

Le chapitre 4 introduit les différentes mesures utilisées pour évaluer les performances des systèmes. Au-delà des définitions, ce chapitre met en évidence le bruit inhérent à ces mesures ; à l'heure actuelle, les performances absolues n'ont pas un grand sens : seules les performances relatives sont importantes.

Le chapitre 5 montre comment les textes sont transformés pour pouvoir être utilisés par les méthodes d'apprentissage numérique. Ce chapitre met en évidence la nécessité et la difficulté d'effectuer une sélection de descripteurs. Nous proposons une méthode entièrement automatique en deux étapes. La première étape détermine le vocabulaire spécifique des documents pertinents par rapport à l'ensemble du corpus ; la deuxième étape est une procédure d'orthogonalisation selon la méthode de Gram-Schmidt qui présente l'avantage d'être adaptée à la classification.

Le chapitre 6 est une présentation succincte des réseaux de neurones. Ce chapitre insiste sur la notion de surapprentissage pour les problèmes de classification, et montre que les méthodes de régularisation comme le *weight decay* apportent une solution à ce problème tout en ajoutant de nouveaux paramètres appelés hyperparamètres. Ces hyperparamètres peuvent être théoriquement déterminés grâce à l'approche bayésienne qui est présentée avec les approximations nécessaires à sa mise en œuvre.

Le chapitre 7 présente les premières expériences effectuées sur le corpus Reuters ainsi que la description de notre participation à TREC-8. Ce chapitre étudie l'impact des différents paramètres intervenant dans la sélection des descripteurs sur les performances (nombre de documents non pertinents, choix de ces documents, nombre de descripteurs initiaux). L'étude du nombre optimal de neurones cachés montre qu'il est nécessaire d'améliorer la représentation des textes avant de complexifier l'architecture des réseaux de neurones en ajoutant des neurones cachés.

Le chapitre 8 présente une méthode originale pour déterminer automatiquement le contexte caractéristique d'un mot pour effectuer une désambiguïsation partielle de ce mot. L'architecture neuronale est modifiée pour prendre en considération cette nouvelle représentation. Avec celle-ci, l'utilisation d'une méthode de régularisation est indispensable ; malheureusement les résultats de l'approche bayésienne n'ont pas permis d'obtenir des résultats satisfaisants pour la détermination des hyperparamètres. Finalement les résultats obtenus sur le corpus Reuters et sur le corpus de TREC-8, montrent une amélioration notable des résultats par rapport au chapitre 7. Ce chapitre se termine par la description de notre participation à TREC-9.

Enfin, le chapitre 9 montre comment ces résultats sont intégrés dans une application existante de filtrage de dépêches de l'AFP en temps réel. Les méthodes d'apprentissage numérique permettent de proposer de nouvelles fonctionnalités à cette application. Une première application utilise la sortie d'un filtre construit sur des méthodes à base de règles conjointement avec la sortie d'un filtre neuronal pour surveiller le premier filtre. Une deuxième application utilise un moteur de recherche couplé aux méthodes neuronales pour permettre à un utilisateur de définir son propre filtre.

AMORÇAGE PERCEPTIF VERBAL RÔLE DE LA CONGRUENCE SYLLABIQUE

Sandrine BELIER

Laboratoire d'Étude des Mécanismes Cognitifs - Université Lumière Lyon 2

Directeur de recherche : Professeur Olivier KOENIG

5, avenue Pierre Mendès France – F-69676 Bron cedex, France

Mel : sandrine.belier@univ-lyon2.fr

Mots clés : *Amorçage perceptif, recouvrement perceptif, identification perceptive, complètement de trigrammes, structure syllabique, compétition lexicale, asymétrie hémisphérique fonctionnelle.*

Au cours de ces quinze dernières années, la mémoire implicite a suscité un intérêt expérimental considérable dans le cadre des études en psychologie cognitive. Dans ce travail de thèse, nous nous sommes plus particulièrement intéressés à l'amorçage, phénomène de récupération implicite, automatique de l'information en mémoire. L'effet d'amorçage s'exprime par une influence, sur le traitement d'un stimulus (*cible*), résultant du traitement antérieur du même stimulus ou d'un stimulus lié au premier (*amorçe*).

Les principaux objectifs de ce travail étaient :

- 1) d'étudier le rôle de l'étendue du recouvrement perceptif entre amorçe et cible sur l'effet d'amorçage dans des tâches de complètement de bi- et trigrammes et d'identification perceptive (expériences 1, 2, 3, 6),
- 2) d'évaluer l'effet, sur l'amorçage, de la structure syllabique des mots utilisés dans ces mêmes tâches (expériences 1, 2, 4, 5, 6).

Ces tâches sont des outils utilisés dans le domaine de la neuropsychologie cognitive afin d'évaluer l'intégrité de la mémoire implicite chez des patients ayant des troubles de mémoire. Leur utilisation s'est avérée critique dans ce domaine, puisque ces tâches ont permis de mettre en évidence, dans les années 1980, une intégrité de la mémoire implicite chez les patients amnésiques. La tâche de complètement de trigrammes est celle dont nous nous sommes plus

particulièrement servis dans notre recherche (i.e., cinq de nos six expériences l'utilisent). Elle se déroule en deux parties. Dans une première partie, on demande au participant d'effectuer un traitement sur des mots (e.g., sur le mot GARAGE), comme compter le nombre de voyelles. Dans une seconde partie, appelée phase de test, mais qui pour le participant n'a aucune relation avec la première partie (i.e., on trouve ici le caractère implicite, inconscient du test), on présente au participant des trigrammes (i.e., groupes de trois lettres), dont la moitié correspond aux trois premières lettres des mots vus lors de la première partie (e.g., GAR). La tâche du participant est alors de compléter ces trigrammes par le premier mot qui lui vient à l'esprit. L'amorçage se manifeste par une forte tendance, de la part des participants, à compléter les trigrammes par les mots rencontrés lors de la première partie, plutôt que par d'autres mots (jamais rencontrés) (e.g., GARDIEN). Dans la tâche d'identification perceptive, on présente au participant un mot pendant une durée très brève (de l'ordre de 30 millisecondes) sur un écran d'ordinateur. La tâche non aisée du participant est alors d'essayer d'identifier le mot. L'identification du mot est plus ou moins facilitée selon le type d'amorce qui le précède.

Les objectifs de cette recherche de thèse nous ont inévitablement amenés à nous intéresser à la littérature concernant deux grands domaines de recherche de la psychologie cognitive sans relation directe apparente, à savoir non seulement la littérature sur la mémoire et plus spécialement sur l'amorçage, mais également à la littérature sur la syllabe et son rôle dans le traitement langagier.

Concernant le premier objectif, les expériences 1 et 2 ont souligné le rôle du recouvrement perceptif amorce-cible, puisque plus d'amorçage était obtenu pour les trigrammes (i.e., présentation des trois premières lettres d'un mot-amorce) que pour les bigrammes (i.e., présentation des deux premières lettres d'un mot-amorce). Ce résultat était compatible avec les recherches montrant que l'amorçage est sensible au partage perceptif entre amorce et cible. Cependant, dans ces expériences, si les bigrammes comportaient moins d'information que les trigrammes d'un point de vue perceptif, ils pouvaient également être complétés par davantage de mots que les trigrammes, ce qui pouvait engendrer une baisse de l'effet d'amorçage compte tenu du plus grand nombre de candidats lexicaux compatibles dans ce cas. Lorsque le nombre d'entrées lexicales commençant par les lettres proposées était équivalent pour les bi- et trigrammes (expérience 3), l'amorçage ne différait pas selon l'ampleur du recouvrement perceptif, suggérant que la quantité d'information disponible au test joue un moins grand rôle sur l'amorçage que la compétition lexicale. Plus la compétition lexicale est forte, moins le mot-amorce a de chance d'être préférentiellement réactivé parmi la cohorte de candidats compatibles de fréquence plus ou moins grande par rapport à celle du mot-amorce. L'expérience 6, utilisant l'identification perceptive, a révélé que l'étendue du recouvrement perceptif amorce-cible ne jouait pas de rôle sur l'amorçage.

Concernant le second objectif, les expériences 1, 2, et 4 ont révélé un plus fort amorçage lorsque l'information-test (bigramme ou trigramme) correspondait à la première syllabe du mot-amorce que lorsque ce n'était pas le cas. Par exemple, lorsque le trigramme MAR était proposé au test, plus d'amorçage était obtenu pour le mot-amorce MAR-TEAU dont la première syllabe correspondait au trigramme que pour le mot-amorce MA-RAIS dont la première syllabe ne correspondait pas au trigramme, et inversement lorsque le bigramme MA était proposé au test. L'expérience 4, impliquant une présentation latéralisée des stimuli visant à démontrer le rôle différentiel des hémisphères cérébraux dans une tâche de complèvement de trigrammes, a montré que seul l'hémisphère gauche était sensible à la correspondance syllabique entre amorce et cible, et donc susceptible de représenter les mots sous une forme syllabifiée (i.e., en fonction de leur structure syllabique). L'expérience 5, utilisant le complèvement de trigrammes en modalités croisées avec division de l'attention lors de l'encodage, a révélé que le code orthographique était à l'origine de l'effet de la structure syllabique sur l'amorçage. Ce résultat suggère que les mots écrits sont représentés dans le lexique mental (orthographique) en termes d'unités syllabiques orthographiques. Finalement, l'expérience 6 a montré que l'identification d'un mot-cible était facilitée par la présentation antérieure très brève (33 millisecondes) d'une amorce masquée correspondant à sa première syllabe (MA suivi de MA-RAIS), par rapport à la présentation antérieure d'une amorce masquée correspondant à sa première syllabe plus ou moins une lettre (MAR suivi de MA-RAIS). Ce résultat, issu d'une expérience utilisant une nouvelle tâche, souligne une nouvelle fois l'importance de la structure syllabique du matériel verbal dans les tâches d'amorçage perceptif. La syllabe constituerait ainsi une unité perceptive importante dans le traitement des mots écrits chez le sujet de langue maternelle française. Globalement, l'ensemble de ces résultats suggère que les unités syllabiques pourraient constituer une forme possible de représentations des mots en mémoire et que de telles représentations seraient plus particulièrement représentées dans l'hémisphère gauche.