer la proximité sémantique entre les concepts (auxquels les termes des requêtes et documents sont rattachés). En recherche d'information, les mesures de similarité jouent un rôle important, en particulier dans le processus de désambiguïsation des concepts, la pondération des concepts et l'évaluation de la pertinence. De nombreuses approches ont été proposées pour évaluer la similarité sémantique entre deux concepts. Ces approches se divisent en deux catégories : les approches basées sur les arcs, les approches basées sur le contenu informationnel.

Cette mesure a intéressé diverses communautés de recherche en intelligence artificielle, psychologie et sciences cognitives. Elle a comme principaux champs d'application la recherche d'information et le traitement automatique de la langue.

La similarité sémantique prend en considération tout type de relation entre les concepts. Ainsi deux concepts peuvent être proches sémantiquement par leur similarité (e.g. voiture et automobile), mais aussi par d'autres relations comme partie-de (voiture-roue) ou contraire (guerre-paix), etc.

Deux approches principales sont utilisées pour la mesure de similarité entre concepts dans une ontologie :

- (1) en utilisant la structure arborescente de l'ontologie (nombre d'arcs)
- (2) en utilisant le contenu informatif des différents concepts en intégrant des mesures statistiques.

La plupart de ces propositions portent sur WordNet, elles peuvent cependant être généralisées à une ontologie puisqu'elles exploitent la structure taxonomique.

WordNet possède l'avantage d'être une ressource assez fournie, le grand nombre de synsets (synonymes) ainsi que sa facilité d'accès en ont fait une ressource très utilisée en RI et en TAL.

## **2-Domains d'application :**

Pour motiver l'importance de la mesure de similarité, nous récapitulons son utilisation dans plusieurs domaines d'applications.

### **Traitement du langage naturel (NLP) :**

Plusieurs travaux sur la mesure de similarité ont été motivés par le traitement du langage naturel (NLP). Parmi les travaux dans ce domaine on peut citer : le travail de [Pat 03] qui utilise la métrique de la similarité sémantique pour mesurer la similarité sémantique entre tous les sens de mot d'une paire donnée de mot et les désambiguïser ainsi dans un contexte donné. [McC 04] a combiné l'utilisation d'un thesaurus automatiquement acquis à partir des corpus textuels bruts et Wordnet (basé sur la métrique de la similarité) pour trouver des sens

prédominants des mots dans des textes non structurés. Les auteurs du travail [Gus 04] ont appliqué les mesures de similarité sémantique de WordNet pour évaluer la pertinence des expressions, étant donné un dialogue spécifique, et de construire automatiquement les sommaires du dialogue parlé. Le travail de [Hib 04] étudie l'utilité de la similarité sémantique dans le problème de correction d'orthographe, où des erreurs d'orthographe réelles sont détectées et corrigées automatiquement.

### Bio-informatique :

Une variation de mesure de similarité basée sur le contenu informationnel est adoptée pour trouver une meilleure façon d'organiser et d'interroger les données d'une ontologie de gène (GO). Le travail de [Lor 03] s'intéresse à la similarité sémantique entre les protéines, plutôt que les termes de l'ontologie GO, c'est pour cette raison qu'il a combiné entre trois mesures de similarités [Res 95][Lin 98] [Jiac 97].

### Web Services :

La détermination de la similarité des services sémantiques permet d'obtenir des informations utiles concernant leurs comptabilités. Dans le travail de [Jef 05], il y a une proposition des métriques pour mesurer la similarité des services sémantiques annotés avec une ontologie OWL. La mesure de similarité proposée est basée sur l'intuition que les objets similaires partagent les informations descriptives les plus communes.

### Détection des liens :

Dans le travail de [Che 04], il y a une description de l'amélioration des systèmes à base de détection de lien d'histoire en employant l'information spécifique de source et en combinant un certain nombre de mesures de similarité. Les mesures de similarités adoptées par ce travail sont représentées par cosinus, Hellinger, Tanimoto et clarity. Chacune de ces mesures capture différents aspects de la similarité des termes dans un document.

## 3-Propriétés Des Mesures de la similarité sémantique:

Les propriétés générales de la similarité sémantique aperçues par Lin s'appliquent universellement aux objets arbitraires et ont été théoriquement justifiées (Lin 1998). Elles ne sont pas limitées aux domaines particuliers d'application, aux ressources ou à une source spécifique de la connaissance. Généralement les propriétés à appliquer par les mesures de similarité sont classifiées dans les propriétés spécifiques de structure. Les propriétés de base définies par Lin incluent la propriété des caractéristiques communes, de différence et d'identité. Plus loin il a défini les propriétés spécifiques structurales qui incluent la profondeur. La section suivante explique les propriétés de base, et les propriétés spécifiques de structure.

## Propriétés de base :

### **La Propriété des caractéristiques communes:**

La détermination de la similarité est basée sur les caractéristiques communes possédées par les concepts étant évalués. La similarité entre C1 et C2 est basée sur les caractéristiques communes partagées par les concepts C1 et C2. Plus les caractéristiques qu'ils partagent plus communes plus ils sont semblables.

### **La Propriété de Différence :**

La détermination de la similarité est basée sur les caractéristiques uniques distinctes possédées par les concepts étant évalués. La similarité entre les concepts C1 et C2 est basée sur des caractéristiques distinctes possédées par les concepts C1 et C2. Plus les caractéristiques qu'ils ont plus distinctes, moins ils sont semblables.

### **La Propriété d'identité:**

La détermination de la similarité est basée sur des caractéristiques semblables possédées par les concepts étant évalués. La similarité entre les concepts C1 et C2 est basée sur la similarité des caractéristiques possédées par les concepts C1 et C2. Plus les caractéristiques qu'ils ont plus semblables, plus ils sont semblables.

## Propriété spécifique de structure:

### **La Propriété de profondeur :**

La mesure de similarité entre deux concepts C1 et C2 peut être quantifiée sur la base de l'emplacement des concepts dans la taxonomie. Bulskov indique que la distance représentée par deux concepts est influencée par la profondeur de l'endroit des concepts en taxonomie (Bulskov 2006). Par conséquent la propriété de profondeur influence le degré de similarité.

## 4-Terminologie et notation :

La similarité est la fonction inverse de la distance, plus deux mots sont similaires, moins ils sont distants. Une distance *dist* est une mesure qui respecte les trois propriétés suivantes :

- nullité de la distance d'un concept avec lui-même :  $\text{dist}(a,a) = 0$
- symétrie :  $\text{dist}(a,b) = \text{dist}(b,a)$
- inégalité triangulaire :  $\text{dist}(a,b) \leq \text{dist}(a,c) + \text{dist}(c,b)$

La notion de similarité devrait respecter les caractéristiques suivantes :

La similarité entre deux entités A et B est

- fonction des caractéristiques communes. Plus les entités ont des caractéristiques en commun, plus elles sont similaires.
- fonction de leurs différences. La similarité décroît inversement aux caractéristiques différentes.
- maximale quand A est identique à B.

### Définition :

Une mesure de similarité est une fonction  $sim : S \rightarrow [0,1]$  avec  $S$  l'ensemble de concepts.

Les propriétés que doit généralement respecter par une mesure de similarité :

- $sim(x,x) = 1$  : réflexivité
- $sim(x,y) = sim(y,x)$  : symétrie

## 5-Classification des approches de mesure de similarité :

Dans cette section nous classifions les approches principales de la mesure de similarité.

### 5.1-Approches basées sur les arcs :

La mesure de similarité la plus intuitive des objets dans une ontologie est leurs distances [Rada 89] [Lee 93] [Wup 94]. Evidemment, un objet  $X$  est plus similaire à un objet  $Y$  qu'un objet  $Z$ . Cette similarité est évaluée par la distance qui sépare les objets dans l'ontologie. Ces mesures se servent de la structure hiérarchique de l'ontologie pour déterminer la similarité sémantique entre les concepts. Le calcul des distances dans l'ontologie est basé sur un graphe de spécialisation des objets. Dans chaque graphe, la distance de l'ontologie doit être caractérisée par le plus court chemin qui fait intervenir un ancêtre commun ou le plus petit généralisant, connectant potentiellement deux objets à travers des descendants communs. Parmi les travaux classifiés sous cette bannière on peut citer :

- 1-La mesure de Rada et al [1989]
- 2-La mesure de Hirst et St-Onge[1998]
- 3-La mesure de Wu & Palmer[1994]
- 4-La mesure de Zargayouna
- 5-La mesure de Sussna
- 6-La mesure de Leacock et Chodorow

### 5.2-Approches basées sur le contenu informationnel :

Ces approches adoptent une nouvelle mesure en termes de la mesure entropique de la théorie de l'information [Lin 98] [Res 99]. La probabilité  $P(.)$  Pour l'identification de l'utilisation d'une classe ou de ses descendants dans un corpus désigne l'information de la classe. On définit l'entropie d'une classe par la formule suivante :  $E(c) = -\log(P(c))$  où  $P$  est la probabilité de trouver une instance du concept  $c$ . La probabilité d'un concept  $c$  est calculée en divisant le nombre des instances de  $c$  par le nombre total des instances. En associant des probabilités aux concepts d'une taxonomie, il est possible d'éviter le manque de fiabilité des distances des arcs. Cette caractéristique quantitative de l'information fournit une nouvelle façon de mesurer la similarité sémantique. Plus l'information est partagée par deux concepts,



Ou

T et K sont des constantes, dans la pratique les auteurs fixent  $T = 8$  et  $K = 1$

Chemin (C1,C2) : plus court chemin en nombre d'arc entre c1 et c2

D : le nombre de changement de direction

L'idée est que deux mots sont proches sémantiquement si leurs synsets sont connectés par un chemin qui n'est pas très long et qui ne change pas souvent de direction. S'il n'y a pas de chemin, ce calcul est égal à zéro.

La limite d'utiliser le chemin le plus court est qu'on ne prend pas en considération la position des concepts dans l'ontologie. Intuitivement, deux concepts classés en bas de l'ontologie sont très spécifiques et sont donc à un degré de granularité plus fin que deux concepts classés en haut de l'ontologie, ce qui n'est pas assuré par cette mesure.

[Wu & Palmer, 1994] apporte une réponse à ce problème en comptant la position des concepts par rapport à la racine de l'ontologie.

### 3-La mesure de Wu & Palmer:

[Wu & Palmer, 1994] Le principe de calcul de la similarité de Wu & Palmer est basé sur les distances, en nombre d'arcs,  $N1$  et  $N2$ , qui séparent les nœuds  $C1$  et  $C2$  du nœud  $R$  racine de la hiérarchie passant par le nœud généralisant le plus proche ( $C3$ ) et distance  $N3$  qui sépare le nœud généralisant  $C3$  de la racine  $R$ .

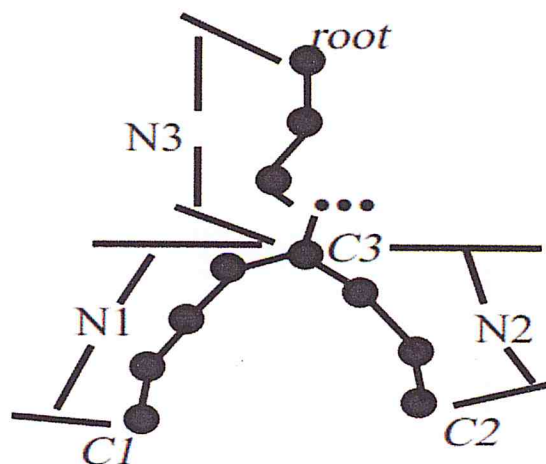


FIG 1.1– Les relations conceptuelles [Wu & Palmer, 1994]

Plus formellement cette mesure devient :

$$\text{simWP}(C1, C2) = 2 * \text{profondeur}(C3) / (\text{profondeur } c3(C1) + \text{profondeur } c3(C2))$$

Ou

$C3$  est le plus petit ancêtre commun entre  $C1$  et  $C2$

Profondeur(C3) est le nombre d'arcs minimal qui séparent C3 de la racine de la hiérarchie.

Profondeur c3(Ci) est le nombre d'arcs minimal qui séparent Ci de la racine de la hiérarchie. En passant par c3.

Cette mesure de similarité est intéressante parce qu'elle prend en considération le plus petit ancêtre commun (elle permet, donc, de prendre en considération la granularité). Plus cet ancêtre est général, moins les concepts sont similaires (et inversement). Néanmoins, cette mesure présente une limite car elle vise essentiellement à détecter la similarité uniquement par rapport à ce ppac. Dans le cas de recherche d'information, il est à notre sens plus intuitif de retourner en priorité les concepts qui sont subsumés par les concepts de la requête avant les concepts voisins. Si l'utilisateur recherche les documents qui contiennent chat, il est plus intéressant de lui retourner les différentes races de chat avant de lui retourner des documents qui contiennent les différents types de félin. La similarité doit de ce fait prendre en considération les liens père/fils et la densité des concepts. Nous voulons obtenir une mesure de similarité qui classe les concepts de la manière suivante :

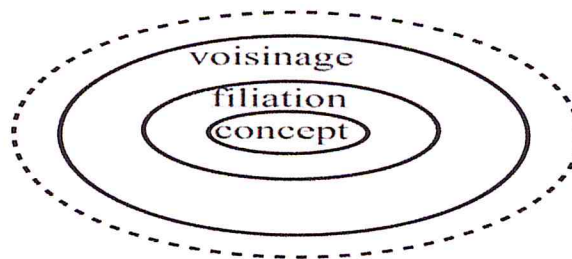


FIG1.2– Relations entre concepts

Ceci n'est pas assuré par la mesure de Wu & Palmer. Si nous calculons la similarité (voir figure entre *feline* et *blue-point-siamese-cat* (qui est un des descendants de *feline*) et entre *feline* et *canine* (qui est un frère) nous obtenons les valeurs rapportées dans le tableau : Or un félin est intuitivement plus proche d'un chat siamois bleu point, parce qu'ils partagent toutes les caractéristiques de félin, que d'un canin où ils partagent toutes les caractéristiques de carnivore qui est plus général que félin.

$sim(mot\#catégorie\#numsens, mot\#catégorie\#numsens)$	Trace d'exécution
$sim_{wp}(feline\#n\#1, canine\#n\#2) = 0.9$	ppac : <i>carnivore</i> #n#1 profondeur = 9 profondeur( <i>feline</i> #n#1) = 10 profondeur( <i>canine</i> #n#2) = 10
$sim_{wp}(feline\#n\#1, blue\_point\_siamese\_cat\#n\#1) = 0.833$	ppac : <i>feline</i> #n#1 profondeur = 10 profondeur( <i>feline</i> #n#1) = 10 profondeur( <i>blue\_point\_siamese\_cat</i> #n#1) = 14

TAB1.1– Résultat de simWP avec trace d'exécution

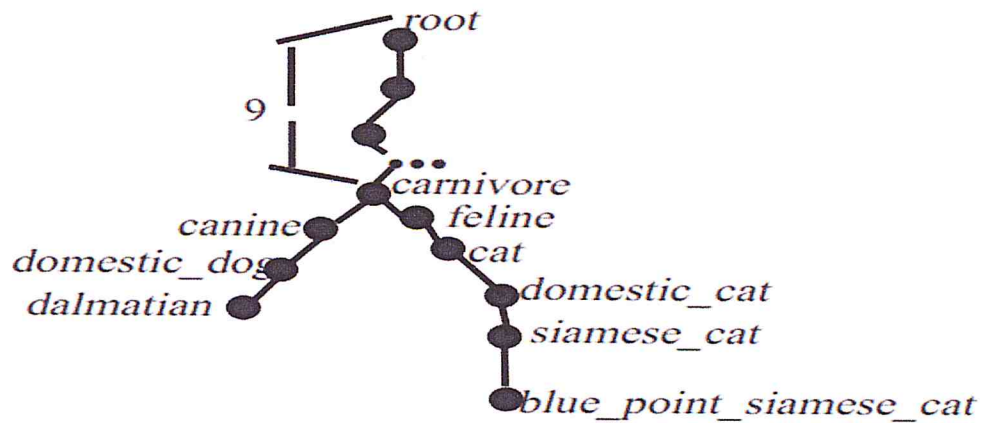


FIG1.3 – Extrait de Wordnet

#### 4-La Mesure deZargayouna :

Elle palie aux limites de la mesure de Wu &Palmer ,elle évite qu'un concept soit plus proche de son voisinage que de ses descendants. Pour cela Zargayouna pénalise les concepts qui ne sont pas de la même lignée ; elle rajoute une mesure *spec* qui prend en considération le degré de spécialisation du concept par rapport au nombre d'arcs qui le séparent de (l'anti-racine). Rappelons que le concept (l'anti-racine) n'existe pas dans l'ontologie, c'est un concept virtuel qui symbolise le concept le plus bas de l'ontologie, il est subsumé par tous les concepts de l'ontologie, nous avons donc distance (racine ,l'anti-racine) est égal à la profondeur de l'ontologie plus 1.

$$spec(C1, C2) = profondeur_b(C) * distance(C, C1) * distance(C, C2)$$

Avec

**profondeur<sub>b</sub>(C)**:le nombre d'arc qui sépare C (le ppac(C1; C2)) de l'anti-racine

**distance(C,Ci)** :la distance en nombre d'arc entre C et Ci.

**profondeur<sub>b</sub>(C)** calcule la spécialisation du plus petit ancêtre commun.

unemesure *spec* compte seulement pour les concepts qui sont frères (ou voisins), quand C=C1 et C2. Dans le cas où C = C1 ou C2, *spec* devient nulle.

Ainsi la mesure de similarité devient :

$$sim (C1, C2) = 2 * profondeur(C) / (profondeur c(C1) + profondeur c(C2) + spec(c1,c2))$$

Ou

C c'est le plus petit ancêtre commun entre C1 et C2

Profondeur(C) est le nombre d'arcs minimal qui séparent C de la racine de la hiérarchie.

Profondeur  $c(C_i)$  est le nombre d'arcs minimal qui séparent  $C_i$  de la racine de la hiérarchie.  
En passant par  $c$ .

### Remarque :

Dans le cas où *spec* est nulle, pour les concepts de même lignée, notre similarité revient à calculer  $\text{sim}_{\text{WP}}$ .

La mesure de Zargayounavérifie bien les caractéristiques suivant :

- La similarité augmente en fonction des caractéristiques communes. En effet, deux concepts Frères situés en bas de l'ontologie, sont plus similaires car ils ont plus de propriétés en communs que deux concepts frères situés en haut de l'ontologie.
- La similarité diminue en fonction des caractéristiques différentes. En effet, plus deux concepts sont éloignés, plus ils ont des propriétés différentes.
- La similarité est maximale (égale à 1) quand les deux concepts sont identiques.

### 5-La Mesure de Sussna:

La Mesure de Sussna est l'une des premières mesures de similarité basée sur WordNet. Pour déterminer la distance sémantique entre deux concepts par la mesure de Sussna, on cherche d'abord le plus court chemin entre leurs synsets dans la hiérarchie de WordNet. Chaque arc de la hiérarchie lui est associé un poids. La distance sémantique entre deux concepts est égale à la somme des poids des arcs du chemin reliant ces deux concepts

Chaque relation (arc) de la hiérarchie des noms de Wordnet est représentée par deux arcs pondérés : un pour la relation et l'autre pour son inverse. A chaque type de relation  $r$  de Wordnet est associé un intervalle de poids compris entre  $\text{min}_r$  et  $\text{max}_r$ . La valeur de la distance sémantique d'une relation de type  $r$  sortante d'un nœud  $c_1$  est :

$$\text{wt}(c_1 \rightarrow r) = \text{max}_r - (\text{max}_r - \text{min}_r) / \text{Edges}_r(c_1)$$

avec  $\text{edges}_r(c_1)$  est le nombre de relation de type  $r$  sortante de  $c_1$ .

Le poids de la distance sémantique d'un arc est également réduit par la profondeur de relation dans la hiérarchie. La valeur de la distance sémantique d'un arc reliant deux nœuds adjacents  $c_1$  et  $c_2$  est donnée par:

$$\text{Dists}(c1, c2) = \text{wt}(c1 \rightarrow r) + \text{wt}(c2 - r') / [2 \times \max(\text{depth}(c1), \text{depth}(c2))]$$

Avec  $r'$  la relation inverse de  $r$ .

La distance sémantique entre n'importe quelle paire de nœuds  $c_i$  et  $c_j$ , est calculée comme étant la somme des distances entre les paires de nœuds adjacents dans le plus court chemin reliant  $c_i$  et  $c_j$ .

### **6-La mesure de Leacock et Chodorow:**

Une méthode intuitive pour mesurer la similarité sémantique des sens de mot en utilisant WordNet, compte tenu de sa structure arborescence, serait de compter le nombre de liens entre les deux synsets. Plus la longueur du chemin entre eux est courte, plus ils sont considérés plus connexe. Une telle mesure avait été expérimentée avec

La mesure de Leacock et Chodorow consiste à trouver le plus court chemin en nombre de nœuds et non pas en nombre d'arcs dans la hiérarchie de Wordnet.

$$\text{Sim}(c1,c2) = -\log(\text{chemin}(c1,c2) / 2^*D)$$

Ou

- Chemin ( $c1,c2$ ) = le nombre de nœuds faisant partie du chemin reliant  $c1$  et  $c2$ ,  $c1$  et  $c2$  exclus.
- La quantité ( $\text{chemin}(c1,c2) / 2^*D$ ) est comprise entre 0 et 1.
- $D$  = la profondeur maximale reliant  $c1$  et  $c2$  à leur racine dans la hiérarchie.

## **B. Approches basées sur le contenu informationnel :**

### **Définition de contenu informationnel :**

Les mesures de Resnik, Jiang & Conrath et lin sont fondées sur la notion de contenu informatif. La notion de contenu informatif (CI) a été introduite pour la première fois par Resnik, dans son article qui décrit une nouvelle méthode pour calculer relations sémantiques. Elle utilise conjointement l'ontologie et un corpus. Le contenu informatif d'un concept traduit la pertinence d'un concept dans le corpus en tenant compte de la fréquence de l'apparition des mots auxquels il se réfère ainsi que de la fréquence d'apparition des concepts qu'il généralise. En bref, le contenu informationnel d'un concept mesure la spécificité ou la généralité de ce concept, Par exemple, un concept comme voiture est un concept très actuel et serait avoir un contenu d'information élevé. D'autre part, un concept plus général tel objet aurait un beaucoup plus faible contenu en information.

Plus précisément le contenu informatif se calcule par la formule suivante :

$$\text{CI}(C) = -\log(p(C))$$

Où  $p(C)$  est la probabilité de retrouver qu'un mot du corpus soit une instance du concept  $C$

(un des mots référés par le concept C ou par un de ses descendants). Elle est monotone quand on remonte dans la hiérarchie ( $p(A) \leq p(B)$ ) si A est un ancêtre B (e.g. dans la figure)  $p(\text{hill}) < p(\text{geological -formation})$ . Dans les expérimentations de Resnik, ces probabilités sont calculées par :  $p(C) = \text{frequence}(C) / N$

où

N est le nombre total de concepts

$\text{frequence}(C) = \sum_{w:\text{instance}(C)} \text{count}(w)$ .

Plus un concept est général, plus son contenu informatif est faible, ainsi  $CI(p(\text{racine})) = 0$  parcequ'il a un contenu informatif nul. A l'inverse, plus le concept est spécifique plus son contenu informatif est important (e.g. le contenu informatif de hill est plus important que celui de geological-formation). L'intuition de la notion de contenu informatif est que la similarité entre deux concepts est la portion d'information qu'ils ont en commun qui, dans le cadre d'une ontologie, peut être déterminée par le concept le plus spécifique qui les subsume (ppac). Cette intuition est indirectement appliquée par les mesures présentées dans la section précédente qui calculent la similarité avec le nombre d'arcs qui séparent deux concepts.

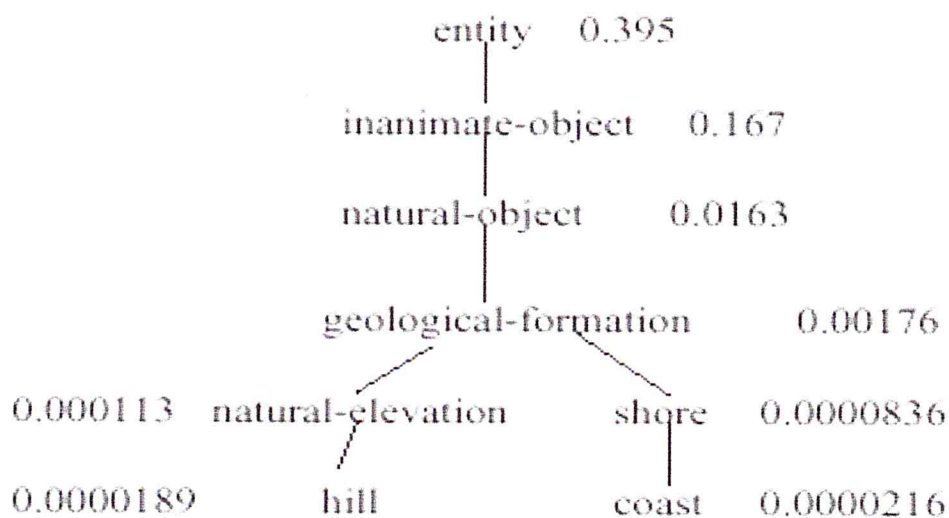


FIG1.4 – Extrait de WordNet avec les probabilités correspondantes aux différents concepts

## 1-La mesure de Resnik:

Resnik définit la similarité sémantique entre deux concepts par la quantité d'information qu'ils Partagent. Cette information partagée est évaluée numériquement par le contenu informatif du plus petit ancêtre commun (PPAC).

Donné deux concepts C1 et C2, la similarité sémantique entre les concepts C1 et C2 est défini comme teneur maximum en information du concept superbe qui englobe les concepts C1 et C2.

Selon Resnik, la similarité dépend de la quantité de l'information part de deux concepts en commun Resnik Cette information partagée est fournie par plus le concept commun d'abstraction (PPAC) de détail qui englobe les concepts. Si le concept de PPAC n'existe pas, alors les deux concepts sont au maximum différents, différemment l'information partagée est égale à la valeur d'IC de leur PPAC. La formule de Resnik est modelée comme suit :

$$\text{simRes}(C1, C2) = CI(\text{ppac}(C1, C2))$$

Ainsi, si deux termes sont très éloignés et ont comme ppac la racine, leur similarité est égale à 0. L'approche de Resnik essaie d'éviter le problème de granularité, cité au-dessus, en diminuant le rôle des arcs dans le calcul de similarité. En effet, les arcs ne sont utilisés que pour retrouver le ppac, elle est de ce fait un peu sommaire car nous pouvons avoir  $\text{ppac}(\text{hill}, \text{shore}) = \text{ppac}(\text{shore}; \text{natural-elevation})$  même si shore et natural-elevation sont plus proches de leur ppac (geological-formation) que hore et hill.

## 2-La mesure de Giang&Conrath:

Donné deux concepts C1 and C2, la similarité sémantique entre deux concepts est la différence de leur masse sémantique (contenu de l'information) s'ils sont sur le même axe. Si les concepts sont sur différentes axe puis la similitude des concepts C1 et C2 est définie comme addition de la teneur en information des concepts C1 et C2 calculés à partir de chaque nœud à un nœud superbe commun des deux concepts.

La mesure de Jiang & Conrath pallie les limites de la mesure de Resnik en combinant le contenu informatif du ppac à ceux des concepts .Il reprend le contenu de l'information des deux concepts, ainsi que celle de leur plus bas subsumer commun. La mesure est une distance de mesure qui indique la mesure de similarité de deux concepts.

La distance est définie :

$$\text{distance}(C1, C2) = CI(C1) + CI(C2) - (2CI(\text{ppac}(C1; C2)))$$

La mesure de similarité revient donc à calculer :

$$\text{sim}_{JC}(C1, C2) = 1/\text{distance}(C1, C2)$$

### 3-La mesure de lin:

Une autre mesure, en fonction du contenu de l'information de concepts, est décrit par [Lin, 1998] il propose une mesure de similarité qui calcule la proportion d'information commune entre deux concepts par rapport à leur description. La théorie de sa mesure repose sur les trois intuitions suivantes :

- Intuition 1: La similarité entre deux objets A et B est lié à leur information commune. Plus l'ensemble de leur intersection est importante, plus ils sont similaires.
- Intuition 2: La similarité entre A et B est liée à leur disparité (différences). Plus il y'a de différences, moins ils sont similaires.

Intuition 3: La similarité entre A et B est maximale lorsque A et B sont identiques

La mesure de lin :

$$\text{Sim}_{\text{Lin}}(c1, c2) = 2 * \text{CI}(\text{ppac}(C1; C2)/(\text{CI}(C1) + \text{CI}(C2)))$$

### -Discussion:

Le calcul de la similarité par nombre d'arcs avec une restriction sur le lien « is-a » pose deux

Problèmes :

– La granularité : Les concepts placés en bas de l'ontologie, donc plus spécifiques sont plus

Similaires que les concepts en haut de l'ontologie (cf. exemple donné plus haut).

– La densité : Pour un concept qui a plusieurs fils, il paraît logique de supposer que sa similarité avec ses fils est plus grande qu'avec les autres concepts. Ainsi dans les parties plus denses de l'ontologie, la similarité doit augmenter.

Ces problèmes peuvent être résolus en associant des poids aux liens. L'affectation de ces poids peut être basée sur : les types de liens présents, la profondeur du lien dans la taxonomie et la densité du concept par ses voisins immédiats Sussna.

La performance des mesures hybrides ainsi que celles fondées sur le contenu informatif dépend de la qualité du corpus sur lequel sont calculés les contenus informatifs.



## -Evaluation des mesures de similarité:

Comment peut-on juger si une mesure de similarité est performante et comment peut-on comparer les performances de deux mesures ?

L'évaluation des mesures de similarité peut se faire en étudiant la corrélation avec les jugements humains. Rubenstein&Goudenough [1965] furent les premiers à proposer de conduire une expérience quantitative avec 51 sujets humains qui ont été amenés à évaluer 65 paires de mots en leur donnant une note allant de 0 à 4. Les paires de mots varient de très similaires (les synonymes) à des couples non reliés. Le but de cette expérimentation était d'expliquer la perception humaine de la synonymie. Miller & Charles [1991] ont reproduit la même expérimentation en étudiant 30 paires de mots extraites de celles proposées par Rubenstein&Goudenough (10 paires ayant un score entre 3 et 4, 10 paires ayant un score entre 1 et 3 et 10 autres ayant un score entre 1 et 0). L'expérimentation a porté sur 38 sujets. La corrélation entre les jugements humains de Rubenstein&Goudenough et Miller & Charles est de 0.95. Les expérimentations sont été effectuées à 25 ans d'intervalle, ce qui peut expliquer la variation de perception de similarités, qui peut aussi être plus simplement due au fait que les sujets ne sont pas les mêmes (voir tableau). Ces paires de mots ont été reprises par d'autres expérimentations ([Resnik, 1995; Lin, 1998; Hirst & St-Onge, 2004; Patwardham, 2003]), le but étant de détecter les corrélations avec les résultats des différentes mesures de similarités. De telles expérimentations ont des limites, nous en dégagons quelques-unes dans ce qui suit.

TAB1.2– Les paires des mots dans les expérimentations avec les jugements humains (extrait

Paires de mots	M&C		R&G		Paires de mots	M&C		R&G	
	score	rang	score	rang		score	rang	score	rang
car-automobile	3.92	1	3.92	2	crane-implement	1.66	15	2.37	14
gem-jewel	3.84	2	3.94	1	journey-car	1.16	16	1.55	15
journey-voyage	3.84	2	3.58	6	monk-oracle	1.10	17	0.91	21
boy-lad	3.76	3	3.82	3	cemetery-woodland	0.95	18	1.18	17
coast-shore	3.70	4	3.60	5	food-rooster	0.89	19	1.09	21
asylum-madhouse	3.61	5	3.04	9	coast-hill	0.87	20	1.26	16
magician-wizard	3.5	6	3.21	7	forest-graveyard	0.84	21	1.00	19
midday-noon	3.42	7	3.94	1	shore-woodland	0.63	22	0.90	22
furnace-stove	3.11	8	3.11	8	monk-slave	0.55	23	0.57	24
food-fruit	3.08	9	2.69	11	coast-forest	0.42	24	0.85	23
bird-cock	3.05	10	2.63	12	lad-wizard	0.42	24	0.99	20
bird-crane	2.97	11	2.63	12	chord-smile	0.13	25	0.02	27
tool-implement	2.95	12	3.66	4	glass-magician	0.11	26	0.44	25
brother-monk	2.82	13	2.74	10	rooster-voyage	0.08	27	0.04	26
lad-brother	1.68	14	2.41	13	noon-string	0.08	27	0.04	26

de [Miller & Charles, 1991].

# -Chapitre 2 : Wordnet

## 1-Introduction :

WordNet est une base de données lexicale pour l'anglais. Cette base de données est organisée en fonction de la signification des mots et non pas sur leur forme lexicale. L'unité de base de Wordnet et le Synset.

Il groupe des mots anglais en jeux de synonymes appelés les synsets ,Un synst est un ensemble de mots qui ont la même signification (synset = ensemble de synonymes).

Fournit sous peu, des définitions générales, et les diverses relations sémantiques entre ces ensembles de synonyme. Le but est double : pour produire une combinaison de dictionnaire et de thesaurus qui est plus intuitivement utilisable, et soutenir des applications d'analyse automatique des textes et d'intelligence artificielle. Les outils logiciels de base de données et ont été libérés sous un permis de modèle de schéma et peuvent être téléchargés et employés librement. La base de données peut également être passée en revue en ligne

WordNet peut être visualisé comme un réseau sémantique, où chaque nœud du réseau représente un concept de monde réel. Par exemple, le concept peut être un objet comme house, ou une entité comme teacher, ou un concept abstrait comme art, et ainsi de suite.

Chaque noeud se compose d'un ensemble de mots, chacun qui représente le concept de monde réel lié à ce noeud. Ainsi, chaque noeud est essentiellement un ensemble de synonymes qui représentent le même concept. Par exemple, le concept d'une voiture peut être représenté par l'ensemble de mots car, auto, automobile, motorcar. Un tel ensemble, dans la terminologie de WordNet, est connu comme synset. Un synset également a associé à lui une définition ou une description courte du concept de monde réel connu sous le nom de gloss.

Les synsets et les gloss dans WordNet sont comparables au contenu d'un dictionnaire Ordinaire.

En date du novembre 2012 la dernière version de WordNetEn ligne est 3.1 (annoncé le juin 2011), mais la dernière version libérée est 3.0 (libéré le décembre 2006). La base de données WordNet 3.0 contient 155.287 mots organisés dans 117.659 synsets, sous la forme compressée, c'est environ 12 méga-octets dans la taille

## 2-Histoire et membres de l'équipe :

WordNet a été créé au laboratoire de la Science cognitive de l'Université de Princeton sous la direction de professeur George Armitage Miller de psychologie. Wordnet est encore maintenu par le laboratoire de la Science cognitive. Le développement a commencé en 1985. Au cours des années, le projet a reçu le placement des organismes gouvernementaux intéressés à la traduction automatique. À partir de 2009, l'équipe de WordNet a inclus les membres suivants du laboratoire de la Science cognitive : George Armitage Miller, Christiane Fellbaum, RandeeTengi, Pamela Wakefield, Helen Langone et Benjamin R. Haskell.

WordNet a été soutenu par des concessions du National Science Fondation, de DARPA, du bureau disruptif de technologie (autrefois l'activité avancée de recherche et développement), et de RÉFLEXE. George Miller et Christiane Fellbaum ont été attribués le prix 2006 d'Antonio Zampolli pour leur travail avec WordNet.

## 3-Wordnet Dans d'autres langues :

Il existe des projets similaires pour d'autres langages.

### Langue arabe

AWN Browser( Arabic wordnet Browser) version 2.0 beta 2007

### Langue française

Le *Wordnet Libre du Français (WOLF)* est une ressource linguistique généraliste libre pour le français issue d'une traduction automatique de Wordnet (depuis la version 0.2.5). Il a été initié le [15 février 2008](#) et est à ce jour encore en version bêta (version 1.0 bêta publiée en octobre 2012). Il est à noter que ce projet bénéficie à présent d'une validation manuelle partielle des littéraux qu'il contient.

### Langues européennes :

Le projet *EuroWordnet* développé entre 1996 et 1999 a permis la constitution de ressources "Wordnet" pour plusieurs langues européennes, le hollandais, l'italien, l'espagnol, l'allemand, le français, le tchèque et l'estonien. Il est à noter que ces "Wordnet" ne sont pas libres et leurs licences sont payantes : l'ELDA/ELRA se charge de les commercialiser.

Les langues sont reliées par l'intermédiaire de l'Index-Inter-Langues. Il est ainsi possible de passer des mots dans une langue aux mêmes mots dans n'importe quelle autre langue. EuroWordNet permet donc une recherche d'information monolingue ou multilingue.

### Langues chinois :

CWN (Wordnet chinois) soutenu par l'[université de Taiwan national](#).

### Langues finlandaise :

FinnWordNet est une version finlandaise du WordNet où toutes les entrées du WordNet anglais original ont été traduites.

### Langues allemande :

GermaNet est une version allemande du WordNet développé par l'université de Tübingen.

### Langues portugaise brésilienne :

OpenWN-PINTE est une version portugaise brésilienne du WordNet original librement disponible pour le téléchargement sous le permis de CC-BY-SA.

### Autres Langues :

La Global Wordnet Association (GWA) recense au total une petite centaine de "Wordnet" dans différentes langues à travers le monde.

Nous citerons ainsi le projet BalkaNet qui a permis la construction de six "Wordnet" pour le bulgare, le tchèque, le grec, le roumain, le turque et le serbe

## 4-WordNet comme ontologie:

Les liens de hypernym/hyponym parmi les synsets de nom peuvent être interprétés comme relations de spécialisation entre les catégories conceptuelles. En d'autres termes, WordNet peut être interprété et employé comme ontologie lexicologique dans le sens de l'informatique.

Une ontologie  $O = \langle S, R, \text{type}, \text{racine}, \rangle$

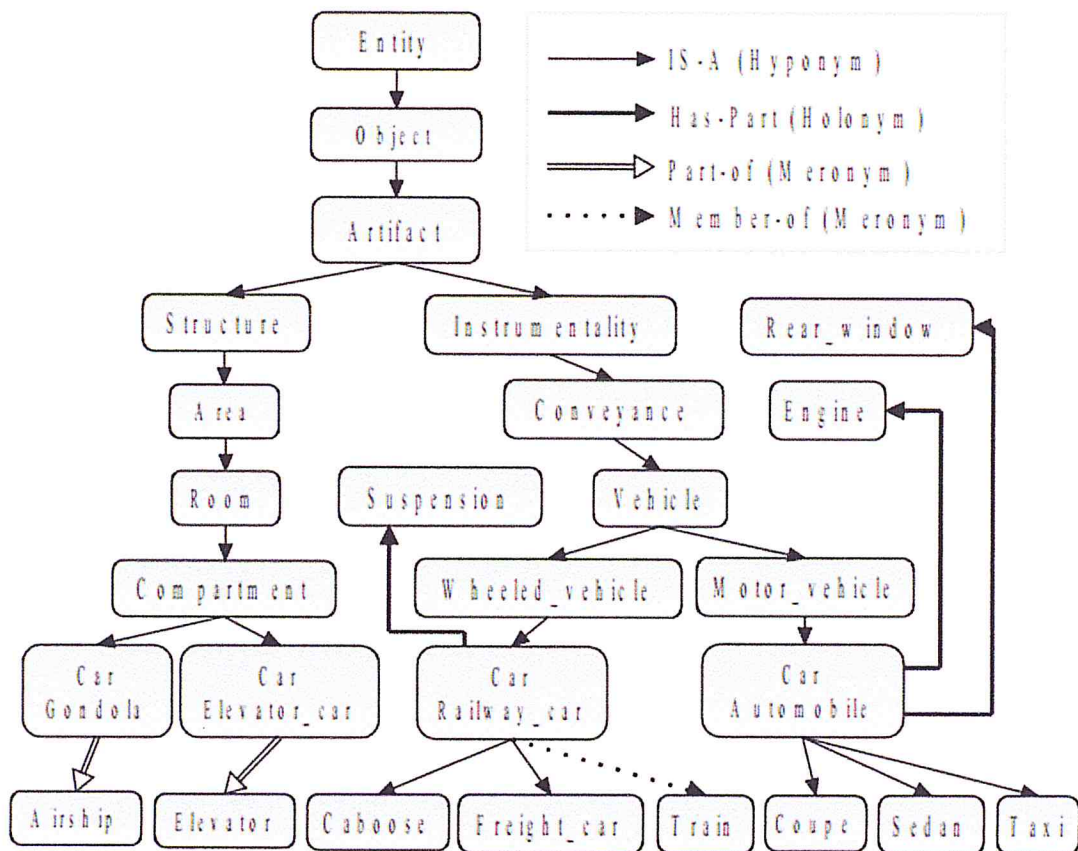
– S : l'ensemble des concepts de l'ontologie(synset).

– R : l'ensemble des relations entre les concepts.

– type : une fonction qui associe un type à une relation (e.g. is-a(R) est la relation hiérarchique de spécialisation).

– >racine : est la racine de l'ontologie.

WordNet peut donc être représenté comme une ontologie, où S est l'ensemble des synsets, R est l'ensemble de relations (hyperonymie, hyponymie, antonymie, etc.).



**Fig 2.1 Extrait de WordNet 2.0**

Racine(root) est un concept que nous ajoutons au-dessus des 9 entrées de WordNet, Pour la position de nom, dans WordNet 2.0 racine est un concept que nous ajoutons au-dessus des 9 entrées suivant (entity#1#1, event#1#4, group#1#3, phenomenon#1#2, possession#2#7, psychologicalfeature #1#1, state#4#8, abstraction#6#6 et act#2#5) comme montre la figure suivante

Pour s'assurer d'avoir toujours un chemin entre deux concepts. La plupart des mesures reposent sur le lien taxonomique is-a(hyperonyme, hyponyme) pour le calcul de similarité .La relation de subsumption "is-a" définit un lien de specialisation/généralisation, elle est utilisée pour structurer les ontologies. Cette relation permet formellement l'héritage de propriétés des concepts.

Chaque concept excepté la racine a un ou plusieurs hyperonyme directs (jusqu'à 6 dans WordNet 2.0).dans la figure suivante le concept atropine #1#1 a cinq hyperonymes directs.

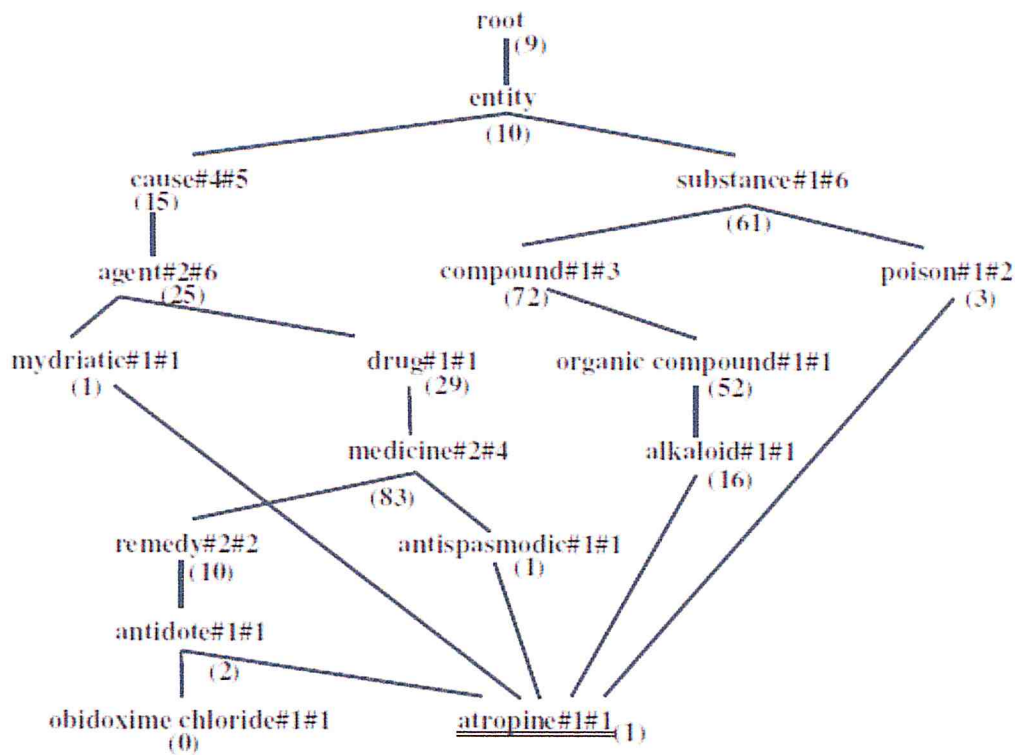


Fig2.2 Extrait de WordNet 2.0

## 5-Structure de la connaissance :

Les noms et les verbes sont organisés en hiérarchies, définies par la relation Hypernym ou la relation est un. Par exemple, le premier sens du mot chien aurait la hiérarchie suivante de Hypernym, les mots au même niveau sont des synonymes de l'un à l'autre : un certain sens de chien est synonyme de quelques autres sens de domestic dog et Canis familiaris, et ainsi de suite. Chaque ensemble de synonymes (synset), a un index unique et partage ses propriétés, telles qu'une définition gloss

Sense 1

**dog**, domestic dog, *Canis familiaris* -- (a member of the genus *Canis* (probably descended from the common wolf) that has been domesticated by man since prehistoric times; occurs in many breeds; "the dog barked all night")

=> canine, canid -- (any of various fissiped mammals with nonretractile claws and typically long muzzles)

=> carnivore -- (terrestrial or aquatic flesh-eating mammal; terrestrial carnivores have four or five clawed digits on each limb)

=> placental, placental mammal, eutherian, eutherian mammal -- (mammals having a placenta; all mammals except monotremes and marsupials)

=> mammal -- (any warm-blooded vertebrate having the skin more or less covered with hair; young are born alive except for the small subclass of monotremes and nourished with milk)

=> vertebrate, craniate -- (animals having a bony or cartilaginous skeleton with a segmented spinal column and a large brain enclosed in a skull or cranium)

=> chordate -- (any animal of the phylum Chordata having a notochord or spinal column)

=> animal, animate being, beast, brute, creature, fauna -- (a living organism characterized by voluntary movement)

=> organism, being -- (a living thing that has (or can develop) the ability to act or function independently)

=> living thing, animate thing -- (a living (or once living) entity)

=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

## 6-La hiérarchie est-un dans WordNet :

Une des relations dans WordNet d'intérêt pour nous, principalement en raison de sa structure et de l'utilité dans la mesure similarité sémantique, est là est une sorte de relation ou tout simplement est un. Cette relation entre synsets est Réservee aux noms et aux verbes. Cette relation organise les noms et des verbes synsets en grandes hiérarchies ou les arbres.

Chaque arbre a un nœud racine unique. Les nœuds plus généraux concept sont les ancêtres de plus spécifique nœuds concepts. Nous disons que les concepts plus généraux subsument les concepts les plus spécifiques. Par exemple, entité est le concept le plus général dans l'une des hiérarchies de noms et est le nœud racine de l'arbre. Elle intègre d'autres concepts plus spécifiques tels que les meubles, vélos, etc., qui sont plus bas dans l'arbre. De même, meubles peut subsumer d'autres concepts tels que ceux de chaise ou une table. Il existe neuf telles hiérarchies dans la partie des noms de WordNet, alors qu'il Ya environ 628 hiérarchies pour les verbes. Le grand nombre de hiérarchies dans les verbes est en raison du fait que les hiérarchies des verbes sont en moyenne beaucoup plus courtes et plus large que les hiérarchies des noms. La profondeur moyenne des hiérarchies de noms est d'environ 12,5 nœuds, tandis que celle des hiérarchies de verbe est d'environ 2,3 nœuds. Chacun des hiérarchies de verbe, par conséquent, il couvre une portion beaucoup plus petite des synsets, par rapport à les hiérarchies de noms. La figure 2 montre un exemple de la hiérarchie est-un WordNet.

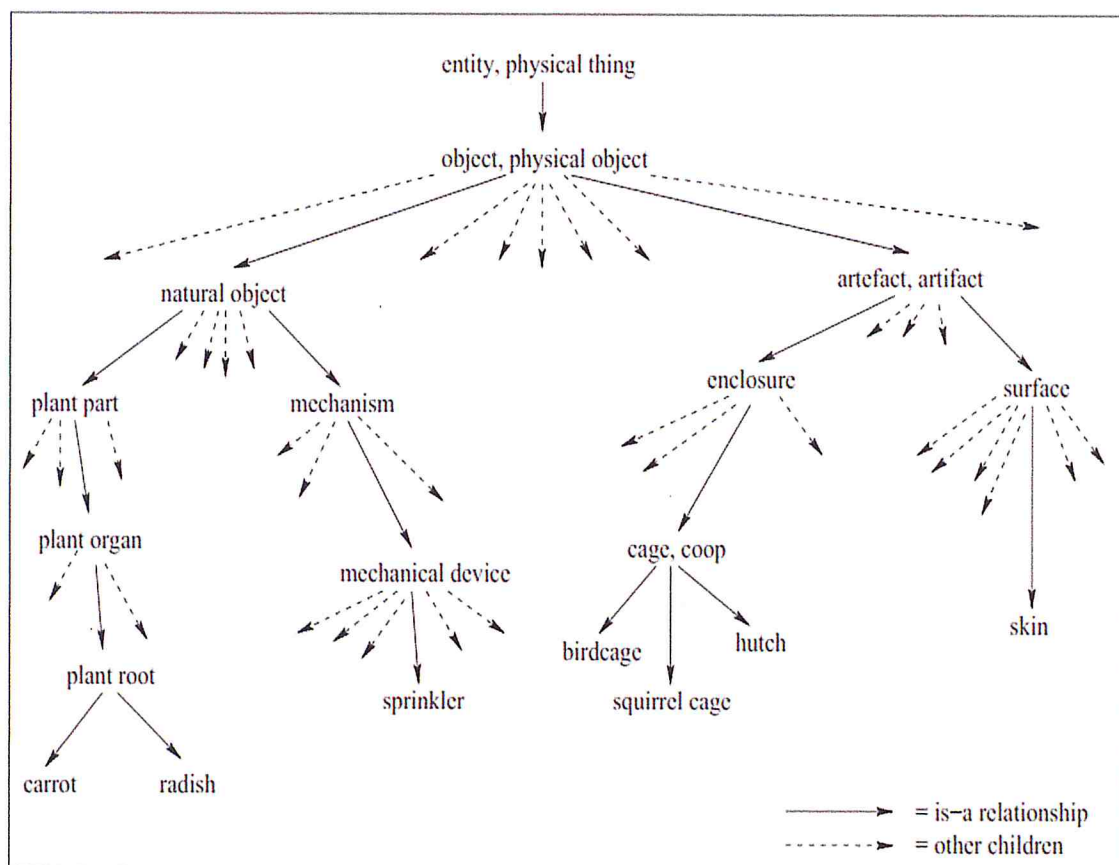


Figure 2.3: Un schéma de la hiérarchie est-un dans WordNet

## 7-Notion de synset :

La composante atomique sur laquelle repose le système entier est le synset (synonym set), un groupe de mots interchangeables, dénotant un sens ou un usage particulier .Un synset est défini d'une façon différentielle par les relations qu'il entretient avec les sens voisins. La version 2.0 de WordNet définit ainsi le nom commun anglais car à l'aide de cinq synsets:

1. *car, auto, automobile, machine, motorcar -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; he needs a car to get to work)*
2. *car, railcar, railway car, railroad car -- (a wheeled vehicle adapted to the rails of railroad; three cars had jumped the rails)*
3. *car, gondola -- (car suspended from an airship and carrying personnel and cargo and power plant)*
4. *car, elevator car -- (where passengers ride up and down; the car was on the top floor)*
5. *cable car, car -- (a conveyance for passengers or freight on a cable railway; they took a cable car to the top of the mountain)*

Chaque synset dénote une acception différente du mot car, décrite par une courte définition. Une occurrence particulière de ce mot dénotant par exemple le premier sens (le plus courant), dans le contexte d'une phrase ou d'un énoncé, serait ainsi caractérisée par le fait qu'on pourrait remplacer le mot polysémique par l'un ou l'autre des mots du synset sans altérer la signification de l'ensemble.

A chaque synset est associé une petite définition (comme celle trouvée dans le dictionnaire) qu'on appelle gloss, par exemple les mots (cat#1#5, auto#1#1, automobile#1#1, motorcar#1#1, machine#4#6) w#i#j représente Le synset numéro i parmi les j synsets du mot w.c'est un synset qui représente par le concepts défini par gloss (a motor vehicle with four wheels, usually propelled by an internal combustion engine) plusieurs gloss avoir des exemple d'utilisation comme "he needs a car to get to work"

Chaque synset dénote un domaine d'utilisation du sens Le mot bank, par exemple, à dix sens dans WordNet 2.0. Trois d'entre eux (BANK#1, BANK#3 et BANK#6) sont regroupés au sein du domaine Economie, tandis que deux (BANK#2 et BANK#7) sont regroupés avec les étiquettes de domaine Géographie et Géologie.

Sens	Synset	(Définition)gloss+ "exemple d'utilisation"	Domaines
#1	depository financial institution, bank, banking concern, banking company	(a financial institution that accepts deposits and channels the money into lending activities; "he cashed a check at the bank"; "that bank holds the mortgage on my home")	Economy
#2	Bank	(sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents")	Geography, Geology
#3	Bank	(a supply or stock held in reserve for	Economy



		future use (especially in emergencies))	
#4	bank, bank building	(a building in which commercial banking is transacted; "the bank is on the corner of Nassau and Witherspoon")	Architecture, Economy
#5	Bank	(an arrangement of similar objects in a row or in tiers; "he operated a bank of switches")	Factotum
#6	savings bank, coin bank, money box, bank	(a container (usually with a slot in the top) for keeping money at home; "the coin bank was empty")	Economy
#7	Bank	(a long ridge or pile; "a huge bank of earth")	Geography, Geology
#8	Bank	(the fund held by a gambling house or the dealer in some gambling games; "he tried to break the bank at Monte Carlo")	Economy, Play
#9	bank, cant, camber	(a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)	Architecture
#10	Bank	(a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning); "the plane went into a steep bank")	Transport

Les synsets (nœud) sont reliés entre eux par les relations sémantiques (arcs) pour former un réseau sémantique. (Dans WordNet 2.0 le nombre total des synsets=79689)

Dans Wordnet, on peut distinguer quatre catégories de synset :

- 1- catégorie des noms
- 2- catégorie des verbes
- 3- catégorie des adjectifs
- 4- catégorie des adverbes

Les noms sont ainsi classés en un système de catégories complet et précis comprenant plusieurs niveaux d'imbrication. On retrouve notamment certaines sections de cette ontologie où la profondeur dépasse 10 niveaux.

## 8-Relations entre les synsets définis dans WordNet :

<b>Relation</b>	<b>description</b>	<b>exemple</b>
Hyperonyme	est une généralisation de	Meuble est un hyperonyme de chaise
Hyponyme	est une sorte de	Chaise est un hyponyme de meuble
Troponym	est une manière de	L'amble est un troponym de promenade
Meronym	fait partie Ou membre de	La roue de la bicyclette est une partie de la bicyclette
Holonym	contient	Bicyclette est un Holonym de roue
Antonyme	Contraire de	Monter est un antonymes de descende
Attribut	est l'attribut de	Lourd est un attribut de poids
Implication	Implique	Le labourage implique le creusement
Cause	à cause de	Offenser a cause de renvoyer
Voir également	<i>verbe</i> liées à	Loger est liée a résider
Semblable à	est semblable à	Mort est semblable a assassiné
Participe de	est le participe de	Stocké (adj) est participe de stoker
Pertainym	Concerne à	Le radial concerne le rayon

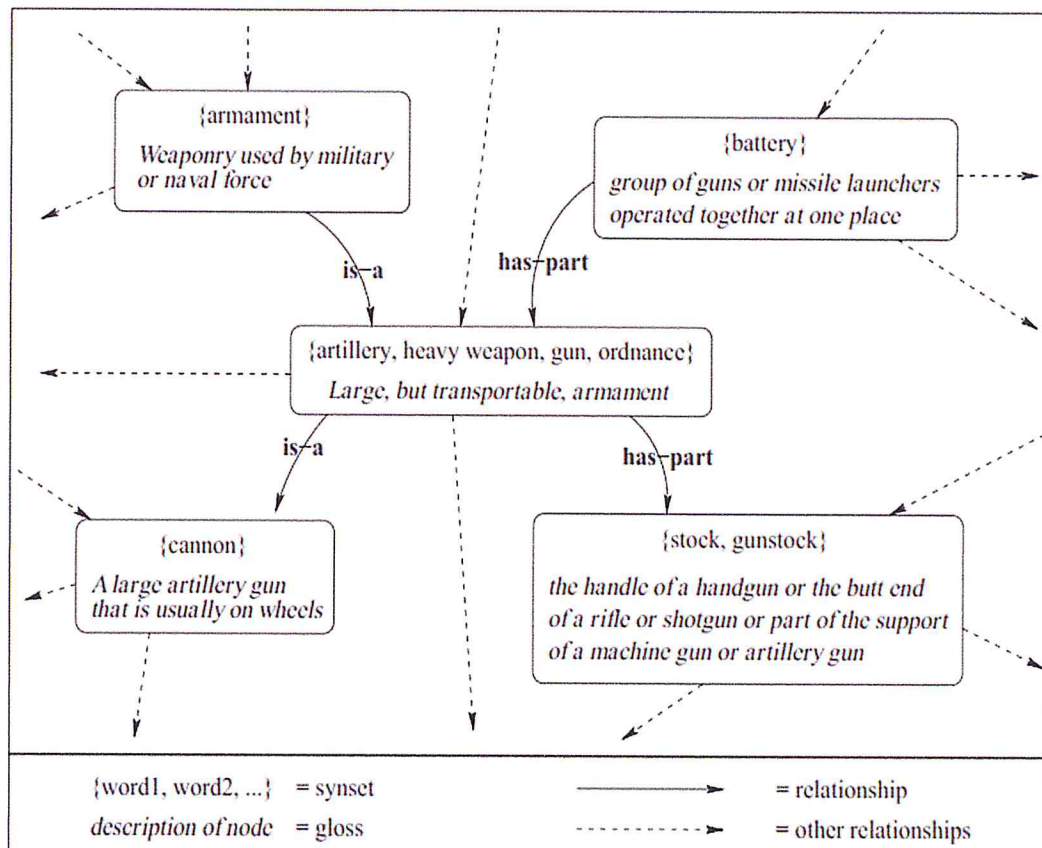
**Tableau 2.1 : Relations entre les synsets définis dans WordNet**

Par exemple, les relations de la forme « véhicule est une sorte de transport » ou « un rayon est une partie d'une roue » sont définis. D'autres relations sont est opposé, est un, membre de, causes, semblable à, etc...

Le tableau montre la liste des relations définies dans WordNet. Le réseau de relations entre les mots sens présents dans WordNet code pour une vaste quantité de connaissances

humaines. Cela donne lieu à un grand nombre de possibilités dans la manière dont il pourrait être utilisé pour divers tâches de Traitement de langage naturel (et autres).

La figure suivante met l'accent sur une petite portion de la structure de WordNet et illustre les nœuds et les arcs du réseau sémantique WordNet.



**Figure 2.4:** Une illustration de *synsets* et *les relations* dans WordNet

Le tableau suivant présente un comptage des relations de WordNet 2.1 par catégorie:

Relation	Entre	Nombre	Exemple
Hypernym/Hyponym	Verbe / verbe	13 124	EXHALE/BREATHE
	Nom / nom	75 134	CAT /FELINE

Instance Hyponym	Nom / nom	8 515	EIFFEL TOWER /TOWER
Part	Nom / nom	8 874	FRANCE / EUROPE
Member	Nom / nom	12 262	FRANCE/EUROPEAN UNION
Attribute	Adjectif / nom	643	INACCURATE /ACCURACY
Entailment	Verbe / verbe	409	DREAM /SLEEP
Cause	Verbe / verbe	219	ANESTHETIZE /SLEEP
Similar	Adjectif / adjectif	22 622	DYING /MORIBUND
SeeAlso	Adjectif / adjectif	2 683	BLACK /DARK
	Verbe / verbe	582	SLEEP LATE /SLEEP
Antonym	Adjectif / adjectif	4 080	GOOD /BAD
	adverbe / adverbe	718	POORLY /WELL
	Nom / nom	2 142	WINNER /LOOSER
	Verbe / verbe	1 089	DIE /BE BORN
Pertainym	Adjectif / nom	4 814	ACADEMIC /ACADEMIA
	adverbe / Adjectif	3 213	BOASTFULLY/BOASTFUL
	Adjectif / adjectif	38	EASY/CASUAL

**Tableau 2.2: un comptage des relations de WordNet 2.1 par catégorie**

Le tableau suivant présente les Indicateurs relationnel de WordNet:

noms	Verbes	adjectifs	adverbes
Antonym	Antonym	Antonym	Antonym
Hyponym	Troponym	Similar	Derivedfrom
Hypernym	Hypernym	RelationalAdj	
Meronym	Entailment	AlsoSee	
Holonym	Cause	Attribute	
Attribute	AlsoSee		

**Tableau 2.3: les Indicateurs relationnel de WordNet**

Le tableau suivant présente les Indicateurs réflexive de WordNet:

Indicateur	Refléter
Antonym	Antonym
Hyponym	Hypernym
Hypernym	Hyponym
Holonym	Meronym
Meronym	Holonym
Similar to	Similar to
Attribute	Attribute

**Tableau 2.4:les Indicateurs réflexive de WordNet**

## 9-Partie des noms de WordNet :

Pour la position des noms, WordNet 2.0 contient 9 racines (entity#1#1, event#1#4, group#1#3, phenomnon#1#2, possession#2#7, psychologicalfeature#1#1, state#4#8, abstraction#6#6 et act#2#5)

La partie des noms de WordNet dispose d'une connectivité assez riche et reste de loin le plus partie développée du réseau. Ses plus de 60 500 synsets, représentant plus de 107 400 sens de noms, sont liés par plus de 150 000 arcs de neuf types correspondant aux neuf relations adoptées par les créateurs de WordNet

La hiérarchie de subsomption (hyperonyme / hyponyme) Constitue l'épine dorsale du sous-réseau de nom, ce qui représente près de 80% des liens. Au sommet de la hiérarchie sont 9 des concepts abstraits, appelés débutants uniques, Tel que

**entity#1#1** (Quelque chose ayant existence concrète, «vivant ou non vivant»),

**event#1#4** (Quelque chose qui arrive à un endroit et de l'heure donné)

**group#1#3** (tout nombre d'entités (membres) considérées comme unité)

**phenomenon#1#2** (tout état ou processus connu par les sens plutôt que par intuition ou raisonnement)

**possession#2#7** (quelque chose possédé)

**psychological feature#1#1** (un dispositif de la vie mentale d'une organisation vivante)

**state#4#8** (la manière quelque chose est en ce qui concerne ses attributs principaux ; « l'état actuel de la connaissance » ; « son état de santé » ; « dans un état financier faible »)

**abstraction#6#6** (Un concept formé par extraction de caractéristiques communes à partir des exemples spécifiques)

**act#2#5** (quelque chose que les gens créent ou font pour produire)

Par conséquent, à proprement parler, la partie des noms se compose de neuf hiérarchies distinctes en WordNet 2.0

## 10-Notion d'instance hyponyme :

La version 2.1 a introduit la notion d' « instance hyponyme », qui désigne une instance (et non une sous-classe) d'un synset (une Entité Nommée). Par exemple, GEORGE WASHINGTON est une instance hyponyme de PRESIDENT OF THE UNITED STATES. De même, le nom TOWER#1 a pour hyponymes SILO, MINARET, PYLON... et TOUR EIFFEL comme instance hyponyme.

## 11-Partie des adjectives de WordNet :

wordnet contient actuellement approximativement 19500 formes de mots adjectifs organisé dans environ 10000 signification de mot (synsets)

wordnet divise les adjectifs en deux grandes catégories : les adjectifs descriptifs et les adjectifs relationnels

adjectifs descriptifs (big,interesting,possible):la plus grande catégorie

adjectifs relationnels i.e., reliés (par dérivation) à un nom (electrical,dental, fraternal, presidential,nuclear):classe beaucoup plus petite

### Les adjectifs descriptif :

Un adjectif descriptif est un adjectif dont le rôle est d'assigner une valeur à un attribut d'un nom

Le nom paquet a pour attribut poids dont la valeur peut être spécifiée par l'adjectif lourd

C'est-à-dire

x est un nom suppose qu'il y a un attribut A telle que  $A(x)=adj$

pour dire the package est heavy suppose qu'il ya une attribut weight tels que  $weight(package)=heavy$ . De même, low et hight sont des valeurs pour l'attribut height

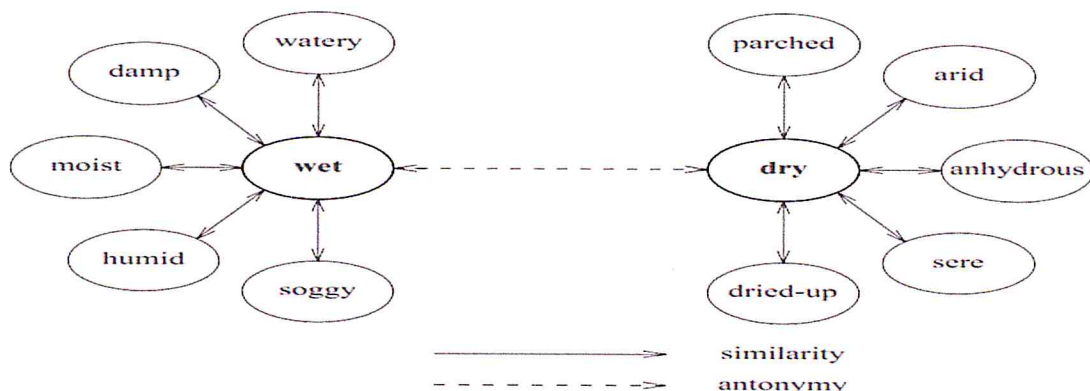
WN associe aux adjectifs descriptifs des pointeurs vers lesNoms

L'organisation sémantique des adjectifs descriptifs est entièrement différente de celle des noms. Rien comme la relation hyponyme,hyponyme qui produit les hiérarchies nominales est disponible pour la classe des adjectifs ; Il n'est pas clair pour indiquer qu'une adjective « est un » « est une sorte de » avec des autres adjectifs.

La relation sémantique de base pour la classe les adjectives descriptives c'est la relation d'antonymie comme exemple (big→small)(interesting→uninteresting)(possible→impossible)

la relation sémantique de synonymie comme(arid→dry)(wet→soggy)

La figure suivante donne la structure d'une hiérarchie des adjectives.



**Figure2.5 : Structure d'adjectifs descriptifs**

### Lesadjectifsrelationnels :

C'est le deuxième genre des adjectifs, une adjective relationnel Signifie quelque chose comme "de, se rapporter/concernant" lié à un certain nom, et joue un rôle semblable à un nom.

Par exemple, fraternel, comme dans les jumeaux fraternel sont rapporté au frère,et dentaire, comme dans l'hygiène dentaire est lié à la dent.

Environ 1700 adjectifs relationnels contenant dans plus de 3000 synsets différents sont actuellement inclus dans wordnet.

Les adjectifs relationnels relient par des pointeurs aux noms dont ils sont dérivés mais ne les structurent pas entre eux

Chaque synset se compose d'un ou plusieurs adjectifs relationnels, suivi d'un indicateur au nom approprié, par exemple l'entrée {stellar, astral, noun.object:star} indique que stellar, astral rapporte au nom star

Sense 2

**stellar**, astral -- (being or relating to or resembling or emanating from stars; "an astral body"; "stellar light")

Pertains to noun star (Sense 1)

## 12-Partie des adverbes de WordNet :

Les adverbes sont le plus souvent définis par les adjectifs dont ils dérivent. Ils héritent donc de la structure des adjectifs.

## 13-Partie des verbes de WordNet :

Les verbes sont organisés en hiérarchies. Des relations d'hyperonymie (« est-un ») et d'hyponymie relient les « ancêtres » des verbes avec leurs « spécialisations ». Au niveau racine, ces hiérarchies sont organisées en types de base. Le réseau des noms est bien plus profond que celui des autres parties du discours

Actuellement, WordNet contient plus de 21.000 formes de mot de verbe et approximativement 8.400 expriment des significations (synsets), inclus les verbes syntagmatiques like look up et fall back.

Les verbes sont divisés en 15 dossiers .En grande partie sur la base des critères sémantiques.

Tout sauf un de ces dossiers correspond à ce que les linguistes ont appelées domaines sémantiques : verbes de Soins du corps et vitalité, changement, Cognition, communication, Compétition, consommation, contact, création, émotion, mouvement, perception, possession, interaction sociale, et Météorologie.

Pratiquement tous les verbes dans ces dossiers dénotent des événements ou des actions. Un autre dossier contient des verbes se rapportant aux états, comme suffire, appartenir, et ressembler, qui ne pourraient pas être intégrés dans les autres dossiers.

## Catégories de verbes :15 grandes familles

**Soins du corps et vitalité:** verbs of grooming, dressing and bodily care

**Changement :** change of size, temperature, intensity, etc.

**Cognition :** verbs of thinking, judging, analyzing, doubting, etc.

**Communication :** verbs of telling, asking, ordering, singing, etc.

**Compétition :** verbs of fighting, athletic activities, etc.

**Consommation :** verbs of eating and drinking, using, ingesting

**Contact :** verbs of touching, hitting, tying, digging, etc.

**Création** :verbs of sewing, baking, painting, performing, etc.

**Emotion** : verbs of feeling

**Mouvement** :verbs of walking, flying, swimming, etc.

**Perception** :verbs of seeing, hearing, feeling, etc.

**Possession** :verbs of buying, selling, owning, and transfer

**Interactions sociales**:verbs of political and social activities and events

**Météorologie** :verbs of raining, snowing, thawing, thundering, etc.

**Etats** :verbs of being, having, spatial relations

## **14-WordNet et les relations sémantiques:**

À l'instar d'un dictionnaire traditionnel, WordNet offre ainsi, pour chaque mot, une liste de synsets correspondant à toutes ses acceptions répertoriées. Mais les synsets ont également d'autres usages : ils peuvent représenter des concepts plus abstraits, de plus haut niveau que les mots et leurs sens, qu'on peut organiser sous forme d'ontologies. Une ontologie est un système de catégories permettant de classer les éléments d'un univers. Le système de catégorisation correspond aux relations sémantiques. Ceci permet de regrouper de manière cohérente toutes les composantes d'un univers linguistique telles que les mots, les sens ou bien les concepts.

La relation sémantique servant de critère pour l'agrégation d'un groupe de concepts définira le type de l'ontologie. WordNet répertorie ainsi une grande variété de relations sémantiques permettant d'organiser le sens des mots (et donc par extension les mots eux-mêmes) en des systèmes de catégories qu'on peut consulter de manière cohérente et uniforme. On pourra ainsi interroger le système quant aux hyperonymes d'un mot particulier. À partir par exemple du sens le plus commun du mot *car* (correspondant au synset *l. car, auto...*) la relation d'hyperonymie définit un arbre de concepts de plus en plus généraux :



## Sense 1

**car**, auto, automobile, machine, motorear -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; "he needs a car to get to work")

=> motor vehicle, automotive vehicle -- (a self-propelled wheeled vehicle that does not run on rails)

=> self-propelled vehicle -- (a wheeled vehicle that carries in itself a means of propulsion)

=> wheeled vehicle -- (a vehicle that moves on wheels and usually has a container for transporting things or people; "the oldest known wheeled vehicles were found in Sumer and Syria and date from around 3500 BC")

=> vehicle -- (a conveyance that transports people or objects)

=> conveyance, transport -- (something that serves as a means of transportation)

=> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)

=> artifact, artefact -- (a man-made object taken as a whole)

=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> whole, whole thing, unit -- (an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit")

=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> container -- (any object that can be used to hold things (especially a large metal boxlike object of standardized dimensions that can be loaded from one form of transport to another))

=> instrumentality, instrumentation -- (an artifact (or system of artifacts) that is instrumental in accomplishing some end)

=> artifact, artefact -- (a man-made object taken as a whole)

=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

=> whole, whole thing, unit -- (an assemblage of parts that is regarded as a single entity; "how big is that part compared to the whole?"; "the team is a unit")

=> object, physical object -- (a tangible and visible entity; an entity that can cast a shadow; "it was full of rackets, balls and other objects")

=> entity -- (that which is perceived or known or inferred to have its own distinct existence (living or nonliving))

Dans cet exemple, le dernier concept, «entity», est le plus général, le plus abstrait. Il pourrait ainsi être le super-concept d'une multitude de concepts plus spécialisés.

On peut également interroger le système quant à la relation inverse de l'hyponymie, l'hyponymie. Les hyponymes associés au sens *car*, *auto*... du mot *car* sont :

- car**, auto, automobile, machine, motorcar -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; "he needs a car to get to work")
- => ambulance -- (a vehicle that takes people to and from hospitals)
    - => funny wagon -- (an ambulance used to transport patients to a mental hospital)
  - => beach wagon, station wagon, wagon, beach wagon, station wagon, wagon -- (a car that has a long body and rear door with space behind rear seat)
    - => shooting brake, estate car -- (another name for a station wagon)
  - => bus, jalopy, heap -- (a car that is old and unreliable; "the fenders had fallen off that old bus")
  - => cab, hack, taxi, taxicab -- (a car driven by a person whose job is to take passengers where they want to go in exchange for money)
    - => gypsy cab -- (a taxicab that cruises for customers although it is licensed only to respond to calls)
    - => minicab -- (a minicar used as a taxicab)
  - => compact, compact car -- (a small and economical car)
  - => convertible -- (a car that has top that can be folded or removed)
  - => coupe -- (a car with two doors and front seats and a luggage compartment)
  - => cruiser, police cruiser, patrol car, police car, prowler car, squad car -- (a car in which policemen cruise the streets; equipped with radiotelephonic communications to headquarters)
    - => panda car -- (a police cruiser)
  - => electric, electric automobile, electric car -- (a car that is powered by electricity)
  - => gas guzzler -- (a car with relatively low fuel efficiency)
  - => hardtop -- (a car that resembles a convertible but has a fixed rigid top)
  - => hatchback -- (a car having a hatchback door)
  - => horseless carriage -- (an early term for an automobile; "when automobiles first replaced horse-drawn carriages they were called horseless carriages")
  - => hot rod, hot-rod -- (a car modified to increase its speed and acceleration)
  - => jeep, landrover -- (a car suitable for traveling over rough terrain)
  - => limousine, limo -- (large luxurious car; usually driven by a chauffeur)
    - => berlin -- (a limousine with a glass partition between the front and back seats)
  - => loaner -- (a car that is lent as a replacement for one that is under repair)
  - => minicar -- (a car that is even smaller than a subcompact car)
    - => minicab -- (a minicar used as a taxicab)
  - => minivan -- (a small box-shaped passenger van; usually has removable seats; used as a family car)
  - => Model T -- (the first widely available automobile powered by a gasoline engine; mass-produced by Henry Ford from 1908 to 1927)
  - => pace car -- (a high-performance car that leads a parade of competing cars through the pace lap and then pulls off the course)
  - => racer, race car, racing car -- (a fast car that competes in races)
    - => stock car -- (a racing car with the basic chassis of a commercially available car)
  - => roadster, runabout, two-seater -- (an open automobile having a front seat and a rumble seat)
  - => sedan -- (a car that is closed and that has front and rear seats and two or four doors)
    - => brougham -- (a sedan that has no roof over the driver's seat)
  - => sports car, sport car -- (a small low car with a high-powered engine; usually seats two persons)
  - => sport utility, sport utility vehicle, S.U.V. -- (a high-performance four-wheel drive car built on a truck chassis)
  - => Stanley Steamer -- (a steam-powered automobile)
  - => stock car -- (a car kept in dealers' stock for regular sales)
  - => subcompact, subcompact car -- (a car smaller than a compact car)

WordNet offre en fait une multitude d'autres ontologies, faisant usage de relations sémantiques plus spécialisées et restrictives. On peut ainsi interroger le système quant aux méronymes d'un mot ou d'un concept, les parties constitutives d'un objet (*HAS-PART*). Les méronymes associés au sens *car*, *auto*... du mot *car* sont :

### Sense 1

- car**, auto, automobile, machine, motorcar -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; "he needs a car to get to work")
- HAS PART: accelerator, accelerator pedal, gas pedal, gas, throttle, gun -- (a pedal that controls the throttle valve; "he stepped on the gas")
  - HAS PART: air bag -- (a safety restraint in an automobile; the bag inflates on collision and prevents the driver or passenger from being thrown forward)
  - HAS PART: auto accessory -- (an accessory for an automobile)
  - HAS PART: automobile engine -- (the engine that propels an automobile)
  - HAS PART: automobile horn, car horn, motor horn, horn, hooter -- (a device on an automobile for making a warning noise)
  - HAS PART: buffer, fender -- (a cushion-like device that reduces shock due to contact)
  - HAS PART: bumper -- (a mechanical device consisting of bars at either end of a vehicle to absorb shock and prevent serious damage)
  - HAS PART: car door -- (the door of a car)
  - HAS PART: car mirror -- (a mirror that the driver of a car can use)
  - HAS PART: car seat -- (a seat in a car)
  - HAS PART: car window -- (a window in a car)
  - HAS PART: fender, wing -- (a barrier that surrounds the wheels of a vehicle to block splashing water or mud; "in England they call a fender a wing")
  - HAS PART: first gear, first, low gear, low -- (the lowest forward gear ratio in the gear box of a motor vehicle; used to start a car moving)
  - HAS PART: floorboard -- (the floor of an automobile)
  - HAS PART: gasoline engine -- (an internal-combustion engine that burns gasoline; most automobiles are driven by gasoline engines)
  - HAS PART: glove compartment -- (compartment on the dashboard of a car)
  - HAS PART: grille, radiator grille -- (grating that admits cooling air to car's radiator)
  - HAS PART: high gear, high -- (a forward gear with a gear ratio giving high vehicle velocity for a given engine speed)
  - HAS PART: hood, bonnet, cowl, cowling -- (protective covering consisting of a metal part that covers the engine; "there are powerful engines under the hoods of new cars"; "the mechanic removed the cowling in order to repair the plane's engine")
  - HAS PART: luggage compartment, automobile trunk, trunk -- (compartment in an automobile that carries luggage or shopping or tools; "he put his golf bag in the trunk")
  - HAS PART: rear window -- (car window that allows vision out of the back of the car)
  - HAS PART: reverse -- (the gears by which the motion of a machine can be reversed)
  - HAS PART: roof -- (protective covering on top of a motor vehicle)
  - HAS PART: running board -- (a narrow footboard serving as a step beneath the doors of some old cars)
  - HAS PART: stabilizer bar, anti-sway bar -- (a rigid metal bar between the front suspensions and between the rear suspensions of cars and trucks; serves to stabilize the chassis)
  - HAS PART: sunroof, sunshine-roof -- (an automobile roof having a sliding or raisable panel; "sunshine-roof" is a British term for "sunroof")
  - HAS PART: tail fin, tailfin, fin -- (one of a pair of decorations projecting above the rear fenders of an automobile)
  - HAS PART: third gear, third -- (the third from the lowest forward ratio gear in the gear box of a motor vehicle; "you shouldn't try to start in third gear")
  - HAS PART: window -- (a transparent opening in a vehicle that allow vision out of the sides or back; usually is capable of being opened)

On peut aussi consulter le système pour les relations de synonymes et d'antonymie. Les synonymes associés au sens *car, auto...* du mot *car* sont :

### Sense 1

- car**, auto, automobile, machine, motorcar -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; "he needs a car to get to work")  
=> motor vehicle, automotive vehicle -- (a self-propelled wheeled vehicle that does not run on rails)

On peut aussi consulter le système pour trouver les termes *du même rang*, les termes *du même rang* associés au sens *car, auto...* du mot *car* sont :

Sense 1

**car**, auto, automobile, machine, motorcar -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; "he needs a car to get to work")

- => motor vehicle, automotive vehicle -- (a self-propelled wheeled vehicle that does not run on rails)
- => amphibian, amphibious vehicle -- (a flat-bottomed motor vehicle that can travel on land or water)
- => bloodmobile -- (a motor vehicle equipped to collect blood donations)
- => car, auto, automobile, machine, motorcar -- (4-wheeled motor vehicle; usually propelled by an internal combustion engine; "he needs a car to get to work")
- => doodlebug -- (a small motor vehicle)
- => four-wheel drive, 4WD -- (a motor vehicle with a four-wheel drive transmission system)
- => go-kart -- (a small low motor vehicle with four wheels and an open framework; used for racing)
- => golfcart, golf cart -- (a small motor vehicle in which golfers can ride between shots)
- => hearse -- (a vehicle for carrying a coffin to a church or a cemetery; formerly drawn by horses but now usually a motor vehicle)
- => motorcycle, bike -- (a motor vehicle with two wheels and a strong frame)
- => snowplow, snowplough -- (a vehicle used to push snow from roads)
- => truck, motortruck -- (an automotive vehicle suitable for hauling)

Ou encore On peut aussi consulter le système quant à la relation inverse, l'holonyme, Les holonymes associés par exemple au sens *wheel* du mot *wheels* sont :

Sense 1

**wheel** -- (a simple machine consisting of a circular frame with spokes (or a solid disc) that can rotate on a shaft or axle (as in vehicles or other machines))

PART OF: wheeled vehicle -- (a vehicle that moves on wheels and usually has a container for transporting things or people; "the oldest known wheeled vehicles were found in Sumer and Syria and date from around 3500 BC")

Le système peut interroger pour trouver les attributs d'un concept par exemples les attributs associés par exemple au sens *weight* du mot *wheels* sont :

Sense 1

**weight** -- (the vertical force exerted by a mass as a result of gravity)

- => heavy (vs. light) -- (of comparatively great physical weight or density; "a heavy load"; "lead is a heavy metal"; "heavy mahogany furniture")
- => light (vs. heavy) -- (of comparatively little physical weight or density; "a light load"; "magnesium is a light metal--having a specific gravity of 1.74 at 20 degrees C")

## 15-Applications :

Les ressources WordNet (dont la ressource originale anglaise fait partie) jouissent d'une bonne popularité auprès de la communauté scientifique et des industriels du traitement de la langue. La Global Wordnet conférence (GWC), organisée tous les deux ans, vise à rassembler les personnes de ces deux communautés afin de partager sur les avancées des Wordnet à travers le monde.

La richesse et la précision de Wordnet en font un outil de choix, susceptible d'être mis à profit par une multitude de techniques et de théories diverses

Son utilisation fait en sorte de procurer aux algorithmes et applications une importante plateforme de connaissances a priori du langage et du monde dans lequel il s'articule.

WordNet a été employé pour un certain nombre de différents buts dans des systèmes d'information, y compris la désambiguïsation de sens de mot, la classification de recherche documentaire et automatique des textes, la récapitulation automatique des textes, la traduction automatique et même la génération automatique de jeu de mots croisé.

Un projet à l'université de Brown commencée par Jeff Stibel, James A. Anderson, Steve Reiss et d'autres laboratoire appelé d'Applied Cognition a créé un désambiguisateur utilisé WordNet en 1998. Le projet plus tard à déplacer dans une compagnie a appelé Simpli, qui est maintenant possédée par ValueClick. George Miller a joint la compagnie en tant que membre du comité consultatif. Simpli a établi un SearchEngine d'Internet qui a utilisé une base de connaissance principalement basée sur WordNet pour désambiguïser et augmenter des mots-clés et des synsets pour aider à rechercher l'information en ligne.

Les algorithmes de réseau neurologique ont recherché le WordNet augmenté des limites relatives pour désambiguïser les mots-clés de recherche (Java, dans le sens du café) et pour augmenter le synset de recherche (café, boisson, Joe) pour améliorer des résultats des moteurs de recherche. Avant que la compagnie ait été acquise, elle a exécuté des recherches à travers des moteurs de recherche tels que Google, Yahoo! , Ask.com et d'autres.

Un autre exemple en avant de l'utilisation de WordNet est de déterminer la similitude entre les mots. Des divers algorithmes ont été proposés, et ceux-ci incluent vu la distance entre les catégories conceptuelles des mots, aussi bien que considérer la structure hiérarchisée de l'ontology de WordNet. Un certain nombre de ces algorithmes WordNet-basés de similitude de mot sont mis en application dans un paquet de Perl appelé WordNet: : La similitude, et dans le python un paquet a appelé NLTK.

## 16-Limites de WordNet :

### Partie des verbes de WordNet :

On retrouve en revanche un système de classification beaucoup moins élaboré pour les verbes, qui sont organisés en un système hiérarchique beaucoup plus « plat » avec moins de niveaux d'imbrication, où on passe très rapidement d'un concept spécialisé (le sens operate, run du verbe running, par exemple) à un concept très général (control, command).

Sense 4

operate, **run** -- (direct or control; projects, businesses, etc.; "She is running a relief operation in the Sudan")

=> direct -- (be in charge of)

=> manage, deal, care, handle -- (be in charge of, act on, or dispose of; "I can deal with this crew of workers"; "This blender can't handle nuts"; "She managed her parents' affairs after they got too old")

=> control, command -- (exercise authoritative control or power over; "control the budget"; "Command the military forces")

Si on examine par conséquent l'ontology générée par la relation d'hyponymie, il est notable qu'elle est la plus complète dans son embranchement nominal.

À ce jour, il n'y a aucune catégorisation hiérarchique définie pour les embranchements des adjectifs et des adverbes. Ce déséquilibre potentiellement problématique se retrouve à l'intérieur même des supercatégories, où il est beaucoup plus apparent dans la branche nominale : certains mots sont ainsi liés à une grande chaîne de concepts finement graduée, tandis que d'autres sont très proches des concepts les plus généraux.

### Informations manquantes :

WordNet ne précise pas l'étymologie, la prononciation, les formes de verbes irréguliers et ne contient que des informations limitées sur l'usage des mots.

### Profusion de sens pour un mot donné

La contrepartie de son importante couverture est que WordNet est très précis dans le sens des définitions. On a une granularité très (trop ?) fine des sens. Par exemple, le verbe to give (« donner ») n'a pas moins de 44 sens. Une telle profusion ne facilite pas une tâche de désambiguïsation lexicale.

## 17-Wordnet et la Mesure de la similarité sémantique :

Compte tenu de grand magasin de la connaissance humaine codée dans WordNet, il a été utilisé par de nombreux chercheurs dans l'élaboration de mesures de la similarité sémantique. Certains utilisent seulement la structure et le contenu de WordNet dans le calcul de la Mesure de la similarité sémantique. D'autres approches combinent des données statistiques à partir de grands corpus de la structure de WordNet pour nous donner un score de la similarité sémantique.

(Pedersen, Patwardhan, Michelizzi, 2004) présentent plusieurs de ces algorithmes de similarité entre mots, et une implémentation basée sur WordNet en Perl appelée WordNet::Similarity.

## 18-Conclusion :

Wordnet Une des rares ressources pour la langue générale (anglais) disponible en ligne l'Organisation de la base lexicale assez riche grâce aux diverses relations représentées

Une utilisation possible de l'ontologie fournie par WordNet est la définition de métriques heuristiques de « distance sémantique » entre les synsets. Cette métrique est basée sur la distance à parcourir dans le graphe, combinée ou non avec le Contenu Informationnel. Elle permet de quantifier la similarité de deux concepts. Elle peut également servir dans un cadre de désambiguïsation lexicale.

# -Chapitre 3 : Les approches de calcul de contenu informationnel

## 1-Introduction:

La similarité sémantique, c'est-à-dire l'appréhension de la liaison entre deux concepts, est une capacité de l'homme que les machines ne savent que très mal reproduire. Ainsi, pour un humain, il est évident que les concepts de crayon et de papier sont liés, beaucoup plus que ceux de parapluie et fer à repasser en tout cas. Mais il est très difficile de le formaliser car rien, en surface, ne permet de le décider. Pour ce faire, il faut utiliser des ressources sémantiques : les ontologies, c'est-à-dire des bases de connaissances. Elles seules permettent de montrer les liens (hypéronymie, antonymie, méronymie, etc.) entre des concepts

La mesure de similarité sémantique entre les concepts est une issue générique pour beaucoup d'applications de linguistique computationnelle et d'intelligence artificielle, comme l'extraction de l'information, l'annotation sémantique et les questions et les réponses.

Un problème récurrent de ces applications est la mesure de proximité entre concepts. Elle a été étudiée par de nombreux auteurs, et deux grandes approches basées sur les arcs (Hirst-St-Onge, 1998), (LeacockChodorow, 1998), (Wu et Palmer, 1994) en exploitant la structure de la ressource sémantique utilisée, et les approches utilisant le contenu informationnel (Resnik, 1995), (Jiang et Contrath, 1997) et (Lin, 1998).

Le contenu informationnel (CI) du concept fournit une évaluation de son abstrait/spécialité. Et il a une dimension importante pour évaluer la similarité entre deux concepts ou de deux mots. Depuis longtemps, il a été appliqué avec succès dans le calcul de la similarité sémantique. Resnik a proposé la première fois une mesure de similarité basée sur le contenu informationnel en 1995. Le contenu informationnel (CI) d'un concept traduit la pertinence de ce dernier dans un corpus en tenant compte de la fréquence d'apparition des mots auxquels il se réfère ainsi que de la fréquence d'apparition des concepts qu'il généralise. Il dénote la spécificité ou généralité d'un concept: plus un concept est spécifique plus son CI est important et plus il est général, plus son CI est petit.

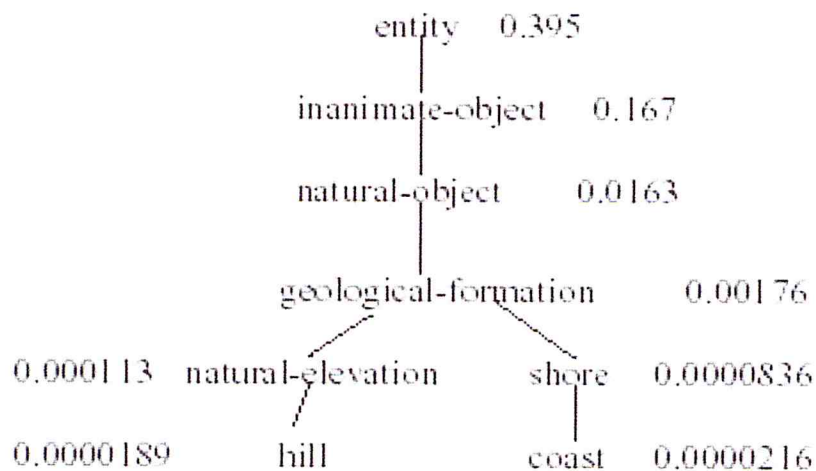
Il existe plusieurs approches pour le calcul de cette valeur et deux grandes approches sont distinguées approches basées sur une ressource externe (corpus) et approches exploitent la structure hiérarchique et significative de la taxonomie « is a » du WordNet sans faire recours à un corpus externe.

## 2-L'Approche basée sur une ressource externe :

Dans cette approche, pour calculer le CI des concepts on utilise conjointement une ontologie (WordNet) et un corpus.

### 2.2-L'Approche de Resnik :

Resnik a proposé d'obtenir le CI suite à une analyse empirique d'un corpus. Les probabilités associées à chaque concept dans l'ontologie WordNet sont basées sur ses occurrences dans un corpus donné. Il a estimé les fréquences de concepts appartenant WordNet (seule la taxonomie des noms est prise en considération avec seulement la relation *is\_a*) en utilisant les fréquences des noms dans le corpus Brown de l'anglais américain. Le corpus de Brown est une collection d'articles de (1000 000 mots). Quant à Lin [Lin, 1998], il a utilisé un corpus de 22 million de mots composé des journaux Wall Street et San Jose Mercury.



**Fig 3.1: les probabilités correspondantes aux différents concepts [Lin, 1998]**

Soit C l'ensemble des concepts d'une taxonomie "is a" qui permet l'héritage multiple. En associant des probabilités aux concepts de la taxonomie, il est possible de représenter le contenu informationnel et éviter le manque de fiabilité introduite par l'utilisation des chemins seulement [Resnik, 1995].

Le contenu informationnel d'un concept c peut se calculer par la formule suivante:

$$CI(c) = -\log(P(c))$$

Ou  $P(c)$  est la probabilité de retrouver qu'un mot du corpus soit une instance du concept C



(Un des mots référés par le concept  $c$  ou par un de ses descendant). Cette probabilité est monotone quand on remonte dans la hiérarchie  $P(A) \leq P(B)$  si  $A$  est un ancêtre de  $B$ .

*Exemple:*

$$P(\text{hill}) < P(\text{geological\_formation})$$

La probabilité d'apparition d'un concept est calculée par :  $P(c) = \frac{\text{frequence}(c)}{N}$

Avec

- $N$  le nombre total de concepts et
- $\text{frequence}(c) = \sum_{w \in \text{instance}(c)} \text{count}(w)$

Où  $\text{instance}(c)$ : l'ensemble des mots subsumés par le concept  $c$  et  $\text{count}(w)$  est une fonction qui renvoie le nombre d'occurrence du concept  $c$  (et des descendants) dans le corpus.

### 3-L'Approche basée sur WordNet:

Dans l'approche précédente, le contenu informationnel d'un concept est obtenu par l'analyse statistique de corpus d'où les probabilités de concepts produites sont déduites.

Dans cette approche, WordNet peut également être utilisé comme une ressource statistique sans avoir besoin à des ressources externes. La taxonomie "is-a" de WordNet peut être exploitée d'une manière innovante pour produire les valeurs de CI nécessaires aux calculs similarité sémantique.

#### 3.1-L'approche de Seco :

Le calcul du CI basé sur les hyponymes a été proposé pour la première fois par Seco [Seco, 2004]. Il considère que dans WordNet, la valeur CI d'un concept peut être écrite en fonction de ses hyponymes. Un concept avec plus hyponymes exprime moins d'informations que les concepts qui en ont moins ont des CI petits. Pour un concept, plus hyponymes il a, le plus abstrait qu'il est. Inversement, moins hyponymes il a, plus il est spécifié. Cela implique que les nœuds feuilles n'ont pas hyponymes et ils sont les plus spécifiques. Donc, l'information qu'ils véhiculent est maximale. Cependant, le nœud racine le maximum d'hyponymes et il est le plus abstrait. Ainsi, il exprime le minimum d'information. Le contenu informationnel d'un concept dépend du nombre des concepts qu'il subsume. La formule de calcul est la suivante :

Avec

$$CI(c) = 1 - \frac{\text{Log}(\text{hypo}(c) + 1)}{\text{Log}(\text{max}_{wm})}$$

- hypo(c) est une fonction qui retourne le nombre d'hyponymes d'un concept donné.

- maxwn est une constante qui représente le nombre maximum de concepts dans le thesaurus WordNet (dans WordNet 2.0 maxwn=79689)

Le dénominateur sert de facteur de normalisation. Il assure que les valeurs de CI sont dans l'intervalle [0.. 1]. La formulation garantie que le contenu informationnel diminue de façon monotone en remontant dans la hiérarchie des nœuds feuilles au nœud racine. Le contenu informationnel de la racine de WordNet donne une valeur de 0 cependant le CI des feuilles vaut 1.

L'approche de Seco ne tient compte que des hyponymes d'un concept dans la taxonomie. Ce qui représente une limite avec cette approche l'exemple suivant donne une illustration de cette limite.

### Limite de L'approche de Seco :

Prenons les extraits suivants du WordNet 2.0 :

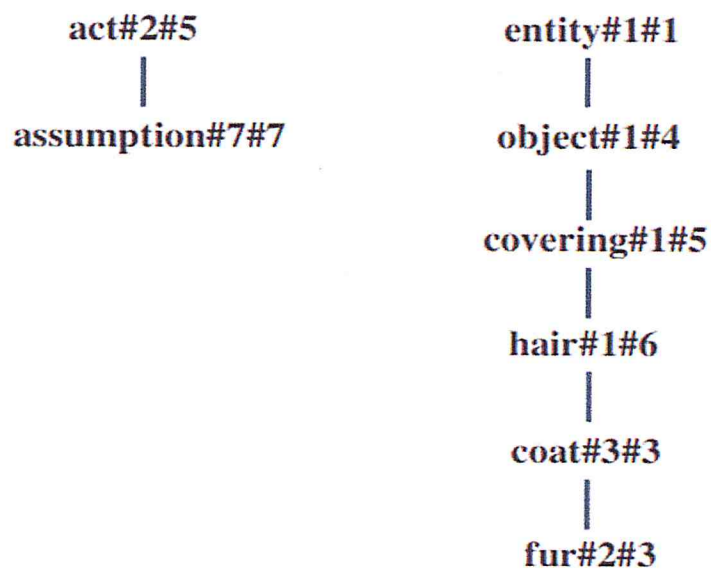


Figure3.2 Extraits du WordNet 2.0

D'après les extraits ci-dessus du WordNet 2.0 nous pouvons remarquer que si nous appliquons la méthode de seco pour calculer le CI du concept assumption et celui du concept fur nous les trouvons égales :

**Remarque :**

Nous avons utilisé la valeur 79689 qui représente le nombre total des synsets en WordNet 2.0.

$$CI(\text{assumption}) = 1 - \frac{\text{Log}(\text{hypo}(\text{assumption}) + 1)}{\text{Log}(79689)} = 1 - \frac{\text{Log}(1 + 1)}{\text{Log}(79689)} = 0.9385$$

$$CI(\text{fur}) = 1 - \frac{\text{Log}(\text{hypo}(\text{fur}) + 1)}{\text{Log}(79689)} = 1 - \frac{\text{Log}(1 + 1)}{\text{Log}(79689)} = 0.9385$$

$CI(\text{assumption}) = CI(\text{fur})$  malgré que les deux concepts ne se trouvent pas à la même profondeur (la différence est de 4 niveaux). Sachant que la profondeur dans la taxonomie « is a » du wordnet est significative car le passage d'un niveau à un autre engendre un transfert de données aussi que l'attribution de nouvelles spécificités à n'importe quel concept. Il en résulte que cette méthode ne prend pas en considération la profondeur d'un concept et ne tient compte que de la partie inférieure c'est-à-dire des hyponymes malgré qu'un concept soit formé de tous les hyperonymes dont il hérite. De plus, il faut donner de l'importance à la partie supérieure.

### 3.2-CI de Zhou :

#### -Approche 1 :

Zhou [Zhou, 2008] a ajouté la profondeur au calcul du contenu informationnel basé sur l'hyponymie.

$$CI_{\text{Zhou\_Wang\_Gu}} = k \left( 1 - \frac{\log(\text{hypo}(a) + 1)}{\log(\text{max\_nodes})} \right) + (1 - k) \left( \frac{\log(\text{depth}(a))}{\log(\text{max\_depth})} \right)$$

Avec :

- $\text{hypo}(c)$ : une fonction qui renvoie le nombre d'hyponymes du concept  $c$ .
- $\text{Max}_{\text{wn}}$ : le nombre maximum de concepts qui existent dans la taxonomie.
- $\text{Depth}(a)$ : correspond à la profondeur du concept  $a$  dans la taxonomie
- $\text{max\_depth}$ : est la profondeur maximale de la taxonomie.
- $K$  est un facteur de réglage qui permet de régler le poids des deux entités impliquées dans l'évaluation IC

Dans son expérience,  $k$  est 0.5.

#### -Approche 2 :

(1) un concept avec plus hyponymes exprime moins d'information qu'un concept avec moins hyponymes.

(2) le concept le plus profond, possède l'information la plus grande.

Par conséquent, le nouveau modèle est présenté par la formule suivante :

$$IC(c) = \log(\text{depth}(c)) / \log(\text{depth\_max}) * (1 - \log(\sum_{a \in \text{hypo}(c)} 1 / \text{depth}(a) + 1 / \log(\text{node\_max})))$$

Pour un concept C donné, ou  $\text{depth}(c)$  c'est la profondeur du concept C dans la taxonomie,  $\text{depth\_max}$  dénote la profondeur maximale de la taxonomie, a est un concept de la taxonomie, ce qui satisfait  $a \in \text{hypo}(c)$ ,  $\text{node\_max}$  représente le nombre maximum de concepts qui existent dans la taxonomie.

Si C est la racine,  $\text{depth}(C)$  égale 1, Alors  $\log(\text{depth}(c)) = \log(1) = 0$ .

Si C est une feuille,  $\text{hypo}(c)$  égale 0, Alors  $\sum_{a \in \text{hypo}(c)} 1 / \text{depth}(a) = 0$

Et

$$IC(c) = \log(\text{depth}(c)) / \log(\text{depth\_max})$$

### 3.3 -L'approche de Sanchez :

#### -Approche 1 :

Le CI basée sur les feuilles a proposé en 2011 par Sanchez, Pour Sanchez Les feuilles représentent la sémantique des concepts les plus spécifiques d'un domaine, et elles sont suffisantes pour décrire et différencier le concept de tous autres, quelle que soit le nombre de concepts internes incorporées dans la taxonomie. Il prétend plus qu'un concept possède de feuilles moins il exprime d'information. En d'autres termes, un concept avec plus de feuilles est plus général.

La profondeur du concept en taxonomie est considérée comme Zhou a proposé .ici, la profondeur est au lieu de la profondeur avec le nombre des concepts qui les subsument.

Le CI d'un concept est défini comme suit:

$$CI(c)_{\text{Sanchez}} = -\log\left(\frac{|\text{leaves}(c)|}{|\text{subsumers}(c)| + 1} \cdot \frac{1}{\max\_leaves + 1}\right)$$

Ou C est un concept donné de l'ensemble des concepts de l'ontologie.

Avec :

1-max\_leaves représente le nombre de feuilles dans la hiérarchie

2- leaves(c): l'ensemble des feuilles de c.

$$\text{leaves}(c) = \{l \in C \mid l \in \text{hyponyms}(c) \wedge l \text{ est une feuille}\}$$

3- subsumers(c) renvoie l'ensemble des nœuds généralisant de c.

$\text{subsumers}(c) = \{a \in C \mid c \leq a\} \cup \{c\}$ ,  $c \leq a$  : c est une spécialisation de a.

### -Approche 2 :

#### **Définition 1 :**

$\text{subsumers}(c) = \{a \in C \mid c \leq a\} \cup \{c\}$ ,  $c \leq a$  : c est une spécialisation de a.

#### **Définition 2 :**

Le commonness d'une feuille l est inversement proportionnel à son nombre de subsumers :

$$\text{Commonness}(l) = \frac{1}{|\text{subsumers}(c)|}$$

#### **Définition 3 :**

Le commonness d'un concept taxonomique intérieur s (c.-à-d., s n'est pas une feuille) est calculé comme la somme des valeurs de commonness de toutes les feuilles englobées par s :

$$\text{Commonness}(S) = \sum_{\text{leaf } l \wedge l < s} \text{Commonness}(l)$$

#### **Définition 4 :**

L'IC d'un concept c est défini comme suit:

$$\text{IC}(c) = -\log \frac{\text{Commonness}(l)}{\text{Commonness}(\text{root})}$$

### 3.4- L'approche de Sebti :

Sebti [Sebti, 2008] présente une nouvelle méthode de calcul de CI des concepts dans une structure hiérarchique basée sur les hyponymes directs. Cette méthode inclut implicitement la notion de profondeur d'un concept cible. Leur méthode de calcul de valeurs CI est basée sur

le fait que la taxonomie « isa » de WordNet est organisée suivant une hiérarchie significative. C'est à dire, les concepts se trouvant en profondeur plus élevée dans la taxonomie « is a » du wordnet) sont plus informatifs et leurs CI sont plus grandes.

Le contenu informationnel de Sebti d'un concept c est calculé de la manière suivante :

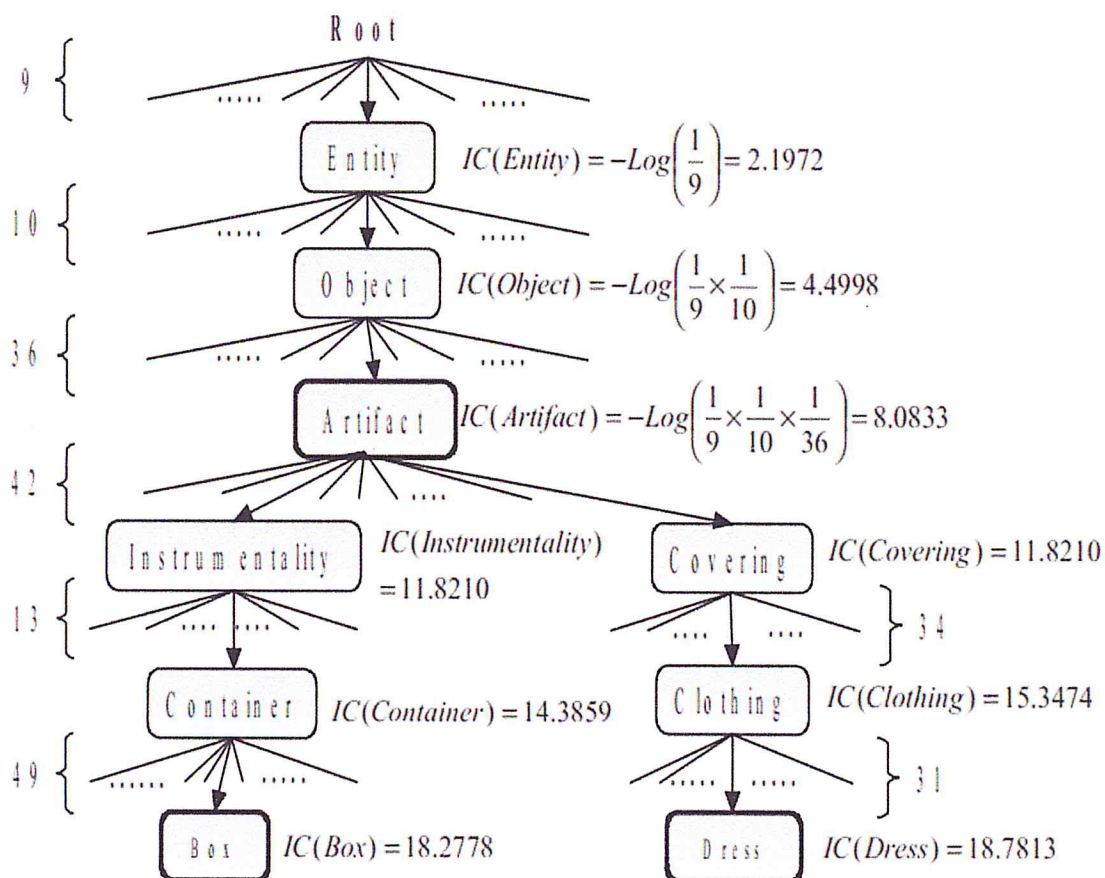
$$CI(c) = \max_{p \in \text{path}(c)} \left( -\log \left( \prod_{c' \in \text{chemin}(c)} \frac{1}{\text{hypo}(c')} \right) \right)$$

Avec

- Chemin(c) représente l'ensemble des concepts figurant dans le chemin le plus court reliant c à la racine de l'ontologie.
- Path(c) représente l'ensemble chemin reliant c à la racine de l'ontologie.
- Hypo(c') retourne le nombre des hyponymes directs du concept c'.

### Exemple d'utilisation :

La figure ci-dessous explique le processus suivi afin de calculer le CI d'un concept donné. Comme nous le remarquons cette méthode est basée sur le nombre d'hyponymes directs de Chaque concept appartenant au chemin débutant de la racine jusqu'au concept cible.



### Figure 3.3: Exemple de calcul d'IC pour quelques concepts avec l'approche de Sebti

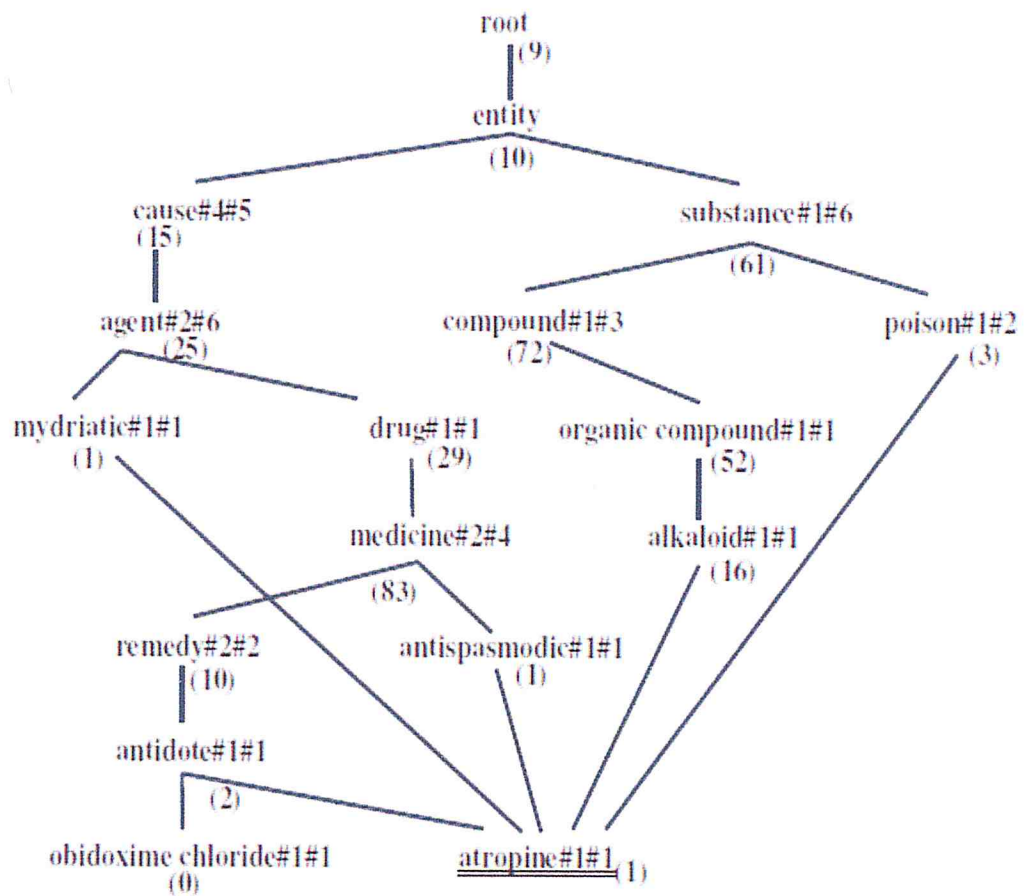
Pour mieux comprendre cette méthode, nous présentons dans l'équation suivante

La méthode avec laquelle le CI est calculé pour Boxen relation avec la figure ci-dessus.

$$CI(Box) = -\text{Log} \left( \frac{1}{9} \times \frac{1}{10} \times \frac{1}{36} \times \frac{1}{42} \times \frac{1}{13} \times \frac{1}{49} \right) = 18.2778$$

#### Limite de L'approche de Sebti :

En appliquant la méthode de Sebti pour calculer le CI des concepts *atropine* et *obidoximechloride* en suivant le chemin qui nous garantit le plus grand CI nous trouvons le suivant :



**Fig 3.4 Extrait de WordNet 2.0**

Le fragment ci-dessus représente un extrait du wordnet et les nombres entre parenthèses sont les nombres des hyponymes directs de chaque concept. Nous pouvons remarquer que si nous suivons la méthode de Sebti. Pour calculer le CI des concepts *atropine* et *obidoximechloride* en suivant le chemin qui nous garantit le plus grand CI nous trouvons le suivant :

$$CI(atropine) = -\text{Log} \left( \frac{1}{9} \times \frac{1}{10} \times \frac{1}{61} \times \frac{1}{72} \times \frac{1}{52} \times \frac{1}{16} \right) = 19.6111$$

$$CI(obidoxime\ chloride) = -\text{Log} \left( \frac{1}{9} \times \frac{1}{10} \times \frac{1}{15} \times \frac{1}{25} \times \frac{1}{29} \times \frac{1}{83} \times \frac{1}{10} \right) \\ = 20.5154$$

D'après le calcul ci-dessus, le contenu informationnel du concept *obidoximechloride* est plus grand que celui du concept *atropine* ce qui n'est pas logique. En effet, le concept *atropine* possède 5 hyperonymes tandis que l'autre a un seul hyperonyme. Le concept qui hérite les données de 5 concepts ne peut pas avoir un contenu informationnel inférieur à celui qui hérite d'un seul concept parmi les cinq du premier.



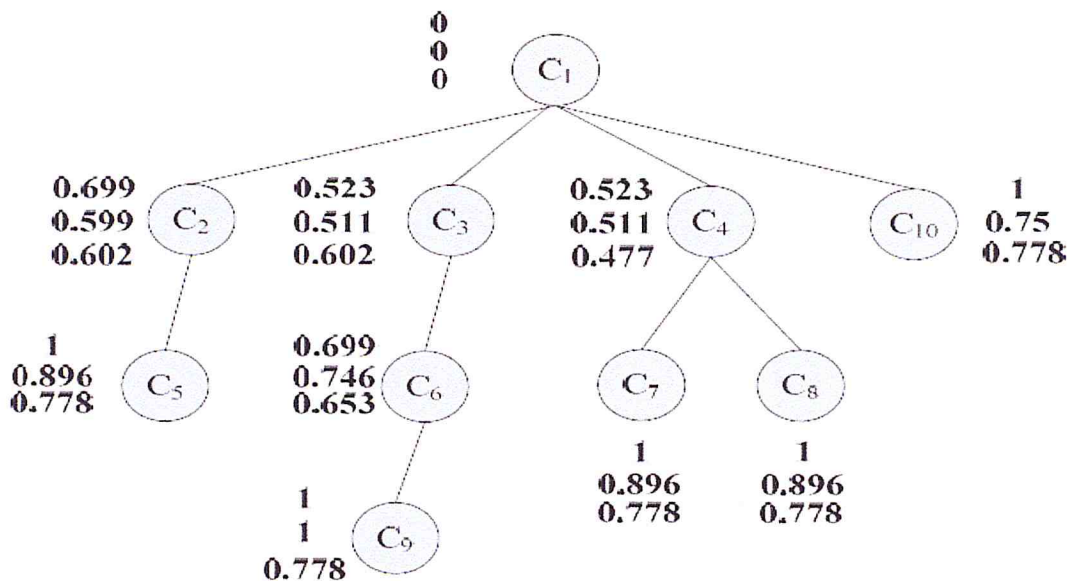
De plus, la présence des concepts ayant seulement un seul hyponyme directe laisse invariant le CI d'un concept donné, c'est-à-dire la participation de ce concept sera neutre.

Exemple

$$CI(atropine) = -\text{Log} \left( \frac{1}{9} \times \frac{1}{10} \times \frac{1}{15} \times \frac{1}{25} \times \frac{1}{29} \times \frac{1}{83} \times \frac{1}{1} \right) = 18.2128$$

La participation du concept *antispasmodicest* neutre dans le calcul du CI du concept *atropine* en suivant un autre chemin.

**Exemple de calcul et d'évaluation d'CI pour (Seco, Zhou et Sanchez,Sebti) :**



**Note:**

Ligne 1 : résultat avec CI basé sur les hyponymes (L'approche de Seco)

Ligne 2 : résultat avec CI basé sur les hyponymes (L'approche de Zhou)

Ligne 3 : résultat avec CI basé sur les feuilles (L'approche de Sanchez)

Figure 3.5 : Un exemple d'une taxonomie des concepts

concept	node max	depth max	depth	le nombre du hypo (c)	CI(c) Seco	CI(c) Zhou	CI(c) Zhou 2	CI(c) Sanchez	CI(c) Sanchez 2	CI(c) Sebti
<i>C1</i>	10	4	1	9	0	0	0	0	0	0
<i>C2</i>	10	4	2	1	0.699	0.599	0,72	0,602	0,5	0,6
<i>C3</i>	10	4	2	2	0.523	0.511	0,84	0.602	0,5	0,6
<i>C4</i>	10	4	2	2	0.523	0.511	0,41	0.477	0,5	0,6
<i>C5</i>	10	4	3	0	1	0.896	0,72	0.778	0,79	0,6
<i>C6</i>	10	4	3	1	0.699	0.746	0,84	0.653	0,79	0,6
<i>C7</i>	10	4	3	0	1	0.896	0,72	0.778	0,79	0,9
<i>C8</i>	10	4	3	0	1	0.896	0,72	0.778	0,79	0,9
<i>C9</i>	10	4	4	0	1	1	0,84	0.778	1	0,6
<i>C10</i>	10	4	2	0	1	0.75	0,54	0.778	0,5	0,6

*Tableau 3.1 : les valeurs d'IC pour chaque concept dans la figure*

### Evaluation:

Avec l'approche de Seco et le tableau les valeurs de CI sont égaux pour les concepts C2 et C6 et de même pour les concepts C7 et C10, C8 et C10, C9 et C5. Malgré que ces concepts ne trouvent pas dans le même niveau et la même profondeur, ce qui n'est pas logique car la profondeur est significative dans la taxonomie et le concept qui possède la profondeur la plus élevée doit posséder la valeur d'IC la plus grande.

Mais Avec la formule de Zhou nous trouvons cela :

Premièrement, l'avantage principal de cette approche est qu'elle a pris en considération la profondeur et aborde le problème qui surgit dans l'approche de Seco, avec la formule de Zhou si deux concepts ont le même nombre de hyponymes, le plus profond est le plus concret et à la valeur la plus élevée d'IC que l'autre, ici,  $CI(C2)=0.599$ ,  $CI(C6)=0.746$ . Ce qui est logique parce que  $profondeur(C2)=2$  et  $profondeur(C6)=3$  et le passage d'une profondeur à une profondeur plus élevée signifie un transfert de donnée et le concept qui se situe dans une profondeur plus élevée possède la valeur d'IC la plus grande  $CI(C6)>CI(C2)$ . Le concept C9 possède la profondeur maximale  $profondeur(C9)=4$  et possède aussi la valeur la plus élevée d'IC  $CI(C2)=1$

Avec la deuxième formule de Zhou nous trouvons cela :

$CI(C2)=0.72$ ,  $CI(C6)=0.84$ . Ce qui est acceptable parce que  $profondeur(C2)=2$  et  $profondeur(C6)=3$  et le passage d'une profondeur à une profondeur plus élevée signifie un transfert de donnée et le concept qui se situe dans une profondeur plus élevée possède la valeur d'IC la plus grande  $CI(C6)>CI(C2)$ .

Les concepts C2 et C5 possèdent le même nombre des feuilles qui est égale à 1 et ses valeurs de CI sont égaux  $CI(C2)=CI(C5)=0,72$  et de même pour les concepts C3 et C6 et C9 possèdent le même nombre des feuilles qui est égale à 1 et ses valeurs de CI sont égaux  $CI(C3)=CI(C6)=CI(C9)=0,84$  ce qui n'est pas logique parce que ces concepts ne trouvent pas dans la même profondeur.

Avec l'approche de Sanchez on trouve que les feuilles possèdent la même valeur de CI  $CI(C5)=CI(C9)=CI(C7)=CI(C8)=CI(C10)$  ce qui n'est pas logique parce que les feuilles qui se situent dans des profondeurs faibles logiquement sa valeur d'IC est moins que la valeur d'IC des feuilles situent dans des profondeurs plus élevées dans la taxonomie des concepts. Ce qui est effectué par l'approche de Zhou d'après le tableau on trouve  $CI(C10)< CI(C8)< CI(C9)$  ce qui est acceptable.

Avec la deuxième formule de Sanchez on trouve que les concepts qui se situent dans la même profondeur possèdent la même valeur de CI

La profondeur 2 :  $CI(C2)= CI(C3)= CI(C4)= CI(C10)= 0,5$

La profondeur 3 :  $CI(C5)= CI(C6)= CI(C7)= CI(C8)= 0,79$

Cette formule ne prend pas en considération les hyponymes des concepts dans le calcul de la valeur de CI.

Avec l'approche de Sebtî le grand problème c'est la présence des concepts ayant seulement un seul hyponyme directe dans le chemin reliant le concept à la racine laisse invariant le CI

d'un concept donné  $CI(C2)= CI(C3)= CI(C4)= CI(C5)=CI(C6)= CI(C9)=CI(C10)= 0,6$  malgré que ces concepts ne trouvent pas dans le même profondeur on trouve que ses valeur d'IC sont égaux  $CI(C2)= CI(C9)$  la profondeur de  $C2=2$  et la profondeur  $C9=4$  et le passage d'une profondeur a une profondeur plus élevée signifie un transfert de donnée et le concept qui situe dans une profondeur plus élevée possède la valeur d'IC la plus grande ce qui n'est pas garantie avec cette approche.

### 3.5- L'approche de Hanif Seddiqui

Hanif Seddiqui propose de calculer le contenu informationnel en se basant sur les relations. En plus du concept et sa hiérarchie, l'approche prend en considération les fonctions de propriété, les restrictions de propriété et d'autres assertions logiques.

Le contenu informationnel basé sur les relations est défini comme suit:

$$ic_{rel}(c) = \frac{\log(rel(c) + 1)}{\log(total\_rel + 1)}$$

où

*rel* représente la relation de propriétés, fonctions de propriété et les restrictions de propriété,

*rel(c)* désigne le nombre de relations d'un concept *c* et

*total\_rel* représente le nombre total de relations disponibles dans l'ontologie.

Du fait que le contenu de l'information d'un concept dépend à la fois sur les hyponymes ou des relations de subsomption d'un concept et des propriétés connexes du concept, nous devons intégrer l'IC<sub>rel</sub>(c) avec mesure de Seco. Cette intégration introduit un coefficient  $\rho$  et l'équation devient aussi:

$$ic(c) = \rho \cdot ic_{rel}(c) + (1 - \rho) \cdot ic_{seco}(c)$$

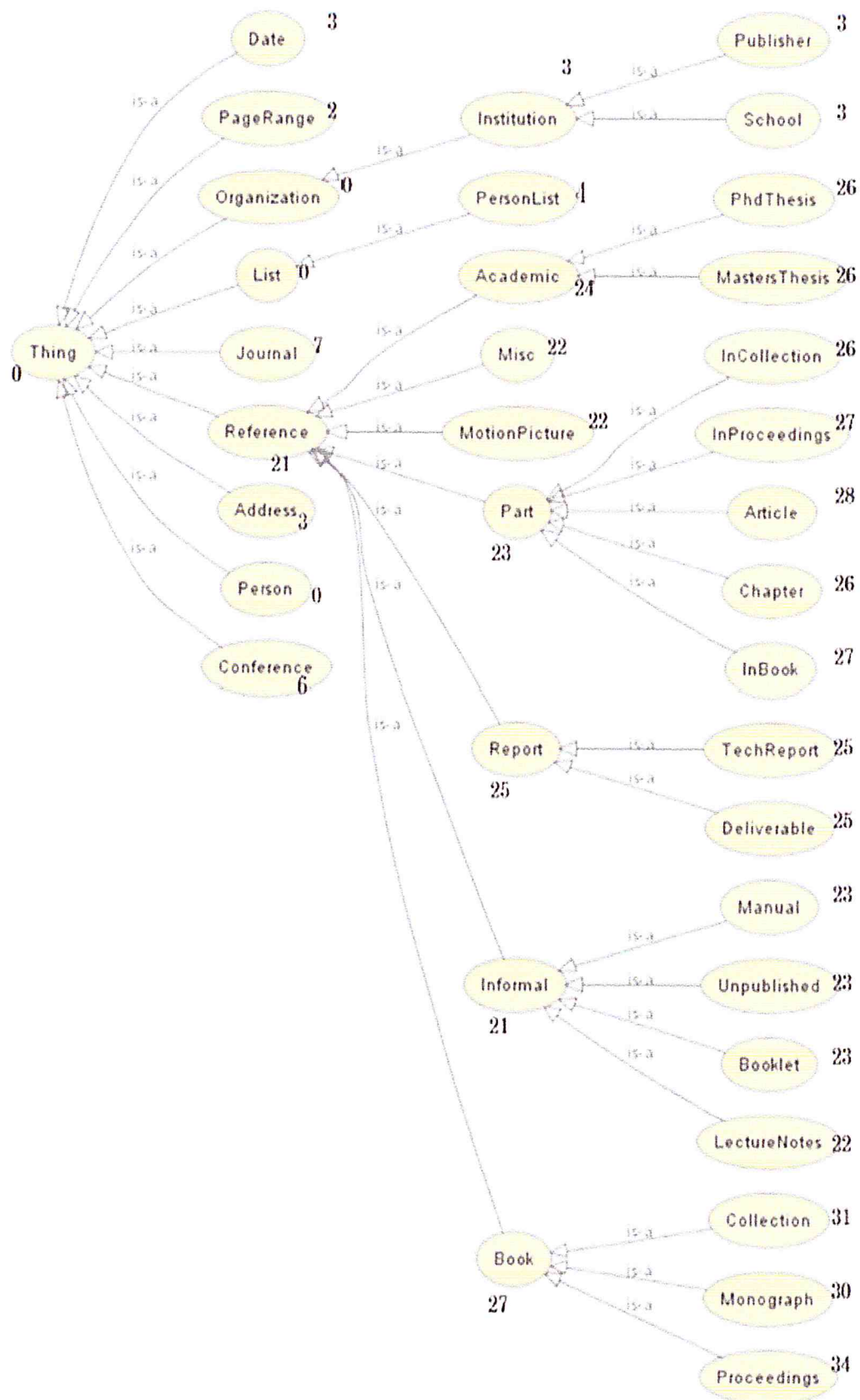
Où le coefficient  $\rho$  est défini par la nature d'une ontologie. Lorsque l'ontologie est de petite taille ou incomplète, le coefficient tend à augmenter pour se concentrer sur les relations. Au contraire, lorsque les relations ne suffisent pas à définir un concept et il y'a un grand nombre de concepts dans la taxonomie,  $\rho$  tend à diminuer.

Cependant, on a absolument besoin d'un compromis pour sélectionner le coefficient  $\rho$  et on le définit comme suit:

$$\rho = \frac{\log(\text{total\_rel} + 1)}{\log(\text{total\_rel}) + \log(\text{total\_concept})}$$

Où

*total\_rel* désigne le nombre maximum de relations et  
*total\_concept* est le nombre maximum de concepts de l'ontologie.



**Fig 3.6 nombre de relations dans une taxonomie d'une ontologie.**

Concepts	Number of Relations	Number of Hyponyms	$ic_{seco}$	$ic_{rel}$	$ic_{Seddiqui}$
Date	3	0	1.000	0.332	0.641
PageRange	2	0	1.000	0.263	0.603
Organization	0	3	0.613	0.000	0.283
Institution	3	2	0.693	0.332	0.499
Publisher	3	0	1.000	0.332	0.641
School	3	0	1.000	0.332	0.641
List	0	1	0.807	0.000	0.373
PersonList	4	0	1.000	0.386	0.670
Journal	7	0	1.000	0.498	0.730
Address	3	0	1.000	0.332	0.641
Person	0	0	1.000	0.000	0.462
Conference	6	0	1.000	0.466	0.713
Reference	21	23	0.113	0.740	0.450
Academic	24	2	0.693	0.771	0.735
PhdThesis	26	0	1.000	0.790	0.887
MastersThesis	26	0	1.000	0.790	0.887
Misc	22	0	1.000	0.751	0.866
MotionPicture	22	0	1.000	0.751	0.866
Part	23	5	0.500	0.761	0.640
InCollection	26	0	1.000	0.790	0.887
InProceedings	27	0	1.000	0.798	0.891
Article	28	0	1.000	0.807	0.896
Chapter	26	0	1.000	0.790	0.887
InBook	27	0	1.000	0.798	0.891
Report	25	2	0.693	0.781	0.740
TechReport	25	0	1.000	0.781	0.882
Deliverable	25	0	1.000	0.781	0.882
Informal	21	4	0.551	0.740	0.653
Manual	23	0	1.000	0.761	0.871
Unpublished	23	0	1.000	0.761	0.871
Booklet	23	0	1.000	0.761	0.871
LectureNotes	22	0	1.000	0.751	0.866
Book	27	3	0.613	0.798	0.713
Collection	31	0	1.000	0.830	0.909
Monograph	30	0	1.000	0.823	0.905
Proceedings	34	0	1.000	0.852	0.920

**Tableau 3.2 : contient des valeurs d'IC mesurées par Seco et Hanif Seddiqui**

### 3.6-Le CI de Hadj Taieb :

Hadj Taieb [Hadj Taieb,2012 ] propose d'exploiter d'une façon plus approfondie la structure WordNet et spécialement la taxonomie « IS A » (Hyperonyme/Hyponyme).

Les deux principales idées participantes à la création de cette nouvelle méthode de calcul du contenu informationnel sont expliquées dans les paragraphes suivants :

#### La relation Hyperonyme/Hyponyme « IS A » :

La relation C2(concept C2) is a C1(concept C1), c'est-à-dire C1est inclus dans C2. Autrement dit, le concept C2est formé par les données de C1plus certaines spécificités qui le différencie de son père. Il en résulte donc qu'un concept est une accumulation de l'information propagée d'un ancêtre un autre en ajoutant quelques spécificités à chaque descendant. D'où, il est trivial qu'un concept dépend fortement de ses parents directs et ses ancêtres. L'ensemble des relations d'hyperonymes directs et indirects d'un concept C forment un sous arbre qui participera à la quantification de son CI.

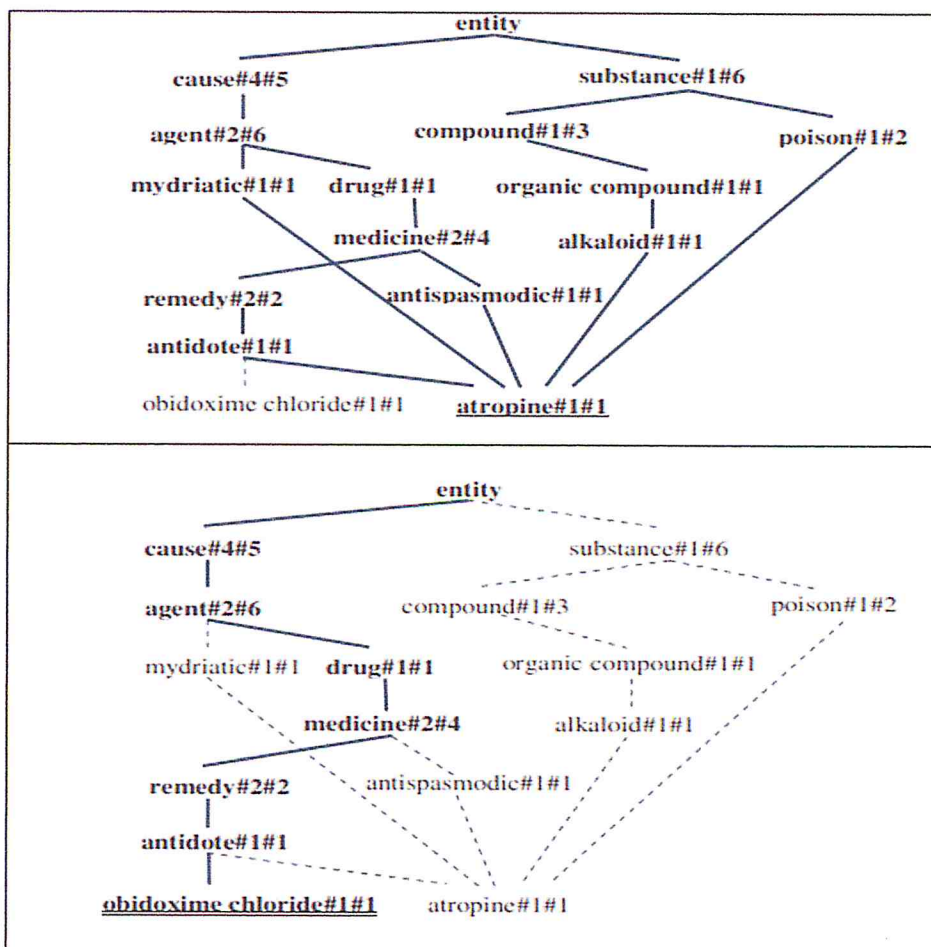


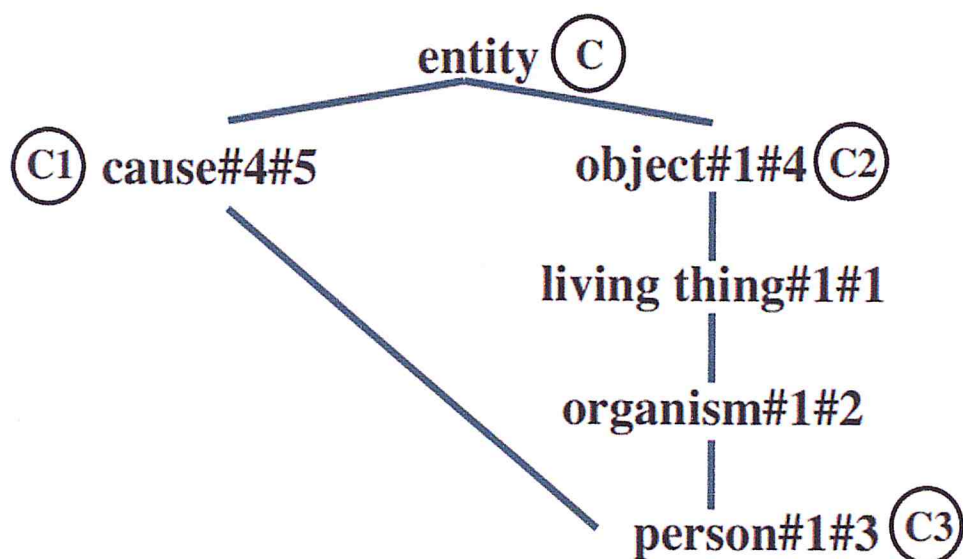
Figure3.7:Les sous arbre définissant CI (atropine) et CI (obidoximechloride).



La figure ci-dessus montre que le sous arbre de CI (atropine) est plus riche que celui du CI (obidoximechloride). Cette richesse ( $\text{hyperonymes\_directs(atropine)}=5$  et  $\text{hyperonymes(obidoximechloride)}=1$ ) doit être transmise au niveau du calcul du CI. Ce qui reste, c'est de quantifier la participation de chaque concept (appartenant au sous arbre) dans la totalité du CI. Cette participation dépend du niveau de profondeur de chaque concept comme il est expliqué dans le paragraphe suivant.

### La signification de la profondeur dans la taxonomie « IS A »:

Chaque concept sauf la racine possède un ou plusieurs hyperonymes (jusqu'à 6 en WordNet 2.0). Ces hyperonymes ne se trouvent pas à la même profondeur dans la taxonomie « IS A ». La profondeur est significative car le passage d'un niveau à un autre plus approfondi engendre le passage des données vers les descendants avec certaines spécificités. La transition d'un concept C1 vers un concept C2 en utilisant la relation d'hyponymie ne signifie pas le passage d'une profondeur  $i$  à une profondeur  $i+1$ . Mais, il se peut que malgré cette relation directe entre ces deux concepts, ils se trouvent à des profondeurs non successives dans l'arbre taxinomique « IS A ».



**Figure 3.8 Extrait du WordNet 2.0**

Dans la figure le concept Person possède deux hyperonymes organism(profondeur=4) et cause(profondeur=2). Vu que le concept organism a été enrichi par les concepts (entity, object, living thing) et le concept cause inclut seulement le concept entity. Il est évident que la part du organism est plus grande que celle du cause dans le CI de person. D'où, Hadj Taieb pense à intégrer la notion de profondeur dans votre méthode.

## Calcul du CI de Hadj Taieb:

Pour le calcul du CI d'un concept Hadj Taieb prend en considération les points suivant :

- 1- L'arbre des hyperonymes d'un concept
- 2- La profondeur de chaque concept de l'arbre des hyperonymes.
- 3- Les hyponymes.
- 4- La profondeur du concept

Il propose de quantifier le sous arbre des hyperonymes d'un concept donné en prenant en considération l'influence de la profondeur de chaque concept. Soit :

- **Hyper(c)**: une fonction qui retourne l'ensemble des hyperonymes directs du concept  $c$ .
- **Hypo(c)** : une fonction qui retourne le nombre des concepts subsumés par le concept  $c$ .
- **Arbre(c)** : L'ensemble de tous les concepts (hyperonymes) participants au sous arbre modélisant le contenu informationnel du concept  $c$  : c'est ensembles des hyperonymes directs et indirects du concept  $c$ .
- **Depth(c)**: une fonction qui retourne la profondeur maximale d'un concept dans le thesaurus WordNet.
- **DepthMoy(c)** : une fonction qui retourne la profondeur moyenne du sous arbre formant le CI du concept  $c$ .

$$\text{DepthMoy}(c) = \frac{1}{\|\text{Arbre}(c)\|} \times \sum_{c' \in \text{Arbre}(c)} \text{Depth}(c')$$

- **score(c)** : la contribution de chacun des hyperonymes dans le calcul du CI du concept  $c$ .

$$\text{Score}(c) = \left( \sum_{c' \in \text{Hyper}(c) \cap \text{Arbre}(c)} \frac{\text{Depth}(c')}{\text{Hypo}(c')} \right) \times \text{Hypo}(c)$$

Le calcul du contenu informatif du concept  $c$  est proposé dans deux version :

**Versión 1** : Une première version est proposée sans la profondeur moyenne : ,

$$\text{CI}(c) = \left( \sum_{c' \in \text{Arbre}(c)} \text{Score}(c') \right)$$

**Version 2** : Elle utilise la profondeur moyenne du sous arbre participante au contenu informationnel du concept c.

$$CI(c) = \left( \sum_{c' \in \text{Arbres}(c)} \text{Score}(c') \right) \times \text{DepthMoy}(c)$$

#### 4-Comparaison de différents modèles d'IC :

	caractéristique résultats				
	le modèle du CI dépendante d'un corpus	le modèle du CI indépendante d'un corpus			
	L'approche de resnik	L'approche de seco	L'approche de Zhou	L'approche de Sanchez	L'approche de Hanif Seddiqui
Augmentation Depth(c)	NON	NON	OUI Augmentation d'IC	OUI Augmentation d'IC	OUI Augmentation d'IC
Augmentation Hypo(c)	NON	OUI Diminution d'IC	OUI Diminution d'IC	NON	OUI Diminution d'IC
Augmentation Leaves(c)	NON	NON	NON	OUI Diminution d'IC	NON
Augmentation Rel(c)	NON	NON	NON	NON	OUI ICrel (c) Augmentation
compter sur un corpus	OUI	NON	NON	NON	NON
Problème clairsemé de données	OUI	NON	NON	NON	NON
Suggestions pour l'Amélioration	-----	Prendre la topologie de chaque concept en considération,comme,l'arrangement des concepts		Tenir compte des nombres des hyponymes, profondeur des feuilles	Prendre en considération le type de lien, structure d'espace du concept

**Tableau 3.3: Comparaison des différents modèles d'IC**

## 5-Discussion et solutions proposées :

Le contenu informationnel est capital dans le calcul de la similarité sémantique il Y a un problème de clairsemé des données dans l'approche de resnik qui utilise une ressource externe (corpus), dans le travail de seco, la profondeur n'est pas pris en considération, Zhou apporte quelques améliorations. Cependant, dans l'approche de Zhou quand les concepts ont la même profondeur et ont le même nombre des hyponymes leurs valeurs d'IC seront égaux (ex.  $CI(C3) = CI(C4)$  dans la Figure) il y a ici une pièce pour l'amélioration, Sanchez dit que les feuilles sont assez suffisantes pour décrire et différencier un concepts de tous les autres mais on le note que les concepts avec le même nombre de feuilles et le même nombre de concepts qui les généralisent auront la même valeur d'IC aussi (ex.  $CI(C2) = CI(C3)$  dans la Figure) pourtant ils sont différents dans le nombre des hyponymes ( $hypo(C2)=1, hypo(C3)=2$ )

Dans le travail de l'amélioration, elle sera une bonne idée de prendre la topologie complètement de chaque concept en considération, comme la profondeur, le nombre de hyponymes, la profondeur de chaque hyponymes dans la taxonomie, l'arrangement des concepts, etc...

## 6-Conclusion:

Ce rapport passe en revue les différentes approches de calcul du contenu informationnel, CI est obtenu grâce à deux méthodes, la première utilise une analyse statistique d'un corpus pour attribuer des probabilités à chaque concept. Tandis la deuxième méthode elle exploite la structure de WordNet sans faire recours à une ressource externe. Elle inclut le modèle de CI basé sur les hyponymes. Le modèles basé les feuilles basée, le modèle basé sur les relations et le modèle qui prend en considération la structure détaillée de l'ontologie Wordnet.

Nous évaluons ces distinctes approches qui calculent cette valeur notre but est fortement efficace distinguer les différentes concepts et obtenir la valeur d'IC la plus adéquate.

# -Chapitre 4 : Implémentation

## 4.1-Réalisation:

### Introduction:

Dans cette partie nous présenterons les principales parties de notre application qui est basée sur le langage java. Nous commencerons par une définition générale de ce langage et après nous exposerons la fenêtre principale de notre application.

### Définition Java:

Le langage Java est un langage de programmation informatique orienté objet créé par James Gosling et Patrick Naughton, employés de Sun Microsystems, avec le soutien de Bill Joy (cofondateur de Sun Microsystems en 1982), présenté officiellement le 23 mai 1995 au *SunWorld*.

La société Sun a été ensuite rachetée en 2009 par la société Oracle qui détient et maintient désormais Java.

La particularité et l'objectif central de Java est que les logiciels écrits dans ce langage doivent être très facilement portables sur plusieurs systèmes d'exploitation tels que UNIX, Windows, Mac OS ou GNU/Linux, avec peu ou pas de modifications. Pour cela, divers plateformes et Framework associés visent à guider, sinon garantir, cette portabilité des applications développées en Java.

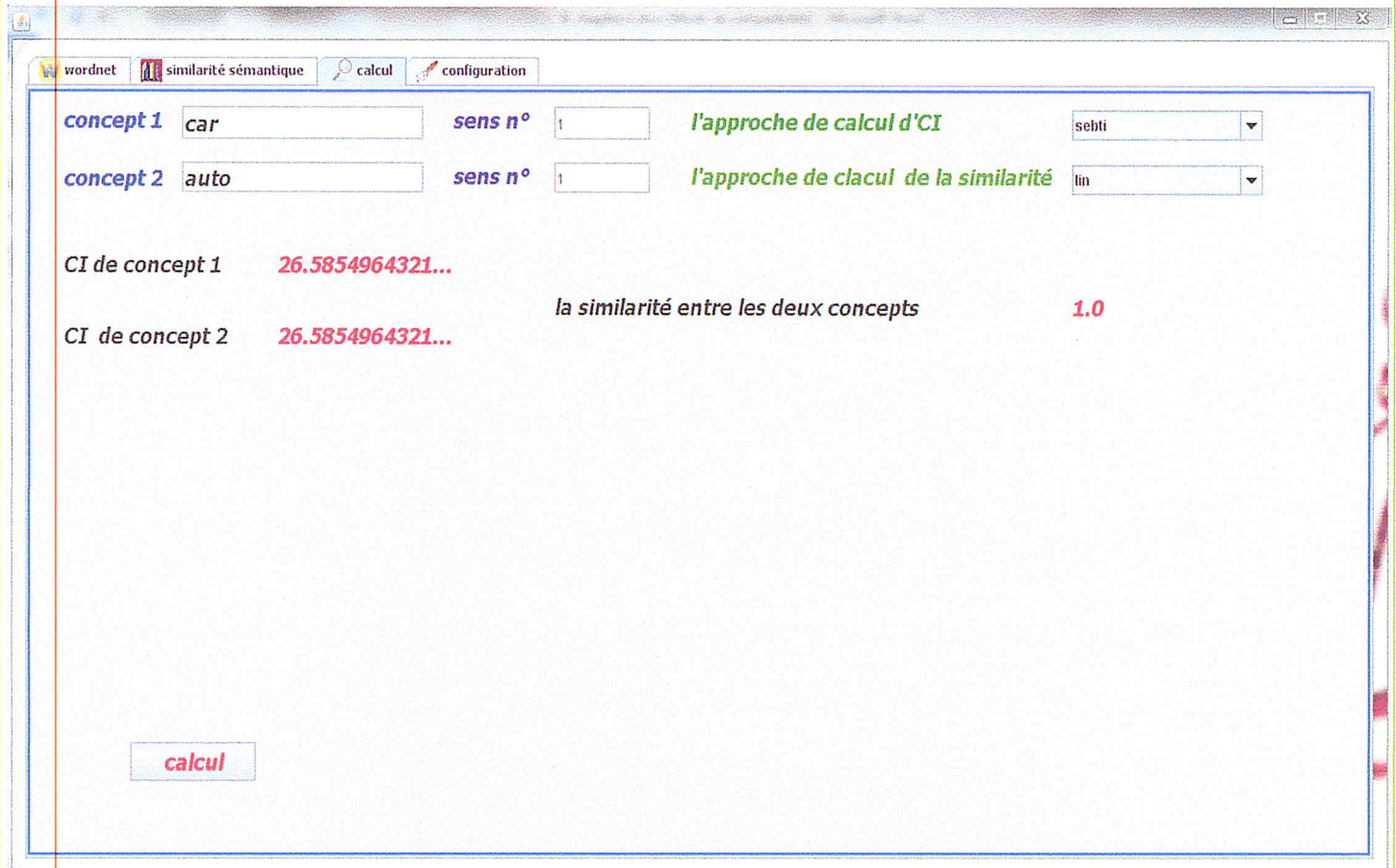
Le langage Java reprend en grande partie la syntaxe du langage C++, très utilisée par les informaticiens. Néanmoins, Java a été épurée des concepts les plus subtils du C++ et à la fois les plus déroutants, tels que les pointeurs et références, ou l'héritage multiple contourné par l'implémentation des 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.)

On peut faire de nombreuses sortes de programmes avec Java :

- Des applications, sous forme de fenêtre ou de console ;
- des applets, qui sont des programmes Java incorporés à des pages web ;
- des applications pour appareils mobiles, avec J2ME ;
- et bien d'autres ! J2EE, JMF, J3D pour la 3D...

## Présentation de l'application :

### Page d'accueil :



**Fig4.1** La Fenêtre principale de notre application

### Explication:

Avec cette interface on peut calculer La valeur de la similarité sémantique entre n'importe quelle couple de deux concepts avec les numéros des sens de ces deux concepts dans la base de données Wordnet avec n'importe quelle approche de calcul de la valeur de contenu informationnel couplée avec une mesure de similarité sémantique basée sur le contenu informationnel.

## 4.3 Testset Résultats:

### Introduction:

Dans cette partie nous donnons les tests et les résultats de notre travail d'évaluation des différentes approches de calcul de contenu informationnel pour faire ce travail il faut premièrement calculer les valeurs d'CI pour chaque concept de la liste des jugements humains de Miller & Charles [1991] de 30 couples des concepts et la liste de Rubenstein & Goudenough [1965] de 65 couples avec les différentes approches puis nous calculons la similarité sémantique entre les différents couples avec les mesures de Resnik [1995], Giang\_Conrath [1997], Lin [1998] et à l'étape suivante nous calculons les corrélations entre les couples de la liste des jugements humains de Miller & Charles [1991] et Rubenstein & Goudenough [1965] avec les différentes approches couplées avec les trois mesures de calcul des similarités sémantiques basées sur le contenu informationnel Resnik [1995], Giang\_Conrath [1997], Lin [1998] et en fin nous traçons Les graphes des comparaisons des Similarités entre les couples de la liste des jugements humains de Miller & Charles [1991] ] de 30 couples des concepts et la liste de Rubenstein & Goudenough [1965] de 65 couples avec les différentes approches d'CI couplées avec les trois mesures de calcul des similarités sémantiques basées sur le contenu informationnel Resnik [1995], Giang\_Conrath [1997], Lin [1998].

## 4.2.1-Miller & Charles [1991] :

Les valeurs des contenus informationnels pour chaque concept de la liste des jugements humains de Miller & Charles [1991] de 60 concepts avec les différentes approches d'CI:

les concepts	ic resnik	ic seco	ic sanchez	ic sanchez2	ic shou	ic shou2	ic sebt1	ic hadjtiebl	ic hadjtieb2
car#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069304	207.2596
automobile#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069304	207.2596
autolum#2	11.416422	0.9381574	4.8023195	4.890783	0.8600344	0.7759406	22.058949	14.732128	73.910649
madhouse#1	11.821887	1.0	4.805924	4.890783	0.9071905	0.81438106	22.058949	15.907128	87.489204
midday#1	10.31781	1.0	4.805924	4.890783	0.9071905	0.81438106	14.876495	25.269593	138.97946
noon#1	10.31781	1.0	4.805924	4.890783	0.9071905	0.81438106	14.876495	25.269593	138.97946
rooster#1	11.416422	0.9381574	4.8115277	4.995518	0.9046369	0.8657543	28.660517	32.15485	209.00456
voyage#1	10.030128	0.9019818	4.7767653	4.6623034	0.87295693	0.84109044	23.592619	22.772133	136.6329
gem#3	0.0	1.0	4.7925596	4.75248	0.8304382	0.6608764	17.445349	12.786334	43.838963
jewel#2	0.0	1.0	4.7925596	4.75248	0.8304382	0.6608764	17.445349	12.786334	43.838963
journey#1	5.735112	0.6509692	4.1775346	3.3624902	0.73267514	0.81426525	22.494005	21.982658	120.90462
voyage#2	11.821887	0.9019818	4.7767653	4.6623034	0.87295693	0.84109044	25.265654	26.280503	157.69542
boy#1	7.4911537	0.73729706	4.4177437	3.7544572	0.72176194	0.7058737	19.843245	12.878388	49.903755
lad#2	10.905597	1.0	4.8023195	4.84939	0.8731143	0.7462286	22.241138	14.067067	60.957294
coast#1	9.470511	0.8564059	4.6218634	4.092428	0.73246455	0.60661805	10.749784	2.2071805	6.6215415
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6805374	0.54553374	9.651173	2.0443897	5.110975
magician#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.784492	49.59416
wizard#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.784492	49.59416
journey#1	5.735112	0.6509692	4.1775346	3.3624902	0.73267514	0.81426525	22.494005	21.982658	120.90462
car#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069304	207.2596
furnace#1	12.515034	0.7242171	4.331829	3.5655966	0.71522194	0.7059061	17.057107	9.897156	39.588623
stove#1	10.569124	0.8144722	4.64414	4.1226997	0.8292022	0.8431092	20.500122	34.592335	207.554
food#3	11.821887	0.9381574	4.7857203	4.75248	0.79951696	0.6839315	10.312147	7.268106	25.42437
fruit#2	11.821887	1.0	4.8023195	4.84939	0.89100575	0.7820115	16.408876	11.849773	58.248967
bird#1	7.655222	0.39591002	2.9817135	2.1563246	0.5710693	0.7462227	18.717808	20.503105	92.26398
cock#5	12.515034	0.9381574	4.8023195	4.890783	0.8600344	0.7759406	21.975904	22.772812	113.86406
bird#1	7.655222	0.39591002	2.9817135	2.1563246	0.5710693	0.7462227	18.717808	20.503105	92.26398
cranes#5	0.0	0.9381574	4.8089504	4.9633336	0.89104474	0.83835214	26.478714	30.130323	180.78194
tool#1	8.238368	0.4781231	3.2682397	2.4582218	0.59217495	0.70621106	17.55218	10.23434	40.93706
implement#1	7.013776	0.41250476	2.907055	2.1129973	0.5366906	0.6608697	14.150982	9.437499	33.031246
brother#5	0.0	1.0	4.805924	4.890783	0.89100575	0.7820115	24.10948	14.981019	71.9089
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
cranes#5	0.0	0.9381574	4.8089504	4.9633336	0.89104474	0.83835214	26.478714	30.130323	180.78194
implement#1	7.013776	0.41250476	2.907055	2.1129973	0.5366906	0.6608697	14.150982	9.437499	33.031246
lad#2	10.905597	1.0	4.8023195	4.84939	0.8731143	0.7462286	22.241138	14.067067	60.957294
brother#2	10.723275	0.9019818	4.805924	4.9633336	0.8419966	0.77904207	24.630735	28.36253	112.14015
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
oracle#1	11.416422	0.87631476	4.738977	4.4514503	0.82916313	0.7799236	24.11991	18.649933	89.516684
cemetery#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.978212	10.988512	49.448303
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771994	0.6600954	9.152182	9.976696	33.255653
food#1	5.324058	0.39405305	2.3662088	1.7515942	0.41243544	0.47081426	6.8665332	0.8296604	1.6593208
rooster#1	11.416422	0.9381574	4.8115277	4.995518	0.9046369	0.8657543	28.660517	32.15485	209.00456
coast#1	9.470511	0.8564059	4.6218634	4.092428	0.73246455	0.60661805	10.749784	2.2071805	6.6215415
hill#2	7.545221	0.7782966	4.48469	3.8554645	0.7195865	0.66028965	15.622789	7.4439406	26.053793
forest#2	10.723275	0.7945633	4.6132636	4.061479	0.72771994	0.6600954	9.152182	9.976696	33.255653
graveyard#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.978212	10.988512	49.448303
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6805374	0.54553374	9.651173	2.0443897	5.110975
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771994	0.6600954	9.152182	9.976696	33.255653
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
slave#2	11.821887	1.0	4.797955	4.8036927	0.85311335	0.70622677	21.679455	10.571398	52.589165
coast#1	9.470511	0.8564059	4.6218634	4.092428	0.73246455	0.60661805	10.749784	2.2071805	6.6215415
forest#2	10.723275	0.7945633	4.6132636	4.061479	0.72771994	0.6600954	9.152182	9.976696	33.255653
lad#1	8.471983	0.9381574	4.797955	4.84939	0.8221921	0.699588	19.843245	12.878388	49.903755
wizard#1	10.030128	0.9381574	4.797955	4.84939	0.8221921	0.699588	21.183018	12.937442	50.132597
chord#2	11.416422	0.8564059	4.697584	4.288723	0.8152087	0.78044355	18.249523	14.928956	74.64423
smile#1	8.877448	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	13.187282	7.8411134	30.164454
glass#1	9.049298	0.73729706	4.2416525	3.4726396	0.64194993	0.5461729	8.946375	1.4552382	3.6380954
magicians#1	10.905597	0.9019818	4.764531	4.5897527	0.8241052	0.74298024	22.610134	19.23327	83.308835
noon#1	10.31781	1.0	4.805924	4.890783	0.9071905	0.81438106	14.876495	25.269593	138.97946
string#1	10.569124	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	17.409079	10.112452	40.449306



Les valeurs des similarités sémantiques pour chaque couple dans la liste des jugements humains de Miller & Charles [1991] de 30 couples avec les différentes approches couplées avec les trois mesures de calcul des similarités sémantiques basées sur le contenu informationnel Resnik [1995],Giang\_Conrath [1997],Lin [1998] :

les couples des concepts	ic resnik			ic zero			ic sanchez		
	resnik	giang	lin	resnik	giang	lin	resnik	giang	lin
car#1--automobile#1	** 6.0298293	0.16595227	1.0	** 0.6708781	1.4905838	1.0	** 4.3044434	0.23201209	1.0
asylum#2--madhouse#1	** 11.416422	0.08458886	0.9825519	** 0.9381974	1.0	0.9690921	** 4.8023195	0.20807654	0.9996249
midday#1--noon#1	** 10.31781	0.0969198	1.0	** 1.0	1.0	1.0	** 4.805524	0.20807654	1.0
rooster#1--voyage#1	** 0.0	0.046627548	0.0	** 0.0	0.5484372	0.0	** 0.0	0.10429385	0.0
gem#3--jewel#2	** 10.435593	-0.095825896	0.0	** 0.7945683	0.8295749	0.7945683	** 4.592835	0.20034923	0.9595347
journey#1--voyage#2	** 5.785112	0.08458886	0.6531335	** 0.6509692	1.1096699	0.8383641	** 4.1775396	0.20934668	0.903079
boy#1--lad#2	** 7.4911537	0.09169604	0.81439966	** 0.73729706	1.0	0.8487864	** 4.4177437	0.2082327	0.9582892
coast#1--shore#1	** 8.851473	0.10559092	0.9662133	** 0.8144722	1.1676706	0.97490317	** 4.50129	0.21636294	0.99678374
magician#2--wizard#2	** 11.821887	0.08458886	1.0	** 0.8144722	1.227789	1.0	** 4.651827	0.2149693	1.0
journey#1--car#1	** 0.0	0.0850272	0.0	** 0.0	0.79691705	0.0	** 0.0	0.117897026	0.0
furnace#1--stove#1	** 2.4785674	0.048524895	0.21451658	** 0.1747368	0.73916396	0.22712421	** 1.5861777	0.13932183	0.35442762
food#3--fruit#2	** 1.402261	0.044961717	0.11864697	** 0.1655667	0.56414604	0.17084959	** 1.4677287	0.12314759	0.30615824
bird#1--cock#5	** 7.655222	0.07990385	0.79590644	** 0.39591002	1.0659199	0.5958282	** 2.8817135	0.2082327	0.7661102
bird#1--crane#5	** 7.655222	0.0	2.0	** 0.39591002	1.0659199	0.5959382	** 2.8817135	0.2079486	0.76548817
tool#1--implement#1	** 7.013776	0.121388265	0.91971016	** 0.41250476	2.0918115	0.8263235	** 2.907055	0.20807654	0.9415119
brother#5--monk#1	** 10.905597	0.0	2.0	** 0.87631476	1.0	0.9340808	** 4.7297688	0.20807654	0.9920137
crane#5--implement#1	** 1.4361894	0.17828901	0.40958386	** 0.08231713	0.788429	0.12189151	** 0.8228895	0.14720795	0.23821597
lad#2--brother#2	** 2.4185324	0.052055909	0.22363926	** 0.21123032	0.59145296	0.22211604	** 1.9007834	0.12974443	0.3956869
monk#1--oracle#1	** 2.4185324	0.050242459	0.12669479	** 0.21123032	0.64876115	0.24104488	** 1.9007834	0.13213356	0.4014959
cemetery#1--woodland#1	** 1.4361894	0.05036208	0.1349016	** 0.08231713	0.6059124	0.09501488	** 0.8228895	0.11780888	0.19612627
food#1--rooster#1	** 0.9090739	0.06316438	0.103605914	** 0.056359116	0.8091588	0.087229006	** 0.59623766	0.15011697	0.14900175
coast#1--hill#2	** 1.4361894	0.06418674	0.16880724	** 0.08231713	0.6441699	0.10071206	** 0.8228895	0.12219471	0.20268777
forest#2--graveyard#1	** 1.4361894	0.05036208	0.1349016	** 0.08231713	0.6059124	0.09501488	** 0.8228895	0.11780888	0.19612627
shore#1--woodland#1	** 1.4361894	0.055131175	0.14678399	** 0.08231713	0.6549997	0.10291859	** 0.8228895	0.12207539	0.20250988
monk#1--slave#2	** 2.4185324	0.04923997	0.21282899	** 0.21123032	0.60097014	0.22515446	** 1.9007834	0.13111417	0.39900053
coast#1--forest#2	** 1.4361894	0.05031179	0.14224072	** 0.08231713	0.63749	0.099719755	** 0.8228895	0.120304614	0.1998659
lad#1--wizard#1	** 2.4185324	0.06217522	0.26143915	** 0.21123032	0.60097014	0.22515446	** 1.9007834	0.12995239	0.39616534
chord#2--smile#1	** 6.439688	0.072180875	0.6346437	** 0.60362655	0.8697997	0.6865682	** 3.8627385	0.17935446	0.8179961
glass#1--magician#1	** 0.9090739	0.052804957	0.09111287	** 0.056359116	0.63174697	0.08760864	** 0.59623766	0.11806451	0.11902211
noon#1--string#1	** 0.0	0.04787682	0.0	** 0.0	0.5257674	0.0	** 0.0	0.10469216	0.0

les couples des concepts	ic sanchez2			ic zhou			ic zhou2		
	resnik	giang	lin	resnik	giang	lin	resnik	giang	lin
car#1--automobile#1	** 3.507341	0.28511629	1.0	** 0.7874051	1.3202974	1.0	** 0.84378666	1.1831337	1.0
asylum#2--madhouse#1	** 4.890783	0.20446624	1.0	** 0.8600844	1.1023043	0.97334534	** 0.7759406	1.2279264	0.9758255
midday#1--noon#1	** 4.890783	0.20446624	1.0	** 0.9071905	1.1023043	1.0	** 0.81438106	1.2279264	1.0
rooster#1--voyage#1	** 0.0	0.10354301	0.0	** 0.0	0.5625582	0.0	** 0.0	0.58597635	0.0
gem#3--jewel#2	** 4.015722	0.18217462	0.84497386	** 0.77039593	1.1229893	0.92769814	** 0.74580885	1.7383762	1.1280609
journey#1--voyage#2	** 3.3624902	0.21448624	0.83902533	** 0.73267514	1.1455319	0.9126314	** 0.81426525	1.1889328	0.9837949
boy#1--lad#2	** 3.7544572	0.20621149	0.87273914	** 0.72176194	1.1453254	0.9051009	** 0.7058737	1.9400719	0.97220933
coast#1--shore#1	** 3.892856	0.2443537	0.9750075	** 0.6805374	1.3652538	0.96325046	** 0.54555374	1.6484839	0.94700074
magician#2--wizard#2	** 4.1665072	0.2400092	1.0	** 0.76034945	1.3151247	1.0	** 0.70526475	1.4179072	1.0
journey#1--car#1	** 0.0	0.14586399	0.0	** 0.0	0.6711048	0.0	** 0.0	0.60311747	0.0
furnace#1--stove#1	** 0.90511346	0.14742385	0.23945289	** 0.36066967	0.84476924	0.46706042	** 0.5466021	0.9975926	0.7057414
food#3--fruit#2	** 0.9080453	0.11502417	0.18913925	** 0.31819227	0.7286875	0.37644246	** 0.47081748	1.0361347	0.6557607
bird#1--cock#5	** 2.1563246	0.20446624	0.61197436	** 0.5710693	1.1626766	0.79005446	** 0.7462227	1.2887584	0.88047656
bird#1--crane#5	** 2.1563246	0.2014775	0.6057382	** 0.5710693	1.1222781	0.7811556	** 0.7462227	1.1928163	0.9418586
tool#1--implement#1	** 2.1129973	0.40762722	0.925491	** 0.5366906	1.6886902	0.9503495	** 0.6608687	1.4160082	0.96683
brother#5--monk#1	** 4.4136618	0.20448624	0.9487212	** 0.81127167	1.1223272	0.95316035	** 0.7440455	1.2787536	0.9751215
crane#5--implement#1	** 0.47880245	0.15150303	0.13447717	** 0.22771572	0.8333197	0.31899868	** 0.3731142	0.8880154	0.49774414
lad#2--brother#2	** 1.0970573	0.11473592	0.22359396	** 0.40987676	0.76614606	0.47795948	** 0.6085224	1.0908112	0.7879205
monk#1--oracle#1	** 1.0970573	0.12873237	0.24749996	** 0.40987676	0.8126394	0.49971724	** 0.6085224	1.0923569	0.79559936
cemetery#1--woodland#1	** 0.47880245	0.11855271	0.10679148	** 0.22771572	0.74504703	0.29009983	** 0.3731142	0.9738366	0.5330285
food#1--rooster#1	** 0.24426073	0.15377899	0.07240463	** 0.145884	0.8538336	0.22152771	** 0.29540886	0.9081335	0.35228586
coast#1--hill#2	** 0.47880245	0.13383137	0.11973047	** 0.22771572	0.8167697	0.31364697	** 0.3731142	1.1188266	0.5890156
forest#2--graveyard#1	** 0.47880245	0.11855271	0.10679148	** 0.22771572	0.74504703	0.29009983	** 0.3731142	0.9738366	0.5330285
shore#1--woodland#1	** 0.47880245	0.13371608	0.119633496	** 0.22771572	0.8470689	0.32340074	** 0.3731142	1.2011508	0.6189432
monk#1--slave#2	** 1.0970573	0.12314912	0.23904326	** 0.40987676	0.79712504	0.49252638	** 0.6085224	1.1880014	0.8391837
coast#1--forest#2	** 0.47880245	0.13024047	0.11670539	** 0.22771572	0.8113796	0.31189993	** 0.3731142	1.1190999	0.5891059
lad#1--wizard#1	** 1.0970573	0.116285775	0.22622584	** 0.40987676	0.8100397	0.49851707	** 0.6085224	1.2647763	0.86982965
chord#2--smile#1	** 3.037981	0.17381194	0.69113153	** 0.63225144	1.0090191	0.77896434	** 0.6607996	1.2158539	0.8910078
glass#1--magician#1	** 0.24426073	0.1279078	0.060592618	** 0.145884	0.75747764	0.19901574	** 0.29540886	0.94899894	0.36521474
noon#1--string#1	** 0.0	0.106485789	0.0	** 0.0	0.58435285	0.0	** 0.0	0.65910846	0.0

les couples des concepts

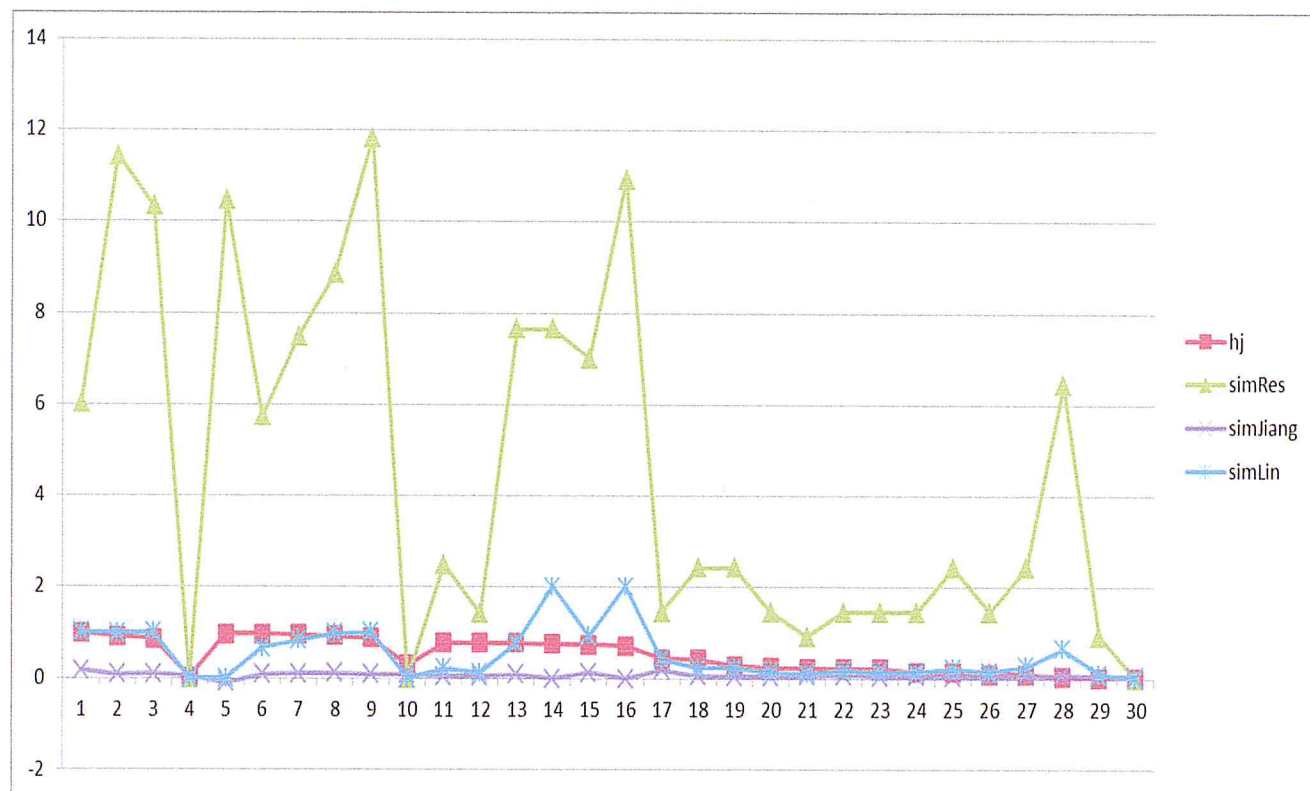
	ic sebti			ic hady taleb1			ic hady taleb2		
	resnik	jiang	lin	resnik	jiang	lin	resnik	jiang	lin
car#1--automobile#1	** 24.515457	0.027614495	1.0	** 24.065924	0.023551417	1.0	** 207.2586	0.0548241500	1.0
asylum#2--madhouse#1	** 22.053549	0.045822076	1.0	** 14.782123	0.0628649	0.96334225	** 73.510245	0.011462951	0.91537
midday#1--noon#1	** 14.376465	0.06722019	1.0	** 25.243591	0.03957419	1.0	** 122.97346	0.0071950074	1.0
rooster#1--voyage#1	** 0.0	0.019107603	0.0	** 0.0	0.018206033	0.0	** 0.0	0.002892057	0.0
gem#1--jewel#2	** 20.591923	0.0699961	1.1902678	** 12.042309	0.07809303	0.94181023	** 54.190525	0.02882152	1.2361238
journey#1--voyage#2	** 22.494005	0.03957795	0.9419412	** 21.822658	0.03850446	0.9109422	** 120.90462	0.006817403	0.8679752
boy#1--lad#2	** 19.342245	0.04490170	0.9400217	** 12.878388	0.07108802	0.9589857	** 49.802755	0.016404927	0.90229277
coast#1--shore#1	** 9.691172	0.09102512	0.946149	** 2.044297	0.4520667	0.96171045	** 5.110975	0.19102224	0.8712495
magician#2--wizard#2	** 18.891642	0.051020110	1.0	** 12.798492	0.0781942	1.0	** 49.59416	0.020160665	1.0
journey#1--car#1	** 0.0	0.020295105	0.0	** 0.0	0.017840397	0.0	** 0.0	0.0000472642	0.0
furnace#1--stove#1	** 7.705227	0.028493697	0.4102212	** 1.2045903	0.024221933	0.14406057	** 8.0114765	0.054811055	0.044932814
food#1--fruit#2	** 1.1783549	0.04247552	0.20786916	** 2.472324	0.060552075	0.2597473	** 4.945665	0.0125421125	0.11631766
bird#1--cock#5	** 18.717803	0.045504884	0.9189961	** 20.501103	0.048912012	0.9475527	** 92.26938	0.00782409	0.8952103
bird#1--crane#5	** 18.717803	0.097770465	0.8282403	** 20.501103	0.031189155	0.8039644	** 92.26938	0.0059115257	0.6758129
tool#1--implement#1	** 14.150932	0.05697239	0.3927174	** 5.407499	0.09771026	0.9349324	** 22.031246	0.024427665	0.8391150
brother#5--monk#1	** 23.003869	0.041479178	0.976634	** 13.991019	0.06675114	0.96547204	** 60.55442	0.013026685	0.91462793
crane#5--implements#1	** 2.7010503	0.02607225	0.13001393	** 0.5816516	0.025617235	0.026674091	** 0.7975209	0.00469449	0.007465973
lad#2--brother#2	** 11.466462	0.02324427	0.49326926	** 8.05916	0.04046683	0.4205769	** 24.174479	0.006714881	0.27921443
monk#1--oracle#1	** 11.466462	0.023143013	0.49359093	** 8.05916	0.040659418	0.4923567	** 24.174479	0.007540543	0.32210284
cemetery#1--woodland#1	** 2.7010503	0.035187004	0.17099788	** 0.5816506	0.048938276	0.05072028	** 0.7975209	0.012209053	0.01922616
food#1--rooster#1	** 1.0936123	0.02904542	0.00394593	** 0.0	0.0201754	0.0	** 0.0	0.004746893	0.0
coast#1--hill#2	** 2.7010503	0.04225733	0.2058686	** 0.5816506	0.10965593	0.110180065	** 0.7975209	0.01816978	0.048914856
forest#2--graveyard#1	** 2.7010503	0.035187004	0.17099788	** 0.5816506	0.048938276	0.05072028	** 0.7975209	0.012209053	0.01922616
shore#1--woodland#1	** 2.7010503	0.062123922	0.28808906	** 0.5816506	0.08702671	0.038467994	** 0.7975209	0.026617615	0.041874677
monk#1--slave#2	** 11.466462	0.010099755	0.9101694	** 8.05916	0.05129716	0.58493304	** 24.174479	0.011236068	0.42721063
coast#1--forest#2	** 2.7010503	0.058160103	0.27219358	** 0.5816506	0.05812075	0.03727609	** 0.7975209	0.02593375	0.039993944
lad#1--wizard#1	** 11.466462	0.02329771	0.5599316	** 8.05916	0.056012654	0.6242305	** 24.174479	0.010191954	0.49321395
chord#2--smile#1	** 9.758127	0.046211612	0.4033545	** 10.403954	0.0323774	0.5260312	** 36.410937	0.014620975	0.69492266
glass#1--magician#1	** 1.0596123	0.01282221	0.06962325	** 0.0	0.04939939	0.0	** 0.0	0.0115022465	0.0
noon#1--string#1	** 0.0	0.010979585	0.0	** 0.0	0.023263401	0.0	** 0.0	0.0559732266	0.0

Les corrélations avec les jugements humains de Miller & Charles [1991] :

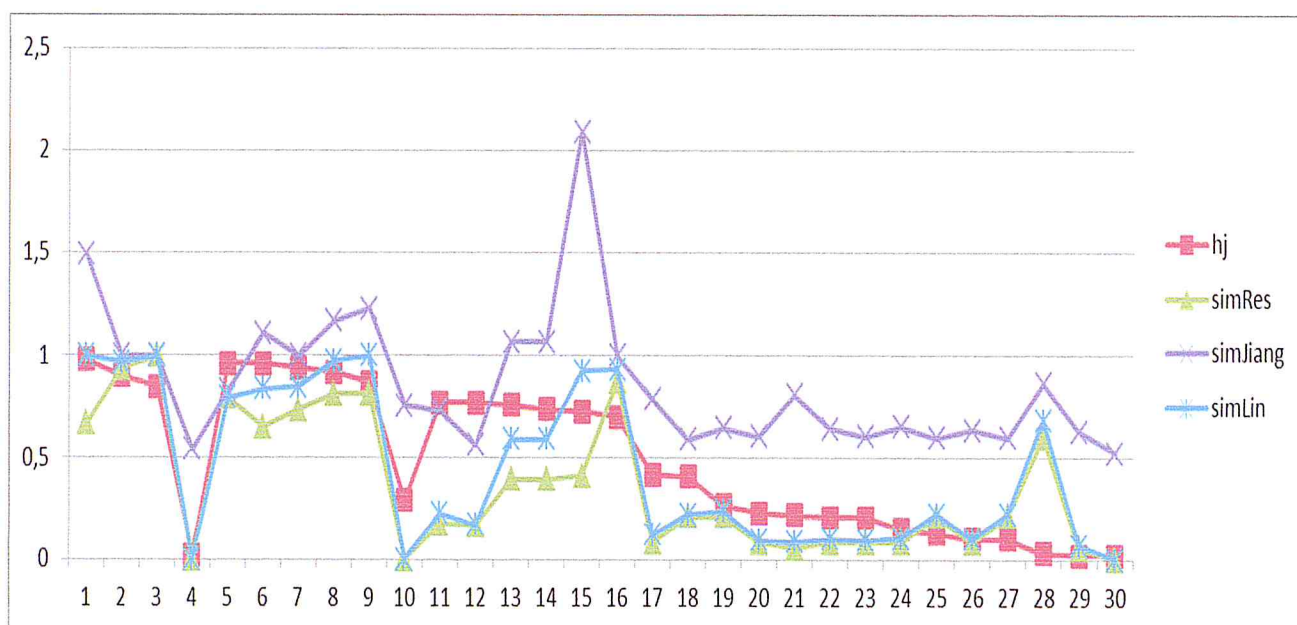
l'approche de calcul d'IC	resnik	jiang	lin
resnik	0.765533638741374	0.10235883039740734	0.5858181369462909
seco	0.7710685483124603	0.6397303124927902	0.8103224276114428
sanchez1	0.7922323236067521	0.7662624159185909	0.813899604193932
sanchez2	0.779777884612171	0.6645814311943056	0.8154396241825151
zhou1	0.7898897774332724	0.740202592302377	0.8092546372539724
zhou2	0.717558330160617	0.6080993675593893	0.697297526654338
sebti	0.7696996751917936	0.4682308834796747	0.8102342275850803
hadjtaieb1	0.6681761970405743	0.22173211776072116	0.7260631379172227
hadjtaieb2	0.6452286198866503	0.21712535919192613	0.7674049655792712

Les graphes des comparaisons des Similarités entre les couples de la liste des jugements humains de Miller & Charles [1991] avec les différentes approches d'CI couplées avec les trois mesures de calcul des similarités sémantiques basées sur le contenu informationnel Resnik [1995],Giang\_Conrath [1997],Lin [1998] :

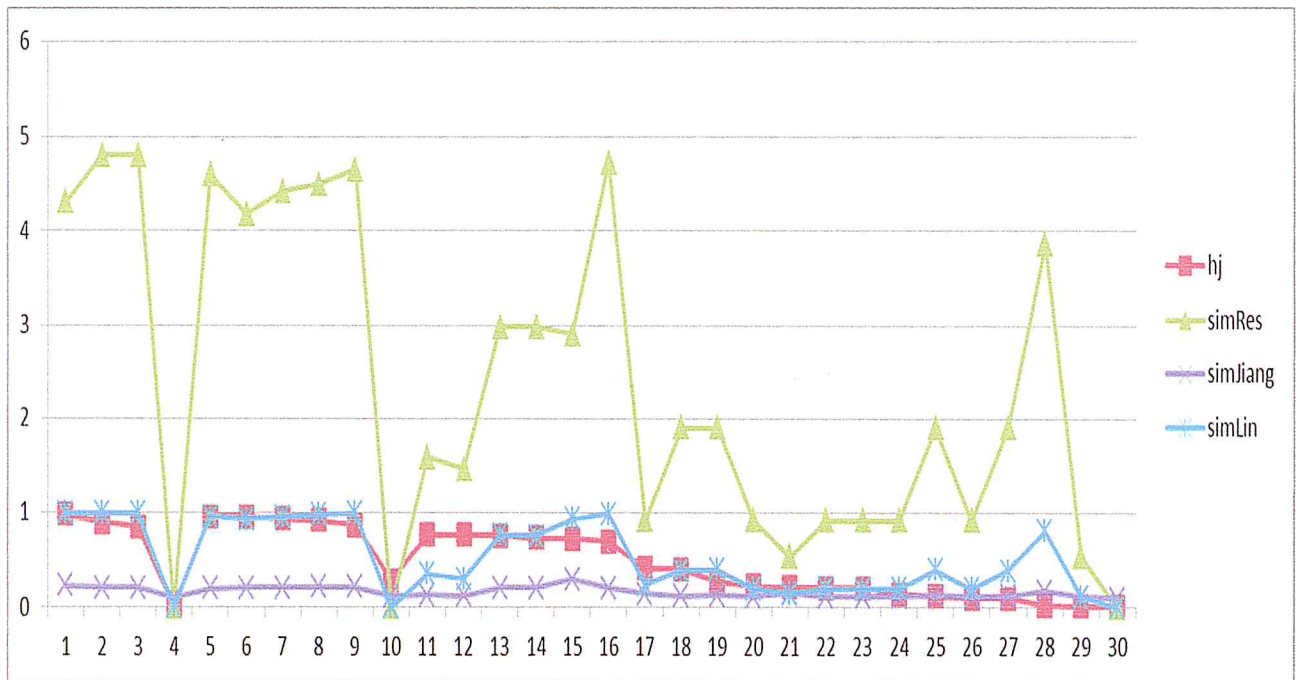
### 1-L'approche de Resnik :



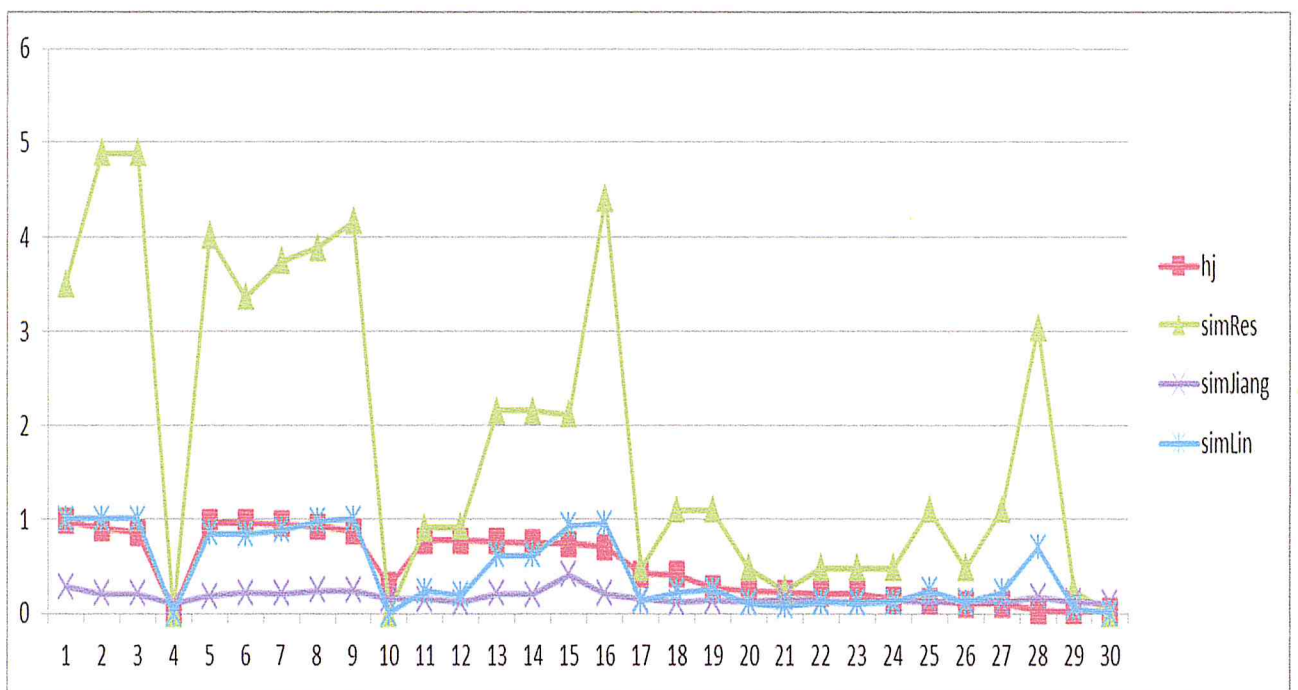
### 2-L'approche de Seco :



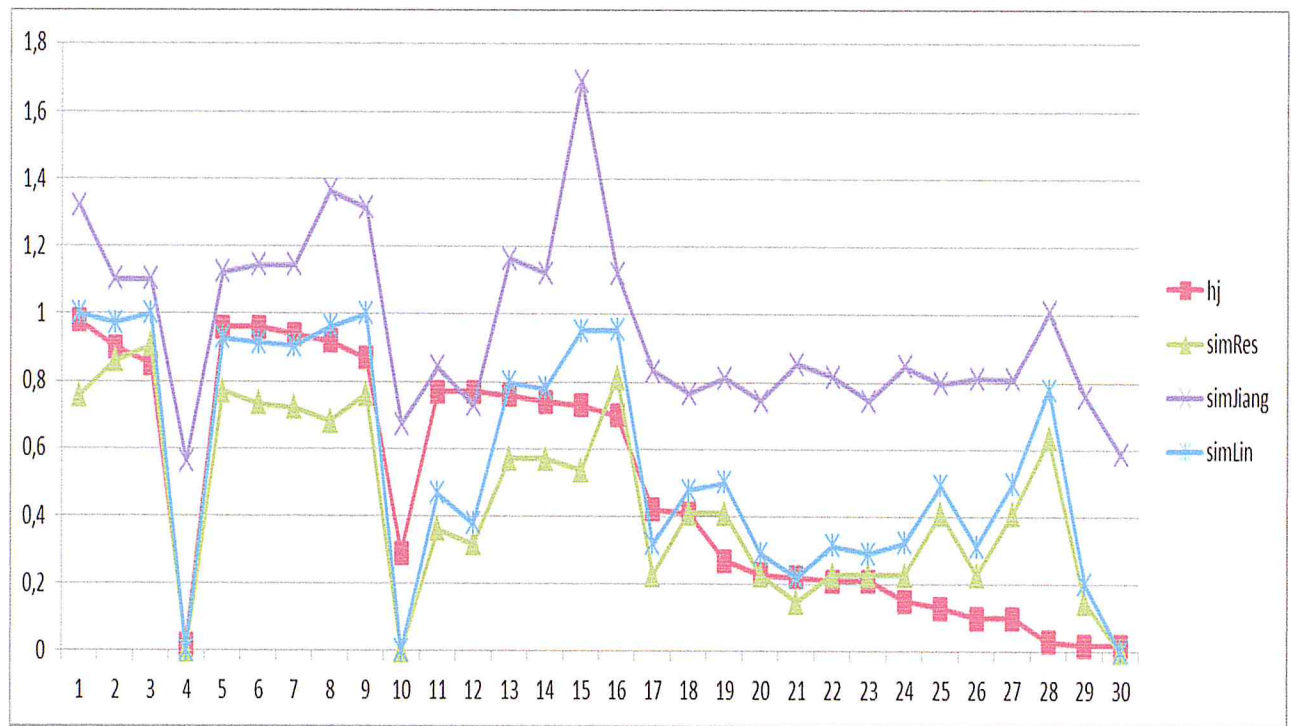
### 3-L'approche de Sanchez 1:



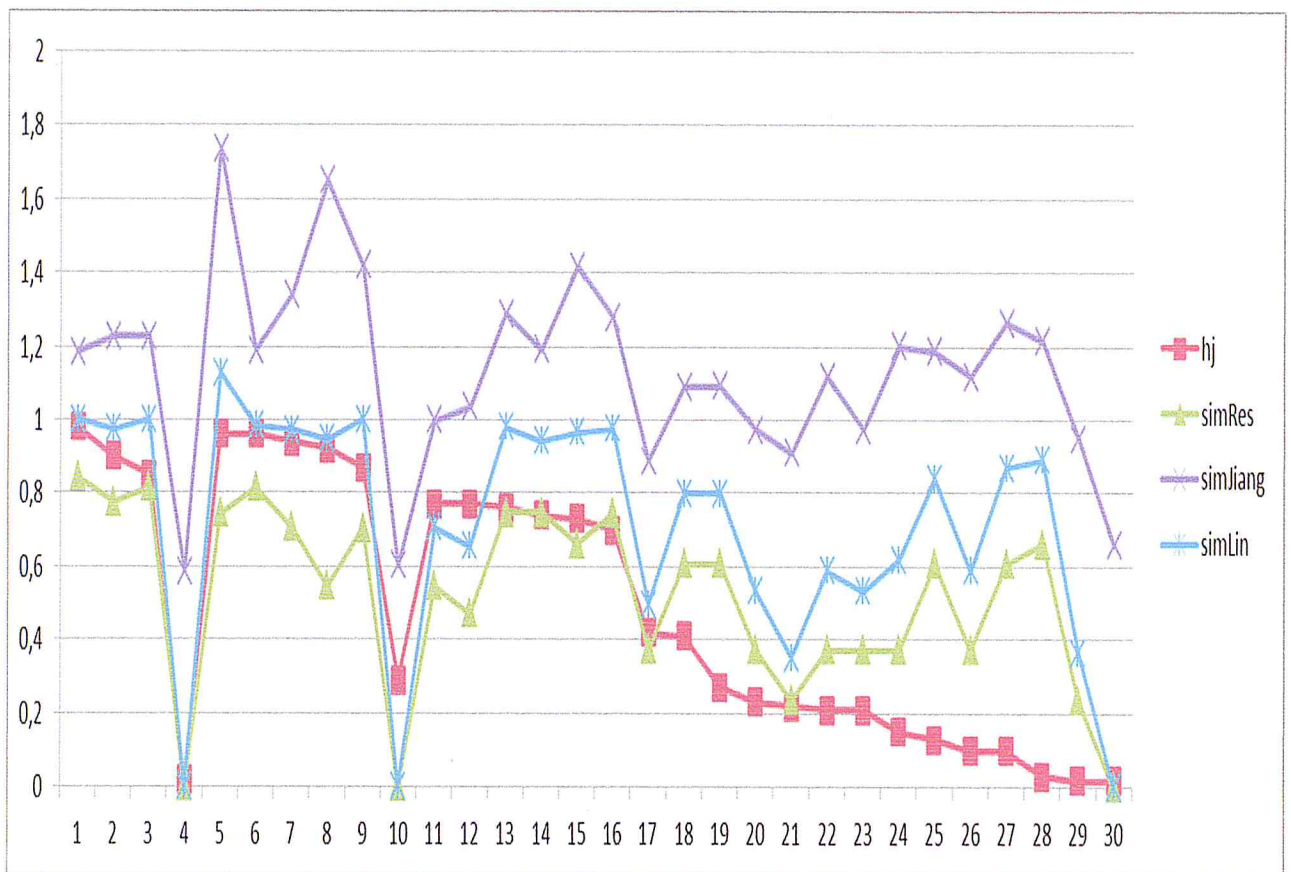
### 4-L'approche de Sanchez 2:



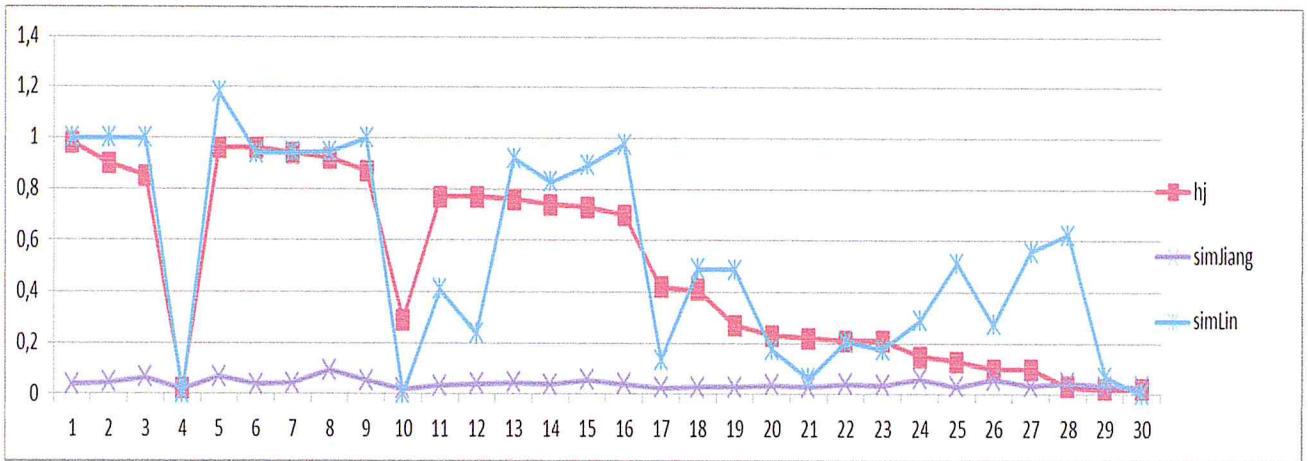
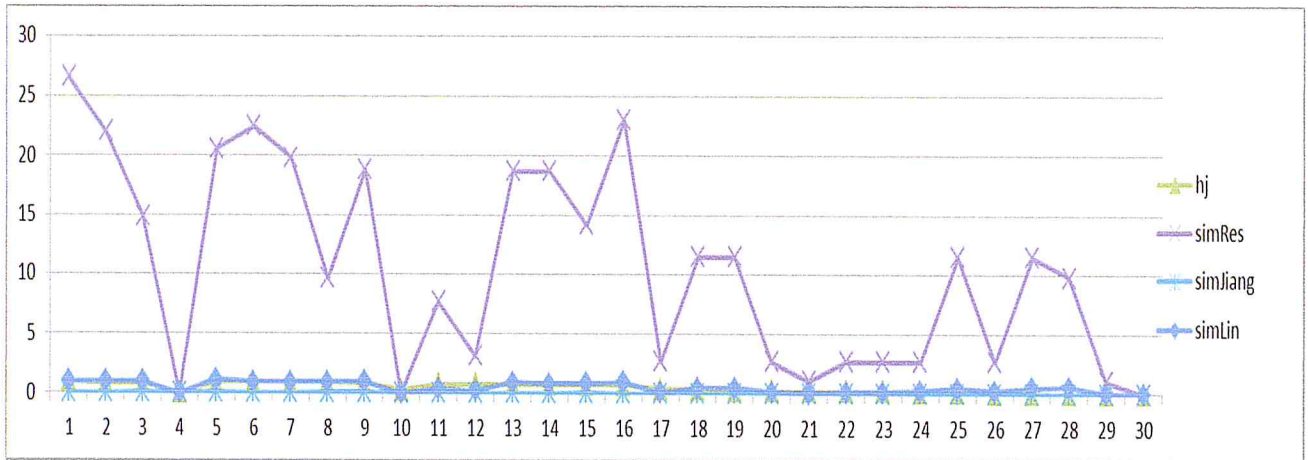
## 5-L'approche de Zhou 1:



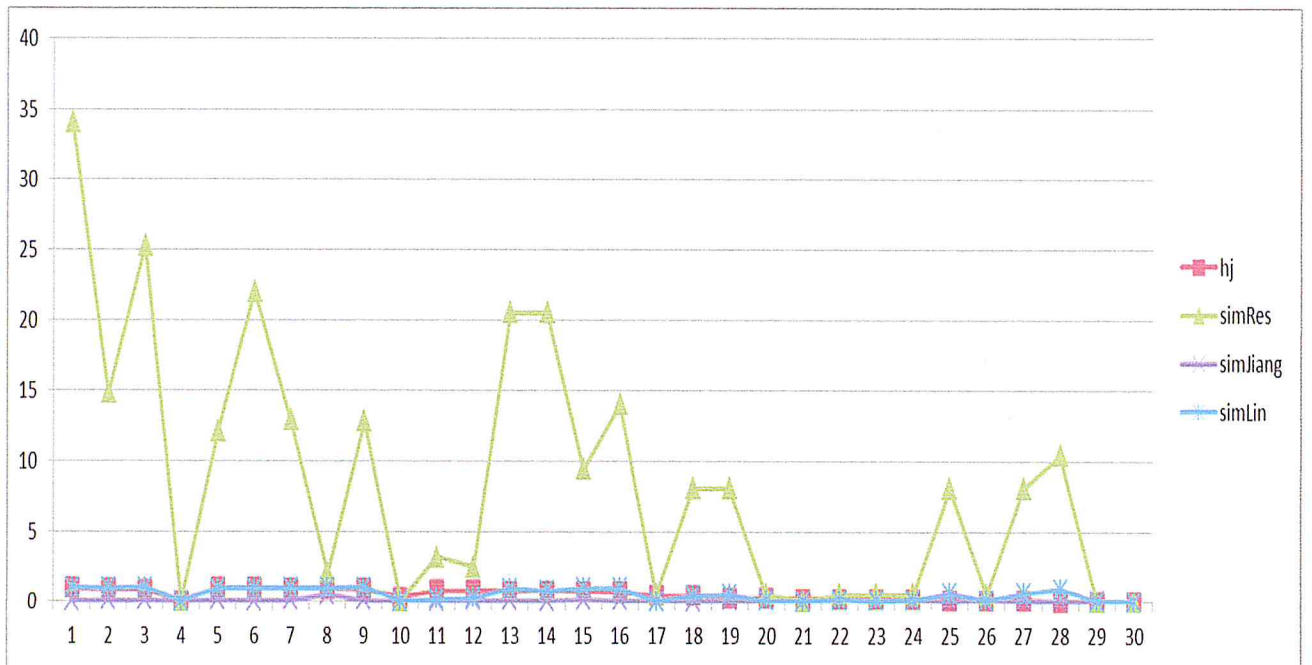
## 6-L'approche de Zhou 2:

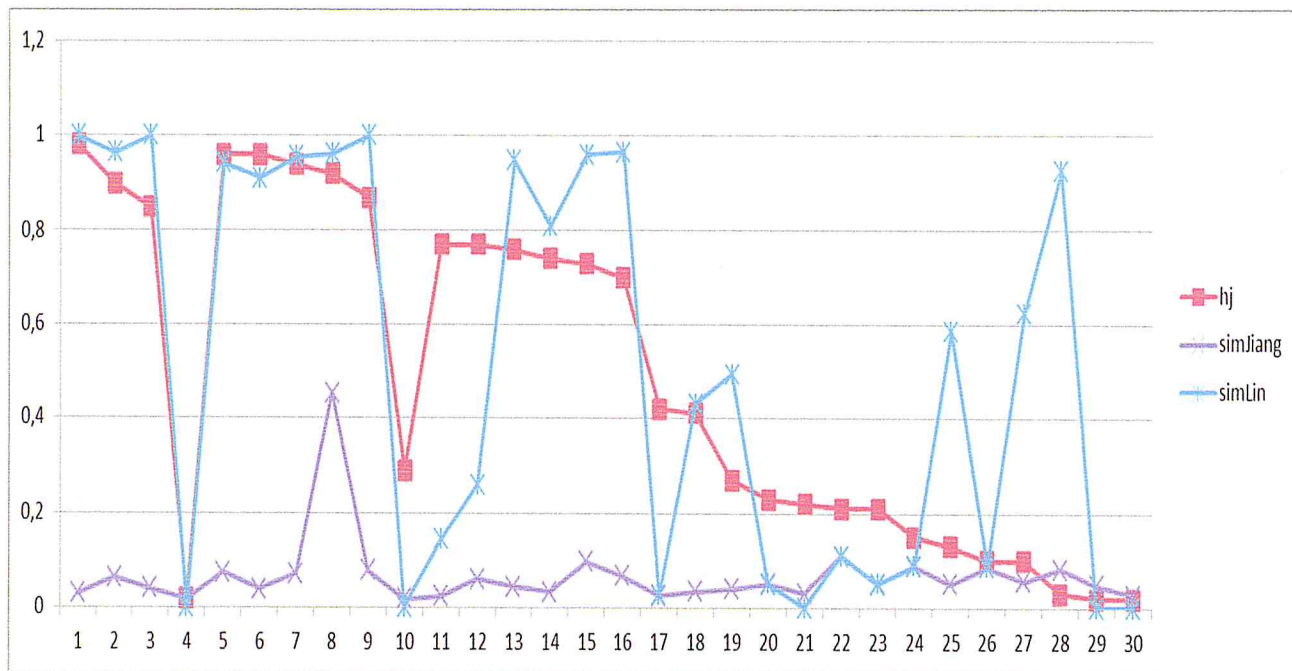


## 7-L'approche de Sebti:

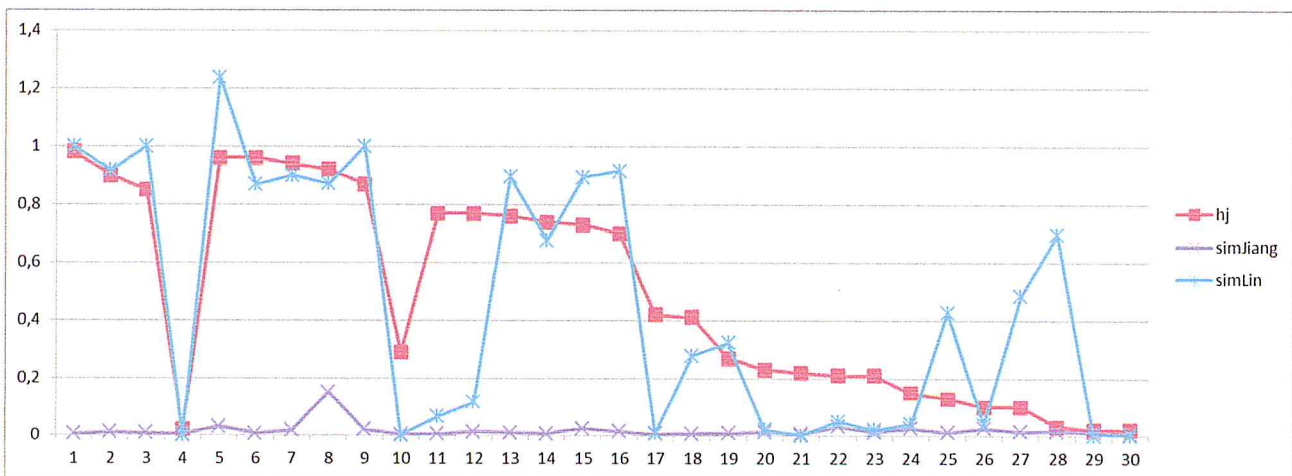
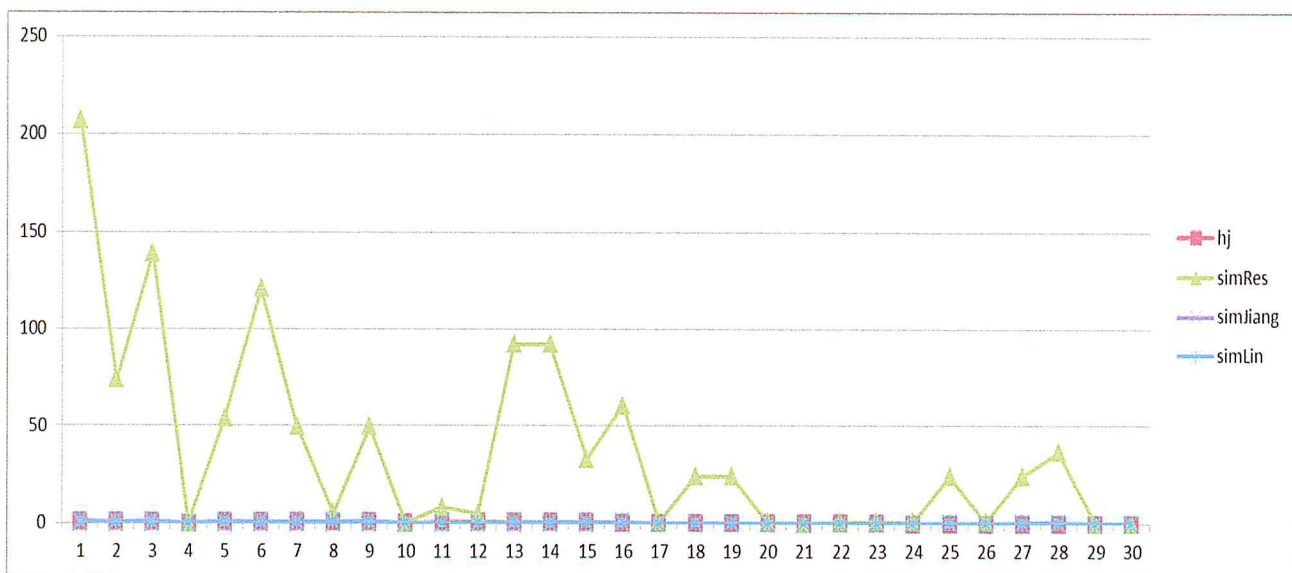


## 8-L'approche de Hadj Taieb 1:





### 9-L'approche de Hadj Taieb 2:



## 4.2.2-Rubenstein&Goudenough [1965]:

Les valeurs des contenus informationnels pour chaque concept de la liste des jugements humains de Rubenstein&Goudenough [1965] de 130 concepts avec les différentes approches d'CI :

les concepts	ic resnik	ic seco	ic sanches	ic sanches2	ic shou	ic shou2	ic sebt1	ic hadjtrieb1	ic hadjtrieb2
asylum#2	11.416422	0.9321574	4.8023195	4.890783	0.8600844	0.7759406	22.058949	14.782128	73.910645
cemetery#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.975212	10.988512	49.448303
asylum#2	11.416422	0.9381574	4.8023195	4.890783	0.8600844	0.7759406	22.058949	14.782128	73.910645
fruit#2	11.821887	1.0	4.8023195	4.84939	0.89100575	0.7820115	16.408876	11.849773	59.248867
asylum#2	11.416422	0.9381574	4.8023195	4.890783	0.8600844	0.7759406	22.058949	14.782128	73.910645
madhouse#1	11.821887	1.0	4.805924	4.890783	0.9071905	0.81438106	22.058949	19.907128	87.489204
asylum#2	11.416422	0.9381574	4.8023195	4.890783	0.8600844	0.7759406	22.058949	14.782128	73.910645
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
autograph#2	0.0	1.0	4.797955	4.8036327	0.8731143	0.7462286	15.863635	10.1113405	45.501034
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6805374	0.54555374	9.651173	2.0443897	5.110975
autograph#2	0.0	1.0	4.797955	4.8036327	0.8731143	0.7462286	15.863635	10.1113405	45.501034
signature#1	12.515034	0.92638586	4.6328588	4.1134367	0.7663063	0.7050914	14.260197	9.498694	37.974777
automobile#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069904	207.2586
cushions#1	11.821887	0.9381574	4.797955	4.84939	0.842193	0.739893	20.085876	12.455474	56.048633
automobile#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069904	207.2586
wizard#2	11.821887	0.8144722	4.651827	4.1668072	0.76034945	0.70526475	19.831642	12.788492	49.59416
bird#1	7.655222	0.39591002	2.9817135	2.1563246	0.5710693	0.7462227	18.717808	20.503105	92.26398
cock#5	12.515034	0.9381574	4.8023195	4.890783	0.8600844	0.7759406	21.975904	22.772812	113.86406
bird#1	7.655222	0.39591002	2.9817135	2.1563246	0.5710693	0.7462227	18.717808	20.503105	92.26398
crane#5	0.0	0.9381574	4.8089504	4.9633336	0.89104474	0.83835214	26.475714	30.130323	180.78194
bird#1	7.655222	0.39591002	2.9817135	2.1563246	0.5710693	0.7462227	18.717808	20.503105	92.26398
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
boy#1	7.4911537	0.73729706	4.4177437	3.7544572	0.72176194	0.7058737	19.843245	12.878388	49.903755
lad#2	10.905597	1.0	4.8023195	4.84939	0.8731143	0.7462286	22.241138	14.067067	60.557294
boy#1	7.4911537	0.73729706	4.4177437	3.7544572	0.72176194	0.7058737	19.843245	12.878388	49.903755
rooster#1	11.416422	0.9381574	4.8115277	4.995518	0.9046369	0.8657543	28.660517	32.15455	209.00456
boy#1	7.4911537	0.73729706	4.4177437	3.7544572	0.72176194	0.7058737	19.843245	12.878388	49.903755
pages#1	12.515034	0.9019918	4.764531	4.692418	0.8241052	0.74298024	20.53639	13.0566	56.578598
brothers#5	0.0	1.0	4.805924	4.890783	0.89100575	0.7820115	24.10484	14.981019	71.9089
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
car#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069904	207.2586
automobile#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069904	207.2586
cemetery#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.975212	10.988512	49.448303
graveyard#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.975212	10.988512	49.448303
cemetery#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.975212	10.988512	49.448303
mound#1	10.905597	1.0	4.8023195	4.84939	0.89100575	0.7820115	22.680103	12.815091	64.075455
cemetery#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739893	21.975212	10.988512	49.448303
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
coast#1	9.470511	0.8564059	4.6218634	4.092428	0.73246455	0.60661805	10.749784	2.2071805	6.6215415
forests#2	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
coast#1	9.470511	0.8564059	4.6218634	4.092428	0.73246455	0.60661805	10.749784	2.2071805	6.6215415
hill#2	7.545221	0.7782966	4.48469	3.8554645	0.7195965	0.66028965	15.622799	7.4439406	26.053793
coast#1	9.470511	0.8564059	4.6218634	4.092428	0.73246455	0.60661805	10.749784	2.2071805	6.6215415
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6805374	0.54555374	9.651173	2.0443897	5.110975
cock#5	12.515034	0.9381574	4.8023195	4.890783	0.8600844	0.7759406	21.975904	22.772812	113.86406
rooster#1	11.416422	0.9381574	4.8115277	4.995518	0.9046369	0.8657543	28.660517	32.15455	209.00456
cord#1	9.296159	0.64410275	3.9898403	3.1499379	0.65248954	0.66074634	14.190982	6.8675537	24.036433
smile#1	8.877448	0.9019918	4.7468023	4.5026026	0.80410427	0.70282	13.187252	7.5411134	30.164454
cord#1	9.296159	0.64410275	3.9898403	3.1499379	0.65248954	0.66074634	14.190982	6.8675537	24.036433
strings#1	10.569124	0.9019918	4.7468023	4.5026026	0.80410427	0.70282	17.409079	10.112452	40.449806
crane#5	0.0	0.9381574	4.8089504	4.9633336	0.89104474	0.83835214	26.475714	30.130323	180.78194
implement#1	7.013776	0.41280476	2.907055	2.1129973	0.5966906	0.6608687	14.190982	9.437499	33.031246
crane#5	0.0	0.9381574	4.8089504	4.9633336	0.89104474	0.83835214	26.475714	30.130323	180.78194
rooster#1	11.416422	0.9381574	4.8115277	4.995518	0.9046369	0.8657543	28.660517	32.15455	209.00456
cushions#1	11.821887	0.9381574	4.797955	4.84939	0.842193	0.739893	20.085876	12.455474	56.048633
jewel#1	10.435933	0.7945633	4.593835	4.015722	0.77038959	0.74550395	20.591528	12.042339	54.190525
cushions#1	11.821887	0.9381574	4.797955	4.84939	0.842193	0.739893	20.085876	12.455474	56.048633
pillow#1	10.31781	0.9019918	4.7468023	4.5026026	0.80410427	0.70282	14.977661	8.205232	32.820927
food#3	11.821887	0.9381574	4.7857203	4.75248	0.79951696	0.6539315	10.312147	7.264106	25.42437
fruit#2	11.821887	1.0	4.8023195	4.84939	0.89100575	0.7820115	16.408876	11.849773	59.248867
food#3	11.821887	0.9381574	4.7857203	4.75248	0.79951696	0.6539315	10.312147	7.264106	25.42437
rooster#1	11.416422	0.9381574	4.8115277	4.995518	0.9046369	0.8657543	28.660517	32.15455	209.00456
forests#2	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653



graveyard#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739883	21.975212	10.988512	49.448303
forest#2	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
fruit#2	11.821887	1.0	4.8023195	4.84939	0.89100575	0.7820115	16.408876	11.849773	59.248867
furnace#1	12.515034	0.7242171	4.301829	3.5655866	0.71522194	0.7059061	17.057102	9.897156	39.588623
furnace#1	12.515034	0.7242171	4.301829	3.5655866	0.71522194	0.7059061	17.057102	9.897156	39.588623
implement#1	7.013776	0.41250476	2.907055	2.1129973	0.5366906	0.6608697	14.150982	9.407499	33.031246
furnace#1	12.515034	0.7242171	4.301829	3.5655866	0.71522194	0.7059061	17.057102	9.897156	39.588623
stove#1	10.569124	0.8144722	4.64414	4.1226897	0.8292022	0.8431092	20.500122	34.592335	207.554
gem#5	10.405593	0.7945633	4.593835	4.015722	0.77039593	0.74550385	20.591928	12.042339	54.190525
jewel#1	10.405593	0.7945633	4.593835	4.015722	0.77039593	0.74550385	20.591928	12.042339	54.190525
glass#2	9.742446	0.7526236	4.4043794	3.6606178	0.7294282	0.7057651	18.063005	10.253076	41.012302
jewel#1	10.405593	0.7945633	4.593835	4.015722	0.77039593	0.74550385	20.591928	12.042339	54.190525
glass#2	9.742446	0.7526236	4.4043794	3.6606178	0.7294282	0.7057651	18.063005	10.253076	41.012302
magician#1	10.905597	0.9019818	4.764531	4.5897527	0.8241052	0.74298024	22.610134	19.22327	83.300935
glass#2	9.742446	0.7526236	4.4043794	3.6606178	0.7294282	0.7057651	18.063005	10.253076	41.012302
tumbler#2	12.515034	1.0	4.797955	4.8066327	0.8731143	0.7462286	20.702063	10.394969	46.777355
graveyard#1	10.569124	0.9381574	4.797955	4.84939	0.842193	0.739883	21.975212	10.988512	49.448303
madhouse#1	11.821887	1.0	4.805924	4.890783	0.9071905	0.81438106	22.059949	15.907128	87.489204
grin#1	8.877448	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	13.187252	7.541134	30.164454
implement#1	7.013776	0.41250476	2.907055	2.1129973	0.5366906	0.6608697	14.150982	9.407499	33.031246
grin#1	8.877448	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	13.187252	7.541134	30.164454
lad#2	10.905597	1.0	4.8023195	4.84939	0.8731143	0.7462286	22.241138	14.067067	60.957294
grin#1	8.877448	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	13.187252	7.541134	30.164454
smile#1	8.877448	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	13.187252	7.541134	30.164454
hill#5	10.905597	1.0	4.8023195	4.84939	0.89100575	0.7820115	22.680103	12.815091	64.075455
mound#1	10.905597	1.0	4.8023195	4.84939	0.89100575	0.7820115	22.680103	12.815091	64.075455
hill#5	10.905597	1.0	4.8023195	4.84939	0.89100575	0.7820115	22.680103	12.815091	64.075455
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
journey#1	5.735112	0.6509692	4.1775346	3.3624902	0.73267514	0.81426525	22.494005	21.982658	120.90462
car#1	6.0258293	0.6708781	4.3044434	3.507341	0.7574051	0.84378666	26.585497	34.069904	207.2586
journey#1	5.735112	0.6509692	4.1775346	3.3624902	0.73267514	0.81426525	22.494005	21.982658	120.90462
voyage#2	11.821887	0.9019818	4.7767653	4.6623034	0.87295693	0.84109044	25.266594	26.280903	157.63542
lad#2	10.905597	1.0	4.8023195	4.84939	0.8731143	0.7462286	22.241138	14.067067	60.957294
brother#5	0.0	1.0	4.805924	4.890783	0.89100575	0.7820115	24.10948	14.981019	71.9089
lad#2	10.905597	1.0	4.8023195	4.84939	0.8731143	0.7462286	22.241138	14.067067	60.957294
wizard#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.798492	49.59416
magician#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.798492	49.59416
oracle#1	11.416422	0.87631476	4.738977	4.4514503	0.82916313	0.7799286	24.11991	18.649933	89.519694
magician#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.798492	49.59416
wizard#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.798492	49.59416
midday#1	10.31781	1.0	4.805924	4.890783	0.9071905	0.81438106	25.269993	25.269993	138.97946
noon#1	10.31781	1.0	4.805924	4.890783	0.9071905	0.81438106	25.269993	25.269993	138.97946
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
oracle#1	11.416422	0.87631476	4.738977	4.4514503	0.82916313	0.7799286	24.11991	18.649933	89.519694
monk#1	10.905597	0.87631476	4.7297688	4.4136615	0.81127167	0.7440455	23.009869	13.981019	60.58442
stove#1	10.569124	0.8144722	4.64414	4.1226897	0.8292022	0.8431092	20.500122	34.592335	207.554
mound#1	10.905597	1.0	4.8023195	4.84939	0.89100575	0.7820115	22.680103	12.815091	64.075455
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6905374	0.54555374	9.651173	2.0443897	5.110975
mound#1	10.905597	1.0	4.8023195	4.84939	0.89100575	0.7820115	22.680103	12.815091	64.075455
slave#2	11.821887	1.0	4.797955	4.8066327	0.85311335	0.70622677	21.679455	13.571399	52.589165
noon#1	10.31781	1.0	4.805924	4.890783	0.9071905	0.81438106	25.269993	25.269993	138.97946
strings#1	10.569124	0.9019818	4.7468023	4.5026026	0.80410427	0.70282	17.409079	10.112482	40.448506
oracle#1	11.416422	0.87631476	4.738977	4.4514503	0.82916313	0.7799286	24.11991	18.649933	89.519694
sage#1	12.515034	0.9019818	4.764531	4.692415	0.8241052	0.74298024	20.53639	13.0566	56.578598
rootster#1	11.416422	0.8981574	4.8115277	4.595518	0.8046369	0.8657543	28.660517	32.15455	209.00456
voyage#2	11.821887	0.9019818	4.7767653	4.6623034	0.87295693	0.84109044	25.266594	26.280903	157.63542
sage#1	12.515034	0.9019818	4.764531	4.692415	0.8241052	0.74298024	20.53639	13.0566	56.578598
wizard#2	11.821887	0.8144722	4.651827	4.1665072	0.76034945	0.70526475	18.831642	12.798492	49.59416
serf#1	0.0	0.9381574	4.805924	4.828571	0.8600844	0.7755406	20.153399	19.495018	93.576096
slave#2	11.821887	1.0	4.797955	4.8066327	0.85311335	0.70622677	21.679455	13.571399	52.589165
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6905374	0.54555374	9.651173	2.0443897	5.110975
voyage#1	10.030128	0.9019818	4.7767653	4.6623034	0.87295693	0.84109044	25.262619	22.772133	136.6328
shore#1	8.851473	0.8144722	4.50129	3.892856	0.6905374	0.54555374	9.651173	2.0443897	5.110975
woodland#1	10.723275	0.7945633	4.6132636	4.061479	0.72771984	0.6600954	9.152182	9.976696	33.255653
tool#1	8.289368	0.4781231	3.2892357	2.4832218	0.59217495	0.7062106	17.58218	10.23434	40.93736
implement#1	7.013776	0.41250476	2.907055	2.1129973	0.5366906	0.6608697	14.150982	9.407499	33.031246

Les valeurs des similarités sémantiques pour chaque couple dans la liste des jugements humains de Rubenstein&Goudenough [1965] de 65 couples avec les différentes approches couplées avec les trois mesures de calcul des similarités sémantiques basées sur le contenu informationnel Resnik [1995],Giang\_Conrath [1997],La mesure de lin [1998] :

les couples des concepts	ic resnik			ic seco			ic sancher		
	resnik	giang	lin	resnik	giang	lin	resnik	giang	lin
asylum2--cemetery#1	** 1.4361894	0.048663026	0.13064851	** 0.08201713	0.55741405	0.08774041	** 0.8228935	0.115242146	0.15226359
asylum2--fruit#2	** 0.0	0.040002393	0.0	** 0.0	0.51595396	0.0	** 0.0	0.10411635	0.0
asylum2--madhouse#1	** 11.416422	0.05458896	0.9825519	** 0.9381574	1.0	0.9690921	** 4.8023195	0.20807654	0.9996249
asylum2--monk#1	** 1.4361894	0.047879353	0.12867917	** 0.08201713	0.5773155	0.09073396	** 0.8228935	0.11615488	0.19963931
autograph#2--shore#1	** 0.0	0.11297555	0.0	** 0.0	0.56112445	0.0	** 0.0	0.107835616	0.0
autograph#2--signature#1	** 12.515034	0.0	2.0	** 0.82688556	1.0	0.90494114	** 4.6329888	0.20942214	0.982494
automobile#1--cushion#1	** 12.515034	0.1875229	1.4024241	** 1.0	1.6419405	1.2423907	** 4.7825996	0.23202723	1.0500324
automobile#1--wisard#2	** 1.4361894	0.06093278	0.16093917	** 0.08201713	0.71274155	0.110803674	** 0.8228935	0.124480665	0.20608891
bird#1--cock#5	** 7.655222	0.07990389	0.75906044	** 0.93991002	1.0659193	0.5935932	** 2.9817135	0.2082327	0.7661102
bird#1--crane#5	** 7.655222	0.0	2.0	** 0.93991002	1.0659193	0.5935932	** 2.9817135	0.2079456	0.76548517
bird#1--woodland#1	** 1.4361894	0.05902394	0.15629019	** 0.08201713	0.60239954	0.1382931	** 0.8228935	0.20807654	0.24002734
boy#1--lad#2	** 7.4511537	0.05163604	0.81493966	** 0.70729706	1.0	0.8497564	** 4.4177437	0.2082327	0.9822352
boy#1--rooster#1	** 2.3446525	0.060375962	0.2480162	** 0.13622475	0.6496756	0.16261229	** 1.4059588	0.12789194	0.30549396
boy#1--sage#1	** 2.4185024	0.056559069	0.24177845	** 0.21123032	0.7002563	0.25771126	** 1.9007894	0.13733451	0.41401145
brother#5--monk#1	** 10.905597	0.0	2.0	** 0.87601476	1.0	0.9340808	** 4.7287654	0.20807654	0.9920107
car#1--automobile#1	** 6.0259293	0.16595217	1.0	** 0.6708781	1.4905938	1.0	** 4.3044434	0.23231809	1.0
cemetery#1--graveyard#1	** 10.569124	0.03461522	1.0	** 0.9381574	1.0659193	1.0	** 4.787955	0.20842214	1.0
cemetery#1--mound#1	** 1.4361894	0.04990396	0.13375628	** 0.08201713	0.59280346	0.0849407	** 0.8228935	0.115242146	0.19226359
cemetery#1--woodland#1	** 1.4361894	0.05006208	0.1349016	** 0.08201713	0.6059124	0.09501488	** 0.8228935	0.11780889	0.19612627
coast#1--forest#2	** 1.4361894	0.05931176	0.14224072	** 0.08201713	0.63749	0.099719755	** 0.8228935	0.120304614	0.1998659
coast#1--hill#2	** 1.4361894	0.06418674	0.16880724	** 0.08201713	0.6441699	0.10071206	** 0.8228935	0.12219471	0.20269777
coast#1--shore#1	** 8.351473	0.10559092	0.9662133	** 0.8144722	1.1676706	0.97490317	** 4.50129	0.21636294	0.98767874
cock#5--rooster#1	** 7.655222	0.061493972	0.6397623	** 0.93991002	0.6784909	0.4220081	** 2.9817135	0.18078104	0.6202956
cord#1--smile#1	** 0.0	0.05502485	0.0	** 0.0	0.64679515	0.0	** 0.0	0.114460446	0.0
cord#1--string#1	** 9.296159	0.03461522	0.92592006	** 0.69410276	1.1086699	0.8332051	** 3.9998403	0.21066913	0.9130578
crane#5--implement#1	** 1.4361894	0.17928901	0.40955336	** 0.08201713	0.788429	0.12189151	** 0.8228935	0.14720795	0.23921537
crane#5--rooster#1	** 7.655222	0.2659726	1.3410895	** 0.93991002	0.6784909	0.4220081	** 2.9817135	0.18080394	0.6198631
cushion#1--jewel#1	** 2.4759674	0.050552253	0.22249407	** 0.1747369	0.6418552	0.20169066	** 1.5861777	0.12811296	0.33779564
cushion#1--pillow#1	** 2.4759674	0.05089505	0.22366769	** 0.1747369	0.6045546	0.18991694	** 1.5861777	0.12568057	0.3323669
food#5--fruit#2	** 1.402631	0.044961717	0.11864697	** 0.1655667	0.56414604	0.17084959	** 1.4677287	0.12014799	0.30615924
food#5--rooster#1	** 0.0	0.040002393	0.0	** 0.0	0.53295964	0.0	** 0.0	0.10419693	0.0
forest#2--graveyard#1	** 1.4361894	0.05036208	0.1349016	** 0.08201713	0.6059124	0.09501488	** 0.8228935	0.11780889	0.19612627
forest#2--woodland#1	** 7.9611573	0.07415492	0.74241847	** 0.8564059	1.3647765	1.0778022	** 4.6474175	0.21893309	1.0074035
fruit#2--furnace#1	** 0.0	0.041089933	0.0	** 0.0	0.57997334	0.0	** 0.0	0.10947928	0.0
furnace#1--implement#1	** 2.4759674	0.039641247	0.23957074	** 0.1747369	1.0395172	0.20740993	** 1.5861777	0.11690642	0.4982082
furnace#1--stove#1	** 2.4759674	0.048524395	0.21451659	** 0.1747369	0.73316302	0.22712421	** 1.5861777	0.13592193	0.35342762
gem#5--jewel#1	** 10.435593	0.039525996	1.0	** 0.7945633	1.2595529	1.0	** 4.593935	0.21769306	1.0
glass#2--jewel#1	** 2.4759674	0.056490567	0.2454121	** 0.1747369	0.72862077	0.2258759	** 1.5861777	0.13491563	0.35255387
glass#2--magician#1	** 1.4361894	0.0520512	0.13911144	** 0.08201713	0.63601327	0.099500254	** 0.8228935	0.121270664	0.20130931
glass#2--tumbler#2	** 9.049299	0.07571065	0.5131467	** 0.73729706	0.99489904	0.841361	** 4.2416525	0.20159519	0.9219644
graveyard#1--madhouse#1	** 1.4361894	0.047721714	0.12828268	** 0.08201713	0.53889946	0.0849407	** 0.8228935	0.1151943	0.19219184
grin#1--implement#1	** 0.0	0.06292791	0.0	** 0.0	0.76075333	0.0	** 0.0	0.13065308	0.0
grin#1--lad#2	** 0.0	0.050548937	0.0	** 0.0	0.5257674	0.0	** 0.0	0.10472167	0.0
grin#1--smile#1	** 8.377448	0.112644985	1.0	** 0.9019919	1.1086699	1.0	** 4.7468023	0.21066913	1.0
hill#5--mound#1	** 10.905597	0.05169604	1.0	** 1.0	1.0	1.0	** 4.8023195	0.2082327	1.0
hill#5--woodland#1	** 1.4361894	0.04952239	0.132803	** 0.08201713	0.5940282	0.09174056	** 0.8228935	0.11774933	0.19603524
journey#1--car#1	** 0.0	0.0850272	0.0	** 0.0	0.78651705	0.0	** 0.0	0.117897026	0.0
journey#1--voyage#2	** 5.735112	0.03458936	0.6533135	** 0.6509692	1.1096699	0.8383641	** 4.1775046	0.20934668	0.933079
lad#2--brother#5	** 2.4185024	0.11782638	0.4435397	** 0.21123032	0.55904947	0.21123032	** 1.9007894	0.12974443	0.3556563
lad#2--wisard#2	** 2.4185024	0.04923937	0.21282899	** 0.21123032	0.6237362	0.2328284	** 1.9007894	0.13239136	0.40210578
magician#2--oracle#1	** 2.4185024	0.048001296	0.20915048	** 0.21123032	0.67587817	0.24986036	** 1.9007894	0.13351099	0.40481906
magician#2--wisard#2	** 11.821887	0.03458936	1.0	** 0.8144722	1.227789	1.0	** 4.651827	0.2149693	1.0
midday#1--noon#1	** 10.31781	0.0969198	1.0	** 1.0	1.0	1.0	** 4.805924	0.20807654	1.0
monk#1--oracle#1	** 2.4185024	0.05024458	0.21669479	** 0.21123032	0.64876115	0.24104399	** 1.9007894	0.13213596	0.4014959
monk#1--stove#1	** 1.4361894	0.04990396	0.13375628	** 0.08201713	0.62170897	0.09737139	** 0.8228935	0.11832897	0.19690689
mound#1--shore#1	** 1.4361894	0.054982592	0.14538496	** 0.08201713	0.5773155	0.09073396	** 0.8228935	0.119321585	0.19939473
mound#1--slave#2	** 1.4361894	0.046697555	0.1263935	** 0.08201713	0.5214627	0.08201713	** 0.8228935	0.115242146	0.19226359
noon#1--string#1	** 0.0	0.04787682	0.0	** 0.0	0.5257674	0.0	** 0.0	0.10468216	0.0
oracle#1--sage#1	** 7.3110276	0.06016692	0.6109973	** 0.55235815	0.81570166	0.62122166	** 3.6429985	0.17069363	0.76666397
rooster#1--voyage#2	** 0.0	0.040002393	0.0	** 0.0	0.5434972	0.0	** 0.0	0.10429385	0.0
sage#1--wisard#2	** 2.4185024	0.04562379	0.19875418	** 0.21123032	0.66438313	0.24612407	** 1.9007894	0.13305703	0.40971946
sex#1--slave#2	** 2.4185024	0.10634803	0.40916194	** 0.21123032	0.57906324	0.21797025	** 1.9007894	0.12891795	0.35589662
shore#1--voyage#1	** 0.0	0.05294614	0.0	** 0.0	0.58239965	0.0	** 0.0	0.10778121	0.0
shore#1--woodland#1	** 1.4361894	0.055131175	0.14673899	** 0.08201713	0.6549997	0.10231859	** 0.8228935	0.12207539	0.20250938
tool#1--implement#1	** 7.013776	0.121393265	0.91971016	** 0.41280476	2.0919115	0.9269235	** 2.907055	0.30595747	0.9419119

les couples des concepts

	ic sanchez2			ic shou			ic shou2		
	resnik	jiang	lin	resnik	jiang	lin	resnik	jiang	lin
asylum2--cemetery#1	** 0.47590245	0.10794041	0.09769897	** 0.22771572	0.6781676	0.26754242	** 0.3701142	0.8751131	0.49229237
asylum2--fruits#2	** 0.0	0.10266758	0.0	** 0.0	0.5710728	0.0	** 0.0	0.64196923	0.0
asylum2--madhouse#1	** 4.890783	0.20446624	1.0	** 0.8600344	1.1023043	0.57334534	** 0.7759406	1.2275264	0.9758285
asylum2--mont#1	** 0.47590245	0.113267705	0.10227423	** 0.22771572	0.6926933	0.27249214	** 0.3701142	0.871597	0.4803442
autograph#2--shore#1	** 0.0	0.11488394	0.0	** 0.0	0.4406449	0.0	** 0.0	0.7741242	0.0
autograph#2--signature#1	** 4.1134067	0.20917578	0.9225983	** 0.7663063	1.1453254	0.59485016	** 0.7050914	1.9400719	0.9716553
automobile#1--cushion#1	** 4.75248	0.27745017	1.1374017	** 0.85311335	1.3396121	1.0666597	** 0.70622677	1.1394754	0.8913865
automobile#1--wizard#2	** 0.47590245	0.1395266	0.12400623	** 0.22771572	0.77517045	0.3000692	** 0.3701142	0.85035555	0.4817025
bird#1--cock#5	** 2.1563246	0.20446624	0.61197436	** 0.5710693	1.1626766	0.79505446	** 0.7462227	1.2887584	0.99047656
bird#1--crane#5	** 2.1563246	0.2014775	0.6057352	** 0.5710693	1.1222781	0.7811556	** 0.7462227	1.1928163	0.9418596
bird#1--woodland#1	** 0.47590245	0.17415531	0.15304918	** 0.22771572	0.9336923	0.35065949	** 0.3701142	0.96786932	0.5306256
boy#1--lad#2	** 3.7544972	0.20621149	0.97273914	** 0.72178194	1.1453284	0.9051009	** 0.7059737	1.3400719	0.97220933
boy#1--rooster#1	** 0.7670323	0.12526709	0.17932216	** 0.34141365	0.7782191	0.41934	** 0.5466023	0.9758952	0.65958704
boy#1--sage#1	** 1.0970573	0.13609784	0.2597847	** 0.40987676	0.8802392	0.5802872	** 0.6085224	1.1900065	0.84000516
brother#5--monk#1	** 4.4136615	0.20446624	0.5487212	** 0.81127167	1.1223272	0.95616935	** 0.7440455	1.2787636	0.9751215
car#1--automobile#1	** 3.507341	0.28511629	1.0	** 0.7574051	1.3202374	1.0	** 0.84078666	1.1551337	1.0
cemetery#1--graveyard#1	** 4.84939	0.20621149	1.0	** 0.842193	1.1873763	1.0	** 0.738983	1.3515651	1.0
cemetery#1--mound#1	** 0.47590245	0.10942485	0.098115936	** 0.22771572	0.66423863	0.2627693	** 0.3701142	0.87048946	0.4903286
cemetery#1--woodland#1	** 0.47590245	0.11855271	0.10679148	** 0.22771572	0.74504703	0.29009993	** 0.3701142	0.9733326	0.5303285
coast#1--forest#2	** 0.47590245	0.13024047	0.11670539	** 0.22771572	0.8113796	0.31199993	** 0.3701142	1.1190659	0.5891059
coast#1--hill#2	** 0.47590245	0.13383137	0.11973047	** 0.22771572	0.8167697	0.31364697	** 0.3701142	1.1185266	0.5891056
coast#1--shore#1	** 3.892356	0.2443537	0.9750075	** 0.6905874	1.3652533	0.96325046	** 0.5455374	1.6484833	0.94700074
cock#5--rooster#1	** 2.1563246	0.1295665	0.49622476	** 0.5710693	0.80776504	0.6472062	** 0.7462227	1.1167293	0.90908913
cord#1--smile#1	** 0.0	0.13067555	0.0	** 0.0	0.6965932	0.0	** 0.0	0.793371	0.0
cord#1--string#1	** 3.1499379	0.22209376	0.8232399	** 0.65248954	1.2436133	0.89581146	** 0.6074834	1.4223394	0.9691444
crane#5--implement#1	** 0.47590245	0.15150303	0.13447717	** 0.22771572	0.8331397	0.31898868	** 0.3701142	0.8880154	0.48774914
crane#5--rooster#1	** 2.1563246	0.1281636	0.48304682	** 0.5710693	0.8165849	0.69604736	** 0.7462227	1.0435691	0.8757935
cushion#1--jewel#1	** 0.90511346	0.12562317	0.20419678	** 0.36066967	0.7987735	0.44731757	** 0.5466021	1.0652013	0.70597026
cushion#1--pillow#1	** 0.90511346	0.11338692	0.1935659	** 0.36066967	0.77730024	0.43915363	** 0.5466021	1.1159458	0.75774723
food#3--fruit#2	** 0.5030493	0.11502417	0.18913925	** 0.31319227	0.7236975	0.37644246	** 0.47031745	1.0361347	0.6557607
food#3--rooster#1	** 0.0	0.10259516	0.0	** 0.0	0.58650147	0.0	** 0.0	0.65800075	0.0
forest#2--graveyard#1	** 0.47590245	0.11855271	0.10679148	** 0.22771572	0.74504703	0.29009993	** 0.3701142	0.9733326	0.5303285
forest#2--woodland#1	** 4.15042	0.25172324	1.0218997	** 0.7586411	1.405135	1.0424907	** 0.65906197	1.5125645	0.9584044
fruit#2--furnace#1	** 0.0	0.11893574	0.0	** 0.0	0.6225763	0.0	** 0.0	0.6720302	0.0
furnace#1--implement#1	** 0.90511346	0.2094812	0.3197814	** 0.36066967	1.1220287	0.87618554	** 0.5466021	1.2152594	0.79994224
furnace#1--stove#1	** 0.90511346	0.14743395	0.23545239	** 0.36066967	0.84476994	0.46706042	** 0.5466021	0.9979526	0.7057414
gem#5--jewel#1	** 4.015722	0.24902125	1.0	** 0.77039593	1.2990338	1.0	** 0.74593885	1.3413637	1.0
glass#2--jewel#1	** 0.90511346	0.14743394	0.23581903	** 0.36066967	0.8778441	0.4804993	** 0.5466021	1.1059731	0.7832721
glass#2--magician#1	** 0.47590245	0.1284245	0.11934096	** 0.22771572	0.75428154	0.29215844	** 0.3701142	0.9256967	0.51503994
glass#2--tumbler#2	** 3.4726396	0.20039611	0.8205427	** 0.64194993	1.041024	0.8011642	** 0.5461729	1.1035711	0.75230753
graveyard#1--madhouse#	** 0.47590245	0.10794041	0.09769897	** 0.22771572	0.65717369	0.26033828	** 0.3701142	0.84663266	0.48011687
grin#1--implement#1	** 0.0	0.18115787	0.0	** 0.0	0.74882624	0.0	** 0.0	0.7933352	0.0
grin#1--lad#2	** 0.0	0.10692909	0.0	** 0.0	0.5962252	0.0	** 0.0	0.690108	0.0
grin#1--smile#1	** 4.5026026	0.22209376	1.0	** 0.80410427	1.2436139	1.0	** 0.70282	1.4223394	1.0
hill#5--mound#1	** 4.84939	0.20621149	1.0	** 0.89100578	1.1223272	1.0	** 0.7820115	1.2787546	1.0
hill#5--woodland#1	** 0.47590245	0.11855271	0.10679148	** 0.22771572	0.7189022	0.28135193	** 0.3701142	0.9254601	0.51745707
journey#1--car#1	** 0.0	0.14896399	0.0	** 0.0	0.6711048	0.0	** 0.0	0.60311747	0.0
journey#1--voyage#2	** 3.3624902	0.21448624	0.93852533	** 0.73267514	1.1455919	0.9126914	** 0.81426525	1.1939529	0.9907949
lad#2--brother#5	** 1.0970573	0.11969916	0.22524446	** 0.40987676	0.7984199	0.46488127	** 0.6085224	1.0878202	0.79637015
lad#2--wizard#2	** 1.0970573	0.12628113	0.24336067	** 0.40987676	0.8172692	0.50194993	** 0.6085224	1.1862205	0.8284777
magician#2--oracle#1	** 1.0970573	0.19296281	0.25495902	** 0.40987676	0.84711928	0.5197269	** 0.6085224	1.1406789	0.8194921
magician#2--wizard#2	** 4.1665072	0.2400092	1.0	** 0.76034949	1.3151247	1.0	** 0.70526475	1.4179702	1.0
midday#1--noon#1	** 4.890783	0.20446624	1.0	** 0.9071905	1.1023043	1.0	** 0.81498106	1.2275264	1.0
monk#1--oracle#1	** 1.0970573	0.12873237	0.24749996	** 0.40987676	0.8126394	0.46971724	** 0.6085224	1.0923565	0.78555936
monk#1--stove#1	** 0.47590245	0.124061026	0.11147677	** 0.22771572	0.70783526	0.27762187	** 0.3701142	0.8236957	0.47016737
mound#1--shore#1	** 0.47590245	0.12097101	0.108851306	** 0.22771572	0.7441432	0.2897999	** 0.3701142	1.0477227	0.562303
mound#1--slave#2	** 0.47590245	0.10896546	0.09958102	** 0.22771572	0.6594951	0.26112404	** 0.3701142	0.89678124	0.5014173
noon#1--string#1	** 0.0	0.10648789	0.0	** 0.0	0.58435255	0.0	** 0.0	0.65910346	0.0
oracle#1--sage#1	** 2.7990096	0.187409	0.6122158	** 0.60661178	0.9542924	0.78284005	** 0.66089485	1.1599993	0.8678957
rooster#1--voyage#2	** 0.0	0.10354301	0.0	** 0.0	0.5625592	0.0	** 0.0	0.58857635	0.0
sage#1--wizard#2	** 1.0970573	0.12883502	0.24767287	** 0.40987676	0.8513697	0.81737267	** 0.6085224	1.1902895	0.840384
serf#1--slave#2	** 1.0970573	0.11580579	0.2254489	** 0.40987676	0.7672707	0.4784692	** 0.6085224	1.1446297	0.8211261
shore#1--voyage#1	** 0.0	0.11688993	0.0	** 0.0	0.6407101	0.0	** 0.0	0.72116594	0.0
shore#1--woodland#1	** 0.47590245	0.13071603	0.119693486	** 0.22771572	0.8470659	0.32340074	** 0.3701142	1.2011509	0.6184942
tool#1--implement#1	** 2.1129973	0.40762722	0.925491	** 0.53669306	1.6826932	0.9808495	** 0.6608697	1.4160092	0.866893

les couples des concepts

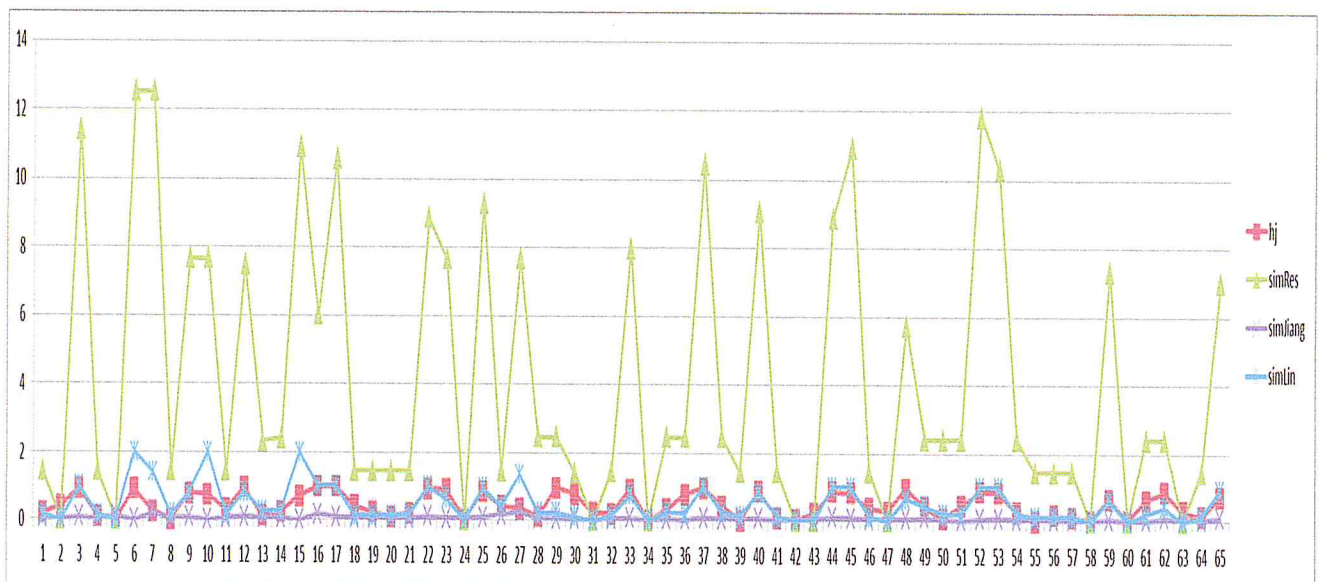
	ic sebt1			ic hadj taieb1			ic hadj taieb2		
	resnik	jiang	lin	resnik	jiang	lin	resnik	jiang	lin
asylum2--cemetery#1	** 2.7080503	0.024197778	0.1229977	** 0.5316906	0.039621286	0.041262504	** 0.7975209	0.008159174	0.012930096
asylum2--fruit#2	** 0.0	0.023999975	0.0	** 0.0	0.02754995	0.0	** 0.0	0.0075097904	0.0
asylum2--madhouse#1	** 22.059949	0.045333076	1.0	** 14.782128	0.0623649	0.96334225	** 79.910645	0.011429981	0.91587
asylum2--monk#1	** 2.7080503	0.02360675	0.120174006	** 0.5316906	0.03542147	0.03656957	** 0.7975209	0.0074795694	0.011939482
autograph2--shore#1	** 0.0	0.039193713	0.0	** 0.0	0.03226973	0.0	** 0.0	0.019769157	0.0
autograph2--signature#1	** 14.260197	0.06301343	0.9469932	** 9.493694	0.03939935	0.9694955	** 37.974777	0.021977523	0.9099993
automobile#1--cushion#1	** 16.127064	0.032739923	0.69109017	** 9.919202	0.027017047	0.42695664	** 39.67291	0.0049719637	0.3013412
automobile#1--wizard#2	** 2.7080503	0.02361422	0.11925235	** 0.5316906	0.02191159	0.022699234	** 0.7975209	0.0099054074	0.006209946
bird#1--cock#5	** 18.717903	0.045504394	0.9199361	** 20.503105	0.049912012	0.9479527	** 92.26999	0.009782403	0.8852103
bird#1--crane#5	** 18.717903	0.037770465	0.8289403	** 20.503105	0.039199155	0.8099644	** 92.26999	0.0059319257	0.6759129
bird#1--woodland#1	** 2.7080503	0.039742567	0.1949345	** 0.5316906	0.039391077	0.049897403	** 0.7975209	0.008017825	0.012770789
boy#1--lad#2	** 19.949245	0.04496173	0.9490217	** 12.878399	0.07109902	0.9599957	** 49.90755	0.016404927	0.90029077
boy#1--rooster#1	** 7.7052627	0.024917079	0.31771815	** 3.6799694	0.024182003	0.16343014	** 9.199674	0.004004667	0.071069105
boy#1--sage#1	** 11.466462	0.034596914	0.56799296	** 8.05916	0.059999337	0.6214123	** 24.174479	0.012149506	0.45409605
brother#5--monk#1	** 23.009969	0.041479179	0.976694	** 13.991019	0.06679114	0.96547204	** 60.99442	0.019906495	0.91452795
car#1--automobile#1	** 26.999497	0.037614495	1.0	** 34.069904	0.029951417	1.0	** 207.2996	0.0049249903	1.0
cemetery#1--graveyard#1	** 21.975212	0.04590592	1.0	** 10.999512	0.09100413	1.0	** 49.448003	0.02022314	1.0
cemetery#1--mound#1	** 2.7080503	0.023999497	0.1212969	** 0.5316906	0.042970236	0.044672277	** 0.7975209	0.00997105	0.01405029
cemetery#1--woodland#1	** 2.7080503	0.035197304	0.17399799	** 0.5316906	0.049999176	0.05072029	** 0.7975209	0.012209093	0.01929616
coast#1--forest#2	** 2.7080503	0.039160109	0.27213995	** 0.5316906	0.039920793	0.03727609	** 0.7975209	0.02599975	0.039999944
coast#1--hill#2	** 2.7080503	0.04225733	0.2059996	** 0.5316906	0.10999999	0.110190065	** 0.7975209	0.03136979	0.049919996
coast#1--shore#1	** 9.651173	0.03002912	0.946149	** 2.0449997	0.4909997	0.9671045	** 5.110975	0.19102224	0.8712495
cock#5--rooster#1	** 18.717903	0.031929694	0.79990216	** 20.503105	0.029099287	0.7469993	** 92.26999	0.004996426	0.5715264
cord#1--smile#1	** 0.0	0.036979903	0.0	** 0.0	0.06940267	0.0	** 0.0	0.01949999	0.0
cord#1--string#1	** 14.190992	0.057441294	0.9967952	** 6.8679997	0.09999799	0.90999997	** 24.036493	0.024719197	0.7494749
crane#5--implement#1	** 2.7080503	0.02697225	0.19391993	** 0.5316906	0.025617235	0.026974991	** 0.7975209	0.004694949	0.007499979
crane#5--rooster#1	** 18.717903	0.02749993	0.6799999	** 20.503105	0.023999996	0.6599999	** 92.26999	0.00999109	0.47949775
cushion#1--jewel#1	** 7.7052627	0.030922964	0.3799999	** 3.2049993	0.04699993	0.26162299	** 8.0114765	0.00979199	0.14549999
cushion#1--pillow#1	** 7.7052627	0.036999912	0.49999997	** 3.2049993	0.057299916	0.31021112	** 8.0114765	0.0129671945	0.19029994
food#3--fruit#2	** 3.1799999	0.04247992	0.37999996	** 2.47299924	0.060999979	0.2597479	** 4.949999	0.0129427129	0.11691766
food#3--rooster#1	** 0.0	0.029999912	0.0	** 0.0	0.02999997	0.0	** 0.0	0.004299999	0.0
forest#2--graveyard#1	** 2.7080503	0.035197304	0.17399799	** 0.5316906	0.049999176	0.05072029	** 0.7975209	0.012209093	0.01929616
forest#2--woodland#1	** 13.169975	0.19461095	1.4999999	** 3.660269	0.061379996	0.36999129	** 12.910921	0.01962194	0.39922996
fruit#2--furnace#1	** 0.0	0.029999912	0.0	** 0.0	0.049999995	0.0	** 0.0	0.010117619	0.0
furnace#1--implement#1	** 7.7052627	0.042999999	0.49999992	** 3.2049993	0.06199999	0.3314997	** 8.0114765	0.01947999	0.22064143
furnace#1--stove#1	** 7.7052627	0.034999997	0.4109212	** 3.2049993	0.024221993	0.14406097	** 8.0114765	0.0049191099	0.064992914
gem#5--jewel#1	** 20.991929	0.04562717	1.0	** 12.042999	0.09304099	1.0	** 94.199929	0.01949994	1.0
glass#2--jewel#1	** 7.7052627	0.03231092	0.39999993	** 3.2049993	0.05291194	0.29746692	** 8.0114765	0.011469927	0.16930933
glass#2--magician#1	** 2.7080503	0.026999997	0.1391616	** 0.5316906	0.03499999	0.036079067	** 0.7975209	0.009996142	0.01299999
glass#2--tumbler#2	** 8.946975	0.039999901	0.4619999	** 1.4992992	0.051029996	0.14099999	** 2.6999994	0.01999992	0.09299992
graveyard#1--madhouse#1	** 2.7080503	0.024197778	0.1229977	** 0.5316906	0.03999997	0.03999996	** 0.7975209	0.007499993	0.011447994
grin#1--implement#1	** 0.0	0.036979903	0.0	** 0.0	0.03999997	0.0	** 0.0	0.019999992	0.0
grin#1--lad#2	** 0.0	0.02922999	0.0	** 0.0	0.046279767	0.0	** 0.0	0.010974929	0.0
grin#1--smile#1	** 13.197292	0.07999991	1.0	** 7.9411194	0.13260642	1.0	** 30.164494	0.039199994	1.0
hill#5--mound#1	** 22.691093	0.044091909	1.0	** 12.919091	0.0799993	1.0	** 64.079495	0.0196066	1.0
hill#5--woodland#1	** 2.7080503	0.034999999	0.17014999	** 0.5316906	0.049999914	0.046699944	** 0.7975209	0.010999999	0.01692779
journey#1--car#1	** 0.0	0.020979109	0.0	** 0.0	0.017940997	0.0	** 0.0	0.0094072642	0.0
journey#1--voyage#2	** 21.494099	0.03997999	0.9419492	** 21.992699	0.039099949	0.9109422	** 120.90462	0.006417409	0.8679792
lad#2--brother#5	** 11.466462	0.02966713	0.4947913	** 8.05916	0.0476419	0.9948192	** 24.174479	0.00920094	0.2699921
lad#2--wizard#2	** 11.466462	0.03976974	0.5999999	** 3.05916	0.0317099	0.9999977	** 24.174479	0.011977199	0.49734999
magician#2--oracle#1	** 11.466462	0.031761099	0.50999994	** 8.05916	0.04279299	0.9124699	** 24.174479	0.009700999	0.2479496
magician#2--wizard#2	** 19.931642	0.039102113	1.0	** 12.799492	0.0791942	1.0	** 49.99416	0.020169499	1.0
midday#1--noon#1	** 14.874999	0.06722013	1.0	** 25.269999	0.03997419	1.0	** 139.97946	0.007199994	1.0
monk#1--oracle#1	** 11.466462	0.02940019	0.49999999	** 8.05916	0.046999419	0.4999997	** 24.174479	0.007940994	0.32210294
monk#1--stove#1	** 2.7080503	0.024999999	0.124479495	** 0.5316906	0.029919299	0.021991962	** 0.7975209	0.007409942	0.009999976
mound#1--shore#1	** 2.7080503	0.039797299	0.16791991	** 0.5316906	0.06979499	0.07196113	** 0.7975209	0.014622294	0.020954297
mound#1--slave#2	** 2.7080503	0.024099794	0.12209945	** 0.5316906	0.039679929	0.040299497	** 0.7975209	0.009609979	0.019622026
noon#1--string#1	** 0.0	0.030979999	0.0	** 0.0	0.029269401	0.0	** 0.0	0.009792226	0.0
oracle#1--sage#1	** 17.449499	0.036799907	0.79191994	** 12.799492	0.05299997	0.8069426	** 40.99999	0.009799991	0.6001294
rooster#1--voyage#2	** 0.0	0.019499949	0.0	** 0.0	0.0171129	0.0	** 0.0	0.002770992	0.0
sage#1--wizard#2	** 11.466462	0.035940277	0.59299996	** 8.05916	0.05619946	0.6239929	** 24.174479	0.012199979	0.49999999
serf#1--slave#2	** 11.466462	0.029991145	0.5422037	** 8.05916	0.03999799	0.4979924	** 24.174479	0.009197941	0.33079293
shore#1--voyage#1	** 0.0	0.030999909	0.0	** 0.0	0.040299799	0.0	** 0.0	0.007999994	0.0
shore#1--woodland#1	** 2.7080503	0.062129922	0.29999906	** 0.5316906	0.0370971	0.039497994	** 0.7975209	0.026617615	0.041979677
tool#1--implement#1	** 14.190992	0.05697299	0.8927174	** 9.407499	0.09771026	0.9994994	** 33.031246	0.02447999	0.8991193

## Les corrélations avec les jugements humains de Rubenstein&Goudenough [1965] :

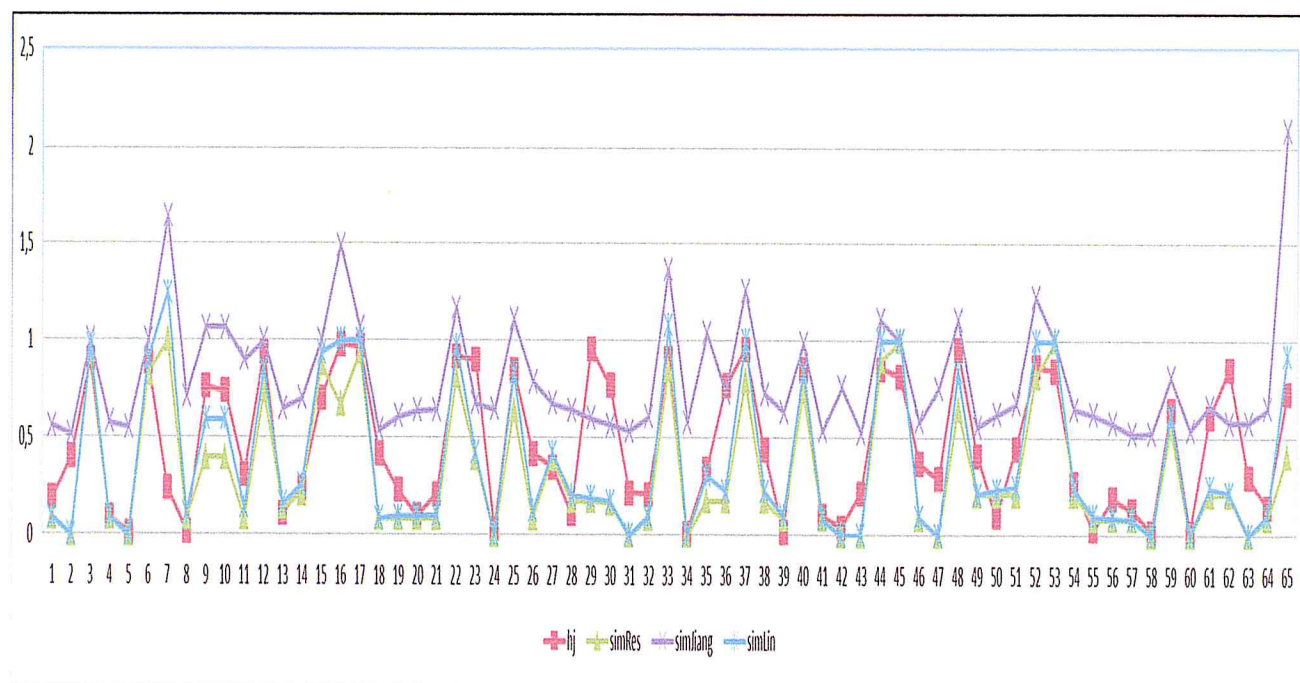
L'approche de calcul d'IC	resnik	jiang	lin
resnik	0.7487554579689379	0.20425302173104773	0.6570091665886055
seco	0.7685107080355473	0.5895215450319463	0.7793578958920863
sanchez1	0.8108955727352161	0.742487383559044	0.8154551217815983
sanchez2	0.7773258476773665	0.6196716976721622	0.7893780718234942
zhou1	0.7934122437599662	0.7119939164509668	0.7986843059931051
zhou2	0.7145324514825404	0.7290344548601826	0.7200209345094257
sebti	0.7649530871605341	0.4392789527557543	0.8243869063497082
hadjtaieb1	0.6393268809907585	0.2916992253364024	0.7906845860581044
hadjtaieb2	0.5958323505612557	0.23401059600276722	0.7930297064997899

Les graphes des comparaisons des Similarités entre les couples de la liste des jugements humains de Rubenstein&Goudenough [1965] avec les différentes approches d'CI couplées avec les trois mesures de calcul des similarités sémantiques basées sur le contenu informationnel Resnik [1995],Giang\_Conrath [1997],La mesure de lin [1998] :

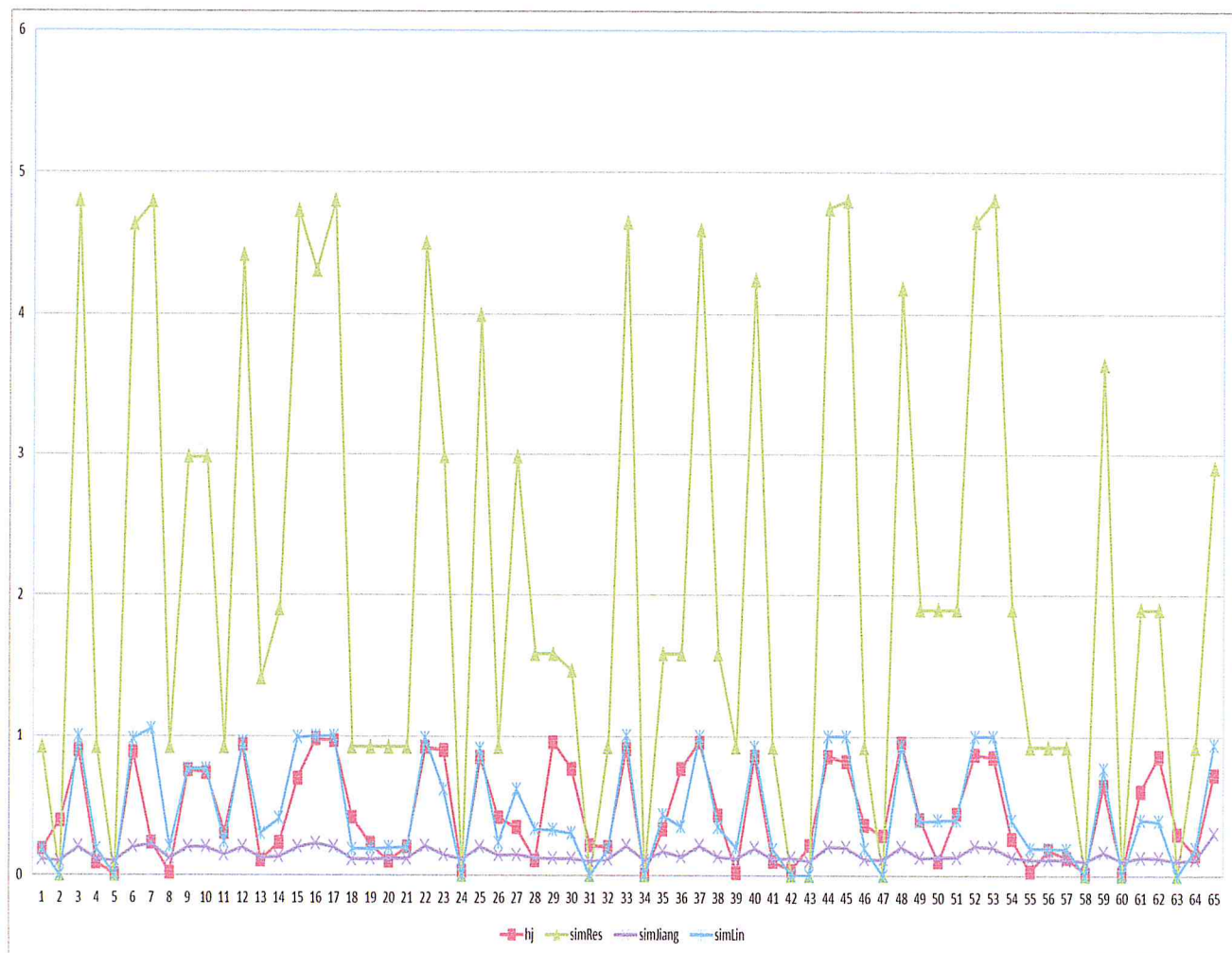
### 1-L'approche de Resnik :



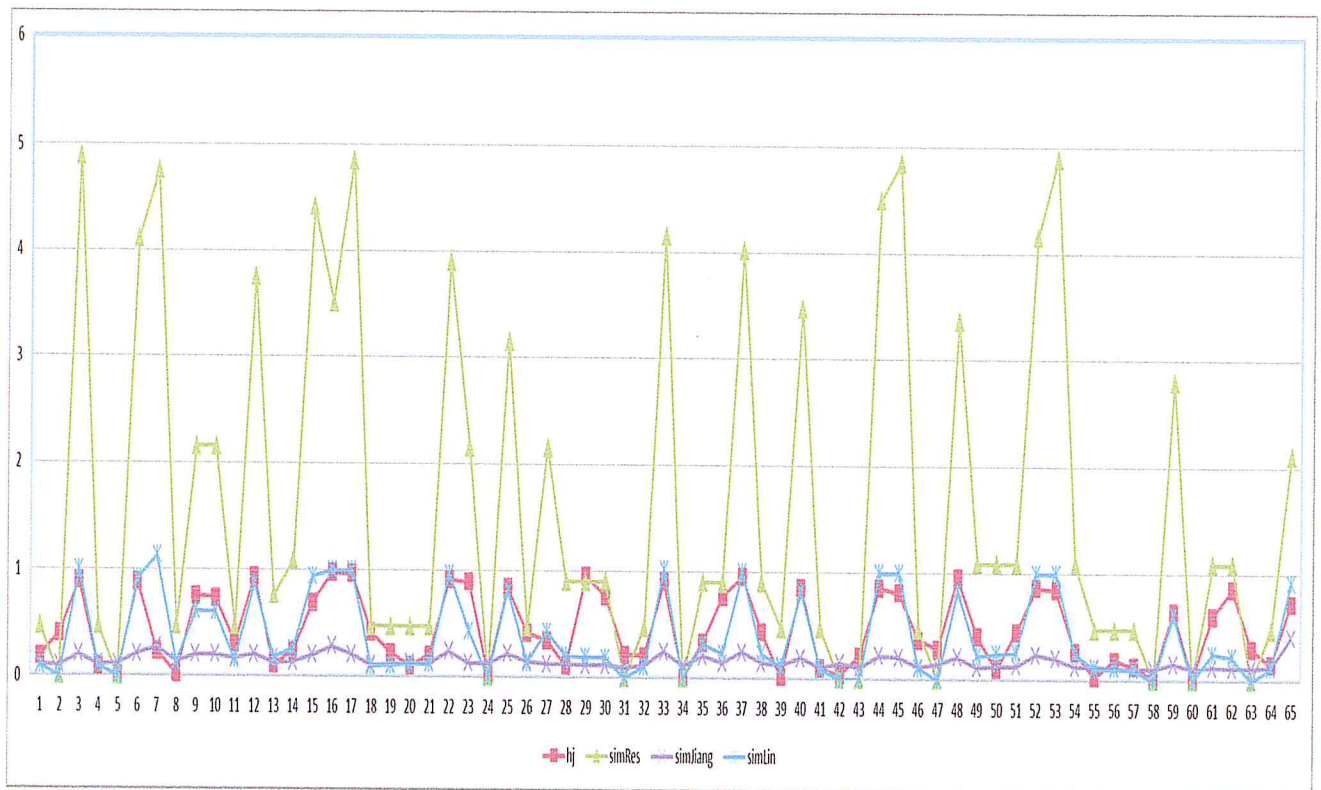
## 2-L'approche de Seco :



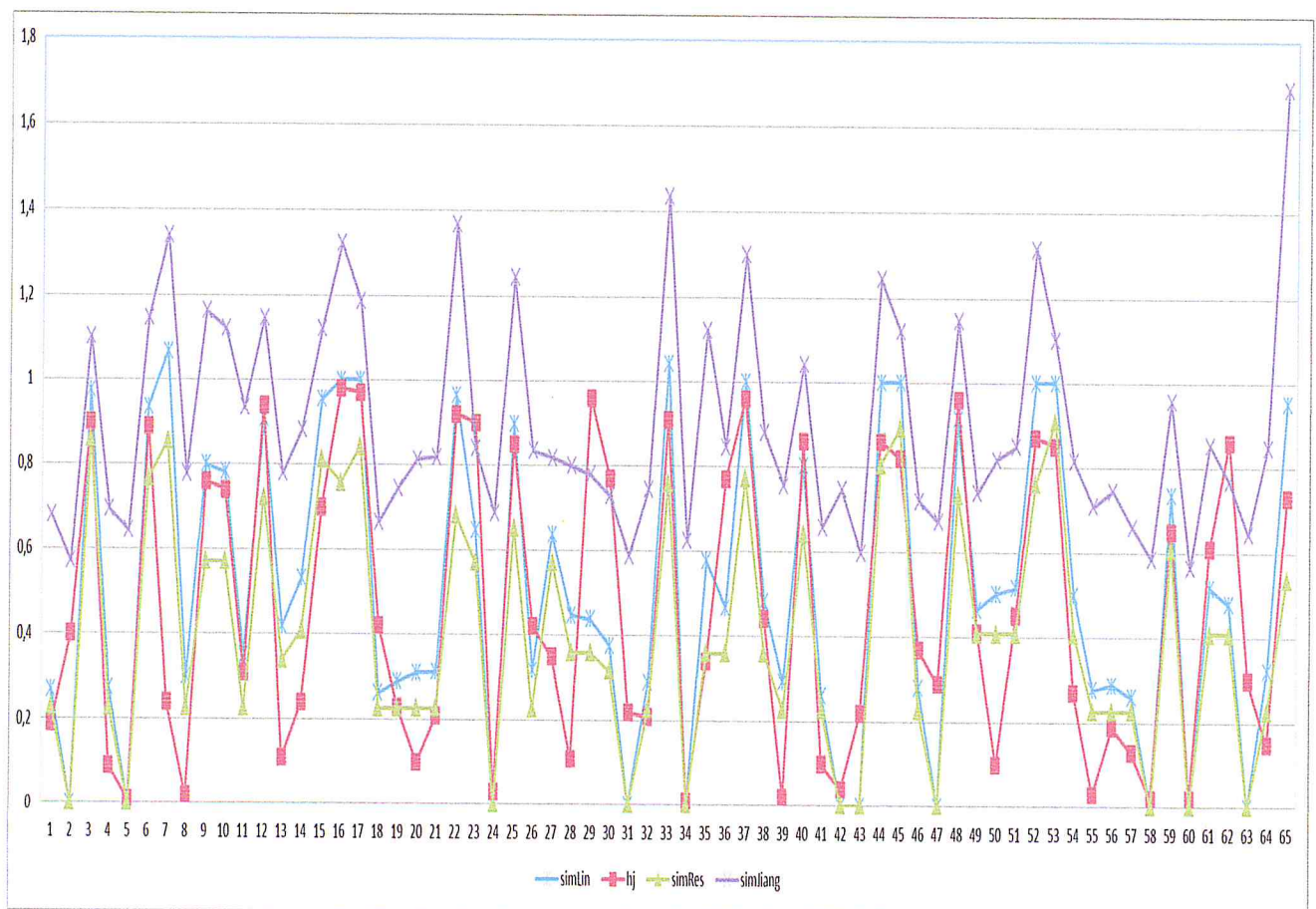
## 3-L'approche de Sanchez 1:



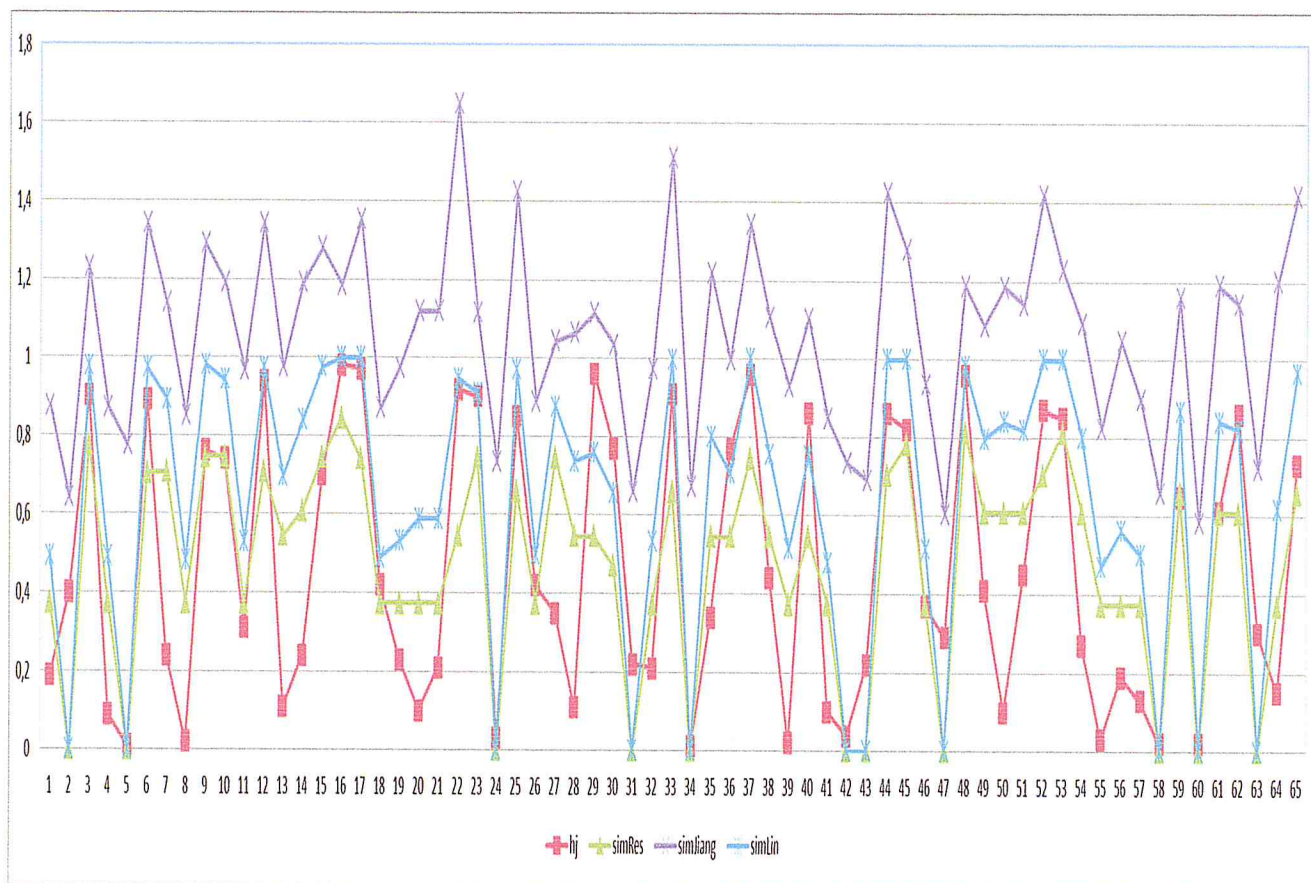
## 4-L'approche de Sanchez 2 :



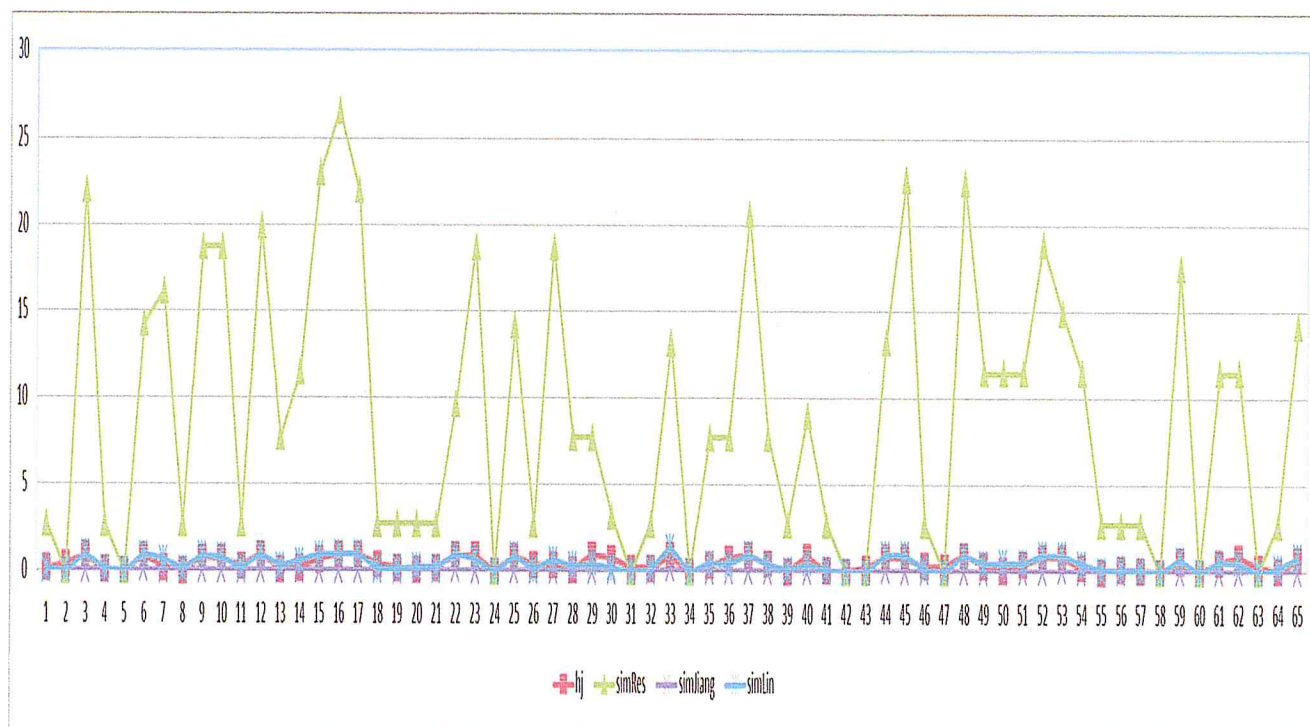
## 5-L'approche de Zhou 1 :



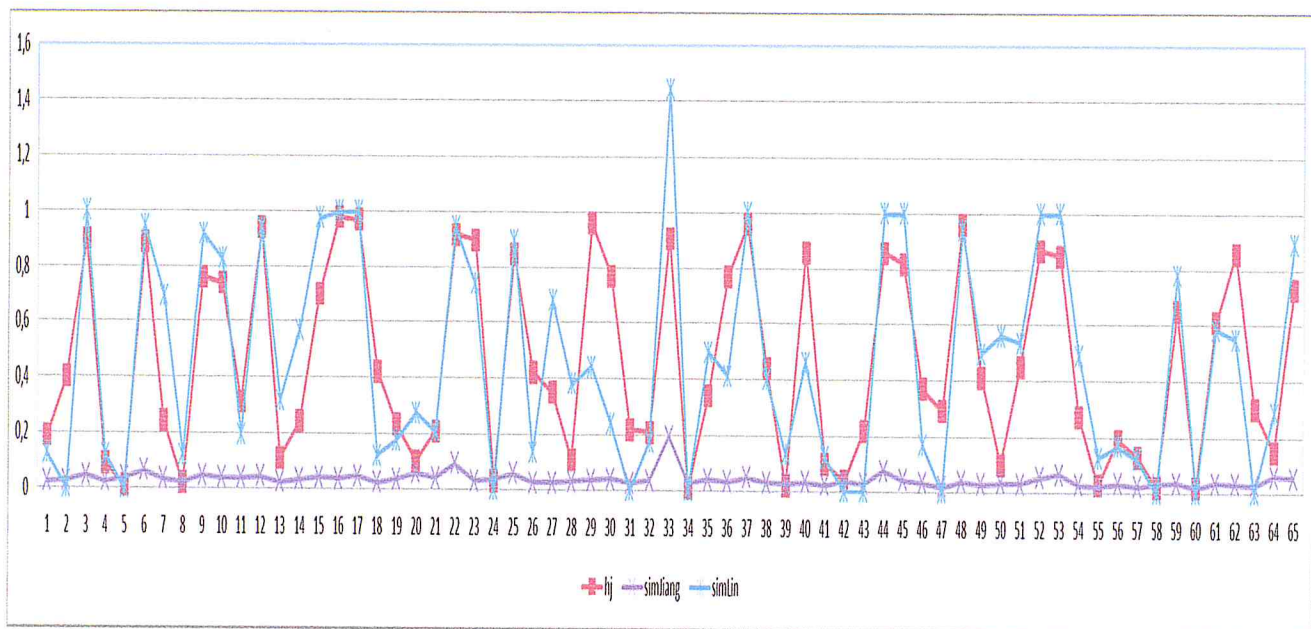
## 6-L'approche de Zhou 2 :



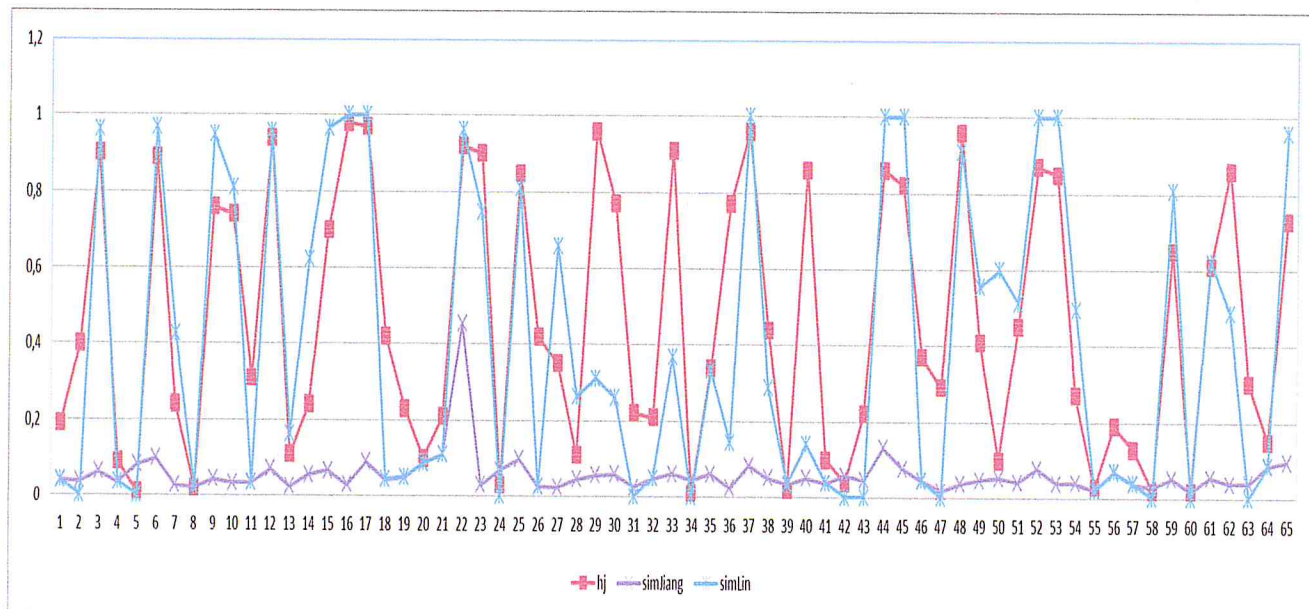
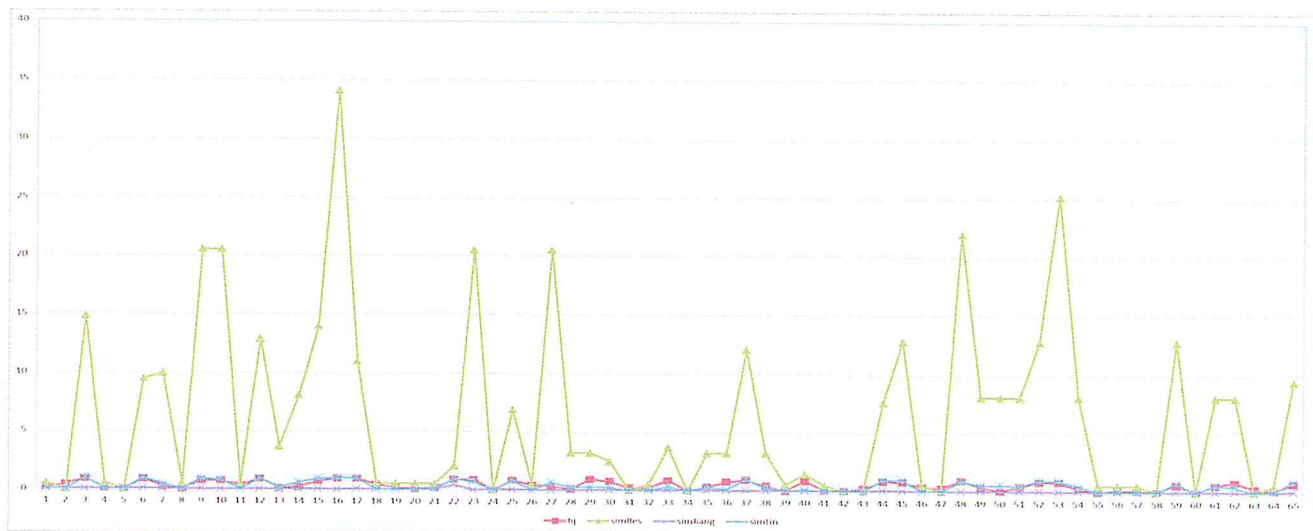
## 7-L'approche de Sebti:



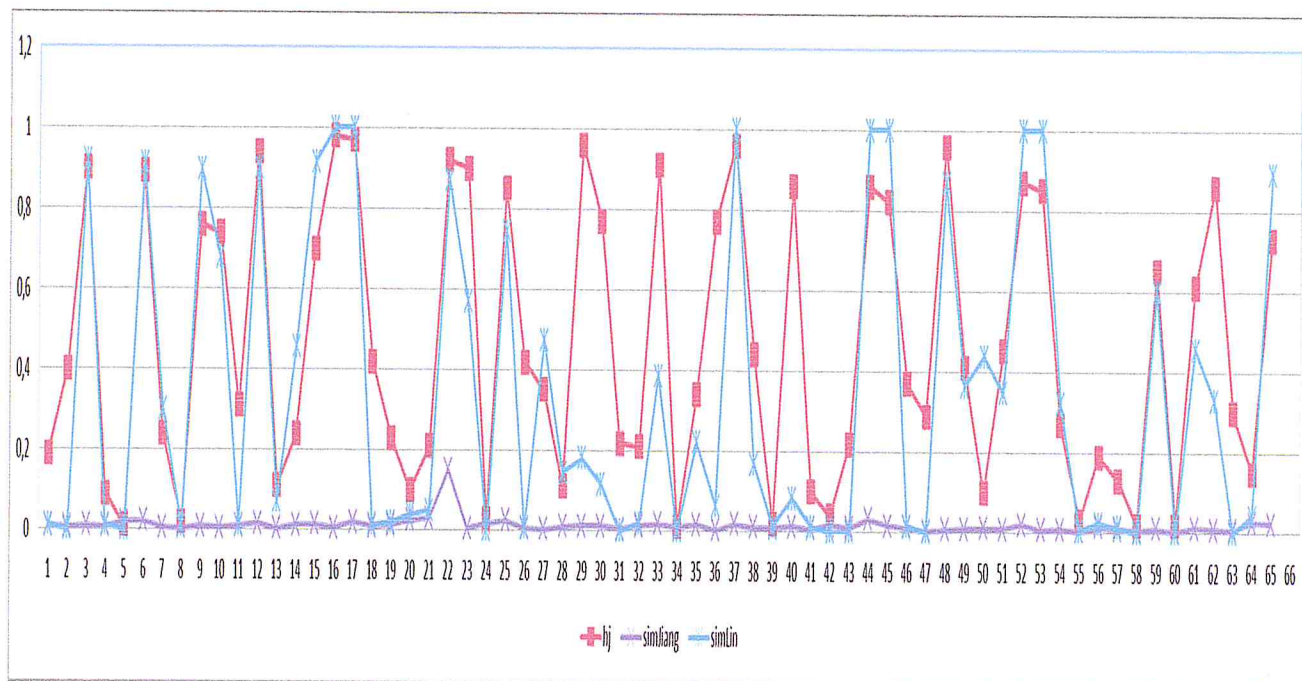
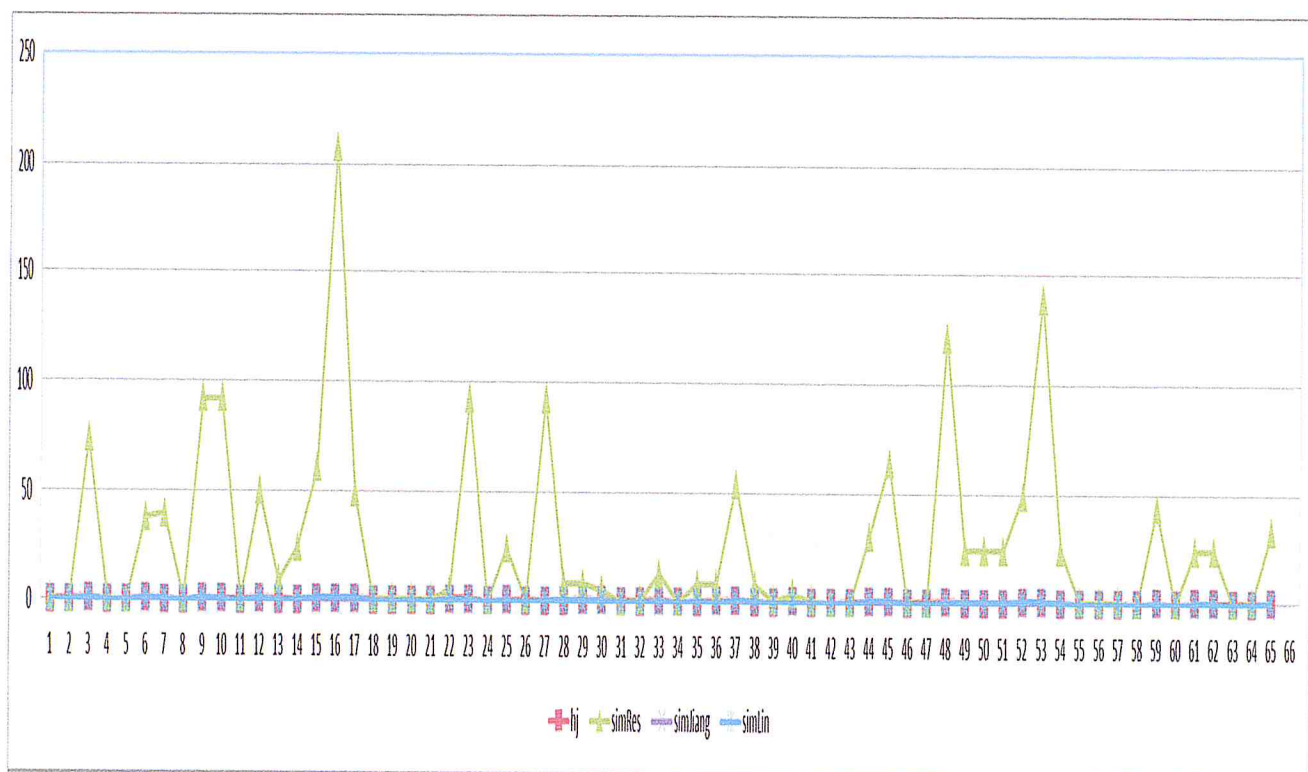




### 8-L'approche de Hadj Taieb 1 :



## 9-L'approche de Hadj Taieb 2 :



## Conclusion:

Avec les résultats précédentes nous remarquons avec la liste des jugements humains de Miller & Charles [1991] nous obtenons le classement des cinq premiers approches de calcul de la valeur de contenu informationnel couplées avec la meilleur mesure de calcul de la similarité sémantique qui est la mesure de lin [1998]qui donnent les meilleurs résultats des corrélations :

classement	L'approche de calcul d'CI	la mesure de similarité sémantique	La corrélation
1	Sanchez 2	Lin	0.8154396241825151
2	Sanchez 1	Lin	0.813899604193932
3	Seco	Lin	0.8103224276114428
4	Sebti	Lin	0.8102342275850803
5	Zhou 1	Lin	0.8092546372539724

Avec les résultats précédentes nous remarquons avec la liste des jugements humains de Rubenstein&Goudenough [1965]nous obtenons le classement des cinq premiers approches de calcul de la valeur de contenu informationnel couplées avec les deux meilleurs mesures de calcul de la similarité sémantique qui sont les mesures de lin[1998]et Resnik [1995]] qui donnent les meilleurs résultats des corrélations :

classement	L'approche de calcul d'CI	la mesure de similarité sémantique	La corrélation
1	Sebti	Lin	0.8243869063497082
2	Sanchez 1	Lin	0.8154551217815983
3	Sanchez 1	Resnik	0.8108955727352161
4	Zhou 1	Lin	0.7986843059931051
5	Zhou 1	Resnik	0.7934122437599662

## Conclusion générale :

Dans ce mémoire et dans le premier chapitre nous avons présenté les différentes approches qui calculent la similarité sémantique celles qui sont basées sur les arcs et celles qui sont basées sur la notion de contenu informationnel la majorité de ces mesures de similarité se basent sur la base de données lexicales Wordnet.

Dans le deuxième chapitre nous avons détaillé cette base de données et dans le troisième chapitre nous avons donné un état de l'art des différentes approches qui calculent la valeur de contenu informationnel.

Et dans le dernier chapitre nous avons donné notre conception UML avec un diagramme de classes qui modélise les différentes classes chacun de ces classes présente une approche qui permet de calculer la valeur de contenu informationnel selon une approche ces classes héritent les trois classes qui permet de calculer la similarité sémantique entre les couples des concepts Dans la base de données Wordnet.

Et ensuite nous avons donné la fenêtre principale de notre application qui est basée sur le langage java avec cette fenêtre on peut calculer la similarité sémantique entre n'importe quel couple de deux concepts dans la base de données Wordnet avec n'importe quelle approche de calcul de contenu informationnel couplée avec n'importe quelle mesure de similarité sémantique basée sur le contenu informationnel.

Et à la fin de notre travail nous avons donné nos tests et résultats pour étudier les performances des approches qui calculent le contenu informationnel nous avons calculé la similarité sémantique entre les couples des concepts de la liste des jugements humains de Miller & Charles [1991] et Rubenstein&Goudenough [1965] avec les différentes approches couplées avec les mesures de Lin,Resnik,Jiang et en dernier nous avons calculé les corrélations de chaque approche avec les trois mesures de Lin,Resnik,Jiang avec les listes des jugements humains de Miller & Charles [1991] et Rubenstein&Goudenough [1965] pour soutenir notre travail nous avons donné les graphes des comparaisons des Similarités entre les couples de la liste des jugements humains de Rubenstein&Goudenough [1965] et Miller & Charles [1991] avec les différentes approches d'CI couplées avec les mesures de Resnik,Giang\_Conrath,lin. Nous avons constaté que l'approche de Sanchez2 couplée avec la mesure de Lin donne la meilleure corrélation avec la liste des jugements humains de Miller & Charles [1991], et avec la liste des jugements humains de Rubenstein&Goudenough [1965] Nous avons constaté que l'approche de Sebtî couplée avec la mesure de Lin donne la meilleure corrélation.

# Références

- [1] M. Hadj Taieb, M. Ben Aouicha, A. Ben Hamadou.: "Une nouvelle approche de calcul du contenu informationnel pour mesurer la similarité sémantique utilisant WordNet". Position paper, INFORSID'11, Lille, 2011.
- [2] A. Sebti. and A. A. Barfrouch: " A new wordsensesimilaritymeasure in WordNet", Proceedings of the International Multiconference on Computer Science and Information Technologie. Poland, 2008.
- [3] H. Zargayouna, "Contexte et sémantique pour une indexation de documents semi-structurés". ACM Conférence en Recherche Information et Applications, CORIA'2004.
- [4] P. Resnik, "Semanticsimilarity in a taxonomy: An information-basedmeasure and its application to problems of ambiguity in naturallanguage". J. Artificial Intelligence Research, vol. 11, pp. 95-130, 1999.
- [5] P. Resnik, "Using information content to evaluatesemanticsimilarity in a taxonomy". In Proceedings of the 14th International Joint Conference on Artificial Intelligence, pages 448–453, Montreal, 1995.
- [6] Lin. 1998. An information-theoreticdefinition ofsimilarity. In Proceedings of the International Conference on Machine Learning, Madison, August.
- [7] Jiang, J.J., Conrath, D.W.: Semanticsimilaritybased on corpus statistics and lexical taxonomy. In: Proceedings of International Conference on Research in Computational Linguistics, Taiwan, pp. 19–33 (1997)
- [8] Seco, N., Veale, T., Hayes, J.: An intrinsic information content metric for semantic similarity in WordNet. In: Proceedings of the 16th EuropeanConference on Artificial Intelligence, Valencia, Spain, pp. 1089–1090 (2004)
- [9] Zhou, Z., Wang, Y., Gu, J.: A new model of information content for semanticsimilarity in WordNet. In: Proceedings of Second International Conference on Future Generation Communication and Networking Symposia, China, pp. 85–89 (2008)
- [10] Sánchez, D., Batet, M., Isern, D.: Ontology-based information content computation. Knowl. -BasedSyst. 24, 297–303 (2011)
- [11] Seddiqui, H., Aono, M.: Metric of intrinsic information content for measuringsemantic similarity in an ontology. In: Proceedings of 7th Asia-Pacific Conference on Conceptual Modeling, Australia, pp. 89–96 (2010)

