

# MODÉLISATION ET SIMULATION D'UNE TÂCHE DE FLUENCE VERBALE SÉMANTIQUE

Catriona RABOUTET, Marc-Michel CORSINI, Bernard CLAVERIE,  
Bernard N'KAOUA

Laboratoire de Sciences Cognitives (EA 487)  
Université Victor Ségalen (Bordeaux 2)  
146, rue Leo Saignat — F33076 Bordeaux Cedex, France  
Mél : [raboutet@scico.u-bordeaux2.fr](mailto:raboutet@scico.u-bordeaux2.fr)

---

## Résumé

*L'objectif de cette étude est de proposer un modèle de simulation informatique afin de tester les théories de la mémoire sémantique et du lexique mental à partir d'une tâche de fluence verbale sémantique. Ce modèle repose sur trois hypothèses principales : 1) l'enchaînement des mots produit dans ce type de tâche dépend des liens sémantiques et phonologiques existant entre les mots ; 2) ces liens sont déterminés par l'organisation de la mémoire sémantique et du lexique mental et 3) la production d'un mot dépend aussi de ses propriétés de typicité et de fréquence d'utilisation. Le modèle proposé repose sur les organisations catégorielle et schématique de la mémoire sémantique et sur des réseaux lexicaux phonologiques de types phonémiques et syllabiques. Le simulateur SysPro a été développé à partir d'un corpus de mots recueillis auprès de 30 sujets sains soumis à la tâche de fluence « supermarché ». Il produit des listes simulées, dépendantes des liens et des listes aléatoires. La pertinence du modèle et des organisations qu'il sous-tend a été déterminée en comparant les productions simulées et aléatoires avec celles produites par les sujets, en termes de nombre de mots communs, de fréquence des liens et de récupération catégorielle. Les résultats obtenus confortent les théories du lexique mental qui supposent une organisation en réseau de plusieurs unités. En particulier, ils montrent la pertinence des réseaux phonologiques et de la syllabe de début comme unité de récupération des mots en mémoire. Concernant la validité de l'organisation de la mémoire sémantique, les résultats soulignent la nécessité de reformaliser l'arbre catégoriel du système et ses modes de récupération.*

## Abstract

*This study proposes a computational modelling of semantic memory and mental lexicon through a semantic fluency task. This model is supported by three hypotheses: 1) fluency productions depend on the number of semantic and phonological links between words 2) these links are determined by the semantic memory and mental lexicon organizations 3) fluency productions also depend on the typicality and frequency word properties. The model is based on categorical and schematic semantic organizations and on phonological (phonemic and syllabic) lexical networks. The computational system was elaborated from an item's lexicon collected through a "supermarket" fluency of 30 subjects. It produces simulated lists, dependent on links, and random lists. The relevance of the model and of semantic and lexical organizations was determined by comparing simulated and random lists with those produced by subjects in terms of common words number, links frequency and semantic categorical exploitation. The main results were in agreement with mental lexicon theories*

*postulating a network organization of several units. In particular they showed the relevance of phonological networks and of the first syllable defined as a unit of words retrieval. As regards semantic memory organization, the results suggest a need to improve formalization of the categorical network and its retrieval processes.*

## Resumen

*El objetivo de este estudio es proponer un modelo de simulación informática que permite poner a prueba las teorías de la memoria semántica y del léxico mental a partir de una tarea de fluencia verbal semántica. Dicho modelo esta basado en tres hipótesis principales : 1) el enlace de palabras producido en este tipo de tarea depende de los lazos semánticos y fonológicos que existen entre las palabras ; 2) la organización de la memoria semántica y del léxico mental determinan dichos lazos ; 3) la producción de una palabra depende también de sus características de tipicidad y de frecuencia de uso. El modelo propuesto se basa en las organizaciones categoriales y esquemáticas de la memoria semántica así como en redes lexicales fonológicas de tipo fonémico y silábico. El simulador SysPro ha sido desarrollado a partir de un corpus de palabras procedentes de 30 sujetos sanos sometidos a la tarea de fluencia de tipo « supermercado ». El simulador produce listas simuladas, las cuales dependen de los lazos, y listas aleatorias. La comparación de producciones simuladas y aleatorias con las producidas por los sujetos (por lo que se refiere al número de palabras en común, frecuencia de lazos y recuperación categorial) permitió definir la pertinencia de dicho modelo así como la pertinencia de las organizaciones subyacentes. Los resultados obtenidos confirman las teorías del léxico mental que suponen una organización en una red de varias unidades. Muestran en particular la pertinencia de redes fonológicas y de la sílaba de comienzo que actúan como unidad de recuperación de las palabras presentes en la memoria. Por lo que concierne a la validez de la organización de la memoria semántica, los resultados ponen de relieve la necesidad de volver a formalizar el árbol categorial del sistema así como sus modos de recuperación.*

## Riassunto

*Lo scopo di questo studio è di proporre un modello di simulazione informatica per verificare le teorie della memoria semantica e del lessico mentale ad operazione di un' attività di fluencia verbale semantica. Questo modello si basa su tre presupposti principali : 1) la concatenazione delle parole prodotta da questo tipo di operazione dipende dai legami semantici e fonologici che esistono fra le parole ; 2) questi legami sono determinati dall' organizzazione della memoria semantica e dal lessico mentale e 3) la produzione di una parola dipende inoltre delle sue proprietà di tipicità e di frequenza d'uso. Il modello suggerito si fonde su delle organizzazioni categoriale e schematiche della memoria semantica e su delle reti lessicali fonologiche di tipo fonemico e sillabico. Il simulatore SysPro è stato sviluppato a partire da un corpus di parole ottenute da 30 individui sani sottoposti all' attività di fluencia « supermercado ». Questo sistema elabora delle liste simulate, dipendenti dei legami e delle liste casuali. L'attinenza del modello e delle organizzazioni che tale modello presuppone è stata determinata comparando le produzioni simulate e casuali a quelle prodotte dagli individui, in termini di numero di parole comuni, di frequenza dei legami e di recupero categoriale. I risultati ottenuti consolidano le teorie del lessico mentale che suppongono un'organizzazione reticolare di parecchie unità. In particolare, mostrano l'attinenza delle reti fonologiche e della sillaba di inizio come l'unità di recupero delle parole in memoria. Per quanto riguarda la validità dell'organizzazione della memoria semantica, i risultati sottolineano la necessità di formalizzare di nuovo l'arborescenza categoriale del sistema e i suoi modi di recupero.*

## Resumo

*A proposta deste estudo é apresentar um modelo de simulação computacional a fim de testar as teorias da memória semântica e do léxico mental a partir de uma tarefa de fluência verbal semântica. Este modelo está ancorado em três hipóteses principais : (1) o encadeamento das palavras produzido nesse tipo de tarefa depende das relações semânticas e fonológicas existentes entre as palavras ; (2) essas relações são determinadas pela organização da memória semântica e do léxico mental e (3) a produção de uma palavra depende também de suas propriedades de tipicidade e de frequência de uso. O modelo proposto está fundamentado nas organizações categorial e esquemática da memória semântica, bem como em redes lexicais fonológicas de tipos fonêmicos e silábicos. O simulador SysPro foi desenvolvido a partir de um corpus de palavras recolhidas de 30 sujeitos sadios submetidos à tarefa de fluência de « supermercado ». Tal corpus produz listas simuladas, dependentes das relações, e listas aleatórias. A pertinência do modelo e das organizações que subjazem ao mesmo foi determinada comparando as produções simuladas e aleatórias com aquelas produzidas pelos sujeitos, em termos de número de palavras comuns, de frequência das relações e de recuperação categorial. Os resultados obtidos estão de acordo com as teorias do léxico mental que supõe uma organização em rede de várias unidades. Em particular, mostram a pertinência das redes fonológicas e da sílaba inicial como unidade de recuperação de palavras na memória. Com relação à validade da organização da memória semântica, os resultados indicam a necessidade de reformular a árvore categorial do sistema e seus modos de recuperação.*

## 1. Introduction

Ce travail s'inscrit dans le cadre des sciences cognitives et vise à proposer un modèle de simulation informatique afin de tester les théories de l'organisation de la mémoire sémantique et du lexique mental à l'aide d'une tâche de production langagière orale.

La mémoire sémantique et le lexique mental sont deux structures cognitives supposées rendre compte de l'organisation des mots en mémoire.

Traditionnellement représentée sous forme de concepts, la mémoire sémantique contient les connaissances abstraites et générales que nous avons du monde. Ces connaissances sont organisées en catégories et définies par des relations subordonnées/surordonnées de type inclusion (organisation hiérarchique proposée par Collins et Quillian, 1969). Dans cette organisation, deux types de concepts sont distingués : les concepts généraux ou labels, situés aux niveaux supérieurs de la hiérarchie (ex. « oiseaux ») et les concepts plus spécifiques, les exemplaires (ex. « canari ») aux niveaux inférieurs. La diffusion de l'activation dépend notamment de la distance à parcourir entre deux concepts et du degré de typicité d'un exemplaire<sup>1</sup> (Rosch, 1975 ; Rosch et Mervis, 1975). Certains auteurs, cependant, ont montré que cette catégorisation des objets n'était pas toujours stable (Ross et Murphy, 1999). Ainsi, les propriétés prises en compte pour catégoriser ces objets, et donc déterminer leur similitude avec d'autres, peuvent aussi dépendre du contexte dans lequel les objets apparaissent. Une organisation contextuelle de ces concepts a donc été également proposée (théorie des schémas). Un schéma peut être décrit

---

<sup>1</sup> Typicité catégorielle : les exemplaires les plus typiques d'une catégorie sont traités plus rapidement que les autres.

comme un regroupement de connaissances organisé à partir d'événements (ex. « le petit déjeuner »). Dans ce cas les éléments sont activés simultanément et proportionnellement à leur degré de typicité contextuelle (plus un élément est typique d'un schéma et plus il sera activé) (Cordier, 1993).

Le lexique mental est la structure qui contient les mots associés à ces concepts. Il est organisé en plusieurs réseaux d'unités syntaxico-sémantiques, phonologiques et orthographiques des mots (Bybee, 1988 ; Caron, 1995). Parmi ces réseaux, nous retenons plus particulièrement ceux des unités phonologiques, les plus traditionnelles étant les unités phonologiques phonémiques et syllabiques (syllabe de début, syllabe de fin, nombre de syllabes) (Dell, 1986, 1988, 1990 ; Levelt *et al.*, 1999 ; Roelofs, 1997 ; Wheeldon et Levelt, 1995). Dans ce cas, chaque mot est alors en relation avec les mots constitués des mêmes unités phonologiques, l'activation de ces unités étant particulièrement dépendante de la fréquence d'utilisation des items (Dell, 1990 ; Griffin et Bock, 1998 ; Jescheniack et Levelt, 1994). Les mots de plus forte fréquence sont alors activés et prononcés plus rapidement que les mots de faible fréquence.

L'organisation des mots en mémoire est donc déterminée à la fois par des liens conceptuels de la mémoire sémantique (catégoriels et schématiques) et des liens phonologiques du lexique mental (syllabiques et phonémiques). Ces liens sont, selon les modèles de la production orale, impliqués dans la génération de mots (Caramazza, 1997 ; Dell, 1986, 1988 ; Humphreys *et al.*, 1988, 1995 ; Levelt, 1999 ; Levelt *et al.*, 1999 ; Roelofs, 1997).

Parmi les différentes tâches de production (dénomination, catégorisation, génération de mots etc.), la tâche de fluence verbale est très souvent utilisée en psychologie et neuropsychologie cognitive. Elle consiste à demander aux sujets de générer selon une consigne particulière le maximum de mots en un temps limité. Une des consignes les plus utilisées est de citer le maximum de mots appartenant à une même catégorie sémantique (ex. « Animaux », « Supermarché », composante de la *Dementia Rating Scale*, Mattis, 1988). Les listes d'items produites sont organisées en regroupements sémantiques et phonologiques des mots rendant compte des différents liens existant entre eux (Raskin *et al.*, 1992 ; Troyer *et al.*, 1997 ; Troyer, 2000). Tröster et collaborateur (1995) se sont intéressés aux liens conceptuels responsables de l'enchaînement de mots et plus particulièrement à leur organisation catégorielle. Ces auteurs ont proposé, dans le cadre de la fluence « supermarché » (citer tout ce qui peut être acheté dans un supermarché) un découpage catégoriel du réseau sémantique. Ce découpage est défini selon trois niveaux : un niveau supérieur (10 catégories prédéfinies par les expérimentateurs : « fruit », « viande »), un niveau intermédiaire de labels (ex. « fruit », « fruit frais », « fruit sec ») et un niveau inférieur d'exemplaires (ex. « banane »). Afin de mesurer l'implication de cette organisation catégorielle dans les listes de mots produites, les auteurs ont déterminé différents indices de mesure. Ainsi, le nombre de catégories visitées, le nombre de labels et d'exemplaires produits et le nombre de mots par catégorie mesurent l'étendue du réseau et les différents niveaux exploités. Le nombre de changements de catégories indique quant à lui, les stratégies d'accès au réseau utilisées. Deux types de stratégies d'accès sont possibles : une stratégie d'accès vertical caractérisée par un petit nombre de changements et une stratégie d'accès horizontal se traduisant par un grand nombre de changements de catégories.

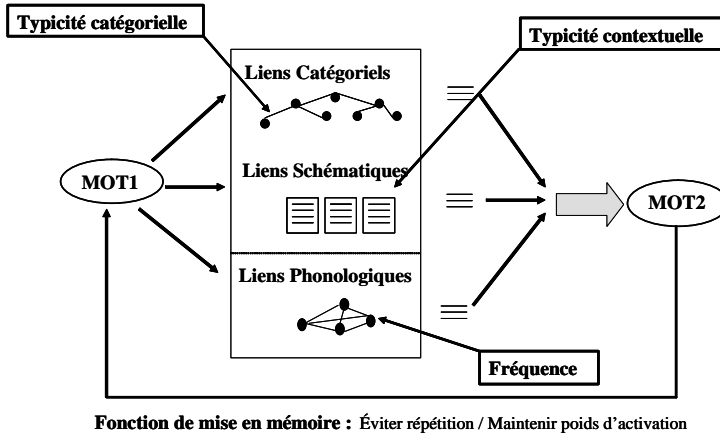


Figure 1 — Modèle de simulation.

Dans cette étude, nous proposons un modèle des théories de l'organisation de la mémoire sémantique et du lexique mental que nous testerons à travers une tâche de fluence verbale. Ce modèle repose essentiellement sur deux hypothèses :

1) les listes de mots produites dans une tâche de fluence dépendent du nombre de liens sémantiques (conceptuels et associatifs) et phonologiques existants entre les mots. Un mot produit activera préférentiellement le ou les mots ayant le plus de caractéristiques communes.

2) La production d'un mot dépend de sa fréquence d'utilisation. Le modèle proposé repose sur un lexique à partir duquel nous avons déterminé pour chaque mot leurs caractéristiques conceptuelles, associatives et phonologiques. Les caractéristiques communes entre deux mots définissent ainsi la nature des liens existant entre eux. Les liens conceptuels ont été formalisés à l'aide d'un arbre catégoriel, les liens associatifs à l'aide de schémas et les liens phonologiques phonémiques et syllabiques à l'aide des transcriptions phonémiques et syllabiques des mots (figure 1). L'enchaînement des mots produits est alors déterminé par une activation indépendante et parallèle de ces liens et par un processus de sélection (système de compétition) dépendant du nombre de liens partagés entre les mots.

La production des mots repose sur le principe suivant : le mot amorce (le premier mot produit par le sujet) active l'ensemble de ses liens et détermine ainsi, en fonction de la fréquence d'utilisation et des degrés de typicité catégorielle et contextuelle, une liste d'items associés. Parmi les items partageant le plus de liens avec le mot amorce, le système sélectionne par compétition le mot qui sera produit et qui deviendra à son tour le mot amorce pour l'itération suivante.

Afin d'évaluer la pertinence de ce modèle, nous avons élaboré le logiciel *SysPro* (Système de Production) qui repose sur ces règles d'organisation. La comparaison des productions simulées (dépendantes des règles) avec des productions aléatoires (indépendantes des règles) et des productions obtenues auprès de sujets sains, nous permettra de tester la validité du modèle et les théories de l'organisation des mots en mémoire sur lesquelles il repose.

## 2. Méthode

Nous avons dans un premier temps, constitué un lexique de base correspondant à l'ensemble des mots recueilli auprès de sujets sains soumis à une tâche de fluence « supermarché ». Par la suite, nous avons formalisé et implémenté les différents liens en déterminant pour chacun des items ses caractéristiques sémantiques (catégorielles, associatives), phonologiques (phonémiques, syllabiques), sa fréquence d'utilisation et ses indices de typicité (catégorielle et contextuelle).

### 2.1. Recueil des données et formalisation

#### 2.1.1. Constitution du lexique

Le recueil des données (mots corrects et erreurs d'intrusion) a été effectué auprès de 30 sujets sains (17 femmes et 13 hommes) âgés de 20 à 30 ans (moyenne 23 ans) de niveau supérieur au Bac. Le lexique comporte 334 items et est composé de mots isolés (ex. « lait », « éponge ») et de mots combinés (ex. « produit d'entretien », « produit surgelé ») (Annexe1).

#### 2.1.2. Fréquence d'utilisation

La fréquence d'utilisation d'un item a été déterminée par son nombre d'occurrences sur l'ensemble des 30 productions recueillies.

#### 2.1.3. Liens conceptuels et indice de typicité catégorielle

Les liens conceptuels sont formalisés au sein d'un arbre catégoriel dans lequel s'inscrivent les 334 items du lexique répertoriés en labels et en exemplaires. Le découpage catégoriel utilisé est inspiré en partie de celui proposé par Tröster *et al.*, (1995) et adapté aux supermarchés français. Il comprend deux grandes catégories « alimentaire » vs. « non alimentaire », elles-mêmes sous divisées en 35 sous catégories : 20 catégories génériques (de niveau1) « fruit », « viande » et 15 catégories spécifiques (de niveau 2) « fruit frais », « fruit sec » (Annexe2). Chaque item est représenté dans l'arbre selon un codage spécifique qui indique son type et son appartenance catégorielle (ex. « pomme » est codé : Sup/A/F/Ff/e où Sup = supermarché ; A = alimentaire ; F = fruit ; Ff = fruit frais ; e = exemplaire).

Afin de déterminer l'indice de typicité des exemplaires, nous avons construit deux questionnaires. Ces questionnaires comprenaient respectivement 11 catégories génériques « alimentaires », 9 « non alimentaires » et leurs exemplaires. Les questionnaires ont été soumis à 80 nouveaux sujets sains (40 par questionnaire) répondant aux mêmes critères de sélection que les sujets précédents. La consigne était de classer par ordre décroissant les 4 exemplaires les plus typiques de la catégorie proposée (4 = le plus typique, 1 = le moins typique). L'ordre de présentation des catégories et des exemplaires était contrebalancé pour l'ensemble des sujets. L'indice de typicité pour un exemplaire donné correspond à la somme des occurrences, multipliée par la valeur des rangs proposés (de 4 à 1).

Nous avons par la suite déterminé deux règles d'activation dans l'arbre. Une règle intracatégorielle favorise l'exploitation verticale de l'arbre. Chaque exemplaire active de ce fait l'ensemble des exemplaires de sa catégorie, le label de sa catégorie et le label directement supérieur. Une seconde règle intercatégorielle a pour but de favoriser l'exploitation horizontale de l'arbre, c'est-à-dire les changements de

catégories. Ainsi chaque label active l'ensemble des exemplaires appartenant à sa catégorie mais aussi l'ensemble des labels de l'arbre.

#### **2.1.4. Liens associatifs et indice de typicité contextuelle**

Nous avons déterminé les liens associatifs à partir de 9 schémas issus du contexte « supermarché » (Ross et Murphy, 1999) (« petit déjeuner », « faire sa toilette »). Ces schémas ont été soumis à 40 nouveaux sujets répondants tous aux mêmes critères de sélection. La consigne était de donner par écrit le maximum d'objets appartenant aux schémas proposés et pouvant être achetés dans un supermarché. La présentation des 9 schémas était contrebalancée pour l'ensemble des sujets. En ne retenant que les items appartenant au lexique étudié, nous avons ainsi obtenu une liste d'exemplaires associés à chacun des schémas, un item pouvant appartenir à plusieurs schémas (ex. l'item « verre » appartient aux schémas « apéritif » et « faire la vaisselle »).

L'indice de typicité contextuelle pour un item a été déterminé par son nombre d'occurrences dans un schéma donné.

#### **2.1.5. Liens phonologiques**

Les liens phonologiques retenus sont des liens de ressemblance phonémique (nombre de phonèmes communs) et syllabique (syllabe de début et syllabe de fin).

La ressemblance phonémique entre deux mots correspond au nombre de phonèmes successifs communs des chaînes phonémiques de début et de fin de mot. La transcription phonémique de chaque item a été réalisée à partir de l'alphabet phonétique international.

Les liens syllabiques sont définis à partir des syllabes de début et de fin déterminées selon les critères de Warrant (1996).

## **2.2. Système de simulation**

Nous avons ainsi répertorié les 334 items selon leurs différentes caractéristiques phonologiques et sémantiques (syllabe de début, syllabe de fin, transcription phonémique, position catégorielle et schémas). Chacun de ces items est associé, dans le système, à un poids d'activation compris entre 0 et 1 déterminé par la somme des activations des caractéristiques phonologiques et sémantiques et du poids d'activation issu des itérations précédentes, le tout soumis à une fonction de normalisation des valeurs à [0,1]. Plus précisément, ce poids s'obtient en additionnant :

- Le poids des connexions de chaque caractéristique phonologique commune avec le mot amorce, multiplié par le score de fréquence d'utilisation de l'item.
- Le poids des connexions de chaque caractéristique sémantique commune avec le mot amorce, multiplié par les scores de typicité catégorielle ou contextuelle.
- Et le poids d'activation de l'item considéré obtenu lors des itérations précédentes.

Le fonctionnement général qui sous-tend cette attribution de poids est le suivant : à chaque mot amorce, le système détermine pour les 333 items restants leur poids d'activation. Ces poids sont stockés dans un système de mémoire à court terme où s'effectue, par compétition entre les items possédant les poids les plus élevés,

l'élection du mot à produire. Après l'élection de ce mot, les poids d'activation des items non élus sont soumis à une fonction mnésique d' « oubli » rendant compte d'une diminution progressive de leur activation au cours du temps, et à nouveau stockés en mémoire pour l'itération suivante.

Pour cela, l'expérimentateur doit fournir plusieurs paramètres indispensables à la simulation tels que :

- Le poids des connexions des différentes caractéristiques. Ce paramètre permet de faire varier l'influence accordée aux différentes caractéristiques sémantiques et phonologiques dans le calcul du poids d'activation d'un item. Cette influence est déterminée par un coefficient compris entre 0 et 1.
- La probabilité d'activation des connexions des différentes caractéristiques. Ce paramètre permet de faire varier la probabilité d'accès à une connexion particulière. Cette probabilité est comprise entre 0 et 1.
- Un seuil de compétition compris entre 0 et 1. Ce seuil permet de faire varier le nombre d'items entrant en compétition afin de déterminer le mot à produire. Plus le seuil est bas (proche de 0) plus le nombre d'items entrant en compétition est important.
- La fonction « oubli ». Plusieurs fonctions sont proposées par le simulateur dont une fonction de remise à zéro des poids entre chaque itération (RAZ), une fonction de perte linéaire, une fonction logarithmique et une de courbe en U rendant compte des effets de primauté et de récence observés lors des études de rappel immédiat (Glanzer et Cunitz, 1966).

D'autres paramètres sont également proposés à l'expérimentateur :

- Une fonction de « typicité catégorielle ». Le logiciel offre deux fonctionnalités différentes de la typicité catégorielle. La première est définie comme une fonction « régulatrice » du poids d'activation d'un item. Dans ce cas, la typicité n'agit que sur les items préalablement activés dans l'arbre catégoriel en modulant leur poids d'activation en fonction de leur indice de typicité. La seconde correspond à une fonction d' « activation étendue » où la typicité permet d'activer, en fonction de leur indice, l'ensemble des exemplaires issus de la même catégorie générique (catégorie utilisée dans les questionnaires de typicité).
- L'utilisation de répétition de mot dans les productions, non = 0 ; oui = 1.
- L'utilisation de la synonymie dans les productions, non = 0 ; oui = 1. Cette fonction permet/interdit l'activation des synonymes d'un item préalablement activé.

Dans cette étude, les poids de connexions et les probabilités d'accès sont fixés à 1 et le seuil de compétition à 0,9. Nous avons appliqué la fonction « oubli » logarithmique ainsi que la fonction « régulatrice » de typicité catégorielle ; les répétitions ainsi que la synonymie ne sont pas autorisées.

Le logiciel produit, pour chacune des 30 fluences recueillies auprès des sujets sains (listes *sujets*), *n* listes *simulées*, dépendantes du poids d'activation des mots, et *n* listes *aléatoires*, indépendantes de ces poids. Ces listes *aléatoires* sont obtenues par tirage au sort d'items parmi les 334 répertoriés dans le système. Ce tirage s'effectue sans remise dans la mesure où les répétitions ne sont pas autorisées.

Les  $n$  listes *simulées* et *aléatoires* générées pour une liste *sujet* ont le même premier mot amorce correspondant au premier mot produit dans la liste *sujet* et comptent le même nombre de mots que la production de référence, soit une moyenne de 27,17 mots par liste. Nous avons fixé dans cette étude une production de 5 listes *simulées* et *aléatoires* pour chaque liste *sujets*.

### 2.3. Indices de comparaison

Les productions *sujets*, *simulées* et *aléatoires* ont été comparées en termes de :

- *Nombre de mots communs.* SysPro détermine pour chaque liste *simulée* et *aléatoire* le nombre de mots commun avec la liste *sujet* associée.
- *Fréquence d'utilisation des liens.* Le système détermine la liste des caractéristiques communes entre deux mots successifs et comptabilise ainsi la fréquence de ces caractéristiques dans les différentes productions.
- *Indices de Tröster et al. (1995).* Ces indices sont définis de la façon suivante. Les nombres de labels et d'exemplaires produits correspondent au nombre d'items produits dans chaque liste répertoriés dans l'arbre catégoriel en label ou exemplaire. Le nombre de catégories visitées est déterminé à partir des 20 catégories génériques de l'arbre et correspond au nombre de catégories nouvelles visitées dans chaque liste. Par exemples pour la série d'items « raisins », « pain », « orange », « boulangerie » le système comptabilisera 2 catégories visitées : « fruit » et « boulangerie ». Le nombre de changements de catégories correspond, quant à lui, au nombre de fois où il y a un changement de catégories entre deux items successifs. Ainsi dans l'exemple précédent, il y a 2 catégories visitées et 3 changements. Enfin, le nombre de mots produits par catégorie correspond au nombre de mots produits dans une liste, divisé par le nombre de catégories visitées.

### 2.4. Analyses statistiques

Afin de tester la validité du modèle, nous avons comparé les listes de mots produites par les *sujets*, les listes *simulées* et les listes *aléatoires*. Les analyses ont été effectuées sur 30 listes *sujets*, 150 *simulées* et 150 *aléatoires* ( $n=5$ ). Les listes *simulées* et les listes *aléatoires* ont été comparées dans un premier temps avec les listes *sujets* en termes de nombre de mots communs (test  $t$  de Student). Puis nous avons comparé ces trois types de production (*sujets*, *simulées*, *aléatoires*) en termes de fréquence des liens et d'exploitation catégorielle (indices de Tröster *et al.*, 1995). Pour cela, nous avons réalisé une analyse de variance multivariée (MANOVA) à 1 facteur « type de production » (à 3 modalités : *sujets*, *simulées*, *aléatoires*) avec pour variables dépendantes les 4 types de liens (syllabe de début, syllabe de fin, ressemblance phonémique et sémantique, comprenant les liens associatifs et conceptuels) et les 5 indices de Tröster (nombre de labels, nombre d'exemplaires, nombre de catégories visitées, nombre de mots par catégories visitées et nombre de changements de catégories). Des analyses de variance univariées (ANOVA) à 1 facteur, suivies de comparaisons de moyennes *a posteriori* (test de Scheffe) ont également été effectuées.

### 3. Résultats

#### 3.1. Nombre de mots communs

Les moyennes et les écart-types du nombre de mots communs entre productions *sujets* et *simulées* d'une part et productions *sujets* et *aléatoires* d'autre part, sont résumés dans le tableau 1. Le test t de Student indique que les productions *simulées* ont significativement plus de mots en commun avec les productions des sujets qu'avec les productions *aléatoires* ( $p < 0.001$ ).

**Tableau 1** — Nombre de mots communs (moyenne et écart-type) des productions simulées et aléatoires avec les productions sujets.

	<i>Productions simulées</i>	<i>Productions aléatoires</i>
Nb. moyen de mots communs avec les productions <i>sujets</i>	6.3	3.4
Écart type	± 4.0	± 2.4

#### 3.2. Fréquence des liens

Le tableau 2 présente la fréquence moyenne et l'écart-type des 4 liens possibles au sein des trois productions. L'analyse multivariée montre un effet significatif du facteur « type de production » sur l'ensemble des variables étudiées [ $F(2,327)=31.8$  ;  $p=0.0001$ ]. Les analyses univariées indiquent que cet effet concerne les liens de syllabe de début [ $F(2,327)=30.76$  ;  $p < 0.0001$ ], syllabe de fin [ $F(2,327)=23.36$  ;  $p < 0.0001$ ], ressemblance phonémique [ $F(2,327)=18.2$  ;  $p < 0.0001$ ] et traits sémantiques [ $F(2,327)=92.1$  ;  $p < 0.0001$ ].

**Tableau 2** — Fréquence (moyenne et écart-type) des différents liens pour les 3 types de productions.

	<i>Productions sujets</i>	<i>Productions simulées</i>	<i>Productions aléatoires</i>
Syllabe de début	2.0 (± 2.0)	1.4 (± 1.8)	0.3 (± 0.5)
Syllabe de fin	1.0 (± 1.5)	0.7 (± 0.9)	0.1 (± 0.4)
Ressemblance phonémique	1.6 (± 0.8)	1.54 (± 0.7)	1.16 (± 0.4)
Traits sémantiques	14.1 (± 6.7)	10.3 (± 7.3)	2.9 (± 2.0)

L'examen des moyennes révèle, de façon générale, l'existence d'une hiérarchie dans la fréquence des liens quel que soit le type de production considéré : les liens sémantiques sont les plus utilisés, suivi de la syllabe de début, de la ressemblance phonémique et enfin de la syllabe de fin. De plus, les productions *sujets* et *simulées* présentent des fréquences de liens plus élevées que les productions *aléatoires*, quel que soit les liens considérés.

Cet examen des moyennes, complété par un test de comparaison *a posteriori* (test de Scheffe), met en évidence des différences significatives entre :

- Les productions *sujets* et *aléatoires* pour les fréquences des liens de syllabe de début ( $p < 0.0001$ ), syllabe de fin ( $p < 0.0001$ ), ressemblance phonémique ( $p < 0.0001$ ) et sémantique ( $p < 0.0001$ ).
- Les productions *simulées* et *aléatoires* pour les fréquences des liens de syllabe de début ( $p < 0.0001$ ), syllabe de fin ( $p < 0.0001$ ), ressemblance phonémique ( $p = 0.0009$ ) et sémantique ( $p < 0.0001$ ).
- Les productions *sujets* et *simulées* uniquement pour la fréquence des liens sémantiques ( $p = 0.0034$ ) : les productions *sujets* ont une fréquence plus importante que les productions *simulées*.

### 3.3. Les indices de Tröster et al. (1995)

Le tableau 3 présente les résultats (moyenne et écart-type) des 5 indices. L'analyse multivariée montre un effet significatif du facteur « type de production » sur l'ensemble des variables étudiées ( $p = 0.0001$ ). Les analyses univariées indiquent que cet effet concerne les indices suivants : nombre de labels ( $p < 0.0001$ ), nombre d'exemplaires ( $p < 0.0001$ ), nombre de mots corrects par catégorie ( $p = 0.0029$ ) et nombre de changements de catégories ( $p < 0.0001$ ).

**Tableau 3** — Indices de Tröster et al. (1995) pour les 3 types de productions.

	<b>Productions <i>sujets</i></b>	<b>Productions <i>simulées</i></b>	<b>Productions <i>aléatoires</i></b>
Nb labels	6.8 ( $\pm$ 3.4)	16.7 ( $\pm$ 10.7)	5.1 ( $\pm$ 2.2)
Nb exemplaires	19.6 ( $\pm$ 9.0)	10.2 ( $\pm$ 9.6)	21.8 ( $\pm$ 7.2)
Nb catégories visitées	8.0 ( $\pm$ 1,5)	8.6 ( $\pm$ 2.5)	8.9 ( $\pm$ 1.4)
Nb mots corrects par catégorie	3.3 ( $\pm$ 1.0)	3.7 ( $\pm$ 2.7)	3.0 ( $\pm$ 0.7)
Nb changement catégories	14.0 ( $\pm$ 3.6)	20.5 ( $\pm$ 8.8)	23.4 ( $\pm$ 7.4)

Les comparaisons *a posteriori* (test de Scheffe) mettent en évidence une différence significative entre :

- Les productions *sujets* et *aléatoires* pour le nombre de changements de catégories ( $p < 0.0001$ ) : les changements de catégories sont beaucoup plus importants dans les productions *aléatoires* que dans les productions *sujets*.
- Les productions *sujets* et *simulées* pour le nombre de labels ( $p < 0.0001$ ), d'exemplaires ( $p < 0.0001$ ) et de changements de catégories ( $p = 0.0002$ ) : il y a beaucoup plus de labels et donc moins d'exemplaires et plus de changements de catégories dans les productions *simulées* que dans les productions *sujets*.
- Les productions *simulées* et *aléatoires* pour le nombre de labels ( $p < 0.0001$ ), d'exemplaires ( $p < 0.0001$ ), de mots corrects par catégorie ( $p = 0.003$ ) et de changements de catégories ( $p = 0.0056$ ) : il y a plus de labels et donc moins

d'exemplaires, moins de changements de catégories et donc plus de mots par catégories dans les productions *simulées* que dans les productions *aléatoires*.

## 4. Discussion

Le but de cette étude était de tester par simulation informatique les théories de la mémoire sémantique et du lexique mental à l'aide d'une tâche de fluence sémantique. La validité de ces théories a été évaluée à partir des liens phonologiques (phonémiques, syllabiques) et sémantiques (associatifs, conceptuels) existant entre les mots. Les principaux résultats ont montré des similitudes entre les productions *sujets* et *simulées* concernant le nombre de mots communs et la fréquence de liens phonologiques. Ils montrent cependant des différences concernant la fréquence de liens sémantiques et plus précisément l'exploitation du réseau catégoriel (indices de Tröster *et al.*, 1995).

Les fréquences des liens sémantiques et phonologiques des productions *sujets* et *simulées* sont supérieures à celles des productions *aléatoires* avec une fréquence de liens sémantiques plus élevée que celle des liens phonologiques. Cette prédominance peut s'expliquer par la nature de la tâche de fluence utilisée (consigne sémantique « supermarché ») (Raskin *et al.*, 1992 ; Troyer *et al.*, 1997 ; Abwender *et al.*, 2001). De plus, la fréquence des liens sémantiques est plus élevée pour les productions *sujets* que pour les productions *simulées*. Ces résultats ont été complétés par les indices de Tröster *et al.*, (1995) mettant en évidence une différence d'exploitation du réseau catégoriel entre ces deux types de productions. Ainsi, les productions *sujets* présentent le même nombre de labels et d'exemplaires que les productions *aléatoires* avec toutefois, un nombre de changements de catégories plus restreint. Ces productions *sujets* présentent également plus d'exemplaires, moins de labels et de changements de catégories que les productions *simulées*. L'ensemble de ces observations montre les limites des modalités de formalisation de l'organisation catégorielle utilisée dans notre système. Tout d'abord, le découpage proposé dans *SysPro* s'inspire du découpage arbitraire de Tröster *et al.* (1995). Même si celui-ci a été adapté aux supermarchés français, certaines de ces catégories peuvent ne pas être représentatives de celles utilisées par les sujets (ex. « snack-dessert »). D'autre part, dans ce découpage, un exemplaire ne peut appartenir qu'à une seule catégorie de l'arbre. Or, de récentes études ont montré qu'un même concept pouvait appartenir à plusieurs catégories (Barsalou, 1982 ; Murphy, 1993 ; Ross et Murphy, 1999). L'exemplaire « lait » peut, par exemple, appartenir aux catégories « produit laitier » et « boisson ». Ce phénomène, appelé « cross-classification », permet ainsi d'augmenter la flexibilité de l'arbre (changements de catégories) tout en restant aux niveaux inférieurs des exemplaires. Or, une des principales caractéristiques de notre système est qu'il exploite peu, comparé aux sujets, les niveaux inférieurs et privilégie davantage les changements de catégories aux niveaux supérieurs (les labels).

Ces observations peuvent être également la conséquence de l'application des deux règles d'exploitation définies dans le système. En effet, la règle intercatégorielle active pour chaque label produit, l'ensemble des labels de l'arbre quelle que soit leur distance, favorisant ainsi les changements de catégories. La règle intracatégorielle active, quant à elle, pour chaque exemplaire produit, l'ensemble des exemplaires de la même catégorie limitant ainsi la diffusion de l'activation à un petit nombre d'exemplaires. Une façon de moduler l'action de ces deux règles est d'y intégrer la notion de proximité sémantique développée par Markman et Winiewski (1997). Selon

ces auteurs, seuls les labels proches sémantiquement (distance minimale à parcourir dans l'arbre) s'auto-activent et l'activation d'un exemplaire induit l'activation de l'ensemble des exemplaires de sa catégorie ainsi que ceux des catégories voisines.

Concernant les liens phonologiques, les résultats montrent des fréquences similaires entre les productions *sujets* et *simulées*. Plus précisément, ces fréquences sont plus importantes pour les liens de syllabe de début que pour ceux issus des phonèmes ou de la syllabe de fin. Ces résultats confortent en partie les prédictions des modèles de l'accès au lexique et les résultats des expérimentations menées dans ce domaine (Dell, 1988 ; Meyer, 1990, 1991 ; Roelofs, 1996, 1997 ; Wheeldon et Levelt, 1995). En particulier, Dell (1988) a proposé un modèle de l'encodage phonologique constitué de deux réseaux interconnectés, l'un phonémique à traitement parallèle et l'autre syllabique à traitement séquentiel. L'ensemble des phonèmes d'un mot est donc activé dans un premier temps, mais la sérialité des traitements syllabiques impose un certain ordre dans leur récupération. Cet effet sériel rendrait compte d'un accès au mot par la syllabe de début plutôt que par ses représentations phonémiques ou par la syllabe de fin. Ces prédictions ont été vérifiées par Meyer (1990, 1991) en tâche de dénomination avec apprentissage de paires de mots. Cet auteur a mis en évidence un effet d'amorçage pour les mots partageant la première ou les deux premières syllabes et une absence d'effet pour les mots ne partageant que la dernière syllabe. De même, Meyer en 1991 a montré un effet d'amorçage plus important pour la première syllabe d'un mot que pour les phonèmes qui la constituent.

Nos résultats mettent en évidence ces mêmes effets dans les productions *sujets* et *simulées* avec des fréquences de liens de syllabe de début plus élevées que celles des liens phonémiques et syllabe de fin. L'intérêt de ces résultats est qu'ici, la tâche de production utilisée est une tâche de fluence qui ne nécessite aucun apprentissage, favorisant ainsi un accès direct à l'organisation interne des mots en mémoire.

L'ensemble des résultats conforte donc les théories du lexique mental postulant une organisation des mots en mémoire sous forme de réseaux à plusieurs unités lexicales. Parmi ces unités, la simulation a montré la pertinence des liens phonologiques dans la récupération des mots en mémoire et souligne l'intérêt des modèles de l'accès au lexique qui posent la syllabe de début comme unité principale de récupération. Pour la mémoire sémantique, notre étude ne permet pas actuellement de valider l'hypothèse d'une organisation catégorielle hiérarchisée. Cependant, les résultats ont permis de mettre en évidence différents points à modifier dans la formalisation des données afin d'obtenir un arbre rendant compte au mieux des propriétés du réseau à tester. La comparaison des résultats issus de ces futures simulations avec ceux déjà obtenus dans cette étude, permettra de tester la pertinence des modifications apportées.

Il est également important de noter que le logiciel offre plusieurs propriétés fonctionnelles permettant de tester plus en détail chacune des théories impliquées. Nous pouvons par exemple, faire varier la probabilité d'utilisation des liens et tester ainsi leur degré d'implication dans la production de mots. Cette manipulation permettra de vérifier les modèles de l'accès lexical postulant ou non une interaction entre les traitements sémantiques et phonologiques des mots : modèles discrets, interactifs ou en cascade de l'accès au lexique (Caramazza, 1997 ; Dell, 1986, 1988 ; Humphreys *et al.*, 1988, 1995 ; Levelt, 1999 ; Levelt *et al.*, 1999 ; Roelofs, 1997). De même, le système permet à l'expérimentateur de tester différentes fonctions telles

que la mise en mémoire à court terme (fonction de la courbe en U, perte linéaire) et la typicité catégorielle (régulatrice des poids d'activation ou fonction étendue).

Cette étude met alors en avant l'intérêt de développer la modélisation et la simulation informatique comme outil de validation de certains concepts et théories issus de domaines aussi vastes que la Psychologie Cognitive et l'Intelligence Artificielle (architectures symboliques ou connexionnistes rendant compte au mieux des théories étudiées).

## Annexes

### Annexe1 : Extrait des productions verbales des sujets

**Sujet 1** : jambon, légume, viande, produit ménager, vêtement, chaussure, chaussette, disque, bouquin, stylo, cahier, bouteille, shampoing, crème, gel douche, savon, lessive, bonbon, gâteau, huile, vinaigre, sucre, sel, poivre, épice, médicament, pansement, brosse, peigne, dentifrice, programme télé.

**Sujet 2** : yaourt, fruit, jambon, viande, coton, produit de toilette, mousse à raser, rasoir, dentifrice, brosse à dent, chaussure, vêtement, fruit, légume, lait, feuille, crayon, cahier, table, chaise, boîte de conserve, petit pois, haricot, tomate, jouet, bonbon, gâteau, pain.

**Sujet 3** : légume, patate, petit pois, carotte, haricot vert, fleur, conserve, poisson, viande, aliment pour animaux, vêtement, dessous, vaisselle, verre, couvert, accessoire pour automobile, ampoule, pile, cassette vidéo, cassette audio, CD, appareil photo, chaîne hi-fi, télévision, meuble, rideau, tringle pour rideau, produit d'entretien, produit de toilette, gel douche, dentifrice, lait, eau, bière, vin.

### Annexe 2 : Découpage catégoriel du système

#### Alimentaire

Découpage Tröster <i>et al.</i>	Découpage utilisé dans l'étude		
	Labels	Exemplaires	
Catégories	Cat. Niveau 1	Cat. Niveau 2	
Fruit	Fruit	Fruit frais Fruit Sec	<i>Banane</i> <i>Raisin sec</i>
Légume	Légume	Légume frais Légume sec	<i>Poireau</i> <i>Lentille</i>
Viande / Poisson	Viande  Poisson	Boucherie Charcuterie Volaille Poisson Crustacé	<i>Bœuf</i> <i>Saucisson</i> <i>Poulet</i> <i>Thon</i> <i>Moule</i>
Féculent	Féculent		<i>Pâtes, Céréales</i>
Produit laitier	Produit laitiers		<i>Pâtes, Céréales</i>
Boisson	Boisson	Boisson alcoolisée Bois. non alcoolisée	<i>Vin</i> <i>Limonade</i>
Snack/dessert	Plat préparé  Boulangerie Sucrierie	Boîte de conserve Produit surgelé	<i>Maïs en boîte</i> <i>Haricots surgelés</i> <i>Pain</i> <i>Bonbon</i>
Condiment	Condiment		<i>Sel, Poivre</i>

## Non alimentaire

Découpage Tröster <i>et al.</i>	Découpage utilisé dans l'étude		Exemplaires
	Labels		
<i>Catégories</i>	<i>Cat. Niveau 1</i>	<i>Cat. Niveau 2</i>	
Objets personnels (hygiène, loisirs)	Hygiène  Vêtement Audio-visuel Librairie presse	Toilette Parapharmacie	<i>Brosse à dent</i> <i>Pansement</i> <i>Jupe</i> <i>Chaîne hi-fi</i> <i>Magazine</i>
Autres objets (ménager, entretien)	Produit ménager Droguerie Ustensile cuisine Blanc textile Bricolage		<i>Javel</i> <i>Allumette</i> <i>Fourchette</i> <i>Drap</i> <i>Marteau</i>

## Références bibliographiques

Abwender D.A., Swan J.G., Bowerman J.T., Connolly S.W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*. 8, 323-336.

Barsalou L.W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition*. 10, 82-93.

Bybee J.L. (1988). Morphology as lexical organization. In Hammond. M., Noonan. M. (éds.), *Theoretical Morphology*. Academic Press, Inc. 7, 119-141.

Caramazza A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology*. 14, 177-208.

Caron J. (1995). Précis de psycholinguistique. 3<sup>e</sup> édition. Paris : P.U.F.

Collins A.M., Quillian M.R. (1969). Retrieval Time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*. 8, 240-248.

Cordier F. (1993). Les représentations cognitives privilégiées : typicalité et niveau de base. Presse Universitaire de Lille.

Dell G.S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review*. 93 (3) 283-321.

Dell G.S. (1988). The retrieval of phonological forms in production: tests of predictions from a connectionist model. *Journal of Memory and Language*. 27, 124-142.

Dell G.S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and cognitive processes*. 5, 313-349.

Glanzer M., Cunitz A.R., (1966). Two storage mechanisms in free recall. *Learning and Verbal Behavior*. 5, 351-360.

Griffin Z.M., Bock K., (1998). Constraint, word frequency and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*. 38, 313-338.

Humphreys G.W., Riddoch M.J., Quinlan P.T. (1988). Cascade processes in picture identification. *Cognitive Neuropsychology*. 5, 67-103.

Humphreys G.W., Lamote C, Lloyd-Jones T.J. (1995). An interactive activation approach to object processing: effects of structural similarity, name frequency, and task in normality and pathology. *Memory*. 3, 535-586.

- Jescheniak J.D., Levelt W.J.M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology*. 20, 824-843.
- Levelt W.J.M. (1999). Models of word production. *Trends in Cognitive Sciences*. 3(6) 223-232.
- Levelt W.J.M., Roelofs A, Meyer A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain sciences*. 22, 1-75.
- Markman A.B., Wisniewski E.J. (1997). Similar and different: The Differentiation of basic-level categories. *Journal Experimental Psychology: Learning, Memory and Cognition*. 1, 54-70.
- Mattis S. (1988). Dementia Rating Scale. Psychological Assessment Resources. Odessa, FL.
- Meyer A.S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language*. 29, 524-545.
- Meyer A.S. (1991). The time course of phonological encoding in language production: phonological encoding inside a syllable. *Journal of Memory and Language*. 30, 69-89.
- Murphy G.L. (1993). A rational theory of concepts. In Nakamura G.V., Taraban R.M., Medin D.L. (éds), *The psychology of learning and motivation — Categorization by humans and machines*. New York : Academic Press. Vol. 29, 327-359
- Raskin S.A., Sliwinski M., Borod J.C. (1992). Clustering strategies on tasks of verbal fluency in Parkinson's disease. *Neuropsychologia*. 30(1) 95-99.
- Roelofs A. (1996). Serial order in planning the production of successive morphemes of word. *Journal of Memory and Language*. 35, 854-876.
- Roelofs A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*. 64, 249-284.
- Rosch E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology : General*. 104, 328-350.
- Rosch E., Mervis C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*. 3, 192-233.
- Ross H., Murphy G.L. (1999). Food for thought: Cross-classification and Category Organization in a Complex Real-World Domain. *Cognitive Psychology*. 38, 495-553.
- Tröster A.I., Warmflash V., Osorio I., Paolo A.M., Alexander L.J., Barr W.B. (1995). The roles of semantic networks and search efficiency in verbal fluency performance in intractable temporal lobe epilepsy. *Epilepsy Research*. 21, 19-26.
- Troyer A.K., Moscovitch M., Winocur G., (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*. 11, 138-146.
- Troyer A.K. (2000). Normative data for clustering and switching on verbal fluency task. *Journal of Clinical and Experimental Neuropsychology*. 22(3) 370-378.
- Warrant L. (1996). Orthographe et prononciation en français. Duculot, coll. Entre guillemets.
- Wheeldon L.R., Levelt W.J.M. (1995). Monitoring in the time course of phonological encoding, *Journal of Memory and Language*. 34, 311-334.