Approche Multi-traces et catégorisation de textes avec Random Indexing

Yann Vigile Hoareau (1) & Adil El Ghali (2)

(1) CHArt – Université Paris 8
2, rue de la Liberté 93526 St Denis Cedex 02
vigilehoarau@gmail.com

(2) Edelweiss – INRIA
2004, route des Lucioles, 06902 Sophia Antipolis
adil.elghali@gmail.com

Résumé – Abstract

Nous présentons le travail réalisé dans le cadre de la participation à DEFT09 pour la tâche 1 en français, en anglais et en italien. L'approche envisagée consiste à appliquer les modèles utilisés en psychologie cognitive pour décrire la mémoire épisodique humaine à la catégorisation automatique de texte au moyen de *Word Vectors*. Nous détaillons un modèle de mémoire épisodique et l'utilisons pour fournir les hypothèses de travail concernant la réalisation des calculs de similarités impliqués dans la catégorisation de texte avec les *Word Vectors*. Le modèle de *Word Vector* utilisé n'est pas basé sur une approche statistique classique, mais relève des projections aléatoires. La chaine de traitements proposée est entièrement automatique.

The paper resumes the work realized in text-mining context DEFT09 for the task 1 in French, English and Italian. The approach consists in applying models used in cognitive psychology to describe human episodic memory on automatic text categorization using Word Vectors. A cogntive model of episodic memory will be detailled. This model provides working hypothesis used to drive the calculus of similarity involved in text categorization with Word Vectors. The Word Vector models used is not based on the classical statistic approach but on ramdon projection. The processing chain described is entirely automatic.

Mots-Clés – Keywords

Random Indexing ; Fouille de textes ; Approche cognitive ; Mémoire épisodique
Random Indexing; Text-mining; Cognitive approach; Episodic memory.

1 Introduction

Le modèle de l'Analyse de la Sémantique Latente (LSA) (Landauer & Dumais, 1997) est le plus connu de la famille de modèles à laquelle il appartient, la famille des *Word Vectors*. Les *Words vectors* ont pour caractéristique de représenter dans un espace vectoriel à grande dimension, la similarité entre mots ou concepts. Pour ce faire, ces modèles utilisent différentes méthodes qui permettent de compter les mots en prenant en compte l'environnement textuel dans lequel ils apparaissent pour construire une matrice qui renseigne sur la co-occurrence des mots dans des contextes donnés, typiquement des documents ou des paragraphes. Ces modèles reposent sur l'hypothèse distributionnelle selon laquelle des mots qui ont un sens similaire apparaissent dans des contextes similaires.

Les *Words Vectors* ont toujours montré une grande aptitude pour la catégorisation thématique de documents. Dans sa première version, LSA, appelé alors « *Latent Semantic Indexing* » (LSI) (Deerwester et *al*, 1990) était proposé spécifiquement pour l'indexation de documents textuels. La catégorisation thématique de textes au moyen des *Words Vectors* repose sur l'application directe de l'hypothèse distributionnelle : des mots qui apparaissent dans des *contextes similaires* partagent des *thématiques similaires*.

Dans le cas de la catégorisation de textes en fonction de l'opinion qu'ils expriment, l'application directe de l'hypothèse distributionnelle montre ses limites. Si l'on prend l'exemple du Défi de Fouille de Texte 2007 (DEFT'07), des articles exprimant une opinion sur des films, des jeux-vidéo ou encore des articles scientifiques devaient être catégorisés en fonction de l'avis qu'ils exprimaient : positif, neutre ou négatif. Dans une telle situation, l'application directe de l'hypothèse distributionnelle n'est pas possible, car les films qui ont reçu une bonne critique n'ont évidemment pas en commun une thématique singulière, particulière. La chose paraît évidente pour les articles scientifiques : il n'y a pas un domaine en particulier qui recevrait des avis positifs et un autre qui n'en recevrait que de négatifs. Autant l'hypothèse distributionnelle favorise les *Word Vectors* pour la catégorisation thématique, autant son application directe fait problème pour la catégorisation de jugement d'opinion.

L'algorithme que nous avons mis en œuvre a pour objectif de pallier les limites de l'application directe de l'hypothèse distributionnelle. Cet algorithme est dérivé de la recherche en psychologie cognitive sur la mémoire épisodique. Il est couplé à un modèle de *Word Vector* appelé *Random Indexing*. Dans la première partie, nous détaillons un modèle de mémoire épisodique comptant parmi les plus fameux de la littérature ; le modèle *MINERVA* 2 (Hintzman, 1986 ;1988). Puis nous exposons les éléments de *MINERVA 2* qui servent de base à l'algorithme de catégorisation de texte. Dans la deuxième partie, nous présentons l'algorithme implémenté, ainsi que le modèle *Random Indexing* (Kanerva et *al* 1998). Dans la troisième partie, nous présentons les résultats des exécutions de la tâche 1 pour les trois langues.

1.1 Un modèle de mémoire épisodique : MINERVA 2

L'intuition sous-jacente aux modèles Multi-Traces est qu'au cours de sa vie, un individu ne rencontre que des exemplaires des objets qui composent le monde et cela, à travers des épisodes discrets. C'est à partir de ces exemplaires isolés que les catégories et les concepts se

structurent. Ainsi, aucun individu ne rencontre jamais le concept de beauté, mais seulement des exemplaires successifs de choses considérées comme belles. Les modèles Multi-traces considèrent que chaque événement de la vie d'un système de mémoire est stocké en mémoire et que chaque événement est appréhendé en fonction de l'ensemble des expériences précédentes disponibles en mémoire.

Le modèle *MINERVA 2* (Hintzman, 1984 ;1986) compte parmi les plus fameux des modèles Multi-Traces. Selon ce modèle, chaque événement ou épisode de la vie du système de mémoire est représenté et stocké sous la forme d'un vecteur à D dimensions dont les valeurs peuvent être 0, (+1) ou (-1). Le modèle a fortement contribué à l'avancement de l'étude de l'effet de fréquence des épisodes sur la capacité mnésique dans des tâches de rappel libre, de rappel indicé, de reconnaissance d'items.

Pour rendre compte de l'activation de la mémoire par un épisode nouveau, appelé sonde, *MINERVA 2* met en œuvre un processus à deux étapes. Dans la première étape, un calcul de similarité est réalisé entre le vecteur-sonde et chaque vecteur-épisode stocké en mémoire (voir Eq 1).

$$S_i = \sum_{j=1}^{N} \frac{P_j T_{i,j}}{N_i}$$

Eq 1 Similirarité d'une trace i, où P_j est la valeur de la coordonnée j de la sonde, et $T_{i,j}$ la valeur de la coordonnée j dans la trace i

Les épisodes les plus similaires à la sonde seront affectés d'une valeur d'activation plus importante que les épisodes qui sont les moins similaires. Dans la deuxième étape, un calcul est réalisé à partir des traits de chaque épisode. Le calcul consiste à reconstituer un vecteur « écho » qui hérite des traits des épisodes ayant précédemment bénéficié des valeurs d'activation les plus élevées, y compris les traits qui n'existaient pas dans la sonde. L' « écho » dispose de deux composants. Le premier composant est l'intensité, appelée *I* (voir Eq 2).

$$\mathbf{I} = \sum_{i=1}^{M} A_i, \text{ où } \quad A_i = S_i^3$$

Eq 2 Intensité de l' « echo »

Le deuxième composant est le contenu. Il est obtenu par la sommation de toutes les traces de la mémoire pondérée par valeur d'activation (voir Eq 3). Le processus d'abstraction réalisé par « echo » est qualifié par Rousset (2000) de « re-création ».

$$C_j = \sum_{i=1}^{M} A_i T_{i,j}$$

Eq 3 Contenu de l' « echo »

1.2 Effet de Fréquence

L'étude de l'effet des différences fréquence des épisodes sur l'intensité de l'écho a montré que lorsqu'une sonde est similaire à de nombreux épisodes stockés en mémoire, alors l'intensité de l'écho est élevé et inversement. Lorsqu'il y a peu d'épisodes en mémoire qui sont similaires à la sonde, alors l'intensité de l'écho est faible. On peut considérer l'intensité de l'écho comme un indicateur de la familiarité de la sonde pour la mémoire. Pour cette raison, le calcul d'écho de *MINERVA 2* a constitué une heuristique pour la mise en place des algorithmes que nous décrivons par la suite.

En considérant, d'une part, la limite décrite plus avant de l'application de l'hypothèse distributionnelle pour la catégorisation de texte d'opinion et, d'autre part, le calcul d'écho de *MINERVA 2*, nous avons transféré une partie du cadre théorique de *MINERVA 2* pour raisonner sur la phase de catégorisation de texte en utilisant le paradigme de la mémoire épisodique. Le raisonnement est le suivant :

1. Il faut se représenter un texte que l'on souhaite catégoriser comme un vecteur-sonde dans le paradigme de la mémoire épisodique. Si cette sonde est d'une catégorie A et qu'elle est comparée à une mémoire épisodique qui regroupe toute les exemplaires de la catégorie A, alors d'après *MINERVA 2*, l'intensité de l' « écho » serait fort. De même, si on compare cette sonde de la catégorie A à une mémoire épisodique qui regroupe tous les exemplaires de la catégorie B, alors l'intensité de l' « écho » sera faible.

2. En poursuivant ce raisonnement, pour approcher le calcul du contenu du vecteur « écho », tous les épisodes (ie, les documents) appartenant à une même catégorie sont sommés au sein d'un même vecteur, appelé vecteur-cible. Pour approcher l'Intensité de l'écho, la sonde est comparée au vecteur « écho » au moyen du calcul du cosinus de l'angle formée par le vecteur-sonde et le vecteur-cible.

3. Le fait de regrouper les textes appartenant à la même catégorie permet d'accroître la saillance du contenu de l'écho de la mémoire ainsi constituée. Les mémoires épisodiques correspondant à chaque catégorie de textes sont homogènes. La catégorie attribuée à une sonde est celle qui correspond à la mémoire qui délivre l'écho le plus fort intense.

1.3 Effet de typicalité des épisodes

Les recherches sur l'activité de catégorisation ont mis en évidence que tous les exemplaires d'une catégorie ne sont pas équivalents et que certains sont plus typiques de la catégorie (Rosh & Mervis 1975 ; Cordier & Tijus, 2000). On définit la typicalité d'un exemplaire pour une catégorie par une forte similarité avec les autres exemplaires appartenant à la même catégorie et une faible similarité avec les exemplaires appartenant à d'autres catégories.

En intégrant le principe de typicalité à la construction des vecteurs-cibles précédemment décrits, nous pouvons considérer que pour une catégorie d'opinion donnée, il

existe plusieurs expressions possibles. Tel qu'il est décrit dans la section précédente, le mode de calcul des vecteurs-cibles les rend sensibles pour la détection des documents les plus typiques. Nous faisons cependant l'hypothèse qu'une catégorie d'opinion est diffuse et que la prise en compte des seuls exemplaires les plus typiques conduirait à rejeter un nombre important de documents qui ne présenteraient pas les critères de typicalité, mais qui appartiendraient effectivement à la catégorie[1].

Pour remédier au problème de rejet de la catégorie pour cause de défaut de typicalité, pour chaque catégorie, nous avons regroupé les exemplaires d'une catégorie en fonction de leur degré de typicalité : à partir des vecteurs-cibles correspondant à chaque catégorie, des sous-vecteurs-cibles homogènes ont été constitués. La méthode retenue pour regrouper les exemplaires les plus similaires a consisté à comparer la similarité de chaque vecteur-exemplaire qui compose un vecteur-cible au vecteur-cible. Les vecteurs-exemplaires sont par suite ordonnée en fonction de leur similarité avec le vecteur-cible. Une partition en *P* éléments de taille identique est réalisée pour réaliser *P* sous-vecteur-cibles homogènes.

1.4 Implémentation des vecteurs-cibles et *des* sous-*vecteurs*-cibles

Les modèles de *Words Vectors* constituent un cadre tout à fait adéquat pour l'implémentation des vecteurs-cibles et des sous-vecteurs-cibles. C'est à partir des représentations vectorielles des mots et des documents issus de la construction d'un espace sémantique que seront construits les vecteurs-cibles et les sous-vecteurs-cibles. Premièrement, nous décrivons le modèle de *Word Vector* auquel nous avons eu recourt. Il est s'agit de *Random Indexing*. Deuxièmement, nous présentons la méthode de construction des vecteurs-cibles et des sous-vecteurs-cibles à partir d'un espace sémantique. Troisièmement, nous présentons la méthode de catégorisation d'un vecteur-sonde à partir des sous-vecteurs-cibles de chacune des catégories.

1.4.1 Un modèle de Word Vector : Random Indexing

Si, l'hypothèse distributionnelle permet de proposer une piste pour résoudre la question de la similarité sémantique entre les mots, le nouveau problème qu'elle met à jour est celui de la circularité entre *mot* et *contexte*. Pour résoudre la question de la circularité, la plupart des modèles de *Word Vectors* font appel à des méthodes statistiques lourdes et complexes, comme c'est le cas pour la Décomposition en Valeurs Singulières avec LSA. Le modèle de *Word Vector* que nous avons utilisé dans le cadre de ce concours est appelé *Random Indexing (RI)* (Kanerva et *al* 1998). Il ne s'appuie pas sur les méthodes mathématiques habituelles de réduction (e.g. calcul de valeurs singulières dans LSA) , mais sur des méthodes de projection aléatoire.

Les *Word Vectors* ont en commun un certains nombre de principes que l'on peut résumer ainsi :

---

[1] Nous serions très satisfait de la capacité des vecteurs-cibles de détecter les exemplaires typiques si cette propriété ne risquait pas se révéler contre-productive par la conséquence du rejet d'un grand nombre de candidats à une catégorie donnée, sous le prétexte qu'ils n'auraient pas « le look de l'emploi ».

- Ils sont basés sur l'hypothèse distributionnelle

- Ils disposent d'une méthode de comptage des mots dans un contexte donné.

- Ils disposent d'une méthode d'abstraction de la signification des mots qui s'appuie sur la prise en compte des contextes dans lesquels ces derniers apparaissent.

- Ils utilisent une méthode de représentation vectorielle pour stocker, puis manipuler le signification des mots.

Le cas de RI est un peu particulier dans la famille des *Word Vectors*. En effet, dans les autres modèles, on peut dire que l'ordre dans lequel nous avons énoncé les principes qui régissent les modèles correspond par ailleurs aux différentes étape de construction des espaces sémantiques. Ce n'est pas du tout le cas pour RI. La méthode de construction d'un espace sémantique avec RI est la suivante :

- Créer une matrice A (*d* x *N*), contenant des vecteurs-Indexes, où *d* est le nombre de documents ou de contextes correspondant au corpus et *N,* le nombre de dimensions (*N* > 1000!) défini par l'expérimentateur. Les Vecteurs Index sont creux et aléatoirement généré. Il consiste en un petit nombre de (+1) et de (-1) et de centaines de 0.

- Créer une matrice B (*t* x *N*) contenant les vecteurs-Termes, où *t* est le nombre de termes différents dans le corpus. Pour commencer la compilation de l'espace, les valeurs des cellules doivent être initialisées à 0.

- Parcourir chaque document du corpus. À chaque fois qu'un terme *t* apparaît dans un document *d,* il faut *accumuler* le *vecteur-index* correspondant au document *d* au *vecteur-terme* correspondant au terme *t.*

À la fin du processus, les *vecteurs-termes* qui sont apparus dans des contextes (ou documents) similaires, auront accumulé des *vecteurs-index* similaires. Ainsi, RI ne fait appel à aucune méthode statistique telle la SVD ou l'analyse de régression (comme c'est le cas d'autres modèles de *Word Vector*) pour réaliser le processus d'abstraction de la signification des mots.

RI dispose par ailleurs d'une option d'apprentissage par cycle. Lorsque tous les documents qui composent le corpus ont été parcourus, la matrice B contient tous les vecteurs-termes finaux. Une matrice A' (*d'* x *N*), avec *d = d'* peut de nouveau être construite, non plus à partir de la génération aléatoire comme cela fut le cas initialement pour A, mais à partir des vecteurs-terms finaux de la matrice B. Le nombre de cycle d'apprentissage est un paramètre du modèle. Le processus d'apprentissage de RI est comparable à ceux décrits dans les approches connexionnistes pour les réseaux de neurones.

Le modèle a démontré des performances aussi convaincantes (Kanerva et *al*, 2000) si ce n'est plus convaincante (Karlgren and Sahlgren, 2001) que LSA pour le test de synonymie du TOEFL (Landaeur & Dumais, 1997).

1.4.2 Construction des vecteurs-cibles et des sous-vecteurs-cibles

Après avoir construit un espace sémantique avec *Random Indexing* à partir de l'ensemble des documents appartenant à toutes les catégories,

- Chaque mot du corpus est représenté par un vecteur à D-dimensions.
- Un vecteur-document est constitué de la sommation des vecteurs-mots dont il est composé.
- Un vecteur-cible est constitué pour chaque catégorie à partir de la sommation des vecteurs-documents qui composent la catégorie.
- Les *P* sous-vecteurs-cibles sont obtenus à partir de la partition de la liste ordonnée par ordre de similarité décroissante de chaque vecteur-documents qui compose une catégorie avec le vecteur-cible de la catégorie (voir la section1.3).

1.4.3 Catégorisation d'opinion avec des sous-vecteurs-cibles

Un fois les sous-vecteurs-cibles constitués pour chaque catégorie, il s'agit de proposer une méthodes pour attribuer une catégorie à un vecteur-sonde (un document que l'on doit catégoriser). Si l'on considère qu'il y a *p* sous-vecteur-cibles pour chacune de *C* catégorie. La méthode que nous avons retenue consiste :

1. à comparer la similarité entre le vecteur-sonde et les *p* sous-vecteurs-cibles de chaque catégorie, pour chacune des *C* catégories.
2. à attribuer à la sonde la catégorie dont les sous-vecteurs-cibles sont les plus similaires.

2 Déroulement du DEFT'09

2.1 Apprentissage

La première étape à consisté à compiler les espaces sémantiques pour les corpus en français, en anglais et en italien. La deuxième étape a été de construire les vecteurs-cibles, puis les sous-vecteurs-cibles.

Afin de pouvoir pré-tester et optimiser les différents paramètres intervenant à différentes étapes du processus de catégorisation, nous avons réservé 10% du corpus d'apprentissage pour les pré-tests.

Les paramètres sur lesquels a porté le travail d'optimisation sont le nombre du dimensions *d* et le nombre de cycles *N* pour *Random Indexing*, ainsi que le nombre de sous-vecteurs-cibles *P*. Les paramètres optimum d'après nos pré-tests sont *d*=5000, *N*= 50, *P*= 5. Ces paramètres ont été retenus pour la phase de test de la tâche 1 pour les trois langues.

Enfin, la chaîne complète de traitement entièrement automatique a été programmée en Java, la

librairie SemanticVectors[2] fournissant une implémentation efficace de RI en Java. Elle comprend la compilation de l'espace avec SemanticVectors, les constructions des P sous-vecteur-cibles pour les C catégories et l'attribution d'une catégorie à chaque document du corpus de test.

2.2 Tests et résultats.

Nous avons compilé l'espace sémantique de test à partir du corpus d'apprentissage et du corpus de test. Les sous-vecteurs-cibles ont été réalisés à partir des documents étiquetés issus du corpus d'apprentissage, mais recalculés à partir de la métrique propre à l'espace sémantique de test. Le résumé des exécutions pour la tâche 1 en français, anglais et italien figure dans le tableau I.

<table>
<tr><td colspan="2"></td><td colspan="2">Français</td><td colspan="2">Anglais</td><td colspan="2">Italien</td></tr>
<tr><td rowspan="2">Nombre de documents</td><td>Appr.</td><td colspan="2">25176</td><td colspan="2">7866</td><td colspan="2">1496</td></tr>
<tr><td>Test</td><td colspan="2">16788</td><td colspan="2">5245</td><td colspan="2">999</td></tr>
<tr><td rowspan="2">Taille (Ko) ~</td><td>Appr.</td><td colspan="2">80000</td><td colspan="2">25000</td><td colspan="2">6000</td></tr>
<tr><td>Test</td><td colspan="2">51000</td><td colspan="2">16000</td><td colspan="2">3500</td></tr>
<tr><td colspan="2">Nombre de dimensions</td><td colspan="2">5120</td><td colspan="2">4096</td><td colspan="2">4096</td></tr>
<tr><td colspan="2">Nombre de Cycles</td><td colspan="2">50</td><td colspan="2">40</td><td colspan="2">40</td></tr>
<tr><td colspan="2">Nombre de sous-vecteurs-cibles</td><td colspan="2">5</td><td colspan="2">5</td><td colspan="2">5</td></tr>
<tr><td rowspan="2">précision</td><td>Obj</td><td rowspan="2">0.740</td><td>0.941</td><td rowspan="2">0.720</td><td>0.515</td><td rowspan="2">0.710</td><td>0.828</td></tr>
<tr><td>Subj</td><td>0.540</td><td>0.925</td><td>0.591</td></tr>
<tr><td rowspan="2">rappel</td><td>Obj</td><td rowspan="2">0.803</td><td>0.869</td><td rowspan="2">0.636</td><td>0.967</td><td rowspan="2">0.723</td><td>0.681</td></tr>
<tr><td>Subj</td><td>0.738</td><td>0.306</td><td>0.765</td></tr>
<tr><td colspan="2">F-mesure</td><td colspan="2">0.771</td><td colspan="2">0.676</td><td colspan="2">0.716</td></tr>
</table>

Figure 1 : Descriptions des corpora, des paramètres et des performances des exécutions de la tâche 1 en trois langues

[2] http://code.google.com/p/semanticvectors/

3 Discussion

La chaîne de traitement que nous avons proposée dans le cadre de DEFT'09 provient de l'utilisation et du transfert de la métaphore de la mémoire épisodique telle qu'elle est modélisée en psychologie cognitive. L'algorithme de catégorisation proposé se couple à un modèle vectoriel de représentation des connaissances. Comme nous l'avons énoncé plus haut, l'application de l'hypothèse distributionnelle classique pour un *Word Vector* dans une tâche de catégorisation de textes exprimant des opinions n'est pas envisageable. Les résultats obtenus avec l'algorithme de catégorisation démontrent cependant que les limites de l'hypothèse distributionnelle ont été dépassées. La capacité de dépasser les limites de l'hypothèse distributionnelle n'est pas à mettre sur le compte de quelconques paramétrages des différentes étapes de traitement : aucun pré-traitement n'a été réalisé sur le corpus.

La capacité pour un modèle de *Word Vector* de performer au-delà de la limite de l'hypothèse distributionnelle provient des traitements qui ont été réalisés lors de la construction des vecteurs-cibles et des sous-vecteurs-cibles. *Random Indexing* est ici utilisé comme un moyen particulièrement efficace de représenter les relations sémantiques entre les mots et les documents. Il n'est pas utilisé pour catégoriser les textes, car nous pensons qu'il n'a pas vocation à cela.

Inversement, l'algorithme que nous avons proposé n'a pas vocation à exprimer les relations sémantiques entre les mots, mais il a vocation à catégoriser des textes en se couplant à un modèle de représentation de la similarité entre les mots. Alors que *MINERVA 2* traite de l'information sub-symbolique, l'algorithme qui s'en inspire est capable de traiter de l'information symbolique d'un niveau d'abstraction très élevé, comme le montre sa capacité à reconnaître le caractère objectif ou subjectif d'un document.

Remerciements

Nous remercions tous les membres du projet Edelweiss et du laboratoire CHArt ainsi Dominique Widdows, le responsable du projet Semantic Vectors. Nous remercions particulièrement Charles Tijus, Denis Legros et Axel Gauvin pour leur soutien.

Références

Cordier F., Tijus C. (2001), Object properties: A typology. *Current Psychology of Cognition,* 20, 445-472

Hintzman, D. L. (1984), MINERVA 2: A simulation of human memory. *Behavior Research Methods, Instruments, & Computers*, 16, 96-101.

Hintzman, D. L. (1986), Schema abstraction in a multi-trace memory model. *Psychological Review,* 93, 411-428.

Karlgren J., Sahlgren M. (2001), From Words to Understanding, In Y. Uesaka, P. Kanerva, & H. Asoh (Eds.) *Foundations of Real-World Intelligence*, CSLI Publications, Stanford.

Landauer, T. K., Dumais, S. T. (1997), A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition, induction, and representation of knowledge. *Psychological Review*, 104, 211-240.

Sahlgren M. (2006), The Word-Space Model: Using distributional analysis to represent syntagmatic and

paradigmatic relations between words in high-dimensional vector spaces. *Ph.D. dissertation*, Department of Linguistics, Stockholm University.

Sahlgren M., Cöster R. (2004), Using Bag-of-Concepts to Improve the Performance of Support Vector Machines in Text Categorization, *Proceedings of the 20th International Conference on Computational Linguistics*, Geneva.

Rosch E. H., Mervis C. B. (1975), Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology,* 7, 573-605.

Rousset, S. (2000), Les conceptions "système unique" de la mémoire: aspect théorique. *Revue de neuropsychologie*, 10(1), 30-56.