

Evaluation de l'effet de l'âge sur la vitesse d'accès lexical dans une population âgée de 45 ans à 60 ans

Auteur : Albin, Sophie

Promoteur(s) : Poncelet, Martine

Faculté : Faculté de Psychologie, Logopédie et Sciences de l'Éducation

Diplôme : Master en logopédie, à finalité spécialisée en neuropsychologie du langage et troubles des apprentissages verbaux

Année académique : 2019-2020

URI/URL : <http://hdl.handle.net/2268.2/10846>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

**EVALUATION DE L'EFFET DE L'ÂGE SUR LA VITESSE
D'ACCÈS LEXICAL
DANS UNE POPULATION ÂGÉE DE 45 À 60 ANS**

Mémoire présenté en vue de l'obtention du grade de master en logopédie

Sous la direction de

Martine PONCELET

Par

Sophie ALBIN

Toute ma reconnaissance va ici au professeur Martine Poncelet qui, avec enthousiasme, a accepté de patronner ce mémoire et, avec bienveillance, en a adapté le sujet.

Puisse cette approche théorique être utile à une réplication ultérieure.

INTRODUCTION GÉNÉRALE

Le manque du mot a fait l'objet de nombreuses études ces dernières années, que ce soit en contexte pathologique (e.g., aphasie) ou chez des personnes tout-venant afin de mieux comprendre le processus et les mécanismes sous-jacents. Le manque du mot est la plainte la plus fréquente chez les personnes plus âgées (voir Verhaegen & Poncelet, 2013). C'est pourquoi, des études se sont particulièrement intéressées à l'effet de l'âge sur cet échec de récupération. Outre les définitions et les caractéristiques du manque du mot, plusieurs hypothèses ont fait l'objet d'étude quant au moment de son apparition et quant à son origine.

Ce mémoire entreprendra, dans un premier temps, de définir ce que les auteurs entendent par manque du mot, quelles sont les caractéristiques et ce qui est utilisé jusqu'à présent pour l'évaluer. Une brève distinction avec d'autres phénomènes d'échec de récupération sera également explicitée.

Sera examiné plus particulièrement la problématique de l'âge. En effet, plusieurs études se sont penchées sur la question en utilisant notamment la tâche de dénomination orale d'images. Le travail présent fera état de ces différentes études et de leurs conclusions qui ont alimenté la problématique.

Puisque ce mémoire a pour cadre la vitesse d'accès lexical, la production orale de mots sera étudiée en exposant les différents modèles qui ont pour objectif de décrire les processus et les étapes de niveau de traitement. Cela permettra également d'exposer les théories mises en avant pour expliquer l'apparition des échecs de récupération, manque du mot ou MBL selon les études.

En 1868, Donders, considéré comme le fondateur de la chronométrie mentale (Duyck, & al., 2008), a écrit un article dans lequel apparaît la notion de *reaction time* (RT). Depuis, les mesures de ce temps de réaction sont utilisées pour étudier davantage les processus cognitifs. La latence a pris dès lors une place essentielle dans la littérature. Elle sera examinée dans la tâche de dénomination d'images, qui sera le sujet central de ce travail, ainsi que ce qui pourrait l'influencer dans la conception de l'étude elle-même et les moyens proposés dans la littérature pour la mesurer.

I. INTRODUCTION THÉORIQUE

1. Manque du mot et son évaluation

a) Le manque du mot : définition et caractéristiques

Le manque du mot ou anomie (en anglais, *word-finding difficulty*) consiste en une difficulté, voire une impossibilité momentanée à récupérer et à produire le mot recherché. Il se retrouve aussi bien dans des pathologies développementales (ex. : dysphasie) ou dégénératives (ex. : Alzheimer) qu'acquises (ex. : aphasie). Par conséquent, sont concernés aussi bien les enfants que les adultes. Ce manque du mot, appelé aussi « trouble de la dénomination » par Bogliotti (2012), ne concerne que le versant expressif du langage dans toutes les modalités (langage spontané, répétition, lecture à voix haute, langage induit comme en dénomination, écriture sous dictée), ce qui implique que « les patients présentant un manque du mot sont capables d'identifier un référent face à divers exemplaires (désignation) mais ont des difficultés à produire le mot cible face à un dessin, une photo (dénomination) ou en production spontanée » (Bragard, 2010, p. 113). Les auteurs ont relevé les caractéristiques qui peuvent indiquer la présence de ce phénomène. Ils rapportent notamment (Bogliotti, 2012 ; Bragard, 2010 ; De Partz & Pillon, 2014) :

- des hésitations
- des pauses - plus longues que la normale- vides ou remplies (ex.: hum, euh)
- des productions de mots « vides de sens » (« chose », « machin », « truc », « bidule »)
- des répétitions de mots ou des conduites d'approche (essais répétés pour produire la cible)
- des circonlocutions ou des périphrases
- des omissions ou des latences, voire des absences, d'évocation
- des substitutions ou paraphrasies qui peuvent être sémantiques ou phonologiques

Le manque du mot est à distinguer d'autres échecs de recouvrement. Ce terme « manque du mot » sera employé dans le cas où l'échec de récupération d'un mot se produit avec une fréquence excessive et pour des mots qu'on considérera comme de haute fréquence (Truffert,

2014) et dans le cas où les personnes sont atteintes d'un trouble neurologique (contexte pathologique).

b) Autres phénomènes : le « mot sur le bout de la langue » et le « Feeling of Knowing »

Un autre phénomène universel similaire qui est bien connu et largement étudié est le « mot sur le bout de la langue » (MBL ou ToT en anglais pour « Tip of the Tongue »). Il s'agit d'un échec momentané de récupération chez les personnes tout-venant. Bonin (2007, p.61) le définit comme une « inaccessibilité ponctuelle à la forme d'un mot alors que certaines informations sémantiques et syntaxiques sont disponibles » mais certains auteurs, comme Dorot et Mathey (2013, p.2), précisent que le MBL « se produit quand un locuteur est momentanément incapable de produire un mot, en dépit d'un fort sentiment de connaissance de ce mot ». En effet, Brown et MacNeill (1966 ; cités par Brown, 2012) explique le MBL, dans l'ensemble des échecs de récupération, soit sur un continuum allant d'une absence de connaissance à un fort sentiment d'accessibilité à cette connaissance qui caractériserait le MBL ou soit en dichotomie où le MBL représenterait un sentiment unique d'accessibilité et de certitude sur l'information à récupérer. Ce phénomène, dès lors, n'est pas à confondre avec le « Feeling of Knowing » (FoK) qui est semblable au MBL en ce qui concerne l'impossibilité à accéder à l'information mais qui est différent dans le sens où il n'y a aucune certitude de récupérer l'information pour le FoK. Le manque du mot se différencie de ces notions par l'absence de ressentis subjectifs. La distinction clairement établie, la revue des études suivantes prendra en considération à la fois le MBL ou le manque du mot en fonction de l'objet des études mentionnées.

c) Méthodes d'observation et d'évaluation dans la littérature

L'étude du MBL ou du manque du mot revêt une importance essentielle pour mieux comprendre le processus de la production du langage, comme la nature des représentations mobilisées, ainsi que pour concevoir les mécanismes qui expliqueraient, voire favoriseraient, leur apparition. C'est pourquoi, plusieurs méthodes ont été mises en œuvre dans la littérature pour récolter des données probantes. Le MBL peut être évalué par des observations « naturelles », c'est-à-dire par un carnet de bord ou un questionnaire rempli par le participant. Par exemple, Burke et al. (1991 ; cité par Bonin, 2007) ont proposé à leurs sujets d'âges

différents des questions variées sur le MBL par rapport à leurs activités quotidiennes. Ils ont pu de cette manière avoir des informations sur le phénomène étudié. Bonin (2007, p.27) en rapporte quelques-unes :

- les états de MBL sont plus fréquents chez les personnes plus âgées que chez les personnes plus jeunes
- certains mots sont plus enclins à provoquer un état de MBL que d'autres : les noms propres sont plus sujets aux MBL, ainsi que les mots de basse fréquence. On peut ajouter également à cette liste les mots abstraits (Schwartz & Metcalfe, 2014)
- en état de MBL, les individus sont capables de rapporter des informations partielles sur le mot qu'ils ne peuvent produire, en particulier les informations syntaxiques

Ces éléments sont complétés par Dorot et Mathey (2013) qui soulignent que les personnes plus âgées restitueraient moins d'informations phonologiques sur la cible manquante et moins de mots alternatifs, à savoir des mots liés mais considérés incorrects par rapport à la cible, que les personnes plus jeunes.

La variable « âge » a pris une place centrale dans les recherches et dans les débats sur ces phénomènes (MBL ou manque du mot). Nous y reviendrons un peu plus tard.

Outre ces moyens, le MBL peut être évalué par des études expérimentales. En effet, le MBL, après sa première apparition dans la littérature (James, 1893 ; cité par Schwartz & Metcalfe, 2014), a fait l'objet d'une première étude expérimentale par Brown et MacNeill (1986 ; cité par Schwartz & Metcalfe) qui met en place une tâche de récupération d'un mot rare à partir de sa définition. Les participants devaient identifier s'ils étaient confrontés à un MBL et, dans le cas positif, donner des informations sur le mot non retrouvé (par exemple, la lettre initiale ou un mot qui lui serait lié). Une vérification de consensus entre le mot pensé par le sujet et celui conçu par l'expérimentateur a été aussi prise en compte. Cette étude est considérée comme pionnière dans le domaine.

Depuis, les études suivantes ont continué abondamment à explorer le sujet. Les tâches utilisées pour évaluer le MBL ou le manque du mot peuvent être appréciées en contexte de discours (appelé « connected speech » par Kavé & Goral, 2017) ou en contexte isolé, à savoir la production de mots simples, comme dans une dénomination sur base de définition, une

dénomination d'images ou une épreuve de fluence verbale. Ces tâches expérimentales, davantage en mots isolés, ont l'avantage d'induire le MBL et elles peuvent évidemment être combinées avec une méthodologie plus écologique (questionnaires, carnet de bord). Par ailleurs, les épreuves demandant la récupération d'un mot en isolé ont l'avantage d'être hors contexte. En effet, selon Kavé et Goral (2017), le recouvrement d'un mot simple demande un appariement exact entre le niveau sémantique et le code phonologique, tandis que la récupération en contexte est plus flexible, c'est-à-dire que la personne peut utiliser d'autres mots pour remplacer le mot cible. De plus, dans ce cas de figure, la personne peut s'aider du contexte. Les tâches de dénomination (sur définitions ou sur images) évaluent le recouvrement des entrées lexicales d'une cible spécifique et prédéfinie par l'expérimentateur. La fluence verbale, quant à elle, permet, elle aussi, une plus grande variabilité en terme de choix de mots, contraint la récupération selon une règle donnée et selon un temps délimité et, néanmoins, est fortement liée aux fonctions exécutives. La tâche de dénomination d'images, épreuve canonique dans la littérature, fera l'objet d'une description plus détaillée dans la suite de ce mémoire.

2. La tâche de dénomination : la problématique de l'âge

La tâche de dénomination sur images, qui est largement utilisée dans l'étude de la production orale, consiste à demander au sujet de dénommer l'image qui lui est présentée le plus rapidement possible. Elle permet aussi à l'expérimentateur de contrôler, autant que faire se peut, les facteurs qui pourraient influencer la performance, comme nous le verrons dans le chapitre suivant. La tâche de dénomination permet d'évaluer la réponse donnée, réponse correcte ou non (précision), qui donne des informations sur la sélection en mémoire, et le temps de récupération de la cible (vitesse, c'est-à-dire temps de latence). Les deux n'ont pas toujours été pris en considération et la dernière s'est avérée cruciale pour la sensibilité de l'épreuve notamment face à l'effet de l'âge sur les performances des sujets. C'est pourquoi, le débat où il est question de l'âge, pour lequel une diminution des performances en dénomination s'observe, a pris une place importante dans la littérature.

En 1994, Goulet et al. (cités par Verhaegen & Poncelet, 2013) ont mentionné, dans les études revues, des conclusions contradictoires sur le sujet en raison de plusieurs facteurs liés à la conception de l'étude : par exemple, l'âge des groupes qui diffère d'une étude à l'autre ou

le type de matériel utilisé (notamment des items avec une familiarité trop haute). C'est pourquoi, les auteurs ont conclu que les études sur la précision en dénomination ne fournissent aucune preuve solide que les difficultés de récupération des mots sont liées à l'âge (Goulet & al., cités par Feyereisen, 1997).

Dans son article, Feyereisen (1997), lui, réexaminant les études passées en revue par Goulet et al. (*op. cit.*) par une procédure statistique méta-analytique (décrite par Rosenthal, 1995, et Hedges & Olkin, 1985, cités par Feyereisen), a constaté par contre un déclin significatif lié à l'âge qui commencerait après 70 ans concernant la précision en dénomination. Toutefois, l'auteur souligne également que cet effet peut différer selon les études de leur sélection : deux d'entre elle obtiennent un effet très large, alors que trois autres obtiennent un effet modéré. Feyereisen l'explique soit par la sélection des items proprement dite ou par les différences dans les échantillons des participants (état de santé, niveau d'éducation, niveau socio-économique ou encore (ancienne) profession).

Dans leur étude, Connor et al. (2004) se sont fondés sur des données longitudinales. Ils ont établi quatre groupes de 236 participants d'âge différents (30-39, 50-59, 60-69 et 70-80 ans) en contrôlant les variables liées aux individus (santé, genre, éducation). Les auteurs – dans la lignée de Nicholas et al. (1998, cités par Verhaegen & Poncelet, 2013) - ont relevé de légers signes allant vers une diminution de la précision chez les personnes âgées de 50 ans mais ils ont souligné que ce déclin est plus affirmé chez celles de 70 ans. Les auteurs ont aussi constaté que l'éducation affecte la dénomination et que le genre, bien que les changements dans les capacités de dénomination n'en sont pas dépendant, interagit avec l'éducation. Le genre et l'éducation sont donc tous les deux un prédicteur de performance. Les auteurs ne donnent pas d'explications dans leur étude et en appellent à d'autres recherches ultérieures. Toutefois, ils ne remarquent aucune interaction entre l'âge et l'éducation.

En 2007, Goral et al. ont mis en place une étude longitudinale avec des participants âgés de 30 à 94 ans. Ils ont démontré différents profils de déclin lié à l'âge au niveau de la récupération lexicale. En effet, ils ont fait passer deux types de tâches, celles qui demande de récupérer un seul item précis (deux tâches de dénomination) et celles qui permettent de récupérer divers items (*list-generation tests* évaluant les fluences sémantiques et phonémiques). Ils constatent que le vocabulaire ne diminue pas, il augmente même avec l'âge. De leurs résultats, ils obtiennent un déclin non linéaire, c'est-à-dire qu'il est plus rapide en augmentant en âge. Les tâches de dénomination se montrent plus sensibles que les autres épreuves car elles

demandent de récupérer un seul item. Les performances dépendent donc de la demande de la tâche de récupération. Contrairement à l'étude précédente, en dénomination, un effet significatif pour le genre s'est révélé mais une interaction entre le genre et l'éducation est également présente. De plus, les auteurs ne rencontrent pas en dénomination un effet d'éducation isolé – contrairement aux autres épreuves –, ce qui, selon eux, est possible puisque les résultats des études antérieures ne sont pas constants. Ils expliquent cette conclusion par la tâche de dénomination qui demande un appariement strict comparé aux autres et laisse peu de place à une compensation.

Les études mentionnées se sont centrées sur la précision en dénomination (taux de réponses correctes). Concernant la latence en dénomination d'images, quelques rares études s'y sont intéressées. Une des premières d'entre elles est celle de Thomas et al. en 1977. Les auteurs avaient plusieurs objectifs : premièrement, évaluer si un effet d'âge éventuel est présent dans la vitesse de récupération ; deuxièmement, savoir si cette différence peut être minimisée par une pratique et/ou une dénomination avec indice (mots écrits présentés avant). A cette fin, les auteurs ont recruté 60 participants de sexe masculin homme (aucune femme dans l'échantillon) en bonne santé venant de différentes professions répartis en cinq groupes (25-35, 36-45, 46-55, 56-65, 65+ ans). Cette expérience s'intègre dans un projet d'étude interdisciplinaire longitudinale. Les résultats obtenus de cette tâche de dénomination d'images (images coloriées) démontrent un effet d'âge significatif qui est atténué par une diminution de la difficulté de la tâche au moyen d'une pratique continue ou d'un amorçage. Les auteurs attribueraient beaucoup de ralentissements liés à l'âge en dénomination à des différences perceptuelles ou motrices suite à leurs analyses des erreurs produites (en grande partie plus d'erreurs perceptuels chez les personnes plus âgées) et face à un effet d'âge toujours significatif dans la condition d'amorçage concordant (mot écrit = image). Les sujets auraient besoin d'un temps de présentation plus long pour identifier correctement les images. Cette étude n'a pas exploité la différence entre les différents groupes d'âge et pourraient être approfondie.

Morrison et al. (2003) ont évalué la précision et la latence en dénomination de verbes sur images et par écrit de deux groupes de participants en bonne santé : un de 18 à 27 ans et l'autre plus âgé de 65 à 87 ans. Ils ont également conclu que les temps de réaction étaient plus rapides chez le groupe plus jeune que celui plus âgés, conclusions cohérentes avec les études antérieures (Thomas & al., 1977 ; Mitchell, 1989, Morrison & al., 2002, cités par Morrison). Cependant, cette étude laisse la tranche de la cinquantaine de côté.

C'est pourquoi, dans une nouvelle étude, Verhaegen et Poncelet (2013) ont analysé la précision mais aussi la latence – nécessaire pour détecter une diminution plus subtile - dans une tâche de dénomination d'images chez quatre groupes de sujets qui reprennent différentes tranches d'âge (25-35, 50-59, 60-69 et au-delà 70 ans) ; une attention particulière a été portée au groupe de la cinquantaine conformément aux difficultés de dénomination subtiles évoquées précédemment. Leur premier objectif était de déterminer si un déclin des performances en dénomination lié à l'âge est présent chez les personnes âgées, spécialement si des difficultés, même subtiles, apparaissent chez les participants dans la cinquantaine. Ils ont émis l'hypothèse que, pour cette tranche d'âge, un temps de latence plus long mais sans une diminution de la précision lors de la tâche serait constaté et que, pour la septantaine, à la fois des latences plus grandes mais aussi des performances plus basses en dénomination d'images apparaîtraient. Les résultats prouvent leur hypothèse initiale : chez les cinquantenaires, une baisse légère est constatée dans les performances à l'épreuve de dénomination par l'augmentation de la latence – le taux d'erreurs n'étant pas plus haut que pour les plus jeunes participants - ; le groupe de 60 ans a montré une diminution de la précision mais également une latence qui augmente ; enfin, dans le groupe des 70 ans, ce déclin, comme souligné par les études précédentes, se marque beaucoup plus, aussi, bien dans la précision que dans la latence lors de la tâche de dénomination.

Cette étude a été répliquée et étendue dans un mémoire (Truffert, 2014). La mémorante en master de logopédie a choisi de déterminer si les difficultés en dénomination observée dès l'âge de 50 ans, dans une moindre proportion, pouvaient arriver plus tôt (avant 50 ans). A cette fin, elle a sélectionné trois tranches d'âge : 30-39, 40-49 et 50-59. Ces trois groupes successifs ont l'avantage également de permettre d'apprécier une éventuelle évolution progressive. La tâche de dénomination a été étendue : en plus de noms communs (NC), l'épreuve comprend des noms propres (NP), catégorie qui est particulièrement sensible aux effets de l'âge et, comme mentionné ci-dessus, plus sujette aux échecs de récupération. Les hypothèses de ce mémoire étaient notamment les deux suivantes : elles concernaient premièrement l'« observation d'un effet de vieillissement sur les performances en dénomination de NC et NP dès la tranche 40-49 ans par rapport au groupe 30-39 ans ; deuxièmement, l'« observation d'un effet de vieillissement se traduisant chez les 40-49 ans par une augmentation du temps de latence uniquement, en dénomination de NC, et par une augmentation du temps de latence et du taux d'échecs de production en dénomination de NP, et essentiellement des MBL » (Truffert, 2014, p.40). Les résultats de cette étude ont démontré

un ralentissement linéaire pour la catégorie des noms communs. Par contre, elle constate une augmentation significative, considérant le temps de latence, dès le groupe 40-49 ans mais qui se stabilise dans le groupe de 50-59 ans.

Par ailleurs, l'étude des différences entre deux groupes d'âge a également été examinée au-delà de la tâche de dénomination. En effet, dans un récent article, Madden, Sale et Robinson (2019 a) ont mis en place une étude sur les différences entre deux groupes d'âge (18-30 ans et 60-80 ans) au niveau de la génération des idées et de la sélection dans le cadre de la production du discours complexe. Les auteurs se sont intéressés aux mécanismes de préparation conceptuelle (génération d'idées et sélection) qui contribueraient aux déficits de langage liés à l'âge (notamment la dénomination). Ils sont partis de preuves qui suggéreraient un déclin dans les mécanismes de la préparation conceptuelle. En effet, il a été observé dans la littérature chez les personnes plus âgées un taux de disfluences plus élevées qui est plus prononcé quand le sujet de la conversation est difficile et moins familier, ce qui indiquerait un déficit lié à l'âge dans la construction du langage au niveau conceptuel (voir Madden & al., 2019, p.488). Une diminution a été constatée également chez les personnes plus âgées en fluence, ainsi que dans un contenu propositionnel nouveau (*op. cit.*). Outre des épreuves pour évaluer préalablement les capacités cognitives et langagières, deux tests sont proposés : un nouveau test, le *Concept Selection Test*, pour la sélection conceptuelle avec trois conditions (condition « haut sélection », condition « bas sélection », condition neutre) et un test, le *Concept Generation test*, repris de Robinson et al. (2005, cité par Madden & al.) qui est une tâche de complétion de phrases avec trois conditions à nouveau (simple, multiple, contrôle). Les résultats ont apporté des données suggérant un déclin lié à l'âge dans les mécanismes de préparation conceptuelle. En effet, les personnes plus âgées étaient plus lentes en général sur les deux tests. Les auteurs soulignent que la différence est plus prononcée là où les demandes sont plus grandes que ce soit au niveau de la génération ou de la sélection. Concernant le *Concept Generation Test*, le temps de réponse était similaire aux deux groupes à travers les conditions « simple » et « multiple ». Toutefois, en regardant les erreurs commises – taux étant significativement plus bas que le groupe était plus jeune – par le groupe des participants plus âgés, les auteurs ont constaté que les participants avaient passé des items – en particulier des items plus difficiles – dans la condition « multiple », ce qui a diminué artificiellement la moyenne des latences, attitude que l'on ne retrouve pas dans le groupe des plus jeunes. Les auteurs ont veillé avant d'arriver à ces conclusions à écarter toute autre explication (ralentissement général dû à l'âge ou préférence différentielle de catégories). Ils ont, enfin,

constaté, entre les deux groupes, des performances réduites dans les fluences sémantiques et équivalentes dans les fluences phonémiques. Cependant, les auteurs tiennent à souligner qu'un facteur peut rentrer en considération dans ces résultats, à savoir le niveau d'éducation qui peut jouer comme un facteur protecteur. Selon l'étude de Tombaugh et al. (1999, cités par Madden & al.), la fluence sémantique est plus sensible à l'âge, alors que la fluence phonémique, elle, est plus affectée par l'éducation. L'échantillon de participants (15 ans d'éducation) est considéré dans l'étude de Tombaugh et al. comme le plus haut, ce Madden et al. avaient constaté dans les épreuves préalables. Cela a pu donc influencer les résultats. En plus de cette remarque sur l'éducation, cette étude mériterait d'être répliquée également sur des tranches intermédiaires qui n'ont pas été prises en compte et qui pourraient, à l'instar de la tâche de dénomination, apporter des nuances sur l'évolution.

La revue de ce débat montre combien le temps de latence est important car elle apporte des nuances dans les conclusions des études et donne une sensibilité accrue de la tâche de dénomination. C'est pourquoi, un chapitre sera consacré par la suite aux moyens de mesurer cette latence de la façon la plus efficace et efficiente possible.

3. Pourquoi l'âge influence-t-il le manque du mot ?

a) Accès lexical : mécanismes et modèles cognitifs

Le manque du mot étant un échec de recouvrement, une difficulté, voire une impossibilité à accéder temporairement à la cible, il est important, pour répondre à cette question, d'établir ce qu'on entend par accès lexical, de le replacer dans son contexte, à savoir la production verbale d'un mot, et d'exposer les modèles explicatifs émis dans la littérature.

Trois caractéristiques principales se retrouvent, selon Bonin (2007), dans l'accès lexical à l'oral : la rapidité (100/200 mots par minute), l'efficacité (taux d'erreurs de 1/1000 mots produits) et l'impénétrabilité (mécanismes impliqués inaccessibles à la conscience).

La production verbale d'un mot comprend plusieurs étapes (énoncées notamment par les travaux de Levelt et al. 1999, 2001) qui va de la préparation conceptuelle en passant par la sélection lexicale à l'encodage phonologique. Le premier niveau est donc conceptuel, c'est-à-

dire que le locuteur va chercher, dans son lexique mental, le concept précis de ce qu'il souhaite exprimer. Le deuxième est lexical et consiste à sélectionner et à récupérer la forme lexicale qui correspond au concept choisi. Enfin, le niveau articulatoire comprend la production du mot, de l'encodage phonologique du mot (récupération du code phonologique) à sa production en tant que telle. Roux et Bonin (2011) mentionnent une distinction aussi communément admise entre un niveau lexical holistique, appelé niveau L par Rapp et Goldrick (2000, cité par Roux & Bonin) reprenant les deux premiers niveaux, à savoir conceptuel et lexical *stricto sensu*) et entre un niveau phonologique. Si ce cadre est admis par les auteurs avec consensus, il en va moins du nombre et de la nature des traitements et, particulièrement, du décours temporel entre ces différents niveaux : comment l'information se propage-t-elle d'un niveau à l'autre, à savoir du niveau lexico-sémantique (*lemmas*) et phonologique (*lexèmes*). En effet, sur ce sujet, un débat vif et de longue haleine s'est engagé et se poursuit entre les auteurs.

Il y a trois types de modèles qui se confrontent sur base d'études à méthodologie variée. Les auteurs proposent une architecture fonctionnelle avec une organisation sérielle et discrète, en cascade et/ou interactive.

Le premier modèle consiste en une activation d'un niveau vers un autre subséquent une fois que le premier est totalement activé. Comme le soulignent Roux et Bonin (2011), il ne faut pas le voir comme absolu. Dans la conception première de Levelt (1999), une transmission en cascade peut se produire entre certains niveaux de traitement. Par ailleurs, dans le modèle revu par la suite, appelé *WEAVER ++* (Roelofs, 2004 ; cité par Roux & Bonin), une activation en cascade est possible entre les concepts et les représentations lexicales (appelées *lemmas* pour ces auteurs). Cependant, l'effet cascade ne peut se produire que dans le niveau L (entre concepts et lemmas) et non entre le niveau lexical et phonologique. L'information passe d'abord par le niveau conceptuel et ensuite par le niveau lexical avec une co-activation éventuelle entre ces deux niveaux (cf. fig. 1). Par contre, une seule représentation lexico-sémantique est sélectionnée, celle qui sera la plus activée par rapport aux autres (mécanisme de compétition lexicale), et va activer le niveau suivant, à savoir phonologique. Cette théorie a été vivement critiquée en raison d'arguments « négatifs » (Roux & Bonin) et de flottement face à certaines erreurs dans les productions pathologiques (ex. : erreurs mixtes). Ce modèle a continué à être plusieurs fois mis à jour, notamment face aux critiques émises à son égard, encore tout récemment (Roelofs, 2018).

Le deuxième modèle (notamment défendu par Humphreys & al., 1988 ; cités par Roux & Bonin, 2011) suppose qu'un niveau subséquent peut être activé alors que le niveau antécédent n'est pas encore complètement activé. C'est pourquoi, l'activation phonologique n'est pas restreinte à la représentation lexico-sémantique sélectionnée au niveau antécédent contrairement au modèle sériel (cf. fig. 1). Dans ce modèle, des liens excitateurs et inhibiteurs sont présents pour réguler ces activations : les représentations adéquates sont fortement activées et celles concurrentes sont inhibées. A noter qu'ici, il s'agit bien de mouvements en cascade et non de rétroaction comme il sera décrit dans le troisième type de modèle.

Un autre modèle en cascade est celui de Caramazza (1997, cité par Bonin, 2007). Il est appelé « modèle en réseaux indépendants de l'accès lexical » (en anglais, *Independent Network Model of Lexical Access*, INM). Dans ce modèle, il y a des réseaux indépendants inter-reliés avec différents sous-ensembles - dans lesquels sont organisées les connaissances lexicales - et des nœuds – qui possèdent des connexions inhibitrices. Le réseau sémantico-lexical contient la signification sous forme de propriétés, traits ou prédicats sémantiques, le réseau syntaxico-lexical les traits syntaxiques et les réseaux lexémiques phonologiques et orthographiques les représentations spécifiques des modalités de sortie respectivement orale et écrite. L'activation, provenant du niveau sémantique, se propage simultanément et indépendamment vers les réseaux syntaxiques et lexémique (Bonin, 2007, p.223-224). Ce modèle a l'avantage d'expliquer la production verbale orale mais aussi écrite.

Enfin, le modèle interactif (Dell, 1986, cité par Roux & Bonin, 2011) propose une architecture en cascade mais avec un effet de rétro-activation (bidirectionnalité). Une interactivité se produit entre les différents niveaux de traitement grâce à ces connexions bidirectionnelles et excitatrices. Dès lors, les représentations lexico-sémantiques qu'elles soient adéquates à la cible ou proches de la cible (distractrices) sont activées et transmises au niveau phonologique, qui active les représentations phonologiques correspondantes. En retour, les représentations phonologiques renvoient l'activation aux représentations lexico-sémantiques associées (effet de rétroaction). La récupération se fait par l'activation la plus élevée de toutes les représentations.

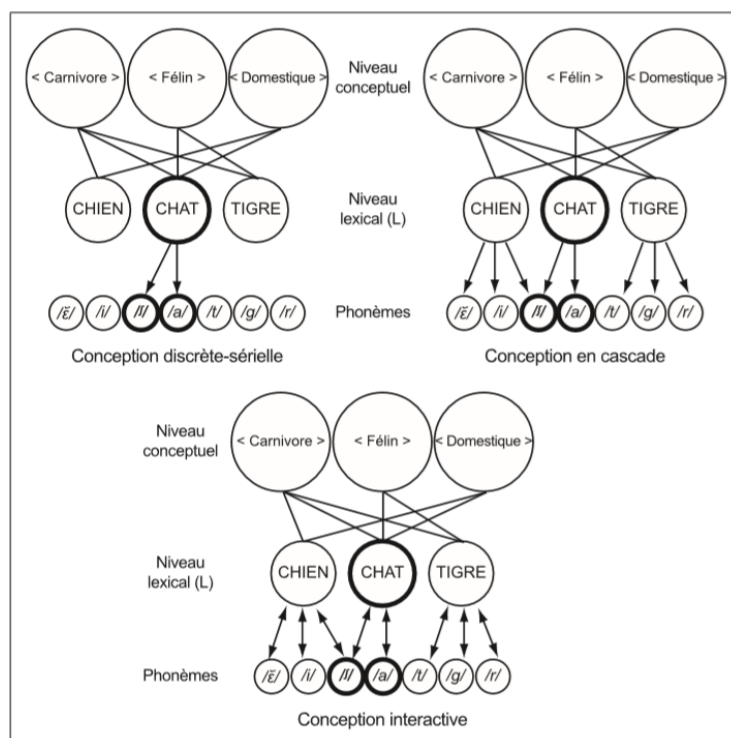


Figure 1 : Comparaison des trois types de modèles cognitifs repris dans Roux et Bonin (2011, p.148)

Un autre modèle, le *Node Structure Theory* (NST) – nom provenant de la conception de cette théorie avec comme élément de base, le *nœud* -, a été proposé par MacKay (1987, cité par Hoyau, 2018). Ce modèle a l'avantage de concilier les modèles avec une organisation sérielle et ceux avec une organisation interactive. En effet, une interaction est présente entre les différents niveaux de traitement mais une perspective sérielle est conservée dans le sens qu'il faut atteindre un certain seuil d'activation pour qu'un nœud active les nœuds qui lui sont associés (activation unidirectionnelle). Les nœuds sont reliés entre eux formant un réseau où la communication se passe non seulement *intra* mais aussi *inter* niveaux. Ce réseau de nœuds est réparti sur deux niveaux (cf. fig. 2) : un niveau conceptuel ou sémantique qui reprend les représentations sémantiques (nœuds propositionnels, sous forme de propositions) et les représentations lexico-sémantiques (nœuds lexicaux, sous forme d'unités lexicales) - ce niveau contient donc des unités, les « lemmes », référant aux informations lexico-syntaxiques (Caramazza & Miozzo, 1997 ; cités par Truffert, 2014) - et un niveau phonologique qui contient des nœuds syllabiques, des nœuds phonologiques et des nœuds « traits », c'est-à-dire des unités, « lexèmes », référant à la forme phonologique du mot syntaxique (*op. cit.*). Un troisième niveau, prévu dans ce modèle, est articulatoire, c'est-à-dire les mouvements musculaires nécessaires à la réalisation motrice du mot. Sans rentrer dans le détail de toutes

les propriétés des nœuds, deux propriétés sont à mentionner plus particulièrement dans leur fonctionnement : l'*amorçage* et l'*activation*. Le mécanisme de sélection est sériel, c'est-à-dire que l'activation entre les nœuds se passe donc, comme mentionné ci-dessus, de façon unidirectionnelle, niveau après niveau. Par ailleurs, il faut qu'un certain seuil d'activation, dit le seuil d'amorçage, soit atteint pour se propager aux nœuds suivants. Par contre, un élément important est souligné par Hoyau (2018, p. 59) :

« De manière intéressante, la quantité d'amorçage diminue avec la distance entre deux nœuds : un nœud activé transmet de l'amorçage vers les nœuds auxquels il est directement relié, puis les nœuds ainsi amorcés transmettent de l'amorçage vers les nœuds auxquels ils sont connectés mais de manière moins importante, etc. L'amorçage est dit sommatif. Enfin, la force de connexion entre deux nœuds (et ainsi, la quantité d'amorçage nécessaire pour l'activation d'un nœud) dépend de la fréquence d'usage et de la récence d'usage. Plus la force de connexion entre deux nœuds est importante et plus la transmission de l'amorçage est rapide. »

Cette théorie est considérée comme la plus explicative des difficultés d'accès au mot (mot sur le bout de la langue ou manque du mot) et de leur fréquence d'apparition.

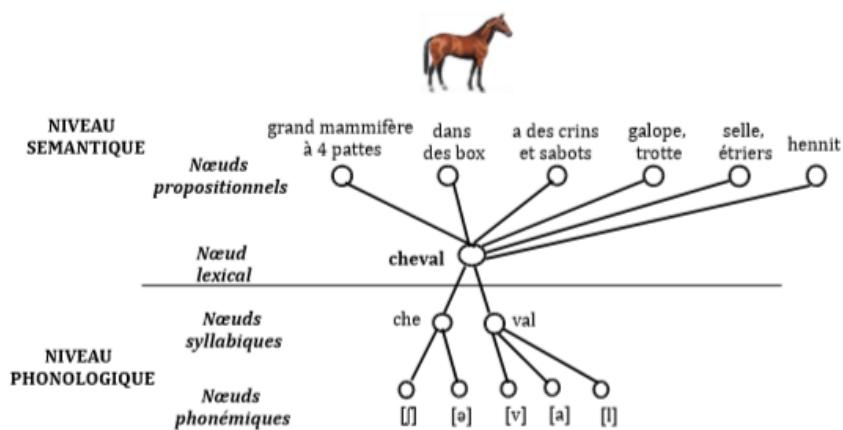


Figure 2 : Schéma synthétique de la NST réalisé par Truffert (2014, p.16)

b) Hypothèses de l'effet de l'âge sur le manque du mot

Les recherches se sont interrogées sur les raisons qui pourraient expliquer l'effet lié à l'âge. Trois hypothèses principales ont été émises dans la littérature.

Premièrement, a été formulée l'hypothèse du déficit de transmission, TDH, fondé sur la NST décrite ci-dessus. Sur cette base, elle émet l'hypothèse d'un affaiblissement dû à l'âge des connexions entre les deux systèmes, à savoir le système lexico-sémantique (représentations sémantiques) et le système phonologique (représentations phonologiques) qui conduit *in fine* à une réduction du seuil d'activation d'un mot donné et, ainsi, à plus d'épisodes du manque du mot chez les personnes âgées. Cela vaut d'autant plus pour les nœuds qui n'ont qu'une seule connexion - qui seront plus sensibles aux effets de l'âge – que pour ceux qui en ont plusieurs. En effet, selon cette structure, le système sémantique comprend plusieurs connexions lexico-sémantiques convergeant vers une représentation lexicale qui laisse place à la compensation, tandis que le système phonologique contient une connexion entre la représentation lexicale et la représentation phonologique qui rendrait la compensation difficile et entraînerait une récupération incomplète. L'effet de fréquence (plus le mot est utilisé, plus la connexion sera forte) et l'effet de récence (la connexion entre les nœuds est plus importante lorsque celle-ci a été récemment sollicitée) module les connexions. Cette hypothèse expliquerait le déficit d'accès aux représentations phonologiques en avançant en âge. Ces affirmations seront développées ci-dessous dans les études qui les ont étayées.

Deuxièmement, alors que la première hypothèse mentionnée est plus liée au langage, des théories prônent un ralentissement général (Myerson & al., 1992 ; Salthouse, 1996, cités par Verhaegen & Poncelet, 2013), c'est-à-dire que les différences qu'on peut observer chez les plus jeunes sont amplifiées chez les personnes plus âgées et que tous les processus cognitifs seraient touchés de façon égale.

Enfin, la troisième hypothèse examinée est un déficit d'inhibition (Hasher & Zacks, 1988, cités par Feyereisen & al., 1998) à savoir de ne pas sélectionner correctement un mot particulier face à des mots « concurrents ». Avec l'âge, on emmagasine de plus en plus de mots qui sont autant de concurrents potentiels. Par conséquent, par ce constat, les personnes âgées seraient plus sujettes à l'interférence. Dans le même ordre d'idées, Dahlgren (1998, cité par Shafto & al., 2017) conclut qu'une augmentation de connaissance en vocabulaire est liée à

l'augmentation du taux de MBL. L'accroissement de ce phénomène avec l'âge viendrait donc d'interférences sémantiques qui exacerbent les difficultés de récupération.

c) Quelques études examinant les différentes hypothèses

Feyereisen et al. (1998) ont consacré une de leurs études à tester ces différentes hypothèses. Ils s'intéressaient particulièrement au constat d'une augmentation de la latence dans une tâche de dénomination d'images. Ils reprennent donc les trois hypothèses développées ci-dessus pour les mettre à l'épreuve dans deux expériences mettant en œuvre différentes tâches : dénomination d'images, prononciation de mots écrits et catégorisation mots-images. Leurs résultats vont plutôt dans le sens d'un ralentissement général. Il faut, cependant, remarquer que les participants étaient répartis en deux groupes : 18-22 ans et 65-80 ans. On peut souligner que l'âge intermédiaire n'est pas représenté.

Depuis, d'autres études se sont emparées du problème. La théorie du déficit d'inhibition, par exemple, a été réfutée (Abrams & al., 2003 ; Cross & Burke, 2004 ; White & Abrams, 2002 ; James & Burke, 2000, cités par Truffert, 2014), notamment à cause de leurs preuves jugées « limitées et hétérogènes » par Shafto et al. (2017).

Nicholas et al. (1985, cité par Verhaegen & Poncelet, 2013) avaient introduit dans leur étude, en tâche de dénomination, un indice phonologique quand les participants (groupes entre 30 et 70 ans) étaient en difficulté pour récupérer la cible. Ils ont constaté que le besoin de cet indice augmente avec l'âge mais le bénéfice était le même quel que soit l'âge. Les auteurs concluent que la difficulté viendrait plutôt de l'accès que de la représentation. Cette étude a été affinée par Au et al. (1995, cité par Verhaegen, 2013) qui, eux, ont trouvé un bénéfice moindre de l'indice au-delà de 70 ans et qui, par conséquent, ont émis l'hypothèse d'un déficit de connexion entre les deux systèmes lexico-sémantique et phonologique et d'une dégradation lexico-sémantique.

C'est dans ce sens que Barresi et al. (2000) vont également dans leur étude : les personnes de 50-60 ans auraient une altération de l'accès phonologique et les personnes de plus de 70 ans auraient, eux, cette altération mais aussi une dégradation sémantique.

Verhaegen et Poncelet (2013) ont évalué, dans leur étude, les capacités sémantiques des participants (test des Pyramides et des Palmiers) pour tenter d'apporter des arguments en

faveur de l'hypothèse d'une dégradation sémantique que Barresi et al. (2000) ont relevée chez les septuagénaires. Les auteurs ont pu de cette manière tester l'hypothèse d'une dégradation liée au langage. Par ailleurs, ils ont également évalué avec une tâche parallèle (jugement de chiffres pairs/impairs) la vitesse du traitement cognitif général pour contrôler ainsi un éventuel ralentissement qui affecterait toutes les tâches cognitives de façon égale. Leurs résultats suggèrent un ralentissement spécifique à la tâche de dénomination d'images, du moins jusqu'à 60 ans. A partir de la soixantaine et au-delà de 70 ans, un ralentissement significatif dans la tâche de jugement de chiffres pourrait signifier qu'un ralentissement général (des facteurs autres que langagiers) et des facteurs liés au langage entraînent en compte dans les difficultés de dénomination.

La réplication dans le mémoire de Truffert (2014), qui inclut aussi les noms propres, arrive également au constat d'un lien du type langage spécifique. Elle a également mis en évidence que « le contrôle de la vitesse de traitement général et le maintien des performances sémantiques à travers les trois décennies témoignent du fait que les ralentissements observés en dénomination et consécutifs au vieillissement sont de type langage-spécifique » (Truffert, 2014, p. 75) et a souligné que le type d'échecs (pour notamment les noms propres) peut dépendre d'un groupe en particulier, ce qui reste un point de vigilance et d'amélioration dans une perspective future.

Dorot et Mathey (2013) ont testé, dans le cadre de la théorie de déficit de transmission, la fréquence des MBL dans deux groupes d'âge (« adultes jeunes » - « adultes âgés ») en reprenant la procédure d'analyse de Gollan et Brown (2006, cité par Dorot & Mathey) qui a pour but d'examiner l'effet de l'âge sur les représentations sémantiques (étape 1) et sur les représentations phonologiques (étape 2) en fonction des réponses métacognitives (« je ne sais pas », « je suis en MBL » et « je sais ») des participants. Les auteurs ont fait attention de contrôler le niveau de vocabulaire avant et d'apparier les deux groupes sur ce niveau. La tâche était un test de reconnaissance sur définitions (noms communs). Ils sont arrivés à la conclusion, suite aux résultats obtenus, que, une fois contrôlé, le niveau de vocabulaire ne peut pas influencer l'apparition d'échec de récupération contrairement à ce qui a été émis par les théories des interférences (en l'occurrence, Dahlgren, 1998 ; voir ci-dessus). En revanche, leurs résultats corroborent la théorie du déficit de transmission. En effet, les auteurs (Dorot & Mathey, p. 11) concluent grâce aux modifications apportées par rapport à l'étude antérieure de Gollan et al. :

« (...) à niveau de vocabulaire comparable, les adultes âgés éprouvent moins de difficulté à récupérer la forme sémantique du mot que les adultes jeunes » et « à niveau de connaissance comparable et avec un accès plus efficace aux représentations sémantiques, les adultes âgés semblent avoir plus de difficultés à accéder à la forme phonologique du mot que les adultes jeunes » (Dorot & Mathey, 2013, p. 11)

Dans ce cas-ci, l'accumulation de connaissances sémantiques – et ainsi des connexions sémantiques plus développées – est un effet positif de l'âge. Ces conclusions sont en accord avec d'autres études notamment celle de Taylor et Burke (2002, cités par Dorot & Mathey).

Une récente étude de Shafto et al. (2017) s'est de nouveau penchée sur les raisons qui pourraient expliquer l'augmentation des échecs de récupération (ici, le MBL). L'étude se focalise sur les noms propres qui sont plus sensibles à ce phénomène que les noms communs, contrairement à Dorot et Mathey (2013). Les auteurs reprennent le protocole de la cohorte Cam-CAN (*Cambridge Centre for Ageing and Neuroscience* de Shafto & al., 2014). Dans leur travail, ils ont pour objectif d'examiner la relation entre l'intelligence cristallisée (évaluée par le Spot-the-Word Test) – dont le vocabulaire fait partie – et la récupération d'un mot (évaluée par une tâche induisant le MBL et une tâche de dénomination sur images). Leurs résultats corroborent l'hypothèse PRD (de l'anglais, *phonological retrieval deficit*, à savoir déficit d'accès phonologique lié à l'âge, parallèle au déficit de transmission). En effet, cette étude arrive au fait que, comme auparavant Dorot et Mathey, le vocabulaire est associé positivement aux meilleures performances en tâche de dénomination. Cela s'explique par le fait que le développement des connaissances en vocabulaire crée un voisinage phonologique plus dense qui peut aider à la récupération lors de la production. Cela infirme, par conséquent, l'hypothèse de l'interférence sémantique liée à l'augmentation du vocabulaire avec l'expérience de l'âge (voir ci-dessus). Les auteurs concluent en soulignant qu'il ne faut pas sous-estimer une compensation potentielle grâce à l'augmentation de connaissances lexicales dans le processus de recouvrement.

Paolieri et al. (2018), dans leur étude, avaient un double objectif. Ils voulaient, d'abord, étudier l'effet de l'âge en regard des deux hypothèses émises, à savoir le déficit d'inhibition et le déficit de transmission (phonologique). Ils voulaient, ensuite, observer si le niveau d'éducation modulait ces déficits chez les personnes plus âgées. Pour évaluer l'interférence, ils ont mis en place une tâche de dénomination avec un *semantic-blocking paradigm* qui comprend deux conditions, un contexte homogène (items provenant d'une même catégorie

sémantique) qui suscite l'interférence et un contexte hétérogène (items provenant de différentes catégories). Les participants étaient répartis en deux groupes : un groupe plus âgé (de 60-79 ans) avec deux niveaux d'éducation (un de 0 à 9 et un de 10-20 ans) et un groupe plus jeune (de 18-23 ans). Seize participants de langue espagnole étaient dans chaque groupe. Les auteurs ont également introduit des noms propres en plus des noms communs. Ils concluent (fig. 3) que les sujets plus jeunes étaient plus précis et plus rapides que les sujets plus âgés (effet de l'âge). Les sujets plus âgés avec un niveau d'éducation plus bas étaient plus lents que leur homologue avec un niveau plus haut (modulation du niveau d'éducation). Pour tous les groupes, la récupération des noms propres était plus lente et la condition homogène également (effet d'interférence). Pour les auteurs, les deux hypothèses ensemble permettent d'expliquer la totalité des résultats. L'hypothèse du déficit d'inhibition expliquerait la plus grande vulnérabilité des personnes plus âgées avec un niveau d'éducation plus bas face à l'interférence ; l'hypothèse du déficit de transmission expliquerait, elle, l'effet d'interférence principalement limité aux noms propres qui ont moins de connections que les noms communs. Les auteurs soulignent également qu'un effet de la demande de la tâche est présent.

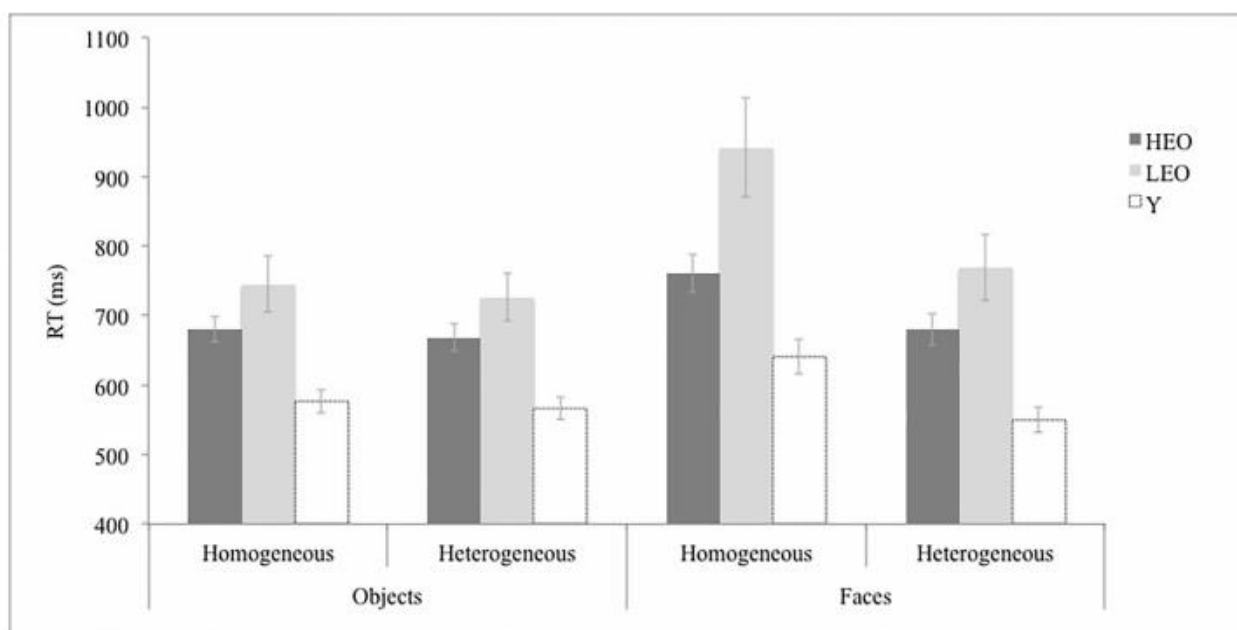


Fig 1. Mean naming latencies (in milliseconds) as a function of the different experimental conditions: Group (LEO; HEO; Y), stimulus type (objects vs. faces) and context (homogeneous vs. heterogeneous).

Fig.3 : Résultats obtenus par Paolieri et al. (2018, p.7). S'y retrouvent, en blanc, le temps de réponse obtenu par les groupes plus jeunes (Y), en gris clair, celui obtenu par le groupe plus âgé avec un niveau d'éducation plus bas (LEO) et, en gris foncé, celui obtenu par le groupe plus âgé avec un haut niveau d'éducation (HEO)

Enfin, nous avons mentionné que, dans leur étude, Madden, Sale et Robinson (2019 a) avaient écarté le ralentissement général qui ne pouvait expliquer complètement leurs résultats. En effet, le ralentissement était spécifique aux conditions de l'expérience, celles, en l'occurrence, qui demandaient un plus grand degré de demandes au niveau de la génération d'idées et de la sélection. Aucune preuve probante de ce ralentissement n'est constatée dans la condition « contrôle » ni dans les autres tests préalables dans le domaine du langage.

4. Problématiques

L'étude du manque du mot – et autres échecs de récupération – est un enjeu important pour mieux comprendre le fonctionnement du langage (production orale en l'occurrence) et les différentes natures de représentations mobilisées – et ainsi étayer les modèles élaborés tout au long de ces décennies -, pour mieux appréhender le phénomène en question et pour pouvoir *in fine* apporter des preuves probantes afin d'établir au mieux un cadre, une aide, une thérapie en contexte pathologique.

Beaucoup d'auteurs se sont penchés sur la problématique, que ce soit pour établir le modèle le plus adéquat avec les résultats des études ou pour tenter d'expliquer les phénomènes et les effets observés. Tous ces débats, qui n'aboutissent pas toujours à des conclusions convergentes mais, au contraire, hétérogènes, voire contradictoires, met en exergue l'importance de la construction des épreuves et des protocoles d'étude. En effet, nous avons vu, pour ne citer qu'un exemple, que la procédure d'analyse en deux étapes de Gollan et Brown (2006, cité par Dorot et Mathey, 2013), même si elle était tout à fait pertinente, a apporté une interprétation jetant une ombre sur l'hypothèse du déficit d'accès phonologique émise par plusieurs auteurs (cet accès ne déclinerait pas avec l'âge). La correction du manque de contrôle des caractéristiques des participants dans la réplique de Dorot et Mathey a infirmé cette conclusion divergente. C'est pourquoi, il est primordial de tenir compte de tout facteur qui pourrait biaiser les conclusions. On peut citer à la vue de ce cadre théorique notamment les tranches d'âge représentées, le matériel utilisé (noms propres/noms communs) et l'appariement des sujets (niveau de vocabulaire, niveau d'éducation).

De même, la prise en compte de la latence a permis de préciser, de nuancer, d'affiner les données récoltées de la tâche de dénomination. La tâche de dénomination a bénéficié d'une sensibilité plus importante de cette manière concernant principalement l'effet de l'âge. C'est pourquoi, les moyens de mesurer la latence sont d'une importance capitale. Les réponses orales demandent, dans ce domaine, une attention particulière.

C'est de ces réflexions que le présent travail trouva son objectif principal, à savoir l'évaluation de la vitesse d'accès lexical. Ce mémoire se focalisera sur la tâche de dénomination orale d'image, épreuve canonique dans ce domaine, et sur les noms communs comme initialement prévu. Étant donné que la latence apporte une sensibilité à l'épreuve et des nuances dans les résultats obtenus, la première problématique soulevée est la conception même de la tâche de dénomination. Elle pourrait être reformulée sous ces questions de recherche : chez les personnes tout-venants (P), quelles variables sont à prendre en considération dans l'élaboration des stimuli (I) afin de disposer d'une tâche de dénomination bien contrôlée permettant d'obtenir des données fiables (O) ? La deuxième problématique portera sur les moyens d'obtenir des mesures de latence les plus précises possibles, et ce, de manière efficiente. Elle pourrait être reformulée ainsi : Quels dispositifs sont proposés dans la littérature (I') afin de mesurer la latence de manière précise et de façon plus efficiente (O) autre que l'analyse manuelle (C) ? Les chapitres suivants étudieront donc, d'une part, les facteurs qui peuvent influencer et qui sont à prendre en considération dans une tâche de dénomination - ainsi que les ensembles de stimuli disponibles dans la littérature - et, ensuite, d'autre part, les moyens les plus efficaces et efficients pour mesurer le temps de latence autre que l'analyse manuelle. Des propositions de réponses à ces deux problématiques seront émises suite à ces deux revues de la littérature. Ces propositions porteront à la fois sur la conception de la tâche – principalement le choix des images - qui semble être la plus idéale par rapport à l'objectif initial et le(s) dispositif(s) qui ressort(ent) le(s) plus prometteur(s).

1. Il s'agit moins d'une intervention mais plutôt des moyens de la recherche. Cependant, il prend bien place dans l'étiquette de *intervention* de la question de recherche. La population reste toujours les personnes tout-venant.

II. DÉNOMINATION ORALE D'IMAGES

1. Tâche de dénomination orale d'images : préambule

La tâche de dénomination orale d'images, qui consiste à donner le nom de la cible présentée, est une épreuve expérimentale fréquemment utilisée. Dénommer un objet prend en moyenne entre 600 et 1200 ms (Bonin, 2007). Cette tâche permet d'évaluer la production orale en recueillant des données qualitatives (erreurs produites qui sont cruciales en contexte pathologique) et des mesures quantitatives (temps de réponses et précision des réponses). Ces erreurs de production sont rares (moins de 1/1000 selon Butterworth, 1992 ; cité par Hoyau, 2018) et peuvent être complexes (e. g., erreur à la fois phonologique et sémantique). L'analyse qualitative des erreurs permet de récolter des données pour étudier les processus et les représentations mis en œuvre en production verbale et, ainsi, pour inférer les étapes de traitement. Ces informations qualitatives, cependant, n'apportent pas des informations sur le dérouls temporel. Pour cela, il faut considérer aussi le temps de réponse, la latence, de la tâche de dénomination, ce qui renforce l'importance d'obtenir une mesure précise de cette latence.

La tâche de dénomination d'images demande de dénommer oralement, le plus rapidement possible, l'image qui se présente au centre de l'écran de l'ordinateur. Cependant, elle peut se dérouler selon différents paradigmes. Mettre en place différents paradigmes permet d'étudier « la nature des représentations mobilisées en production orale de mots (ou de groupes de mots) ainsi que le dérouls temporel de leur activation » (Bonin, 2007, p. 31).

Le paradigme d'interférence consiste à lier l'image à un mot distracteur présenté visuellement ou auditivement avant ou pendant la dénomination. Le délai peut être manipulé par l'expérimentateur selon son objectif. La personne doit dénommer l'image le plus rapidement possible en ignorant le distracteur. Ce dernier peut avoir une relation sémantique ou une relation phonologique avec la cible. Cela permet de mettre en évidence un effet d'interférence sémantique (i.e., augmentation des temps de réponse lors de la présentation d'un mot sémantiquement relié à la cible) ou un effet de facilitation phonologique (i.e., diminution des temps de réponse lors de la présentation d'un mot phonologiquement relié à la cible) (Hoyau, 2018).

Un second paradigme possible est le paradigme en blocs cycliques qui est utilisé pour étudier l'effet d'interférence sémantique (Hoyau, 2018). Il consiste soit à proposer des images qui appartiennent à une même catégorie sémantique (blocs homogènes), soit à répéter les blocs à plusieurs reprises (blocs mixtes).

Un troisième paradigme qui est assez proche de celui de l'interférence et qui évalue aussi la nature des représentations activées, ainsi que leur dérouls temporel, est le paradigme de l'amorçage (en anglais, *priming paradigm*). Dans ce paradigme, sont choisies une cible et une amorce. L'expérimentateur a le choix de la durée (e.g., intervalle entre l'amorce et la cible ; durée subliminale ou durée permettant une identification consciente) et des conditions de présentation (e. g., de façon visible, de façon masquée) de l'amorce qui précède toujours la cible. Ce modèle a été utilisé pour étudier la reconnaissance visuelle des mots ou l'accès lexical en production verbale orale (Bonin, 2007).

Enfin, la tâche de dénomination orale peut être continue. Les images sont présentées aux participants de façon continue, à savoir les unes après les autres. Ce paradigme est souvent utilisé pour mettre en évidence les troubles de production lexicale dans des populations à risque comme les personnes âgées ou les patients ayant subi un AVC (Hoyau, 2018). Ce paradigme est également utilisé pour étudier l'effet de l'âge sur les performances en dénomination comme c'était le cas de certaines études dans le chapitre précédent.

La tâche de dénomination d'images se déroule selon des étapes bien précises. La production orale d'un mot passe par un niveau conceptuel, un niveau verbal et un niveau articulatoire (Bonin, 2007). Plus précisément, il est communément accepté que la dénomination d'objets comporte au moins trois principaux niveaux de processus (Bonin & al., 2013). Premièrement, la compréhension de l'objet consiste en une analyse perceptuelle, en une identification visuelle (forme, couleur, taille ou autres détails extérieurs) qui permet d'accéder aux représentations structurelles stockées en mémoire à long terme pour pouvoir *in fine* reconnaître l'objet. Deuxièmement, la récupération du nom de la cible peut être considérée selon les points de vue en deux étapes : la sélection lexicale - sélection et récupération de l'information lexico-sémantique - et l'encodage phonologique. Troisièmement enfin, l'articulation du mot demande sa planification motrice et, ensuite, son exécution. Certains autres auteurs ont affiné cette distinction avec ces niveaux de traitement : les niveaux conceptuel, lemma (sémantico-syntaxique), lexème (forme phonologique du mot)

et articulatoire. Nous voyons à nouveau qu'il n'y a pas de consensus parmi les auteurs quant au déroulement temporel, au nombre ou à la nature des traitements successifs dans la littérature.

Dans ce contexte, il est nécessaire que les facteurs qui peuvent influencer sur les performances soient contrôlés au mieux. Il revient donc à l'expérimentateur de les prendre en considération pour éviter toutes variables confondues et pour étudier celles qui ont un effet tangible sur les performances. C'est pourquoi, de nombreuses études se sont penchées sur les différentes dimensions qui peuvent influencer les résultats de diverses tâches cognitives (e.g., en dénomination orale ou écrite, en reconnaissance visuelle). Ce mémoire se focalisera sur la tâche de dénomination orale (continue) d'images. Concernant cette épreuve, des variables ont été étudiées particulièrement par rapport à la vitesse et parfois également à la précision dont un des objectifs était d'offrir un ensemble d'images normé. En effet, chaque aspect de l'objet présenté peut affecter les différentes étapes durant la dénomination (Shao & Stiegert, 2016). Alario et al. (2004) ont identifié et étudié neuf prédicteurs qui peuvent influencer la vitesse en dénomination : la complexité visuelle, l'accord sur l'apparence de l'image (la canonicité de l'image), la familiarité conceptuelle, la valeur de l'image et la variabilité de l'image (l'« imageabilité »), l'âge d'acquisition, la fréquence, l'accord sur le nom (la « codabilité ») et le nombre de phonèmes ou de syllabes. Ces auteurs ont relié ces prédicteurs à des *loci* (fig. 4) même si, pour certains d'entre eux (comme l'âge d'acquisition), la question reste toujours débattue. Ce sont ces facteurs, ainsi que l'apport d'études complémentaires et/ou ultérieures, qui feront l'objet de la section suivante.

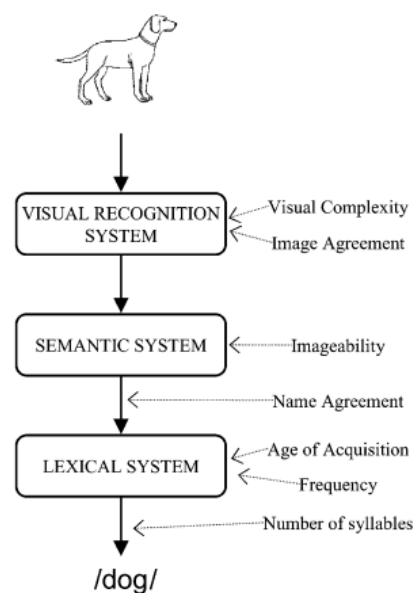


Fig. 4 : Modèle général de dénomination d'images repris par Alario et al. (2004, p.141) avec les *loci* supposés des différents prédicteurs

2. Facteurs influençant la tâche de dénomination orale d'images

a. Facteurs visuels

Les stimuli à la base de la tâche de dénomination sont les images. Ces dernières sont particulières. Comme le mentionne Bonin (2007, p.113), elles sont « par nature idiosyncrasiques. Elles peuvent donc varier sur un nombre important de dimensions ». Dans la littérature, ont été relevées, par conséquent, des caractéristiques intrinsèques à l'image, telles que la complexité visuelle ou l'accord sur l'apparence de l'image. En effet, lorsque l'image est présentée à l'écran, la première étape est de reconnaître ce qui est représenté. Les caractéristiques physiques de l'image peuvent influencer la reconnaissance de l'objet et, par la suite, la dénomination de celui-ci.

La complexité visuelle (*visual complexity*) renvoie au nombre de lignes et à la quantité de détails dans le dessin (Alario & al., 2004 ; Bonin, 2007). Elle peut être évaluée soit subjectivement par une échelle de x points, souvent 5 (1 = dessin très simple à 5 = dessin très complexe), soit objectivement en compressant les fichiers (Rabovsky & al., 2016) ou en prenant la taille du fichier (Székely, 2000 pour les différents types). Les résultats n'ont pas été constants dans les différentes études sur le sujet (Sirois & al., 2006). En effet, les premières études sont arrivées à deux affirmations différentes : les unes que des stimuli plus complexes étaient plus difficiles à traiter, tandis que d'autres n'ont pas obtenu de différences entre les deux stimuli (cf. Alario & al., 1999, 2004). Les études plus récentes vont dans le sens d'une influence de cette caractéristique sur la précision en dénomination. En effet, Székely et Bates (2000) ont démontré par des mesures objectives – ce qui n'a pas toujours été le cas dans les études antérieures – que la complexité visuelle n'avait des effets significatifs que sur la précision et qu'elle n'était pas liée à la fréquence ou l'âge d'acquisition.

Ensuite, l'accord sur l'image (*image agreement*), c'est-à-dire l'accord sur l'apparence de l'image, est « relatif au degré avec lequel les images mentales, que les participants génèrent par rapport aux concepts dénotés par les images, s'accordent avec leur apparence visuelle » (Bonin, 2007). Cela renvoie à ce que l'image représentée s'accorde bien - ou non - à la représentation canonique de l'objet cible stockée en mémoire. Le degré d'accord peut être évalué grâce à une échelle en x point, souvent 5 (1 = correspondance très faible à 5 = très

bonne correspondance). Les études ont démontré qu'une image avec un degré élevé entraîne une latence plus courte. Cela suppose que cet accord influe sur la reconnaissance de l'objet : une image plus proche de notre image mentale de l'objet est dénommée plus rapidement (Alario & al., 2004).

Parmi les autres caractéristiques de l'image, ont également été étudié l'effet de la couleur et des détails de surface sur la reconnaissance de l'objet et, par conséquent, par la suite, la performance en dénomination. Cela prend place dans la discussion sur la reconnaissance des objets, dans laquelle deux visions s'affrontent. Certains auteurs ont une vue « edge-based », à savoir que la forme de l'objet prime dans la reconnaissance. D'autres auteurs ont une vision « surface + edge-based », à savoir que la couleur, la texture et le point de vue, en plus de la forme, contribuent à la reconnaissance.

Rossion et Pourtois (2004) ont comparé les images d'objets dessinés au trait non seulement à des mêmes images avec des nuances de gris et texturés (permettant des détails supplémentaires) mais aussi aux mêmes images en couleur (cf. fig. 5). Une légère amélioration avait déjà été décelée en précision avec les dessins texturés et en nuances de gris. L'apport de la texture seule est plus limité. Il interviendrait particulièrement au niveau de l'accord sur le nom et améliorerait les items qui étaient déjà dénommés plus lentement dans les dessins aux traits. Par contre, les auteurs ont remarqué que pour les items déjà dénommés rapidement, l'ajout de la texture seule ralentirait plutôt et aurait un effet inhibiteur.

Une amélioration plus importante s'est retrouvée à la fois en précision et en vitesse (réduction presque de 100 ms.) avec l'ajout de la couleur, ainsi qu'une amélioration des scores d'accord entre les sujets. Cet apport est pour toutes les catégories mais il est plus large dans le contexte d'une couleur appelée « couleur diagnostic », par exemple l'orange orange et la banane jaune, (Rossion & Pourtois, 2004 ; Paesen & Leijten, 2019) et d'objets similaires au niveau de la structure (e.i., fruits/légumes) – ceci concernant aussi les objets manufacturés.

D'autres études ont également relevé cette influence de la couleur sur le temps de dénomination. Theriault et al. (2009) ont mis en œuvre une expérience avec des images ayant une couleur hautement « diagnostic » comme la citrouille. Ils ont démontré que la couleur influence notamment la dénomination d'objets. En effet, la condition d'incongruence, à savoir une couleur ne correspondant pas du tout à l'objet présenté (cf. fig. 5), ralentissait le temps de dénomination comparé à la condition congruente et à la condition neutre (noir et blanc).



Fig. 5 : Exemples de stimuli manipulés dans les conditions pour tester l'effet de la couleur (à droite, Rossion & Pourtois, 2004 ; à gauche, Therriault & al., 2009).

Plus récemment, Heuer (2016) met en place une expérience pour comparer les photographies *versus* les dessins au trait en noir et blanc avec, comme mesure, les mouvements oculaires (*eye tracking*). Elle obtient des résultats corroborant l'hypothèse que la couleur facilite la reconnaissance de l'image.

Même si l'apport de la couleur a été démontré dans de nombreuses études récentes, elle fait, néanmoins, l'objet de résultats mixtes dans la littérature concernant la couleur « diagnostic ». Dans son étude, Heuer (2016) a démontré, comme Rossion et Pourtois (2004) et d'autres études ultérieures (Bramão, Faísca & al., 2010 ; Bramão, Inácio & al., 2010 ; Laws & Hunter, 2006 ; Mohr, 2010 ; Uttl & al., 2006 ; cités par Heuer, 2016) qu'il n'y avait aucun effet significatif du degré de la couleur diagnostic (élevé ou bas) sur la reconnaissance des images, contrairement à d'autres études (Tanaka & Presnell, 1999, cités par Heuer ; Therriault & al., 2009 partant de cette affirmation) pour lesquelles le degré avait une influence. En effet, ces dernières études avaient conclu que les items avec un degré de couleur diagnostic élevé (associés aux items biologiques) sont reconnus plus rapidement et plus précisément que ceux avec un degré faible (associés, eux, aux objets manufacturés). Cette affirmation provient au départ de l'hypothèse que les items biologiques sont plus difficiles à traiter que les items manufacturés en raison de traits et de détails extérieurs plus nombreux, dont la couleur fait partie (Paesen & Leijten, 2019). Ces résultats hétérogènes peuvent être expliqués notamment par la différence de méthodes pour déterminer la « diagnosticité » de la couleur.

Les photographies en couleur (permettant plus de détails comme la lumière, la texture) se sont révélées également avoir un avantage sur les dessins au trait en diminuant les latences, lors de la tâche de dénomination, notamment dues au supplément de détails externes, d'autant plus si ceux-ci apportent des informations congruentes avec nos représentations en mémoire (e.i., une surface lisse et brillante pour le miroir) et dans le cadre d'objets qui auraient une forme semblable (Heuer, 2016). Cependant, nous y reviendrons lors de la comparaison des différents formats proposés dans les pools, cet apport est mitigé selon les études. Les photographies apporteraient, enfin, pour certains auteurs, une validité écologique dans le sens où elles donnent une description plus réaliste des concepts de la vie quotidienne (Brodeur & al., 2014 ; Moreno-Martinez & Montoro, 2012).

b. Facteurs sémantiques

Parmi les facteurs sémantiques, les auteurs placent la familiarité conceptuelle et la valeur, ainsi que la variabilité, de l'image. La familiarité conceptuelle correspond « au degré avec lequel les participants estiment être en contact avec l'objet représenté par l'image ou dénoté par le mot, ou estiment penser au concept en question » (Bonin, 2007, p.115). Elle peut être évaluée par une échelle en x point, souvent 5 (1 = objet très peu familier à 5 = objet très familier). L'effet de la familiarité conceptuelle ne s'est pas avéré systématique et robuste à travers les études (Alario & al., 2004 ; Bonin, 2007). Par ailleurs, un effet de familiarité peut être minimisé en raison de sa relation très étroite avec d'autres variables comme la fréquence et l'âge d'acquisition, puisqu'un concept qui nous est très familier est souvent un concept appris à un jeune âge et rencontré fréquemment dans la vie quotidienne (Brysbaert & al., 2014 ; cité par Paesen & Leijten, 2019).

La valeur de l'image est « la facilité avec laquelle un mot évoque une image mentale correspondante » (Bonin, 2007, p. 114). En tâche de dénomination, aucun effet robuste n'a été obtenu à travers les études et/ou les analyses statistiques utilisées comme pour la familiarité (Bonin). La variabilité de l'image (*imageability*), elle, renvoie à ce que le mot génère peu ou beaucoup d'images mentales différentes. Elle peut être évaluée aussi avec une échelle à x points, souvent 5 (1 = peu d'images à 5 = beaucoup d'images). Cette variable s'est montrée, par contre, avoir un effet significatif sur les latences en dénomination (Alario & al., 2004) et avoir une contribution importante et significative sur les temps de dénomination que ce soit à

l'oral ou à l'écrit (Bonin, 2007). Cette variable a été expliquée par le nombre de traits sémantiques, c'est-à-dire que les mots avec une « imageabilité » plus haute sont plus « riches » au niveau sémantique (Plaut & Shallice, 1993 ; cités par Alario & al., 2004). C'est pourquoi, a été suggéré un locus sémantique de cet effet (Ellis & Morrison, 1998 ; cités par Alario & al.).

Cette hypothèse de la richesse sémantique a fait l'objet de quelques études. Par exemple, Borman (2011, cité par Bonin & al., 2019) n'a pas trouvé une influence fiable du nombre de voisins sémantiques sur la tâche de dénomination. Récemment, Rabovsky et al. (2016) ont examiné l'influence de la richesse sémantique et de la densité sémantique sur la production orale. Ils ont démontré, dans leur étude, un effet de facilitation sur les temps et la précision de dénomination pour les objets avec un nombre plus important de traits sémantiques. Les auteurs l'expliquent par une augmentation de l'activation de représentations lexicales correspondante. Par contre, la densité de traits inter-corrélationnels, qui reflète la co-activation des groupes lexicaux plus directement, diminue la dénomination. Cela est dû probablement à la co-activation d'un plus grand nombre de compétiteurs qui entraîne une compétition lexicale.

c. Facteurs lexicaux

Parmi les facteurs lexicaux, Alario et al. (2004) reprennent la fréquence, l'âge d'acquisition et l'accord sur le nom. Ce locus ne fait pas unanimité. L'âge d'acquisition fait l'objet de débat. Certains auteurs lui attribuent, comme Alario et al., un locus sémantique tandis que, pour d'autres, ces effets seraient liés à la connexion entre les représentations sémantique et lexicale ou, pour d'autres encore, ils seraient en partie localisés dans le niveau perceptuel (cf. Stadthagen-Gonzalez, 2009, p.1587). De même, pour la fréquence, un locus phonologique, au lieu de lexical, a été proposé par plusieurs études (cf. Boukadi & al., 2016, p.588). L'accord sur le nom aussi peut être considéré sur deux *loci* différents dépendants de la cause explicative d'un accord faible : si l'accord est faible en raison d'une mauvaise identification de l'image, un *locus* possible est le niveau de l'encodage structurel (étape de la reconnaissance de l'objet) ; si l'accord est faible en raison de la disponibilité de plusieurs noms corrects pour le même objet, le *locus* est au niveau de la sélection du lemma (Boukadi & al., p.587), ce qui explique les latences plus longues.

La fréquence peut être objective et subjective. La fréquence objective « correspond pour un mot donné au nombre d'occurrences de ce mot par rapport à un corpus de mots déterminés » (Bonin, 2007, p.115), tandis que la fréquence subjective (appelée aussi familiarité lexicale) est une « estimation subjective de la fréquence de rencontre de tel ou tel mot dans le langage parlé et/ou écrit, estimation réalisée également à l'aide d'échelles » (Bonin, p.116). En contexte de dénomination orale, Oldfield & Wingfield (1965, cités par Bonin, 2007) ont démontré, dans leur étude, un effet de fréquence qui était présent en tâche de dénomination d'images et qui n'était pas présent, en revanche, en tâche de désignation. Par après, d'autres auteurs ont mis également en évidence un effet de fréquence (cf. Székely & al., 2000, p.4). Toutefois, certains ont souligné le fait que ces études n'ont pas pris en compte une autre variable, l'âge d'acquisition des mots (cf. *op. cit.*). C'est pourquoi, ces deux variables ont été prises en compte ensemble, nous y reviendrons après avoir expliqué ce que l'on entend par l'âge d'acquisition.

L'âge d'acquisition (*Age of Acquisition*, abrégé AoA), qui se retrouve de manière solide dans différentes tâches, fait référence à l'âge auquel un mot a été appris et acquis. Il peut être mesuré de deux manières : soit par une estimation de cet âge en question faite par les adultes (on parlera d'AoA subjectif), soit par une analyse des productions recueillies auprès d'enfants (on parlera dans ce cas d'AoA objectif). Les deux mesures sont corrélées de façon élevée (Łuniewska et al., 2016), bien que l'AoA subjectif est intercorrélé avec les autres variables dans une plus large mesure que l'AoA objectif et apporte une mesure moins « pure » (Stadthagen-Gonzalez & al., 2009, p.1588). La conclusion que les mots appris précocement sont traités plus rapidement et plus précisément que ceux appris plus tardivement est ressortie d'un nombre important d'études. Cela pourrait s'expliquer par des connections plus fortes pour autant que les mots acquis plus tôt soient utilisés plus fréquemment (Paesen & Leitjen, 2019).

La fréquence et l'âge d'acquisition sont deux variables fortement liées et elles ont fait l'objet de controverses notamment sur leur relation respective. Quand est prise en compte la corrélation entre ces deux variables, des études ont trouvé des effets solides de l'âge d'acquisition avec aucun effet de fréquence (cf. Sirois & al., 2006), tandis que d'autres ont rapporté des effets significatifs et indépendants des deux variables (Alario & al., 2004 ; Sirois & al., 2006). Des interactions entre les deux ont également été observées par d'autres auteurs :

l'AoA aurait un effet plus prononcé pour les images à basse fréquence (Barry & al., 1997, cité par Alario & al., 1999) ; un effet de fréquence peut être trouvé si l'AoA est plus bas et que la latence de réponse est meilleure avec des mots qui ont un AoA précoce et une fréquence élevée (cf. Paesen & Leitjen, 2019).

La fréquence du mot a également été liée à d'autres variables comme l'accord sur le nom. En effet, LaGrone et Spieler (2006) ont obtenu une relation significative entre ces deux variables et le temps en dénomination (voir ci-dessous) quand nous abordons l'accord sur le nom. LaGrone et Spieler ont observé que la fréquence influence de façon similaire les performances des personnes plus jeunes et des plus âgées mais ils ont également trouvé que la fréquence influencerait légèrement (non significatif) plus les personnes plus âgées (LaGrone et Spieler, 2006, p. 807).

L'accord sur le nom (*name agreement* ou *codability*) est le « degré avec lequel les participants s'accordent sur un label pour référer à une image » (Bonin, 2007, p.113). Il est mesuré en tenant compte du nombre de noms qui peuvent être donnés pour l'image présentée. Si chaque participant donne le même nom à la cible, le taux est de 100% (nombre de participants maximal). Plus des noms différents sont donnés, plus le pourcentage diminue. En plus du pourcentage de participants, les auteurs calculent la statistique H qui correspond à la mesure de l'incertitude (Shannon, 1949 ; cité par Bonin, 2007, p.113) et qui prend en considération la proportion des sujets produisant chaque alternative (Székely & al., 2000, p.5). Alario et al. (1999) ont évalué l'accord sur le nom dans leur étude, par exemple, en demandant de donner le premier mot qui vient à l'esprit.

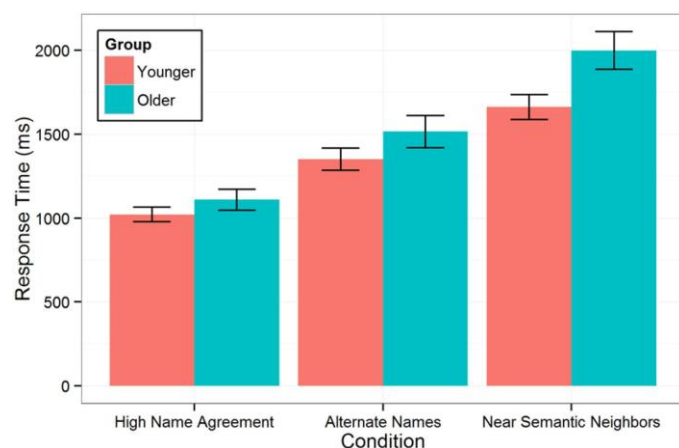
L'accord nom-image s'est avéré dans les études avoir non seulement une influence sur la latence de dénomination – plus l'accord nom-image est bas, plus les latences augmentent - mais aussi être impacté par l'âge.

LaGrone et Spieler (2006) se sont intéressés à l'influence de l'âge sur la récupération, en particulier l'influence de deux variables, l'accord sur le nom et la fréquence du mot. Pour ce faire, ils ont mis en place une tâche de dénomination qu'ils ont proposé à des participants jeunes et plus âgés. Les auteurs observent, comme évoqué ci-dessus pour la fréquence, une relation significative entre l'accord sur le nom, la fréquence du mot et le temps de dénomination que ce soit chez les participants plus jeunes ou chez les personnes plus âgées.

Cependant, ces dernières sont plus sensibles aux variations concernant l'accord sur le nom. Cela s'expliquerait, selon les auteurs, par le fait que la compétition pour la sélection lexicale est relativement haute en l'absence de contexte.

En 2016, Britt et al. avaient comme objectifs de répliquer les effets observés de l'accord sur le nom, d'étudier à quel niveau prend place la sélection (niveau lexico-sémantique ou niveau lexical ou niveau sémantique) et d'examiner l'effet de l'âge dans ce contexte. Ils ont comparé une condition où l'accord est élevé avec deux conditions où l'accord est en bas en raison soit d'images avec des noms multiples (« Alternate Names »), soit des images pour lesquelles les sujets ont produit des réponses liées sémantiquement - mais pas équivalentes, pas de synonymes comme la condition *Alternate Names* – (*Near Semantic Neighbors*). L'expérience 1 (participants de 18-20 ans), qui consistait en une tâche de dénomination, a confirmé que les images avec un accord plus bas étaient nommées plus lentement et moins précisément que celles avec un accord élevé. Les auteurs ont observé que, dans la condition *Alternate Names*, les performances étaient substantiellement plus rapides et plus précises que dans la condition *Near Semantic Neighbors*. L'expérience 2 (participants jeunes de l'expérience 1 comparés à des participants plus âgés de 50-81 ans) a pour but, elle, d'examiner l'effet de l'âge sur l'expérience 1. Les personnes plus âgées sont plus lentes en général mais les auteurs constatent que la différence liée à l'âge augmente « monotonically with increasing selection demands » (Britt & al., 2016, p.9), c'est-à-dire qu'elle est très petite dans la condition d'accord élevé, plus grande dans la condition *Alternative Names* et très grande dans la condition *Near Semantic Neighbors* (cf. fig. 6). Les personnes plus âgées auraient donc, en dénomination, plus de difficultés quand la sélection lexicale et sémantique augmente. Les auteurs remarquent qu'en plus d'un effet principal de l'âge, s'ajoute une interaction entre le groupe d'âge et la condition. La présente étude, selon les auteurs, va dans le sens d'une sélection par modèles de compétition (les réponses alternatives font concurrence et entraînent des réponses plus lentes) et montre des processus de sélection distincts aux niveaux lexical et sémantique (Britt & al., pp.14-15).

Fig. 6 : Histogramme des résultats obtenus par Britt et al. (2016)



L'étude récente de Madden, Sale et Robinson (2019 b) a l'objectif de répliquer deux conclusions mises en avant dans les études antérieures. La première affirmation est donc le fait que la vitesse de dénomination est impactée par le degré d'accord nom-image qui viendrait d'une compétition plus élevée dans la sélection du nom cherché. Plus particulièrement, les auteurs se sont intéressés à l'étude de Vitkovitch et Tyrell (1995, cité par Madden & al., p. 1374) dans laquelle ils ont étudié les sources possibles de désaccord provenant soit de noms corrects mais parfois multiples (ex. : divan, sofa, canapé), soit de l'utilisation d'abréviations et d'élaborations (ex. : télé, TV, télévision), soit de l'usage incorrect de certains noms (ex. : araignée pour fourmi). Ces auteurs avaient trouvé que chaque source n'avait pas le même poids, ce qui n'a pas été repris dans certaines études ultérieures sur le sujet. Les abréviations et les élaborations n'avaient pas un poids conséquent sur le ralentissement malgré le taux d'accord bas. Par contre, la dénomination d'images avec des noms multiples ou des noms incorrects s'est révélée plus lente, le premier cas de figure augmentant la compétition dans la sélection alors que le deuxième augmentant le temps de reconnaissance de l'objet. Pour cette première hypothèse, Madden et al. ont confirmé les résultats des études antérieures en comparant deux groupes (18-35 ans et 60-80 ans), c'est-à-dire les latences augmentant pour les images avec un accord faible contrairement à celle avec un accord haut. Néanmoins, ils n'ont pas obtenu les mêmes résultats que Vitkovitch et Tyrell concernant la source des divergences : les images avec des noms multiples entraînent des performances plus lentes mais celles faisant l'objet d'abréviations et d'élaborations aussi dans une moindre mesure.

La deuxième conclusion reprend la sensibilité de cette caractéristique à l'effet de l'âge que LaGrone et Spieler (2006) ont observé dans toutes les conditions de dénomination conformément aux études précédentes (voir le premier chapitre) mais cet effet varie d'une condition à l'autre : les personnes plus âgées ont montré des performances semblablement plus lentes pour les items avec un accord élevé et pour les items avec un accord faible ayant plusieurs noms possibles ; l'effet est plus marqué pour les items avec un accord bas dû aux abréviations et aux élaborations.

Cette étude montre l'importance de contrôler non seulement le degré d'accord sur le nom mais aussi la source qui explique le désaccord.

d. Facteurs phonologiques

La longueur du mot est mesurée par le nombre de phonèmes ou le nombre de syllabes. Celle-ci n'a pas été clairement établie comme jouant une influence sur la production du mot. En effet, des études ont démontré qu'il s'agissait d'un facteur pertinent, alors que d'autres n'ont rapporté aucun effet. L'encodage phonologique d'un mot implique de récupérer les différents segments ou syllabes qui le composent (Alario & al., 2004). Ceci a mené à l'hypothèse que les mots les plus longs prennent plus de temps à être encodés que les mots les plus courts. Les résultats des études sont inconsistants. Alario et al. (2004) n'ont obtenu aucun effet significatif concernant le nombre de phonèmes sur la vitesse de dénomination et un effet très léger concernant le nombre de syllabes. Les auteurs émettent l'hypothèse que l'effet du nombre de syllabes dépende des conditions de l'expérience et de la fréquence (à une fréquence basse, un effet est observé). C'est également ce que rapportent LaGrone et Spieler (2006) : les segments phonologiques pour les mots fréquents sont assemblés plus rapidement et plus précisément que les segments phonologiques pour les mots moins fréquents (LaGrone et Spieler, p.804).

e. Autres variables

D'autres variables ont été ajoutées dans quelques études de manière éparse. Elles ont été moins prises en considération dans la littérature.

Les variables concernant les dimensions motrices de l'objet représenté ont été peu étudiées. Parmi celles-ci, la *manipulability* et l'interaction corps-objet peuvent être citées.

La *manipulability* a été ajoutée dans certaines études. Toutefois, sa définition varie selon les auteurs qui ont choisi de l'étudier. Guérard et al. (2015) ont voulu affiner et regrouper les définitions en quatre dimensions de la *manipulability*, à savoir la facilité avec laquelle les objets peuvent être saisis, déplacés et mimés, ainsi que le nombre d'actions possibles dans leur exécution. Les auteurs ont aussi examiné l'effet de ces dimensions sur les latences en dénomination. Antérieurement, la base des études était des dessins au trait. Dans leur étude, Guérard et al. sont partis de photographies par souci de validité écologique. Ces quatre dimensions peuvent être regroupées en deux catégories. Il y a, d'une part, les actions

volumétriques de l'objet, c'est-à-dire saisir et déplacer l'objet ; d'autre part, les actions fonctionnelles, c'est-à-dire la facilité de mimer l'usage de l'objet. Les normes ont été collectées avec la participation d'étudiants français canadiens sur des échelles de 7 points (à part le nombre d'actions qui allait de 0 à 6+). Les auteurs ont constaté que les objets qui étaient faciles à saisir et à mimer, ainsi que ceux qui ont un nombre d'actions plus grand, étaient reconnus plus rapidement, résultats qui sont constants avec d'autres études antérieures (Guérard & al., 2015, p.448).

Très récemment, Lorenzoni et al. (2018) ont répliqué l'étude de Guérard et al. (2015) en analysant les résultats avec une autre procédure statistique afin d'obtenir des résultats plus précis. Ils ont obtenu des effets significatifs de la *manipulability* pour trois dimensions (saisir, mimer et déplacer). Ils n'ont aussi observé aucun effet de cette variable sur la catégorie des objets naturels. Les auteurs concluent de leurs données que la *manipulability* est une variable critique qui affecte la tâche de dénomination d'objets.

Dans leur étude, Bonin et al. (2013) ont ajouté, en plus d'autres variables, l'interaction corps-objet (body-object interaction, BOI), à savoir la facilité avec laquelle un corps humain peut physiquement interagir avec le mot référent (Bonin, p.733). Les participants de langue russe devaient estimer les mots sur une échelle de 7 points (1 = interaction corps-objet faible à 7 = interaction corps-objet haute). Les résultats de leur étude n'ont pas révélé une contribution fiable de cette variable sur le temps de dénomination.

Dans leur pool en langue espagnole, en plus d'autres variables récoltées, Moreno-Martínez et Montoro (2012) ont ajouté la *typicality* qui est comment un membre est représentatif de sa catégorie. Les auteurs l'ont évaluée avec une échelle de 5 points (1 = pas prototypique du tout à 5 = très prototypique). Dans le même ordre d'idées, Brodeur et al. (2014 ; déjà en 2010 cité par Brodeur & al.) ont ajouté l'accord sur la catégorie dans leur construction d'ensembles de photographies (voir ci-dessous). L'accord sur la catégorie est d'évaluer si l'item présenté est représentatif de leur catégorie. Les participants devaient classer les objets dans la catégorie qu'ils estimaient la plus appropriée. Les auteurs ont aussi ajouté une autre variable peu prise en compte, à savoir l'accord sur le point de vue, pour laquelle les participants devaient dire sur une échelle de 5 points si l'objet présenté avait la position qui était associée à leur vision de cet objet. Cette dernière appelée « orientation » par Heuer (2016) qui souligne, en effet, qu'elle a été très peu prise en considération.

3. Ensembles d'images normés

a) Au préalable

Les auteurs, bien conscients des influences des variables, ont proposé des ensembles de stimuli pour lesquels certains facteurs ont été étudiés. Cependant, ces déterminants peuvent être différents à travers les langues, les cultures et même les générations (Sirois & al., 2006). Par exemple, un nom renvoyant à une image peut varier selon les langues (Sanfeliu et al. 1996 ; cité par Bonin, 2003) ou un objet dans une culture donnée peut être jugé différemment dans une autre (Sirois & al.). C'est pourquoi, des recherches se sont attelées à construire un ensemble d'images normés. La normalisation des images se fait par un échantillon représentatif d'individus qui évaluent les images et leur nom sur un certain nombre de variables (Brodeur & al., 2014). Même si les variables ont été examinées dans des langues séparées, des études se sont penchées sur une comparaison des normes obtenues entre les langues ou ont comparé leurs résultats (cf., par exemple, Bonin, 2003, 2013 ; Alario & Ferrand, 1999 ; Bates & al., 2003 ; Sirois, 2006 ; Brodeur & al., 2012 ; Łuniewska, Haman & al., 2016, Łuniewska, Wodniecka & al., 2019 pour l'AoA) et des bases de données multilingues existent comme la European Standardized Picture Pool for Oral Naming (Protocole européen de dénomination orale d'images, PEDOI ; Sirois & al.). Les pools d'images proposées collectent principalement les données dans une population de personnes jeunes et instruites.

Ces pools d'images ont donc été effectués pour différentes langues (anglais, espagnol, néerlandais, italien, japonais, chinois, français, ...), différentes populations (enfants-adultes) et différentes cultures. L'ensemble d'images de référence et pionnier est celui de Snodgrass et Vanderwart en 1980 (appelé SV ; cité notamment par Bonin, 2003). Cet ensemble comprend 260 dessins en noir et blanc (représentant des objets communs) normés pour l'anglais américain sur des jeunes adultes en prenant en compte quatre variables, à savoir l'accord sur le nom, l'accord sur l'apparence de l'image, la familiarité conceptuelle et la complexité visuelle. A leur suite, des ensembles d'images en noir et blanc ont fait l'objet de normes dans différentes langues. Cependant, les études actuelles se penchent également sur la possibilité de mettre à la disponibilité des ensembles d'images en couleur et, plus récemment, des photographies. C'est l'ambition de Brodeur et al. (2014) qui ont proposé un nouvel ensemble

de 930 photographies en couleur – qui complète celui de 538 photographies publié en 2010 – normalisées sur plusieurs variables (accord sur le nom, familiarité, complexité visuelle, accord sur la catégorie, accord sur l'image, *manipulability*). Cela porte la *Bank of Standardized Stimuli* (BOSS) à 1468 photographies en couleur normées. Cette banque de photographies comprend des concepts variés et provenant de différentes catégories. Elle propose aussi 1179 photographies non-normées avec d'autres exemples des concepts normés ou de concepts normés sur un autre point de vue. Les auteurs ont choisi de prendre en considération non seulement les différences entre les catégories mais aussi les différences entre les hommes et les femmes, ce qui n'avait pas été pris en compte jusque-là dans les études. Ceci s'est avéré intéressant, puisque les auteurs observent que, pour certaines catégories, une différence entre les genres sont constatées. Par exemple, les femmes ont plus de difficultés à reconnaître et à dénommer les outils (même si cette catégorie est moins bonne aussi pour les hommes, la différence est significative entre les deux sexes). Les variables aussi ont subi l'influence du genre. Par exemple, les femmes ont des scores plus haut pour la complexité visuelle ou la *manipulability*, alors que les hommes ont des scores plus élevés pour le point de vue de l'objet (peut-être parce que, selon les auteurs, le choix des images a été fait préalablement par un homme). Les participants sont des locuteurs anglais qui sont âgés de 18 à 55 ans. Cette banque d'images a été normée aussi en français canadien dans le but de comparer les normes dans les deux langues (mais dans le même contexte culturel, à savoir canadien). Les auteurs obtiennent une bonne concordance en général entre les deux normes. D'autres ensembles de photographies ont été étudiés en espagnol (Moreno-Martínez & Montoro, 2012) et en néerlandais (Shao & Stiegert, 2016).

b) Pools disponibles en français

Concernant le français, Alario et Ferrand (1999) ont proposé un ensemble de 400 dessins aux traits repris d'autres ensembles. En effet, leur set vient d'une part de celui de Cycowicz et al. (1997), qui a été établi sur des échantillons d'enfants de 8-10 ans, et d'autre part a été repris en partie de celui de Snodgrass et Vanderwart (1980). Des images, cependant, ont été remplacées en raison du contexte culturel (américain *versus* français). Cet ensemble d'images a été standardisé sur l'accord sur le nom, l'accord sur l'apparence de l'image, la familiarité, la complexité visuelle et l'âge d'acquisition. La fréquence objective du mot est également disponible, ainsi que le mot correspondant. Leur ensemble a été proposé à 173

étudiants en psychologie dont la langue maternelle est le français. Ils ont suivi la procédure des deux études sources. Selon leur estimation, le meilleur prédicteur est l'AoA.

En 2003, Bonin et al. ont proposé un ensemble de 299 dessins aux traits noirs et blancs avec des normes pour les recherches avec des adultes de langue française. Leur objectif était notamment de compléter la base d'images proposée par Alario et Ferrand (1999) afin que les chercheurs ne soient pas pris de court et ne doivent pas aller chercher des images supplémentaires dont les normes ne seraient pas disponibles. Les variables contrôlées sont l'accord sur le nom, l'accord sur l'apparence de l'image, la variabilité de l'image, la familiarité, la complexité visuelle et l'âge d'acquisition. Dans l'optique de combiner les deux banques d'images, les auteurs ont suivi les procédures utilisées par Alario et Ferrand (1999). Leur ensemble a été normé sur 120 étudiants en psychologie (de 17-33 ans ; 8 hommes et 114+28 femmes) dont la langue maternelle est le français.

A leur suite, en 2004, Rossion et Pourtois ont proposé une base de 260 dessins repris de SV normés sur l'accord sur le nom, la familiarité, la complexité et les jugements d'imagerie. Ils ont comparé cet ensemble avec ces images-mêmes d'une part en nuances de gris et avec des détails en surface, d'autre part en couleur. Les participants étaient des étudiants en psychologie (18-22 ans), initialement 214 répartis en quatre groupes (pour chaque variable). Vingt étudiants étaient recrutés pour chaque condition (parfois moins en raison d'écartement de sujets non valides).

En 2006, Sirois et al. ont proposé 388 dessins en noir et blanc adaptés du protocole européen de dénomination orale d'images (PEDOI) qui est un pool d'images multilingue standardisé (le français est également disponible). Cette banque d'images est proposée aux chercheurs qui ont pour participants des personnes dont la première langue est le français (canadien). Contrairement à leurs prédécesseurs qui avaient comme participants des étudiants, les auteurs ont choisi des participants de tout âge répartis en groupes (18-39, 40-59 et 60-85) et de deux niveaux d'éducation (bas/haut). Les variables étudiées sont l'accord sur le nom, la familiarité de l'objet, la complexité visuelle et l'âge d'acquisition.

Une comparaison a été faite par Sirois et al. (2006) entre les pools français, à savoir celui de Alario et Ferrand (1999), celui du PEDOI (Kremin et al., 2003) et le leur, afin d'évaluer l'impact culturel sur les normes. Les auteurs ont pris 155 items communs. Ils n'ont pas pu

intégrer l'ensemble de Bonin et al. (2003) car ils n'avaient pas assez d'items en commun (seulement 41 items communs). Les résultats (cf. fig. 7) obtenus ont montré que la culture a un impact significatif sur l'accord du nom et qu'elle aurait un impact différent sur chaque item. Cependant, la structure corrélationnelle reste similaire d'un groupe culturel à l'autre.

Table 5
Significant Correlations Among the Measured Variables in the French and Canadian French Samples

	NA	AoA	Fam	VC
Sirois vs. A&F	.376	.829	.898	.758
Sirois vs. Kremin	.527	.836	.928	.783
Kremin vs. A&F	.485	.867	.953	.700

Note—NA, name agreement; AoA, age of acquisition; Fam, familiarity task; VC, visual complexity; A&F, results from Alario & Ferrand (1999); Kremin, results from Kremin et al. (2003).

Fig. 7 : Tableau avec les données des corrélations effectuées sur les différents ensembles en français (Sirois & al., 2006, p.304)

c) Différences entre les formats

Parmi les différents formats proposés pour les images (dessins noir et blanc, dessins en couleur et photographies), certaines conclusions émises dans la littérature soulignent que la précision en dénomination orale est plus basse face à des photographies que face à des dessins des mêmes objets (Shao & Stierger, 2016). Une explication avancée serait que l'accord sur le nom en général est plus haut dans les ensembles en noir et blanc, comme ceux de SV et de Bonin et al. (2003), que dans la banque de photographies. En effet, les photographies apportent, comme déjà évoqué, plus de détails et de la couleur poussant les participants à se fier aux traits idiosyncratiques et, par conséquent, facilitant le nombre de noms différents (O'Sullivan & al., 2012 ; cité par Bonin & al., 2019).

En tâche de dénomination, Bonin et al. (2019) démontrent que les dessins en noir et blanc et les dessins colorés permettent également d'apprécier les déterminants les plus importants de la vitesse de dénomination. Les dessins en couleur augmentent tout de même légèrement cette vitesse en général. Par contre, les dessins en nuances de gris est moins adapté à la tâche de dénomination en temps réel, les résultats étant moins affirmés et stables (en particulier, accord sur l'image). Ils représenteraient les objets de façon plus atypique.

4. Conclusions

Parmi les neuf prédicteurs évalués par Alario et al. (2004), la variabilité de l'image, l'accord sur l'apparence de l'image, l'âge d'acquisition et l'accord sur le nom influencent significativement les latences en dénomination que ce soit à l'oral ou à l'écrit (Bonin & al., 2002). L'accord sur le nom s'est révélé un facteur déterminant, prédicteur (Alario et Ferrand, 1999) et constant dans plusieurs études. Les résultats vont dans le sens d'un facteur clé dans le temps de latence. L'âge d'acquisition, quant à lui, est également un facteur déterminant pour la vitesse de dénomination. Dans les études plus récentes, des variables supplémentaires sont utilisées mais elles n'ont pas encore fait l'objet d'études spécifiques.

Différents formats d'images sont disponibles pour les expériences qui mettent en place une tâche de dénomination d'images. Les auteurs peuvent choisir entre de simples dessins en noir et blanc, des dessins sur des nuances de gris et avec des détails plus fournis, des dessins en couleur ou des photographies. Les ensembles normés sont de plus en plus nombreux et disponibles pour plus en plus de langues. Brodeur et al. (2014) en avait recensé au moins douze entre 2000-2009 et neuf depuis 2010. Néanmoins, les auteurs doivent sélectionner les images en fonction des objectifs poursuivis et tenir compte des différents apports de chaque format. En effet, les dessins en noir et blanc peuvent être choisis dans l'objectif d'avoir une description simple et prototypique des concepts sans détails supplémentaires (Brodeur, 2014) qui, nous l'avons vu, peuvent influencer sur la latence. Les photographies peuvent être choisies dans une optique plus écologique.

III. MOYENS POUR MESURER LA LATENCE

1. Revue des moyens proposés dans la littérature

La latence en dénomination prend une place centrale dans l'interprétation des résultats obtenus, notamment, comme décrit dans le premier chapitre, face à l'effet de l'âge. Elle permet, en effet, d'accroître la sensibilité de la tâche de dénomination. C'est pourquoi, les dispositifs pour la mesurer de la façon la plus précise et la plus fiable sont devenus un enjeu important dans la littérature.

Plusieurs moyens ont été examinés au fil des années. Rastle et Davis (2002) ont fait un bref historique de ce qui avait été mis en place jusque-là. La fin des années 1800 mettait en place des dispositifs tels qu'une « clé labiale » (« lip key ») qui détecte les mouvements des lèvres ou qu'une « clé respiratoire » (« breath key ») qui, elle, détecte l'expulsion d'air. De tels moyens ont été abandonnés dans le début des années 1900 à cause non seulement des difficultés pratiques de mise en place mais surtout à cause d'une détection non égale des différents phonèmes. Depuis, plusieurs autres moyens, que ce soit par un système de clé vocale « traditionnelle », considérée comme des méthodes *hardware-based* (Roux & al., 2017) ou par le biais de programmes plus développés, appelés pour la plupart des solutions *software-based* (Roux & al.), ont été proposés ou revus par les auteurs.

Les principaux dispositifs conçus vont être passés en revue séparément dans la suite de ce chapitre afin d'avoir une perspective temporelle, une filiation de certains dispositifs entre eux et l'apport de chacun (nouveau ou adaptation). Dans un second temps, la comparaison de certains moyens de mesure entre eux sera développée.

Mais auparavant, il faut souligner un point important dans la conception des dispositifs pour mesurer la latence : mesure-t-on selon un mode *online* ou selon un mode *offline*. Cela signifie que la mesure du temps de réponse se fait soit pendant l'expérience (online), soit en dehors de l'expérience à l'aide d'un fichier audio (offline). Certains programmes proposent les deux.

a) Analyse manuelle

La procédure appelée « gold standard » (Jansen & Watter, 2008 ; Roux & al., 2017) et considérée comme la plus précise et la plus fiable jusqu'à présent est l'analyse manuelle (appelée aussi, selon les auteurs, codage manuel, annotation manuelle, évaluation humaine, inspection manuelle etc.). Elle consiste à examiner la forme d'onde visuellement (parfois aussi auditivement) grâce à des programmes comme Audacity (Mazzoni) ou comme Praat (Boersma & Weenink) pour établir très précisément le début de la production du mot. Cette technique demande donc d'inspecter l'onde de la réponse vocale et d'isoler manuellement l'intervalle entre le début de l'apparition du stimulus jusqu'à la réponse vocale du participant. Même si l'apport des programmes a facilité les problèmes techniques (enregistrement, visualisation, manipulation des données) qui pouvaient être rencontrés, elle reste encore coûteuse en temps. Par ailleurs, Schillingmann et al. (2018) soulignent que la fiabilité de cette analyse manuelle est souvent incertaine. En effet, des études se sont penchées sur des biais subjectifs et autres erreurs de mesure (Green & Swets, 1966 ; Morrow & al., 2016 ; cités par Roux & al., 2017) qui peuvent exister, même si les niveaux d'accord entre les évaluateurs sont habituellement relativement hauts (Roux & al., p.1864).

Truffert (2014), dans l'intention de réplication de l'étude de Verhaegen et Poncelet (2013), a utilisé cette méthode. L'analyse des items cibles a été réalisée, en effet, via le programme Audacity afin de relever le temps de latence pour produire chaque item sur un spectrogramme. Le temps de latence a été défini de la fin d'un signal sonore émis jusqu'au début de la production de la cible. Cet intervalle est noté en millisecondes.

b) Clé vocale

Certaines études utilisent un dispositif appelé « voice key » (« clé vocale ») pour mesurer le temps de réponse. Rastle et Davis (2002) ont calculé qu'entre 1995 et 1999, 95% des articles publiés dans le *Journal of Experimental Psychology ; Human Perception and Performance* ont utilisé la clé vocale. Notamment dû à son large usage, cette clé vocale a fait couler beaucoup d'encre dans la littérature. En effet, elle présente des avantages, son caractère automatique, pratique et peu « énergivore » en temps, mais, malheureusement, également des désavantages, notamment sa sensibilité au bruit (environnant ou ne relevant pas de la parole

du participant, comme les bruits des lèvres ou les raclements de gorge) et, ainsi, le risque de faux départ. Des études se sont penchées sur ces inconvénients pour pouvoir les limiter au maximum.

Avant de détailler les conseils émis, un petit détour est nécessaire pour expliquer en quoi elle consiste sous sa forme la plus basique et la plus traditionnelle. En effet, nous le verrons, elle a fait l'objet d'adaptations et/ou d'améliorations au fil du temps. La clé vocale est, selon Kessler et al. (2002, p.145), « *a device that determines voice onset in real time* ». Le principe de ce moyen électronique en temps réel est le suivant : elle est connectée à un microphone et elle convertit la pression du son en volt ; après avoir équipé l'ordinateur avec la clé vocale, le microphone est contrôlé pour que, dès que la pression du son atteint un niveau prédéfini par l'expérimentateur, la clé vocale s'enclenche ; enfin, le signal est envoyé à l'ordinateur qui enregistre le temps en milliseconde. Ainsi, la clé vocale permet de mesurer la latence grâce à l'intervalle entre la présentation du stimulus et la réponse vocale du sujet qui met en marche la clé.

Cependant, elle a des inconvénients qui ont été étudiés en détails notamment par Kessler et al. (2002). Dans leur étude, ils ont demandé aux participants de lire des mots le plus précisément et le plus rapidement possible dans une tâche de dénomination rapide. Selon les auteurs, plus que de faux départs dus à un bruit extérieur ou même dus au bruit de la respiration en raison d'un mauvais réglage du seuil, la précision du temps mesuré peut varier en fonction du premier phonème – voire du deuxième phonème – du mot, ce qu'on appelle le « biais phonétique ». Par exemple, les auteurs soulignent les différences entre les fricatives qui entraînent une pression du son moindre que les voyelles. Les auteurs ont conseillé des alternatives pour y remédier : utiliser une tâche de dénomination différée ou une tâche similaire contrôlée (p.148) ; balancer dans le design de l'étude les stimuli par des critères phonétiques (p.148) ; adapter l'attitude du participant en lui demandant de parler à voix haute de façon claire et la plus constante possible, en ne bégayant pas, en essayant de faire le moins de bruit annexe, en dénommant le plus rapidement possible sans traîner sur le premier phonème (p.166) ; recorriger le temps de réponse en fonction du phonème initial (par exemple, dans leur étude, le /s/ est détecté 39 ms plus tard) (p.166) ou revérifier sur l'onde quand l'expérimentateur estime un faux départ (p. 168).

Auparavant, d'autres auteurs s'étaient déjà penchés sur la question. Kawamoto et Kello (1999 ; cité par Jansen & Watter, 2008) avaient conclu dans leur étude que les mots avec un début complexe, c'est-à-dire commençant par un groupe de consonnes, sont produits plus rapidement que ceux avec un début simple, c'est-à-dire commençant par une consonne simple. Par contre, Jansen et Watter indiquent que l'étude antérieure de Frederiksen et Kroll (1976, cité par Jansen et Watter) a démontré l'inverse. La différence entre ces deux études réside justement dans le moyen employé pour mesurer la latence : les premiers ont utilisé un logiciel qui la mesure à l'aide d'un algorithme offline et les seconds une clé vocale.

Face aux résultats des études antérieures, notamment concernant le premier phonème ou la classe du phonème, Rastle et Davis (2002) ont décidé de répéter l'expérience de Kawamoto et Kello, ainsi que d'évaluer plus particulièrement le critère du premier phonème. Pour ce faire, ils ont mis en place une épreuve de lecture à voix haute de mots monosyllabiques. Ils ont recueilli les réponses en utilisant le DMDX avec deux types de clés : une clé vocale fonctionnant pendant l'expérience selon un seuil déterminé (*simple threshold voice key*) et une autre fonctionnant à partir du fichier enregistré et prenant en compte aussi l'intensité minimum et l'intensité plus basse (*integrator voice key*). Ils ont vérifié leur mesure en examinant l'onde (analyse manuelle). Suite aux résultats obtenus, ils proposent plutôt d'apparier les conditions expérimentales sur le début d'une syllabe complète, c'est-à-dire tous les segments phonologiques précédant la voyelle (p.313), qui contrôle de façon plus adéquate les erreurs dues à la clé vocale.

L'impact de ces variations dépend aussi de la tâche et de son objectif. Par exemple, dans leur étude sur les symboles hiragana en Japonais, Yamada et Tamaoka (2003) ont aussi exposé leur problème face à l'utilisation d'une clé vocale et le biais phonétique qui s'ensuit puisque leurs items sont composés de syllabes et non d'unités linguistiques d'un niveau plus haut comme un mot.

Abrams et Jennings (2004) avec « VoiceRelay » proposent un complément, plus qu'une adaptation, à la clé vocale. Le souci des auteurs était principalement d'offrir une alternative moins chère pour enregistrer le temps des réponses vocales obtenues par ces clés. En effet, il faut souvent acheter une « external response box » ou des programmes qui créent des expériences supplémentaires et qui s'avèrent coûteux (ex. : E-Prime), alors que VoiceRelay fonctionne avec le système Windows d'un ordinateur standard (microphone et

carte son). Ce logiciel permet non seulement de garder la précision des clés vocales traditionnelles – dépendant du processeur, de la carte son ou encore du système d'exploitation comme les auteurs ont pu le démontrer dans leur test de VoiceRelay - mais aussi d'avoir une flexibilité de volume, ainsi qu'un contrôle de la sensibilité. Plusieurs paramètres sont utilisés pour régler la clé vocale : le volume du micro, le volume du seuil – important dans la détection de la parole - et la durée maximale pour la recherche audio en millisecondes (le programme se fondera sur cette valeur, à défaut 5000 ms, pour rechercher l'input audio avant d'indiquer que la réponse est manquante). Les modifications apportées permettent d'enregistrer les réponses vocales réelles et le temps du début de la parole. Les auteurs sont tout même bien conscients des imprécisions dans les mesures des clés vocales relevées dans les études antérieures. Cependant, ils considèrent l'analyse manuelle coûteuse en temps et sujette à l'erreur humaine. Pour eux, l'imprécision des clés vient de la limite des ordinateurs qui peuvent être « lents ». Foster et Foster (2003) ont également soulevé le problème du rafraîchissement des ordinateurs dans leur présentation du DMDX (cf. ci-dessous). Abrams et Jennings ont calculé qu'une interface Windows standard ou un Visual Basic standard est limité approximativement à 55 ms. C'est pourquoi, il faut un timer de haute résolution comme le QueryPerformanceCounter utilisé par Voice Relay. Enfin, Abrams et Jennings ont conçu une interface assez simple avec une fonctionnalité « plug and play » (Abrams & Jennings, 2004), c'est-à-dire « prête à l'emploi », qui facilite l'utilisation du logiciel.

En 2005, Tyler et al. ont également analysé le problème de la clé vocale. Etant conscients aussi que l'analyse manuelle consomme beaucoup de temps, ils mettent en avant dans leur étude une clé vocale améliorée qu'ils appellent « delayed trigger voice key » (p.139), abrégée en DTVK. Celle-ci a deux paramètres de plus par rapport à la clé traditionnelle : « minimum signal duration » et « minimum silence duration » (*op. cit.*). Le premier paramètre, la durée minimale du signal, a été intégré pour pouvoir enregistrer un seuil très bas, juste au-dessus de celui du bruit ambiant et, ainsi, éviter tous les bruits qui n'appartiennent pas au « discours ». Le deuxième paramètre, la durée minimum du silence, permet d'établir un seuil au-dessus duquel la clé est déclenchée, ce qui permet de tenir compte de l'amorce du mot par exemple. Un autre point intéressant de ce paramètre est que la fréquence fondamentale de la personne (homme/femme) est prise en compte. Les auteurs ont testé cette nouvelle clé dans deux expériences mais seulement dans des mots monosyllabiques. Donc, dans le cas des mots polysyllabiques, il faut encore établir les

différents seuils idéaux. Ce modèle de clé proposé rejoint celles que l'on appelle intégrées (cf. ci-dessus).

En 2008, enfin, Duyck et al. ont proposé une autre clé, appelée NEVK (pour *noise elimination voice key*). Les auteurs sont partis de deux groupes de problèmes liés aux clés vocales. Le premier est la variabilité dans le cadre de l'expérience, la comparabilité entre les participants et les études. Il s'agit principalement de problèmes liés aux seuils prédéfinis et niveaux d'amplitude difficiles à mesurer et quantifier de façon très précise, aux conditions matérielles (e.g., type de microphone, distance entre le micro et le participant) ou aux caractéristiques du participant (e.g., genre, voix, capacités) qui peuvent interagir et peuvent influencer différemment la détection des temps. Le second groupe est celui, déjà mentionné, des biais phonétiques où les auteurs soulignent que le problème du délai n'est pas limité à quelques phonèmes particuliers mais qu'il est aussi lié à une grande variabilité dans les clés vocales pour des phonèmes différents. Ils remarquent que l'effet de ces biais peut varier à travers les participants. Pour palier les problèmes soulevés, les auteurs ont proposé une nouvelle clé qui va à la fois procéder à une amplification massive du signal de la clé (appelé *Analogue Part*) et rejeter la plupart des sons amplifiés qui ne sont pas de la parole et qui ne sont caractérisés par aucune oscillation après 50 ms du départ du signal (*Digital Part*), partie qui évite les faux départs. Néanmoins, les auteurs indiquent que l'expérience doit se passer dans un environ calme, bien isolé acoustiquement, car les bruits électroniques risquent d'augmenter le bruit du circuit électrique de la clé, dépasser le seuil fixé et, ainsi, produire des faux départs.

c) DMDX, CheckVocal et Visual DMDX

En 2003, Foster et Foster ont présenté un système plus développé qui mesure notamment le temps de réponse, le DMDX. Il s'agit d'un programme qui fonctionne sur Windows. Ce programme est une extension du DMASTR développé à l'Université Monash en Australie en 1975 et continuant à évoluer depuis. Le nom DMDX a été créé en référence d'une part au système DMASTR (DM) et d'autre part au DirectX (DX) qui est nécessaire pour le déroulement de l'expérience. Les auteurs définissent leur programme comme étant conçu pour « precisely time the presentation of text, audio, graphical, and video material and to enable the measurement of reaction times » (Foster & Foster, 2003, p.116). Ce programme

permet donc de présenter des stimuli de différentes modalités pour des expériences, de collecter les réponses des participants et de mesurer leur temps de réponse. Les latences sont enregistrées par le composant « Vox » intégré dans le programme pour éviter de recourir à un dispositif externe. L'expérimentateur peut spécifier le seuil d'intensité pour le déclenchement (comme pour une clé vocale traditionnelle). Un fichier audio est également disponible pour permettre la vérification par la suite.

Protopapas (2007) reconnaît les qualités d'un tel programme : la flexibilité pour les différents types d'expérience et la simplicité pour enregistrer les réponses vocales qui se font par un microphone et un fichier audio individuel. Néanmoins, pour lui, ce genre de dispositif se prête moins à des expériences de dénomination. En effet, il indique que l'expérimentateur doit, pour ce type de recherche, réécouter chaque fichier audio des participants pour encoder manuellement la précision des réponses (correctes, incorrectes ou manquantes). Cela est très énergivore en temps et cela entraîne un risque d'erreurs plus important – le chercheur doit associer les items du fichier audio et la réponse encodée. Il doit également vérifier que le mécanisme de déclenchement s'est fait correctement.

Garaizar et Reips (2015) mettent en avant aussi la flexibilité de ce programme, ainsi que sa fiabilité, sa précision en millisecondes, sa puissance et sa disponibilité sans coût. La justesse et la précision dans la présentation des stimuli et dans la collecte des réponses est optimisée par des technologies multimédias les plus appropriées et des mécanismes pour mesurer le temps, comme l'un des moyens supplémentaires disponibles, le TimeDX permettant des ajustements. Paradoxalement, la puissance dont dispose le DMDX rend le cadre de l'expérience assez lourd en raison de son interface. En effet, le DMDX a été défini sur une syntaxe compatible avec son programme d'origine (le DMASTR), syntaxe assez compliquée et assez difficile à appréhender pour des néophytes, particulièrement dans la configuration des fichiers.

Le premier, Protopapas (2007), afin de remédier au caractère chronophage du DMDX pour les expériences mettant en œuvre une tâche de dénomination, propose un logiciel qu'il a appelé « CheckVocal » qui doit permettre de faciliter le contrôle de la précision et le temps des réponses dans le cadre de ces expériences. CheckVocal fonctionne également sur Windows. Ce logiciel analyse les fichiers provenant du DMDX et en retire les informations nécessaires. L'interface propose la réponse correcte écrite, la réponse enregistrée avec une

forme d'onde et avec un spectrogramme correspondant, plusieurs boutons de commande disponibles (zoom, notation, écoute ...) et deux marques pour indiquer le temps enregistré (une rouge nette pour celle détectée par CheckVocal et une grise estompée pour celle enregistrée par le DMDX à l'origine). Pour faciliter la tâche, CheckVocal montre la réponse enregistrée non au début du fichier mais au début de chaque essai grâce à des « marques de temps de réponse ». Cela offre, pour la latence, à l'expérimentateur de corriger manuellement beaucoup plus vite en sautant d'essai en essai. Il doit cliquer sur l'onde ou le spectrogramme pour vérifier (il a la possibilité de zoomer ou de réécouter le son), replacer le marqueur et demander à CheckVocal de prendre ces ajustements en considération pour les essais suivants qui seront recalculés automatiquement. Ce procédé est un procédé de mesure *offline* qui corrige la « Vox » du DMDX qui, elle, fonctionne *online*. Pour la précision, parallèlement, l'expérimentateur doit cliquer avec la souris pour indiquer de quelle nature est la réponse. Le travail est facilité par le découpage en essai et par la comparaison de la réponse indiquée par écrit et la lecture de la réponse enregistrée. Une autre possibilité qui s'offre à l'expérimentateur est le fait de pouvoir enregistrer son travail et de le reprendre plus tard. En effet, les données sont sauvegardées à plusieurs endroits. Enfin, la fenêtre de travail a été conçue pour pouvoir s'ajuster à la durée des réponses enregistrées, aux préférences personnelles et aux circonstances individuelles (Protopapas, 2007, p.862). C'est pourquoi, l'auteur a, en plus de l'alphabet latin, prévu une option pour encoder l'alphabet grec par intérêt personnel. Comme le DMDX, CheckVocal est disponible librement sur le site de l'auteur. Comme le DMDX, son utilisation a été largement étendue (plus de deux cents études ont jusqu'à présent utilisé ce moyen pour faciliter l'analyse des temps de réponse).

Dans un deuxième temps, pour palier le manque d'une interface intuitive du DMDX, Garaizar et Reips (2015) l'ont adapté. Ils ont conçu le « Visual DMDX » qui fournit un service web pour faciliter la manipulation du DMDX. En effet, ce programme doit aider à créer les fichiers « items », étape plus difficile à concevoir pour les non-initiés au programme (voir ci-dessus). Cette application est compatible avec plusieurs systèmes d'exploitation (Windows, Mac OS, Linux, IOS, Android ...) et facilement accessible (sans compte ni mot de passe). Le Visual DMDX est utilisé depuis par des utilisateurs novices ne connaissant pas la syntaxe des fichiers « items » utilisés dans le DMDX. Pour ce faire, les auteurs ont choisi de proposer une interface où la personne peut définir des paramètres généraux avec des commandes simples et créer de nouveaux items avec des « drag-and-drop editor » (Garaizar et al., 2014), à savoir des « glisser-posers ».

d) SayWhen

« SayWhen » est un programme mis en œuvre en 2008 par Jansen et Watter pour répondre à ces problèmes de mesure de la latence. Les auteurs ont deux objectifs avec ce programme. Premièrement, ils veulent développer un algorithme qui peut détecter le début de la parole et, ainsi, mesurer la latence. Deuxièmement, ils souhaitent développer une interface d'utilisateur graphique qui permet de faciliter et d'optimiser l'inspection manuelle pour déterminer le début de la réponse orale du sujet. Cela permet de garder la rapidité de mise en œuvre de la clé vocale en évitant les imprécisions et les biais de cette dernière et d'avoir une précision améliorée sans l'effort du codage manuel. Par conséquent, le programme permet trois opérations : avoir des données générées automatiquement par le programme, vérifier chaque essai manuellement avec l'aide offerte par le logiciel (comme pour CheckVocal) ou un état « hybride » (Jansen & Watter, 2008, p.745), c'est-à-dire que le programme génère automatiquement les résultats et signale les essais jugés mal mesurés afin qu'ils soient vérifiés par l'expérimentateur. C'est dans ce dernier cas que le programme apporte le plus d'utilité selon les auteurs et qu'il permet d'obtenir le meilleur équilibre entre la précision et le temps de codage.

SayWhen fonctionne sur Windows. C'est un logiciel qui travaille en *offline*. Il faut également ajouter que, pour permettre de détecter le début de la parole, les auteurs fournissent un bip qui sert de marqueur du début de l'essai et qui doit être incorporé à la présentation de l'expérience. Ce signal doit être sur un des deux canaux du fichier audio stéréo afin que l'algorithme puisse détecter le début de la réponse. SayWhen a été conçu également pour recevoir un fichier audio d'une session expérimentale entière. Après analyse, le programme découpe en essai avec un marqueur sur chaque début d'essai détecté. SayWhen se fonde sur un algorithme utilisant différentes heuristiques pour examiner à la fois en avant et en arrière dans le temps le signal auditif selon la transformation de Fourier, ce qui lui permet d'avoir une performance plus précise et plus sélective dans la détection du début de la parole. L'interface montre l'onde en zoom pour permettre à l'expérimentateur de vérifier et corriger aisément. Comme CheckVocal, il peut passer d'essais en essais qui sont tous marqués (numéro de l'essai accompagné de son temps de réponse, abrégé RT).

e) Chronset

En 2017, Roux et al. ont créé un autre programme qu'ils ont appelé « Chronset ». Il s'agit d'un programme complètement automatisé. Les auteurs avaient l'intention de trouver une possibilité d'améliorer encore la robustesse et la précision d'une détection automatisée du début de la parole. S'aidant des travaux sur les chants des oiseaux (Tchernichovski & al., 2000 ; cités par Roux & al.) qui ont montré que les sons vocalisés ont une structure harmonique riche et absente dans le bruit ou les sons non-vocaux, les auteurs sont arrivés à la conclusion que les sons des bruits et les sons du début de la parole ont des signatures spectrales différentes grâce auxquelles ils pourront être distingués les uns des autres (Roux & al., p.1865). Chronset utilise les différentes caractéristiques acoustiques du signal (amplitude, entropie Wiener, changement spectral, modulation de l'amplitude, modulation de la fréquence, hauteur d'harmonie ; voir Roux & al., p.1867) pour détecter le début de la parole qui s'effectue quand quatre des six caractéristiques excèdent le seuil fixé pour au moins 35 ms. Même si les auteurs reconnaissent que cet outil n'atteint pas la précision d'une évaluation humaine, leur étude a démontré les preuves d'une très grande précision de leur logiciel (< 50 ms) et une estimation d'erreur qui a peu d'impact sur la puissance statistique des données expérimentales avec une taille d'échantillons standards. Les auteurs conseillent même de le combiner avec SayWhen pour avoir une précision encore plus proche du taux des expérimentateurs. Roux et al. offrent deux possibilités à l'utilisateur : soit il télécharge Chronset dans un environnement Windows 7 (disponible sur le site des auteurs), soit il envoie les fichiers audio à l'analyse (les fichiers doivent être déposés sur leur site).

f) AlignTool

En 2018, Schillingmann et al. ont proposé un logiciel qu'ils ont appelé AlignTool. Les auteurs voulaient continuer sur la lancée des résultats encourageant et prometteur de Chronset. AlignTool est différent des autres dans le sens où le programme détecte non seulement le début de la parole (*onset time*) mais aussi ce qu'ils appellent le *offset time* pour les phonèmes, et les mots (même dans des énoncés comprenant plusieurs mots). Cet outil permet donc de procéder à une analyse semi-automatique d'une structure temporellement complète, ce qui n'était possible auparavant que par l'analyse manuelle coûteuse en temps.

Selon les auteurs, AlignTool permet de faire le pont entre deux traditions majeures dans la recherche de la production du langage, à savoir la tradition du « language-as-product » dans un contexte de laboratoire et celle du « language-as-action » dans un contexte de dialogue naturel (Schillingmann & al., 2018, p.467).

Ce programme fonctionne dans un environnement Linux ou sous Windows avec une machine virtuelle Linux (VMWare Workstation Player) qui est payante. Il peut aussi bien travailler avec des essais isolés mais aussi avec des blocks complets d'une expérience, contrairement à SayWhen qui se limitait à l'expérience entière. Cela permet donc d'analyser de simples mots, voire plusieurs mots jusqu'à des discours semi-spontanés. AlignTool fonctionne avec Praat, qui permet à l'utilisateur une éventuelle vérification manuelle, et avec le système MAUS (pour *Munich Automatic Segmentation*, Schiel 1999, 2015 ; cité par Schillingmann & al., 2018) de reconnaissance automatique du discours. Ces deux logiciels sont combinés : Praat permet à AlignTool d'effectuer une pré-segmentation du signal audio et le système MAUS (via WebMAUS, Kisler & al., 2016 ; cités par Schillingmann & al.) pour forcer l'alignement des transcriptions des énoncés avec le signal du discours. Pour ce faire, l'utilisateur doit fournir une transcription orthographique qui est transformée en une transcription phonétique.

Les auteurs insistent sur la qualité de l'enregistrement qui peut influencer les résultats : si un enregistrement de haute qualité permet d'obtenir un excellent « ratio signal-to-noise » et de travailler de manière automatique sans aucune vérification, un enregistrement de moins bonne qualité donne des résultats moins exacts – tout en restant quand même hautement fonctionnel en réduisant le temps de l'annotation manuelle - et demande plus de corrections manuelles. Un enregistrement avec beaucoup de bruits environnants ne fonctionnera pas ou pas bien avec AlignTool. Cet outil permet toujours la vérification et la correction manuelle.

Pour le moment, AlignTool a traité comme langues, l'allemand, le néerlandais et l'anglais (Grande Bretagne). Mais le système MAUS peut supporter d'autres langues. Le logiciel est disponible sur le site des auteurs, ainsi qu'un article et un manuel détaillé pour les procédures à suivre dans la mise en œuvre de ce programme.

2. Mise à l'épreuve des moyens et techniques dans la littérature

Les différents dispositifs ont été testés et comparés entre eux dans des études émises soit par les auteurs eux-mêmes par souci d'étayer leur proposition en comparant à d'autres dispositifs antérieurs, soit par d'autres chercheurs.

Les clés vocales, nous l'avons vu, ont déjà fait l'objet d'études sur les différentes erreurs de précision et biais, pour tenter d'apporter des améliorations ou des solutions aux problèmes soulevés. Rastle et Davis (2002), déjà mentionnés ci-dessus, se sont penchés également sur la question de l'influence des techniques mises en place pour mesurer la latence. Comme évoqué ci-dessus, dans la littérature, l'étude de l'effet de la complexité a apporté un exemple de cette problématique. En effet, des divergences de résultats sont apparues, divergences qui peuvent s'expliquer en raison du type de mesure utilisée. Ceci comprend d'importantes implications pour l'interprétation des résultats dans les études, puisque les différences dans la latence en dénomination ne reflèteraient pas les propriétés des processus cognitifs étudiés mais des contradictions provenant des moyens pour récolter les réponses. Rastle et Davis ont choisi d'étudier cette question dans le cadre de l'effet de la complexité en comparant deux types de clés vocales avec l'analyse manuelle. Les dispositifs utilisés pour mesurer la latence sont donc, d'une part, les clés vocales et, d'autre part, l'analyse manuelle. La clé vocale dite simple est celle du DMDX qui servait à la fois pour la présentation des stimuli et pour l'enregistrement des données pendant l'expérience. Celle-ci fonctionne comme la clé traditionnelle avec un seuil d'amplitude. La deuxième clé dite intégrée est celle de l'université de Cambridge (pas de référence donnée) qui prend, elle, en compte à la fois l'amplitude des signaux et leur durée (une intensité minimum et une intensité plus faible).

Les auteurs ont fait passer à 22 participants (initialement 24) âgés de 18 ans à 30 ans une lecture de quarante mots monosyllabiques à voix haute. Ces mots étaient répartis en deux groupes : la moitié commençait par le phonème /s/, qui pose particulièrement problème à la détection des clés vocales, accompagné d'un autre phonème (/p/ ou /t/), l'autre moitié par le seul phonème /s/.

Les fausses alarmes (déclenchement trop tôt de la clé) ou les manquements de la clé ont été écartés des données. Cela correspond respectivement, dans l'ensemble des données, d'une

part, pour la clé simple, à 0,35 % (trois fausses alarmes) et à 1,73 % (quinze manquements), et, d'autre part, pour la clé intégrée, 1,62 % (quatorze fausses alarmes) et aucun manquement.

Après avoir codé le début de l'énergie acoustique des items cibles manuellement (repérable sur la forme d'onde de la parole par une augmentation nette dans l'amplitude suivant une période de silence) afin, par la suite, de les comparer aux données provenant des clés vocales, les auteurs observent dans leur analyse statistique par item et par participant une interaction significative des deux facteurs indépendants, à savoir la complexité initiale du mot et le type de mesure, un effet significatif pour le type mais aucun effet pour la complexité. L'effet de complexité sur la latence en dénomination apparaît dès lors dépendre de la méthode utilisée pour la mesurer et les trois méthodes de mesure donnent une moyenne différente des latences (fig. 8).

Table 1
Naming Latencies by Participants as a Function of Onset Complexity and Method of Measurement

Onset complexity	Hand-marking	Simple threshold key	Integrator key
Simple	371	500	449
Complex	362	511	447

Fig. 8 : Les latences en dénomination par participants en fonction de la complexité initiale du mot et le type de mesure (Rastle & Davis, 2002, p.310)

Les auteurs ont obtenu, bien que tous les items ont le même phonème initial, un point de détection pour chaque clé qui s'étend du début de l'énergie acoustique à environ, voire dans le début de la parole (*onset voicing*). La distribution des points de détection varie donc en fonction de la clé vocale utilisée (Rastle & Davis, p.310). Comme le montre l'histogramme (fig. 9), la majorité des détections pour la clé simple était simultanée ou postérieure au début de la parole, tandis que la majorité des détections de la clé intégrée avait lieu pendant les segments initiaux de la syllabe (en moyenne environ 80 ms après le signal acoustique). Ces résultats peuvent s'expliquer par le fonctionnement interne des clés décrit ci-dessus. Vu l'importance de l'amplitude acoustique dans les performances des clés vocales, les auteurs tiennent à souligner l'importance des facteurs qui peuvent influencer l'amplitude comme les caractéristiques individuelles des participants ou la fatigue des participants (Rastle & Davis, p.311).

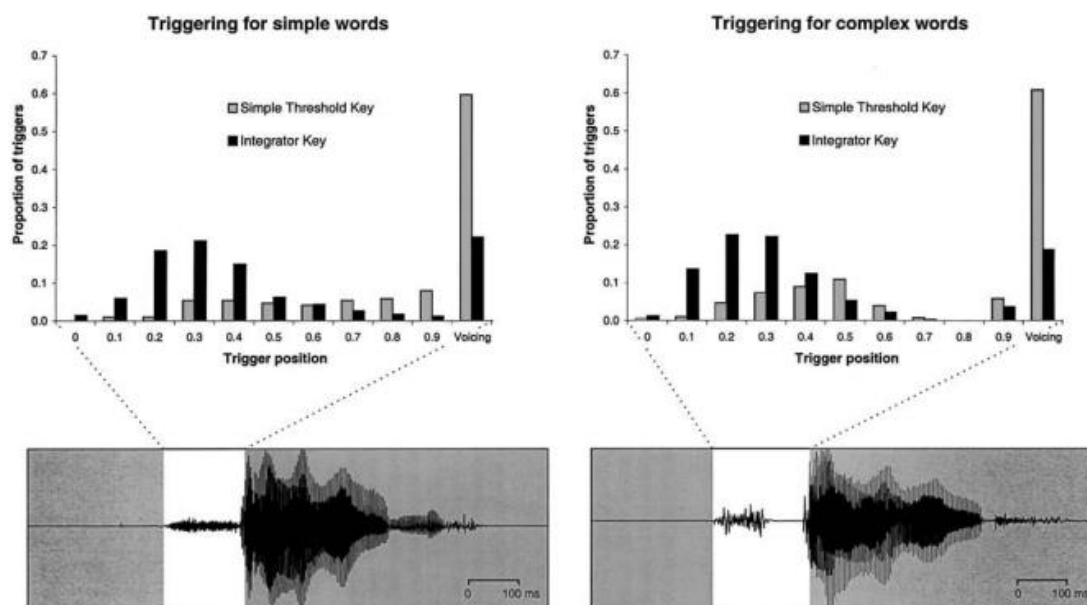


Fig. 9 : Histogramme illustrant les résultats obtenus dans l'étude avec les positions de déclenchement pour les mots simples et le mots complexes en fonction des deux types de clés. Le début de l'énergie acoustique (*onset of acoustic energy*) et le début de la parole (*onset of voicing*). *Voicing* fait référence à tout déclenchement qui viendrait après le début de la parole

Les auteurs ont essayé de calibrer autrement les clés pour augmenter la sensibilité. Cela pose, toutefois, un autre problème car cela augmente les fausses alarmes en détectant notamment des sons qui ne sont pas de la parole. Par exemple, pour la clé intégrée, les auteurs ont obtenu une amélioration dans la détection du début acoustique (45 ms au lieu de 82 ms). Néanmoins, les fausses alarmes sont passées de 1,6 % à 8,5 %, ce qui représente une perte d'environ 10 % des données.

Duyck et al. (2008) ont comparé, dans leur article, deux clés vocales, la leur (NEVK) et la clé simple (CTVK) avec l'analyse manuelle par deux personnes entraînées, appelées « juges ». Ils ont procédé à cinq expériences différentes (cf. fig. 10) afin d'évaluer différents éléments. Dans leur première expérience, ils ont fait passer une tâche de dénomination (noms très fréquents de quatre lettres commençant par différents phonèmes). Les résultats de cette première expérience démontrent, selon les auteurs, que la NEVK combine une précision de temps de réponse excellente avec un taux de fausses alarmes acceptables (2,55 % de fausses alarmes contre aucune pour la CTVK). La NEVK a une déviation moyenne absolue entre elle et l'inspection visuelle de seulement 5,6 ms, tandis que la CTVK de 72,33 en moyenne. De

plus, la NEVK corrèle de façon importante avec l'analyse manuelle des deux « juges » (.98 et .97) avec une fiabilité inter-juge de .97 (la CTVK est à .78). Les auteurs remarquent que la clé simple utilisée a de meilleurs résultats que les dispositifs du même type utilisés dans d'autres études. Dans leur deuxième expérience, les auteurs ont repris exactement la première expérience mais avec une femme, tandis que la première étaient composées d'hommes exclusivement. L'objectif était d'examiner si le genre influait sur la précision des clés vocales comme l'avait évoqué Tyler et al. (2005). Ils ont obtenu des résultats un peu moindres (fausses alarmes de 1,91 %, toujours aucune pour la CTVK ; déviation de 4,1 ms) mais la corrélation avec les juges reste haute. Ils constatent également que la NEVK a de meilleurs résultats que la CTVK, puisque cette dernière, se maintenant plus ou moins dans la déviation, chute dans la corrélation (.51). Enfin, les auteurs en combinant les deux expériences, concluent que les performances varient selon le phonème initial que ce soit pour la NEVK que pour la CTVK. Néanmoins, la NEVK est plus stable pour l'ensemble – ce qui n'est pas le cas pour la CTVK - mais il garde plus de difficultés pour les fricatives. Dans leur troisième expérience, les auteurs ont repris le même type d'expérience (avec un seul juge) pour confirmer les résultats des deux premières et tester la variabilité sur un petit groupe (genre mélangé). Ils ont obtenu des résultats similaires aux deux premières pour la NEVK alors que la CTVK s'est révélée plus mélangée (la corrélation varie de .89 à .17). Les auteurs soulignent que la NEVK reste à .95 pour le participant dont la corrélation était la plus basse. Dans leur quatrième expérience, les auteurs ont répliqué l'expérience de Rastle et Davis. Ils sont arrivés à la même conclusion que pour les expériences précédentes (la NEVK reste stable, la CTVK est plus variable). Enfin, dans leur cinquième expérience, les stimuli n'étaient que des mots bisyllabiques accentués sur la deuxième syllabe. Les conclusions restent similaires. De plus, les auteurs remarquent que la CTVK est en grande difficulté avec ce genre de mots, alors que la NEVK non.

Table 1
Performance of the Noise Elimination Voice Key (NEVK) and a Classical Threshold Voice Key (CTVK) in Experiments 1 to 5

Measure	Experiment 1		Experiment 2		Experiment 3		Experiment 4		Experiment 5	
	NEVK	CTVK	NEVK	CTVK	NEVK	CTVK	NEVK	CTVK	NEVK	CTVK
Deviation (ms) ^a	5.60	72.33	4.11	74.61	5.30	45.65	4.69	22.53	5.34	84.01
Accuracy ^b	0.97	0.78	0.98	0.51	0.95	0.53	0.99	0.88	0.98	0.54

Note. ms = milliseconds.

^a Reported values are mean absolute value deviation scores between voice key reactions times (RTs) and visual inspection RTs across judges (Experiments 1, 2, 4, and 5) and across participants (Experiment 3). ^b Reported values are mean Pearson correlations across judges (Experiments 1, 2, 4, and 5) and across participants (Experiment 3) between voice key RTs and visual inspection RTs.

Fig. 10 : Tableau reprenant les différentes expériences et les différentes données obtenues (Duyck & al., 2008, p.1321)

Par ces différentes expériences et par l'introduction d'une nouvelle mesure (les mesures d'association), les auteurs ont voulu démontrer que la NEVK détecte de façon très précise les débuts acoustiques et qu'elle est comparable à travers les participants et les études. Enfin, les auteurs se sont intéressés sur les moyens alternatifs disponibles. Ils ont calculé que pour la clé vocale intégrée reprise par Rastle et Davis (2002) la corrélation était de .70 (moyenne des participants) et que, pour le DMDX, elle était de .51. Ces résultats restent inférieurs à ceux de la nouvelle clé proposée. Ils ont comparé leur clé avec celle proposée par Tyler et al. (2005) au niveau ici du temps réponse : respectivement, 4 à 6 ms *versus* 11,8 ms, ce qui est plus que la NEVK mais bien meilleur que la clé traditionnelle. Néanmoins, les auteurs tiennent à signaler que, pour la clé intégrée de Tyler et al., les fichiers audio étaient de haute qualité, ainsi que leur équipement, ce qui, dans une réplication, pourrait impacter les résultats si les conditions ne sont pas les mêmes.

Jansen et Watter (2008) ont testé leur programme SayWhen dans leur laboratoire avec une expérience qui avait pour cadre la production orale du langage. Ils ont demandé aux 24 participants (initialement 28) de donner à l'oral un mot associé au mot présenté, comme stimulus, de façon libre et rapide.

Les auteurs ont comparé les données provenant d'une clé vocale, de leur algorithme et celles d'un codage manuel sur les enregistrements audio qui a été effectué par des étudiants entraînés. Le programme a été examiné selon trois tests séparés, à savoir de façon complètement automatique (aucune vérification ou modification), de façon automatique mais en excluant les temps de réponse trop courts ou trop longs (e.i., les essais pour lesquels la détection du début de la parole est au-dessus ou en dessous du seuil de latence prédéterminé) et de façon automatique en excluant tous les essais marqués (qui devraient par après faire l'objet d'une vérification) afin d'obtenir une estimation de la meilleure performance du programme. Les auteurs ont rapporté sur un histogramme (fig. 11) la proportion relative et la distribution des différences entre le codage manuel et la clé vocale ou la performance de l'algorithme dans la première et la troisième condition. Plus bas est le score, plus proche est l'accord avec l'analyse manuelle.

On constate que la clé vocale obtient de moins bons résultats : seulement 32,9 % des essais sont indiqués dans les 5 ms du codage, 41,7 % dans 10 ms et 44,8 % dans les 25 ms ou plus. Par contre, l'algorithme de SayWhen a indiqué dans le mode complètement automatique :

59,6 % dans les 5 ms, 69, 5 % dans les 10 ms et 18, 6 % dans les 25 ms ou plus. Dans la deuxième condition, les résultats restent principalement inchangés. Les performances, cependant, augmentent considérablement dans la troisième condition : 77,2 % dans les 5 ms, 86,6 % dans les 10 ms et 9 % pour les 25 ms ou plus. C'est pourquoi, les auteurs conseillent le mode hybride, c'est-à-dire de vérifier seulement les essais marqués. Cela permet d'obtenir les performances de haute précision observée dans la troisième condition.

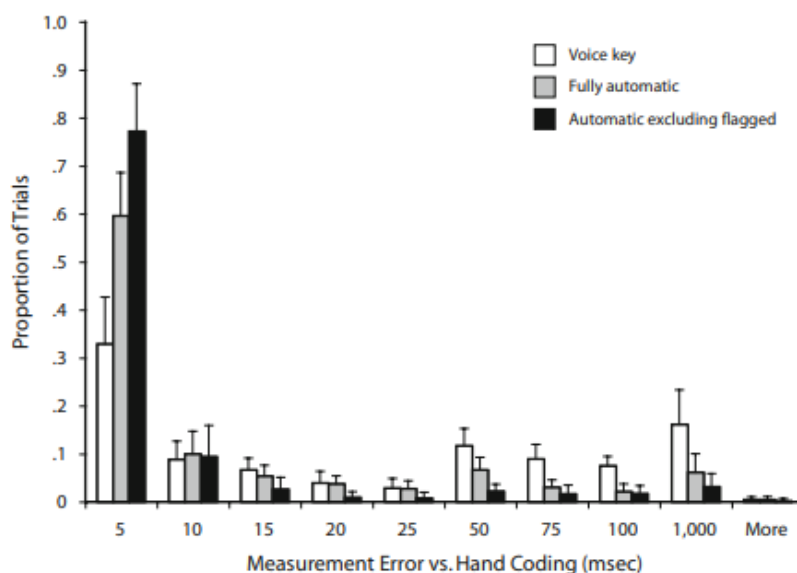


Fig. 11 : Histogramme comparant les différentes méthodes de mesure : la clé vocale (en blanc), l'algorithme de SayWhen dans la première condition (en gris) et l'algorithme de SayWhen dans la troisième condition (en noir). La mesure d'erreur est calculée en tant que la différence en millisecondes entre la mesure de chaque méthode et le « gold-standard », à savoir l'analyse manuelle (Jansen & Watter, 2008, p.749)

Le bémol souligné par les auteurs eux-mêmes est que les essais marqués (principalement les fricatives et les sifflantes qui posent le plus de problème) consistent en près de la moitié des données recueillies. Néanmoins, les auteurs tempèrent ce chiffre en soulignant le côté pratique de SayWhen grâce à l'interface proposée qui permet une inspection et vérification rapide. Cela prend quelques secondes, d'après les auteurs, de vérifier visuellement et de corriger sur le graphique l'essai marqué. Ils ont calculé, pour une personne peu habituée et pour une session expérimentale d'un sujet, environ cinq minutes, ce qui permet *in fine* d'avoir une haute précision dans les données.

Dans leur article, Roux et al. (2017) ont comparé différents programmes, à savoir SayWhen de Jansen et Watter et CheckVocal en version espagnole (Sadat et al., 2012, cité par Roux & al.). Les auteurs ont procédé pour ce faire à une analyse statistique bien précise qu'ils

décrivent en détail dans leur article (Roux et al., p.1868 et p.1877), ainsi que de nombreux schémas non synthétiques de tous leurs résultats. Ils ont utilisé particulièrement les scores de la différence absolue (ADs) entre les latences du début de la parole estimée automatiquement et celle estimée manuellement pour quantifier le biais lié à la mesure et les ajustements par régression, puisque les régressions linéaires ont l'avantage de quantifier la précision d'une détection automatique (notamment avec les résidus de régression). Ils ont choisi non seulement de comparer les programmes mais aussi de comparer l'accord entre les évaluateurs, ainsi que l'influence des phonèmes initiaux sur Chronset et sur SayWhen.

Concernant, d'abord, l'ensemble des données en langue espagnole, les auteurs ont obtenu un très haut niveau d'accord entre les deux évaluateurs. Ce constat peut s'expliquer par l'ajustement proposé par CheckVocal qui permet d'avoir une estimation initiale semblable. Ensuite, une relation linéaire solide entre l'évaluation humaine (*human rating*) et les scores automatiques mesurés par Chronset, et ce sur l'ensemble des participants. Les performances de CheckVocal étaient considérablement plus basses que celles de Chronset.

Concernant les données en anglais, les latences estimées par les évaluateurs se sont avérées également très constantes (fiabilité inter-juge haute). Les auteurs ont obtenu une relation linéaire forte entre l'évaluation humaine moyenne et les latences automatiques estimées par Chronset, alors que les autres algorithmes ont en grande partie une corrélation plus basse.

Enfin, concernant les phonèmes, Chronset est meilleur que SayWhen (pour Chronset, trois phonèmes sont avec un ajustement R^2 en dessous de .9, alors que pour SayWhen seize phonèmes sont dans ce cas). Les auteurs ont donc démontré que les latences identifiées par Chronset sont substantiellement plus solides que celles produites par d'autres algorithmes (CheckVocal ou SayWhen). Chronset évite particulièrement les données aberrantes au détriment d'un degré moindre de précision. En effet, Chronset sacrifie une précision extrêmement fine pour les latences individuelles pour assurer une estimation solide des débuts relativement précise en évitant de trop nombreuses mauvaises estimations (Roux & al., 2017, p.1878), alors que SayWhen offre des estimations extrêmement précises mais davantage de mauvaises estimations. C'est pourquoi, les auteurs conseillent de combiner les deux afin d'obtenir des données optimales.

Chronset a été dernièrement inclus dans une nouvelle étude qui a « testé » le programme et qui a conclu à la précision du programme (Broos et al., 2018, p.1220) :

« A subset of the data that Chronset annotated (...) were also manually annotated while a correlation analyses was performed on these trials. This way, the accuracy of the Chronset package could be objectively measured. The correlation between the handcoded and automatically coded speech was 0.9 meaning that Chronset was quite accurate in determining speech onset. »

Schillingmann et al. (2018), enfin, ont également procédé à des évaluations pour tester le programme. Outre l'évaluation détaillée du programme dans plusieurs langues et plusieurs contextes (mots simples *versus* semi-spontanés), ce qui est intéressant pour notre propos est leur comparaison avec d'autres dispositifs. En effet, les auteurs ont repris les données manuelles de Rastle et Davis (2002) afin de les comparer, en reprenant les mêmes items, à une autre clé vocale (*custom-made hardware voice-key*), à SayWhen, à Chronset et à AlignTool (fig. 12). Pour SayWhen, les auteurs ont intégré les items dans un fichier compatible (fichier avec une expérience entière et avec le bip marqueur d'essai sur un seul canal). Pour AlignTool, ils ont construit un corpus, appelé MTAS (*Manually Temporally Annotated Speech*) qui récolte des données d'étudiants de langues différentes (30 sujets germanophones, 30 néerlandophones et 30 anglophones). Ce corpus comprend une tâche de dénomination d'objets, qui a été contrôlée particulièrement sur la fréquence, l'accord sur le nom et sur la variété de phonèmes initiaux, et une tâche de lecture de mots, qui reprend les mêmes objets combinés en quatre mots (mais en veillant toujours à un appariement strict, voir Schillingmann et al., p.476 pour le détail des critères). Le début du phonème est aussi mesuré par la clé vocale pour la comparaison par la suite. Chaque groupe linguistique avait une qualité d'enregistrement différent (excellent pour le néerlandais, moyenne pour l'allemand et pauvre pour l'anglais). Ces données ont été également soumises à SayWhen et Chronset.

Table 3 Mean utterance onset times (ms) in the simple and complex conditions of the word-naming task reported in Rastle and Davis (2002) as established by a human rater (R&D), by AlignTool, by SayWhen, and by Chronset

	Simple		Complex		M_{diff}	$t1(20)^a$		$t2(38)$	
	M	SE	M	SE					
R&D (incl. all ppts.)	371	13	362	11	9	2.89	**	2.02	(*)
R&D	370	13	362	12	8	2.67	*	1.93	(*)
AlignTool	334	11	332	11	2	0.51	n.s.	0.28	n.s.
AlignTool - R&D	-36	7	-30	6	-6	-1.68	n.s.	-1.10	n.s.
SayWhen	439	16	448	16	-9	-1.67	n.s.	-1.92	(*)
SayWhen - R&D	69	8	86	10	-17	-3.61	**	-4.67	***
Chronset	378	14	506	17	-128	-17.04	***	-24.45	***
Chronset - R&D	8	4	144	8	-135	17.09	***	-32.76	***

(*) $p < .1$, * $p < .05$, ** $p < .01$, *** $p < .001$

^a $df = 21$ for Rastle (incl. all ppts)

Fig. 12 : Tableau récapitulatif des résultats obtenus lors de la comparaison de différents dispositifs

Suite aux résultats de leur comparaison, Schillingmann et al. (p.478-479) formulent les conclusions suivantes. Rastle et Davis avaient obtenu, avec l'annotation manuelle, des temps de réponse 9 ms plus rapides pour les mots complexes (effet significatif par participants et par items) que pour les mots simples. AlignTool marque automatiquement les temps du début dans les deux conditions de mots environ 33 ms plus tôt que l'analyse manuelle. La réduction en temps de réponse était légèrement plus prononcée dans la condition des mots simples par rapport à celle des mots complexes, réduisant l'effet de la condition à 2 ms, ce qui est non significatif (cette réduction n'était pas affectée de façon significative par la condition). SayWhen, par contre, donne des temps initiaux environ 77 ms plus tard que ceux de Rastle et Davis, rapportant une différence entre les conditions dans la direction opposée de ce que ces derniers avaient trouvé. Par ailleurs, l'effet constaté était significatif pour l'analyse par item et non significatif pour l'analyse par participant. Un effet significatif de la condition est également apparu (différence plus importante dans les mots complexes). Chronset rapporte le même problème. Même s'il est meilleur dans la détection du début dans les mots simples, il était imprécis pour les mots complexes. C'est pourquoi, il arrive à des résultats similaires à SayWhen. A noter que ces deux programmes vont dans le sens de ce qu'ont obtenu Rastle et Davis avec la clé vocale simple. AlignTool apparaît dès lors face à ces premières données comme plus précis que les trois autres dispositifs. Pour le deuxième corpus (MTAS), des données ont été perdues en raison de problèmes d'enregistrement, d'erreurs dues aux participants et de mauvais fonctionnement de la clé vocale ou d'AlignTool. Car, en effet, AlignTool n'a pas marqué environ 9 % des essais en raison de la mauvaise qualité d'enregistrement pour les données anglophones. Cependant, les auteurs estiment que cela

reste raisonnable, alors que Chronset et SayWhen, eux, baissent nettement face à ce type d'enregistrement. AlignTool, au contraire des autres dispositifs, rapporte des résultats parallèles pour les deux tâches. On retrouve également dans cette expérience des différences entre AlignTool qui détecte plus tôt et Chronset et SayWhen qui détecte plus tard.

Les auteurs ont également mis en place un test pour le fonctionnement de cet outil dans le cadre d'un discours semi-spontané. Leurs résultats étaient prometteurs. La déviation s'est avérée petite mais il manque un effet systématique de la position des mots dans l'énoncé qui se retrouve, par contre, dans l'analyse d'énoncés de quatre mots (peut-être dû aux silences plus longs).

DISCUSSION

I. Préambule

Le présent mémoire était au départ un mémoire pratique qui portait sur « l'évaluation de l'effet de l'âge sur la vitesse d'accès lexical dans différentes populations de 45 à 60 ans ». Il avait pour objectif premier de répliquer les résultats observés dans l'étude de Verhaegen et Poncelet (2013) et le mémoire qui s'ensuivit de Truffert (2014). C'est pourquoi, les hypothèses de travail et les attentes étaient les conclusions émises dans ces deux études antérieures. La première hypothèse émise alors porte sur le moment d'apparition d'un effet de l'âge sur la récupération. Conformément aux études « modèles », on s'attendrait à des prémices d'une augmentation uniquement de la latence liée à l'âge, déjà avant 50 ans, qui se poursuit chez les sujets de 50 ans et à une diminution de la précision, ainsi qu'à une augmentation de la latence, chez les participants de 60 ans. La deuxième hypothèse concerne la préservation sémantique observée avant la septantaine suivant les assises théoriques et le constat d'un ralentissement linéaire dans le mémoire de Truffert pour les trois groupes mais aussi, par rapport à l'étude de Verhaegen et Poncelet (2013).

Le mémoire pratique devait mettre en place une tâche de dénomination d'images avec comme pool, celui de Bonin et al. (2003) - pool composé de dessins en noir et blanc aux traits – se déroulant sur OpenSesam (Mathôt & al., 2011). Les participants étaient d'âges différents répartis en trois groupes (45-50, 50-55 et 55-60 ans) et auraient été recrutés dans un milieu principalement professoral (en tout cas intellectuel). Les tranches différaient afin de déterminer un autre angle d'étude dans l'espoir d'affiner les résultats. Outre la tâche de dénomination, deux tests de contrôle – également utiles pour avoir des données supplémentaires - étaient prévus, un test de vocabulaire (Mill Hill ; Deltour, 1993) et un test de jugement de chiffres pairs/impairs (Ulg ; Schyns & Poncelet, 2002)

Suite à l'impossibilité de pouvoir faire passer la tâche de dénomination aux personnes volontaires, le mémoire a été transformé, conformément aux dispositions en de pareilles circonstances, en un mémoire théorique. Le sujet reste fondamentalement le même. Il porte également sur l'évaluation de la vitesse d'accès lexical, c'est-à-dire la latence, avec un focus sur l'effet de l'âge.

II. L'effet de l'âge

Dans la première partie, ont été revues les études qui ont examiné la question de l'effet de l'âge sur les performances en dénomination. Les différentes observations concernant son moment d'apparition ont été développées, ainsi que les différents apports pour expliquer son origine.

Concernant le moment d'apparition, la prise en compte de la latence a permis de nuancer les conclusions émises. Dès 40 ans, on observe une légère augmentation de la latence sans pour autant une diminution de la précision, ce qui se poursuit dans la cinquantaine. Vers 60 ans, l'augmentation de la latence est toujours présente mais apparaît également une diminution de la précision. Au-delà de 70 ans, le même profil s'observe mais de façon plus marquée.

Pour l'explication de cet effet, plusieurs hypothèses ont été émises : un déficit d'inhibition, un ralentissement général ou un déficit de transmission.

Le déficit d'inhibition et, dans le même ordre d'idées, les théories sur une interférence sémantique qui augmenterait avec l'âge ont été respectivement réfutés ou contestés par certains auteurs. C'est plutôt le contraire qui a été démontré. Le vocabulaire ne diminue pas avec l'âge mais a tendance à augmenter, ce qui n'interfère pas mais, au contraire, améliore les performances en tâche de dénomination et a un effet positif sur l'âge (Goral et al., 2007 ; Dorot et Mathey, 2013 ; Shafto et al., 2017). Cependant, dans une étude récente, Paolieri et al. (2018) ont mis en évidence des résultats qui pourraient s'expliquer en partie par l'hypothèse d'un déficit d'inhibition. Il s'agissait d'un cas précis : ils ont démontré une plus grande vulnérabilité des personnes plus âgées avec un niveau d'éducation bas à l'interférence. On pourrait supposer que ces personnes aient un bagage moindre de vocabulaire et bénéficieraient moins de l'effet positif. Cela renvoie à la notion de réserve cognitive qui « protégerait » les personnes plus âgées de l'effet de l'âge. En effet, spécifiquement, le niveau d'éducation permet de prédire les scores de vocabulaire lors du vieillissement sain (Hoyau, 2018, p.79). C'est dans ce sens que pourraient s'expliquer les bonnes performances en dénomination des sujets qui ont une éducation haute et un niveau de vocabulaire élevé dans l'étude de Madden, Sale et Gail (2019).

Le ralentissement général, qui se marquerait dans toutes les tâches, a fait l'objet d'infirmer. Verhaegen et Poncelet (2013) n'ont pas constaté de ralentissement dans la tâche de jugements de chiffres (pas avant 70 ans du moins). Mais c'est surtout l'effet de la demande de la tâche qui contrevient à cette hypothèse. Certains auteurs (Goral, 2007 ; Madden, Sale & Gail, 2019 ; Paolieri & al., 2018) ont démontré que l'exigence de la tâche joue un rôle important sur les performances (plus la demande augmente, plus les performances s'amointrissent), ce qui va à contre-sens d'un ralentissement général. C'est d'ailleurs pour cela que la tâche de dénomination est une épreuve plus sensible car elle est exigeante. Elle demande de récupérer un seul item précis, il n'y a pas d'aide du contexte par exemple (Goral & al., 2007). Cet effet de la demande pourrait s'expliquer par le modèle CRUNCH (*compensation-related utilization of neural circuits hypothesis*) de Reuter-Lorenz et Cappell (2008) comme l'explique Hoyau (2018, p.29) :

« D'après l'hypothèse CRUNCH, les personnes âgées ont des ressources de traitement de l'information plus limitées que les jeunes adultes et sont ainsi moins capables de moduler leur fonctionnement cérébral en fonction de la difficulté de la tâche. Lorsque la tâche demande peu de ressources cognitives, les personnes âgées redirigent leurs ressources cérébrales en direction de régions « clés », permettant ainsi de maintenir des performances similaires aux jeunes adultes et de compenser un déficit de ressources de traitement (...). Cependant, lorsque la difficulté de la tâche augmente, l'allocation des ressources cognitives et le recrutement supplémentaire des régions cérébrales ne sont plus suffisants. »

Enfin, les études vont pour la plupart dans le sens d'un déficit de transmission. Les auteurs soulignent la plupart du temps un déficit d'accès à la forme phonologique (Nicholas & al., 1985, Au & al., 1995, cités par Verhaegen & Poncelet, 2013 ; Dorot & Mathey, 2013 ; Shafto & al., 2017). Au-delà de 70 ans, une dégradation sémantique viendrait s'ajouter également (Barresi & al., 2002). Par ailleurs, Paolieri et al. (2018) expliquent une partie de ces résultats avec cette hypothèse.

Les auteurs ont étudié l'effet de l'âge sur certaines variables qui influencent la vitesse en dénomination. Celle qui a fait l'objet d'intérêt particulier est l'accord sur le nom. Plusieurs études l'envisagent. LaGrone et Spieler (2006) ont obtenu un effet de l'âge sur cette variable. Ils ont aussi pris en considération la fréquence du mot qui ne semble pas subir l'effet de l'âge. En effet, les auteurs ont observé des performances similaires dans les deux groupes d'âge (légèrement plus pour les personnes plus âgées mais aucun effet significatif). Madden, Sale et

Robinson (2019) ont étudié l'effet de l'âge sur l'accord sur le nom dans trois conditions qui reprenaient les sources possibles de désaccord (Vitkovitch et Tyrell, 1995 ; cité par Madden & al.), à savoir des images avec des noms multiples, avec des abréviations ou des élaborations. Les auteurs avaient constaté un effet de l'âge pour toutes mais beaucoup plus marqué pour les abréviations et les élaborations (les jeunes ralentissaient aussi pour les noms multiples). L'intérêt de l'accord sur le nom vient peut-être du fait que cette variable prend place dans l'étude de la sélection lexico-sémantique. C'est d'ailleurs dans ce contexte que La Grone et Spieler, ainsi que Britt, ont été amenés à l'étudier par rapport à l'effet de l'âge.

III. La problématique de la latence

Il a été largement discuté dans ce présent travail de l'importance de la latence dans la tâche de dénomination. En effet, elle a contribué de façon significative aux recherches sur l'effet de l'âge en rendant plus sensible la tâche de dénomination et permettant d'obtenir des résultats plus nuancés.

La latence peut être influencée par la construction même de l'épreuve. Différentes variables susceptibles d'influencer la vitesse de dénomination ont été revues. Certaines, comme la fréquence, l'AoA ou encore l'accord sur le nom, ont été amplement étudiées, alors que d'autres ont fait l'objet de moins d'attention et apparaissent de façon éparse dans la littérature comme la *typicality* ou la *manipulability*. Chacune impacte une ou plusieurs étapes de la production orale et, *in fine*, la performance en vitesse de dénomination. Des études sont ressorties les conclusions suivantes. Les variables prédictives et contribuant de façon importante sont l'accord sur le nom, la variabilité de l'image (*imageability*) et l'AoA. Alario et Ferrand (1999) ont observé des liens entre elles : l'accord sur le nom et l'accord sur l'image sont également en relation (des concepts qui ont beaucoup de noms évoquent beaucoup d'images différentes) et des concepts appris plus tôt ont un accord sur le nom plus haut ; les mots appris tôt peuvent avoir plus d'images. L'accord sur l'image s'est révélé aussi avoir une influence sur la vitesse, alors que la complexité visuelle, en tenant compte de mesures objectives, s'est avérée influencer seulement la précision en dénomination. D'autres variables, à savoir la familiarité conceptuelle, la valeur de l'image ou les facteurs phonologiques, n'ont pas obtenu des résultats assez constants à travers les études pour pouvoir affirmer une

influence fiable, même si elles aussi ont des liens avec d'autres variables avérées prédictives : les noms des concepts familiers ont tendance à être appris tôt et, enfin, les images complexes visuellement sont moins familières (Alario et Ferrand, 1999). La fréquence semble liée aux autres variables, même si ce n'est pas toujours évident de savoir quel est leur lien exact. C'est le cas de sa relation avec l'AoA. Par contre, elle semble liée avec l'encodage phonologique (Alario & al., 2004 ; LaGrone et Spieler, 2006) et avec l'accord sur le nom (LaGrone & Spieler, 2006).

Cela montre l'importance de contrôler plusieurs variables pour éviter toute interprétation erronée des résultats en raison d'une influence d'une variable non contrôlée à l'instar du débat entre la fréquence et l'AoA.

En plus des variables liées aux stimuli que ce soit par l'image ou par le nom, des facteurs liés au sujet sont aussi à prendre en considération. L'éducation joue un effet modulateur comme l'ont démontré Paolieri et al. (2018). Les bons résultats observés dans les tâches préalables (vocabulaire et dénomination) par Madden, Sale et Gail (2019) pourraient s'expliquer par ce facteur. Contrairement à Goral et al. (2007), eux, concluent à une influence moindre dans la tâche de dénomination. Ces études n'ayant pas la méthodologie d'étude, les résultats pourraient être différents pour cette raison. Paolieri et al. ont mis en place une tâche de dénomination en blocs notamment pour introduire une condition d'interférence. Néanmoins, l'éducation influence les performances même dans la condition sans interférence. Les résultats obtenus montrent un effet significatif du groupe sur les temps de latence (âge et éducation). Or, dans ce cas de figure, l'éducation n'agit peut-être pas seule, à savoir que le vocabulaire peut aussi intervenir puisqu'on sait que le vocabulaire a tendance à augmenter avec l'âge. Il en va de même pour l'étude Madden, Sale et Gail (2019), même s'ils ont évalué l'effet de l'éducation sur d'autres tâches cognitives (comme la fluence), où le vocabulaire peut intervenir. Malgré ces divergences sur l'effet isolé de l'éducation, ressort une interaction entre le genre et l'éducation (Connor & al., 2004 ; Goral & al., 2007) et probablement une interaction entre l'âge, l'éducation et le niveau de vocabulaire.

Le genre seul ne fait pas non plus l'unanimité. Alors que Connor et al. (2004) suggèrent plus une interaction avec l'éducation, Goral et al. (2007) mettent en avant un effet de genre même seul. Brodeur et al. (2014), dans la normalisation de leur banque de données, ont constaté également des différences de genre en particulier en relation avec la catégorie et les autres variables.

Enfin, d'autres facteurs, comme la fatigue ou la voix féminine *versus* masculine, qui peuvent influencer l'amplitude et, par conséquent, la mesure de la latence, ont aussi été soulignés.

Ce n'est pas uniquement la construction de l'épreuve qui peut influencer la latence. En effet, la mesure de la latence joue un rôle tout aussi important. Un exemple rapporté par l'étude de Rastle & Davis (2002) a démontré les conséquences désastreuses des dispositifs de mesure sur l'interprétation des résultats en obtenant tout et son contraire. Pourtant, le problème épineux de la mesure de réponses orales n'est pas encore parfaitement résolu. Certes, des propositions de plus en plus précises et prometteuses voient le jour, comme a pu le démontrer la revue des moyens proposés dans la littérature. Cependant, des bémols sont toujours présents pour chacun d'entre eux. Le biais phonétique étant leur principal souci, chaque nouveau dispositif lui est attentif mais est encore influencé malgré les dispositions prises (principalement pour les fricatives).

Le choix du pool est important également. Les formats n'ont pas le même impact sur la latence. Les photographies, par exemple, même si elles offrent une alternative écologique aux dessins aux traits ou en couleur, peuvent ralentir les temps en dénomination en raison de l'apport substantiel des détails. Par contre, les auteurs ont largement démontré que la couleur - même si les détails font encore débat - améliore les temps en dénomination (Pourtois & Rossion, 2004 ; Heuer, 2016) par rapport aux dessins en noir et blanc, surtout où la couleur est diagnostic. En effet, j'ai pu le constater en stage, avec des personnes âgées, certains dessins en noir et blanc ne sont pas évidents à reconnaître sans couleur (par exemple, un bec de canard ou un oignon). Ces items étaient constants chez les personnes, ce qui manifeste non un problème individuel mais bien un problème lié à l'item.

Initialement, le mémoire devait se faire avec le pool de Bonin et al. (2003). Dans un souci de réplication parfaite, ce pool serait à garder d'autant plus que l'ensemble est normé sur les variables les plus prédictives. Néanmoins, il serait aussi intéressant de prendre un pool de couleur au vu des observations faites dans ce travail, comme celui proposé par Pourtois et Rossion (2004). Le seul bémol est qu'il est normé sur des étudiants belges. Or, pour certaines variables, l'âge du participant peut changer les scores. L'AoA, par exemple, peut varier en fonction de la génération interrogée (Sirois & al., 2006).

Même si l'ensemble de Sirois et al. (2006) est intéressant pour le fait justement que les auteurs aient normé sur différents âges et apparié en genre, il présente un désavantage évident, à savoir qu'il est normé sur le français canadien. La comparaison des auteurs a mis en avant une structure corrélationnelle globalement semblable. Toutefois, la culture apparaît avoir un impact significatif sur l'accord sur le nom. Alario et Ferrand (1999) en avaient conscience étant donné qu'ils ont changé certains items pour adhérer au contexte culturel français. Or, ce facteur est le déterminant le plus prédictif sur la vitesse de dénomination. De plus, il est sensible à l'âge comme cela a été rappelé ci-dessus.

IV. Les dispositifs

L'aperçu de l'évolution des recherches pour arriver au meilleur outil possible (pour « remplacer » l'analyse manuelle) afin de mesurer de façon la plus précise et la plus efficiente la latence de la réponse vocale témoigne à nouveau à quel point celle-ci est un enjeu crucial.

Outre le souci essentiel d'offrir un outil de haute précision, les auteurs se heurtent à des contraintes plus pratiques. En effet, plusieurs éléments rentrent en considération. Le premier élément est le coût du dispositif qui est un des soucis soulevés par les auteurs. Ce point n'est pas négligeable quand on sait, par exemple, qu'AlignTool, ce récent outil prometteur, demande, si Linux n'est pas possible sur l'ordinateur de l'expérimentateur, une machine virtuelle environ 300 euros.

Ensuite, viennent les problèmes liés à l'équipement (puissance du processus, compatibilité entre les systèmes, environnement du testing, enregistrement de qualité etc.) ou au programme lui-même (opacité du fonctionnement ou perte du bénéfice temporel par rapport à l'analyse manuelle). Pour le DMDX, nous avons vu que Garaizar et Reips (2015) ont créé leur service dans un souci d'aide dans la conception de l'expérience. Protopapas (2007) a proposé CheckVocal pour faire gagner du temps aux chercheurs mettant en place une tâche de dénomination. AlignTool donne un manuel complet, un guide étape par étape pour créer les fichiers nécessaires (ils vont même jusqu'à expliquer comment télécharger la machine virtuelle). Par contre, SayWhen donne quelques explications brèves, ce qui pose problème

pour créer le fichier audio compatible (surtout dans l'intégration du bip marqueur à la présentation).

Enfin, d'autres écueils viennent de la construction de l'expérience-même comme le biais phonétique. Les études ont analysé attentivement ce problème. Quel que soit le dispositif utilisé, il est intéressant de concevoir des stimuli variés (éviter par exemple une trop grande proportion de fricatives qui pourraient fausser les résultats) et, dans la mesure du possible, essayer de les regrouper au début du mot selon la proposition de Rastle et Davis. Cela est important puisque même des algorithmes (SayWhen ou Chronset) ne sont pas assez précis dans le cas de début complexe.

Les moyens proposés ne sont pas encore parfaits. Chacun a des inconvénients et des avantages, ce qui rend difficile le choix de celui qui convient le mieux. Même la clé vocale, qui est sujette à l'imprécision, a toujours été objet d'intérêt pour les auteurs afin de la perfectionner, particulièrement celle de Duyck et al. en 2008 qui offre des améliorations importantes. Les auteurs mettent en avant l'avantage de la clé qui dépend moins de l'ordinateur contrairement aux algorithmes, ce qui la rend plus autonome. La sélection devra se faire en fonction des besoins propres à l'expérience, aux objectifs poursuivis et des moyens dont on dispose. Par exemple, SayWhen ne fonctionne que sur Windows avec une expérience entière, il ne sera pas conseillé si l'expérience est fragmentée sous peine d'une perte de temps substantielle. AlignTool offre un cadre dépassant le simple mot jusqu'à aller à des énoncés de plusieurs mots.

V. Conclusions

Pour mener une expérience répliquable, comme cela était prévu initialement, différents critères devront être pris en considération. Les sujets, comme c'était prévu initialement, devront être de même effectif et appariés en genre, en éducation et en niveau de vocabulaire. Les tranches d'âge devront rester les mêmes car l'âge intermédiaire n'est pas souvent représenté entraînant une rupture dans l'évolution.

Pour le pool d'images, deux solutions sont possibles : privilégier soit la couleur qui, au vu des résultats de la littérature, a démontré ses effets (Rossion et Pourtois, 2004), soit le contrôle de normes plus nombreuses (Bonin et al., 2003). En effet, l'AoA, prédicteur déterminant de la vitesse, n'est pas pris en compte par Rossion et Pourtois. L'idéal serait de prendre celui de Bonin, de vérifier les images pour lesquelles la couleur est diagnostic et de les proposer en couleur dans une tâche de dénomination supplémentaire. Cela permet, par après, de comparer les temps pour ces items.

Il est difficile de dire quel dispositif est préférable. AlignTool reste le programme le plus récent et, par ce fait, il a pu éviter les écueils et récolter les améliorations des autres programmes avant lui. Il est également le seul à donner plusieurs indications temporelles. Néanmoins, la clé mise en place de Duyck et al. (2008) offre une alternative intéressante. Ces deux solutions ont été mises à l'épreuve et testées dans bon nombre de conditions expérimentales.

Bibliographie

Abrams, L., & Jennings, D. T. (2004). VoiceRelay: Voice key operation using Visual Basic. *Behavior Research Methods, Instruments, and Computers*, 36, 771-777.

Alario, F., & Ferrand, L. (1999). A set of 400 pictures standardized for French : Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers*, 31(3), 531-552. <https://doi.org/10.3758/BF03200732>

Alario, F., & Ferrand, L. (2004). Predictors of picture naming speed, *Behavior Research Methods, Instruments, & Computers*, 36(1), 140-155. <https://doi.org/10.3758/bf03195559>

Barresi, B., Nicholas, M. Connor, T. L., Obler, L. K., & Albert, M L. (2000). Semantic degradation and lexical access in age-related naming failures. *Aging, Neuropsychology, and Cognition*, 7(3), 169-178. <https://doi.org/10.1076/1382-5585>

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotchewa, T., ... Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, 10(2), 344-380. <https://doi.org/10.3758/bf03196494>

Bogliotti, C. (2012). Les troubles de la dénomination, *Langue française*, 174(2), 95-110. <https://doi.org/10.3917/lf.174.0095>

Bonin, P., Chalard, M., Méot, A., & Fayol, M. (2002). The determinants of spoken and written picture naming latencies. *British Journal of Psychology*, 93, 89-112.

Bonin, P. (2003). A new set of 299 pictures for psycholinguistic studies : French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers*, 35 (1), 158-167. <https://doi.org/10.3758/bf03195507>

Bonin, P. (2007). *Psychologie du langage. Approche cognitive de la production verbale de mots*. De Boeck.

Bonin, P., Guillemard-Tsaparina, D., & Méot, A. (2013). Determinants of naming latencies, objet comprehension times, and new norms for Russian standardized set of the coloried version of the Snodgrass and Vanderwart pictures. *Behavior Research*, 45, 731-745. <https://doi.org/10.3758/s13428-012-0279-9>

Bonin, P., Méot, A., Laroche, B., Bugaiska, A., & Perret, C. (2019). The impact of image characteristics on written naming in adults. *Reading and Writing*, 32, 13-31. <https://doi.org/10.1007/s11145-017-9727-3>

Boukadi, M., Zouaidi, C., & Wilson, M. (2016). Norms for name agreement, familiarity, subjective frequency, and imageability for 348 object names in Tunisian Arabic. *Behavior Research Methods*, 48(2), 585-599. <https://doi.org/10.3758/s13428-015-0602-3>

Bovart, P. (2014). *Etude des facteurs sous-tendant l'accès lexical au cours de vieillissement normal : Indice de Feeling-of-Knowing (FoK) lors de la récupération de noms de personnes familières* (mémoire en master en logopédie, Université de Liège).

Bragard, A., Schelstraete, M.-A., Collette, E., & Grégoire, J. (2010). Evaluation du manque du mot chez l'enfant : données développementales récoltées auprès d'enfants francophones de sept à 12 ans. *Revue européenne de psychologie appliquée*, 60, 113-127.

Britt, A., Ferrara, C., & Mirman, D. (2016). Distinct Effects of Lexical and Semantic Competition during Picture Naming in Younger Adults, Older Adults, and People with Aphasia. *Frontiers in Psychology*, 7, 813–. <https://doi.org/10.3389/fpsyg.2016.00813>

Brodeur, M., Guérard, K., & Bouras, M. (2014). Bank of Standardized Stimuli (BOSS) Phase II : 930 New Normative Photos. *PloS One*, 9(9), Article e106953–. <https://doi.org/10.1371/journal.pone.0106953>

Brodeur, M., Kehayia, E., Dion-Lessard, G., Chauret, M., Montreuil, T., Dionne-Dostie, E., & Lepage, M. (2012). The bank of standardized stimuli (BOSS): comparison between French and English norms. *Behavior Research Methods*, 44(4), 961–970. <https://doi.org/10.3758/s13428-011-0184-7>

Broos, W., Duyck, W., & Hartsuiker, R. (2018) Are higher level processes delayed in second language word production ? Evidence from picture naming and phoneme monitoring, *Language, Cognition and Neuroscience*, 33 (10), 1219-1234. <https://doi.org/10.1080/23273798.2018.1457168>

Brown, A. (2012). *The tip of the tongue state*. Psychology Press.

Connor, L.T., Spiro, A., Obler, L.K., & Albert, M.L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology : Series B*, 59(5), 203–209.

De Partz, M.-P. & Pillon, A. (2014). Sémiologie, syndromes aphasiques et examen clinique des aphasies. In X. Seron & M. Van der Linden (Eds). *Traité de neuropsychologie clinique de l'adulte* (tome 1, 2nd ed., pp. 249-265). De Boeck.

Dorot, D., & Mathey, S. (2013). Accès aux représentations sémantiques et phonologiques chez des adultes jeunes et âgés : une étude des mots sur le bout de la langue. *Psychologie française*, 58, 1-16. <https://dx.doi.org/10.1016/j.psfr.2012.06.002>.

Duyck, W., Anseel, F., Szmalee, A., Mestdagh, P., Tavernier, A., & Hartsuiker, R. (2008). Improving Accuracy in Detecting Acoustic Onsets. *Journal of Experimental Psychology : Human Perspection and Performance*, 34(5), 1317-1326.

Ferrand, L., Bonin, P., Méot, A., Augustinova, M., New, B., Pallier, C., & Brysbaert, M. (2008). Age-of-acquisition and subjective frequency estimates for all generally known monosyllabic French words and their relation with other psycholinguistic variables. *Behavior Research Methods*, 40(4), 1049-1054.

Feyereisen, P. (1997). A meta-analytic procedure shows an age-related decline in picture naming : comments Goulet, Ska, and Kahn (1994). *Journal of Speech and Hearing Research*. 40(6), 1328-1333.

Feyereisen, P., Demaeght, N., & Samson, D. (1998). Why Do Picture Naming Latencies Increase with Age : General Slowing, Greater Sensitivity to Interference, or Task-Specific Deficits ? *Experimental Aging Research*, 24(1), 21-51. <https://doi.org/10.1080/036107398244346>

Foster, K., & Foster, J. (2003). DMDX : A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers*, 33(1), 116-124.

- Garaizar, P., & Reips, U.-D. (2015). Visual DMDX : A web-based authoring tool for DMDX, a Windows display program with millisecond accuracy. *Behavior Research Methods*, 47(3), 620-631. <https://doi.org/10.3758/s13428-014-0493-8>
- Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech?, *Aging, Neuropsychology, and Cognition*, 24(5), 508-527. <https://doi.org/10.1080/13825585.2016.1226249>
- Guérard, K., Lagacé, S., & Brodeur, M. (2015). Four types of manipulability ratings and naming latencies for a set of 560 photographs of objets. *Behavior Research*, 47, 443-470. doi 10.3758/s13428-014-0488-5
- Heuer, S. (2016). The influence of image characteristics on image recognition : a comparison of photographs and line drawings. *Aphasiology*, 30(8), 943-961. <https://doi.org/10.1080/02687038.2015.1081138>
- Hoyau, E. (2018). *Stratégies de récupération et de sélection de l'information lexicale au cours du vieillissement sain : Evaluation multimodale des mécanismes de réorganisation cérébrale et impact des activités sociales sur les performances de dénomination orale d'objets*. Retrieved from <http://www.theses.fr/2018GREAS033/document>
- Indefrey, P., & Levelt, W. (2004). The spatial and temporal signatures of word production components. *Cognition*, 92(1), 101–144. <https://doi.org/10.1016/j.cognition.2002.06.001>
- Janssen, P., & Watter, S. (2008). SayWhen: An automated method for high-accuracy speech onset detection. *Behavior Research Methods*, 40(3), 744-751. <https://doi.org/10.3758/BRM.40.3.744>
- Kessler, B., Treiman, R., & Mullennix, J. (2002). Phonetic Biases in Voice Key Response Time Measurements. *Journal of Memory and Language*, 47, 145-171. <https://doi.org/10.1006/jmla.2001.2835>
- LaGrone, S., & Spieler, D. (2006). Lexical Competition and Phonological Encoding in Young and Older Speakers. *Psychology and Aging*, 21(4), 804-809.
- Levelt, W., Roelofs, A., & Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), 1–38. <https://doi.org/10.1017/S0140525X99001776>
- Levelt, W. (2001). Spoken word production : A theory of lexical access. *Proceedings of the National Academy of Sciences USA*, 98, 13464–13471. Retrieved from <https://www.narcis.nl/publication/RecordID/oai:repository.ubn.ru.nl:2066%2F148996>
- Lorenzoni, A., Peressotti, F., & Navarrete, E. (2018). The Manipulability Effect in Object Naming. *Journal of Cognition*, 1(1), Article 30, 1-8. <https://doi.org/10.5334/joc.30>
- Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Anđelković, D., & Blom, W. (2016). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*, 48(3), 1154–1177. <https://doi.org/10.3758/s13428-015-0636-6>
- Łuniewska, M., Wodniecka, Z., Miller, C., Smolík, F., Butcher, M., Chondrogianni, V., Hreich, E., Messarra, C., A. Razak, R., Treffers-Daller, J., Yap, N., Abboud, L., Talebi, A., Gureghian, M., Tuller, L., & Haman, E. (2019). Age of acquisition of 299 words in seven languages : American English, Czech, Gaelic, Lebanese Arabic, Malay, Persian and Western Armenian. *PloS One*, 14(8), Article e0220611–. <https://doi.org/10.1371/journal.pone.0220611>

- Madden, D., Sale, M., & Robinson, G. (2019 a). Age-related differences in idea generation and selection for propositional language, *Aging, Neuropsychology, and Cognition*, 26 (4), 486-506. <https://doi.org/10.1080/13825585.2018.1476668>
- Madden, D., Sale, M., & Robinson, G. (2019 b). Differentiating beyond name agreement for picture naming : insight from age-related selection deficits. *Journal of Speech, Language and Hearing Research*, 62, 1373-1380.
- Moreno-Martínez, F., & Montoro, P. (2012). An Ecological Alternative to Snodgrass & Vanderwart: 360 High Quality Colour Images with Norms for Seven Psycholinguistic Variables. *PloS One*, 7(5), Article e37527–. <https://doi.org/10.1371/journal.pone.0037527>
- Morrison, C. M., Hirsh, K. W., & Duggan, G. B. (2003). Age of acquisition, ageing and verb production : Normative and experimental data. *The Quarterly Journal of Experimental Psychology*, 56 (A), 2-26.
- Paesen, L., & Leijten, M. (2019). Name agreement and naming latencies for typed picture naming in aging adults. *Clinical linguistics and phonetics* <https://doi.org/10.1080/02699206.2019.1590734>
- Paolieri, D., Marful, A., Morales, L., & Bajo, M. (2018). The modulating effect of education on semantic interference during healthy aging. *PloS One*, 13(1), Article e0191656–(1-15). <https://doi.org/10.1371/journal.pone.0191656>
- Protopapas, A. (2007). CheckVocal : A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior, Research Methods*, 39(4), 859-862.
- Rabovsky, M., Schad, D., & Rahman, R. (2016). Langage production is facilitated by semantic richness but inhibited by semantic density : Evidence from picture naming. *Cognition*, 146, 240-244.
- Rastle, K., & Davis, M. (2002). On the Complexities of Measuring Naming. *Journal of Experimental Psychology : Human Perception and Performance*, 28(2), 307-314.
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's objet pictorial set: The role of surface detail in basic-level object recognition. *Perception*. 33, 217-236.
- Roux, F., & Armonstrong, B. (2017). Chronset : An automated tool for detecting speech onset. *Behavior Research Methods*, 49, 1864-1881. <https://doi.org/10.3758/s13428-016-0830->
- Roux, S., & Bonin, P. (2011). Comment l'information circule d'un niveau de traitement à l'autre lors de l'accès lexical en production verbale de mots ? Eléments de synthèse, *L'Année psychologique*, 111(1), 145-190. <https://doi.org/10.4074/S0003503311001060>
- Schillingmann, L., Ernst, J., Keite, V., Wrede, B., Meyer, A., & Belke, E. (2018). AlignTool : The automatic temporal alignment of spoken utterances in German, Dutch, and British English for psycholinguistic purposes. *Behavior Research Methods*, 50(2), 466–489. <https://doi.org/10.3758/s13428-017-1002-7>
- Schwartz, B., & Brown, A. (2014). *Tip of the tongue states and related phenomena*. Cambridge University Press.
- Schwartz, B., & Metcalfe, J. (2011). Tip-of-the-tongue (TOT) states : retrieval, behavior, and experience. *Mem Cogn*, 39, 737-749.
- Shafto, M., James, L., Abrams, L., Tyler, L., & Cam-CAN. (2017). Age-Related Increases in Verbal Knowledge Are Not Associated With Word Finding Problems in the Cam-CAN

Cohort : What You Know Won't Hurt You. *Journals of Gerontology : Psychological Sciences*, 72(1), 100-106.

Shao, Z., & Stiegert, J. (2016). Predictors of photo namins : Dutch norms for 327 photos. *Behavior Research*, 48, 577-584. <https://doi.org/10.3758/s13428-015-0613-0>

Sirois, M., Kremin, H., & Cohen, H. (2006). Picture-naming norms for Canadian French : Name agreement, familiarity, visual complexity, and age of acquisition. *Behavior Research Methods*, 38(2), 300-306.

Stadthagen-Gonzalez, H., Damian, M., Pérez, M., Bowers, J., & Marin, J. (2009). Name-picture verification as a contrôle measure for object naming : A task analysis and norms for a large set of pictures. *The Quarterly Journal of Experimental Psychology*, 62(8), 1581-1597.

Székely, A., & Bates, E. (2000). Objective Visual Complexity as a Variable in Studies of Picture Naming. *Center for Research in Language*, 12(2), 1-33.

Therriault, D., Yaxley, R., & Zwaan, R. (2009). The role of color diagnosticity in object recognition and representation. *Cogn process*, 335-342. <https://doi.org/10.1007/s10339-009-0260-4>

Thomas, J., Fozard, J. & Waugh, N. (1977). Age-Related Differences in Naming Latency. *The American Journal of Psychology*, 90(3), 499-509. <https://doi.org/10.2307/1421880>

Truffert, Cl. (2014). *Etude comparative des capacités d'accès aux noms propres et aux communs entre 30-39 ans, 40-49 ans et 50-59 ans, dans un contexte de dénomination d'images* (mémoire de master en logopédie). Université de Liège, Liège, Belgique.

Tyler, M., Tyler, L., & Burnham. (2005). The delayed trigger voice key : An improved analogue voice key for psycholinguistic research. *Behavior Research Methods*, 37(1), 139-147.

Verhaegen, C., & Poncelet, M. (2013). Changes in Naming and Semantic Abilities With Aging From 50 to 90 years. *Journal of The International Neuropsychological Society*, 19, 119-126. <http://dx.doi.org/10.1017/S1355617712001178>

Yamada, J., & Tamaoka, K. (2003). Measurement Errors in Voice-key Naming Latency for Hiragana. *Perceptual and Motor Skills*, 97, 1100-1106.

Liste des abréviations et des figures

a) Abréviations

Aoa = Age of acquisition

CTVK = Classical Threshold Voice Key

FoK = Feeling of Knowing

MBL = Mot sur le bout de la langue

NEVK = Noise Elimination Voice Key

NST = Node Structure Theory

PEDOI = Protocole Européen de Dénomination Orale d'Images

ToT = Tip of the Tongue

b) Figures

Fig. 1 : Modèles cognitifs (Roux et Bonin, 2011)

Fig. 2 : Schéma synthétique de la NST (Truffert, 2014)

Fig. 3 : Résultats de Paolieri et al. (2018)

Fig. 4 : *Loci* des facteurs selon Alario et al. (2004)

Fig. 5 : Exemples de stimuli en couleur (Rossion & Pourtois, 2004 ; Therriault & al., 2009)

Fig. 6 : Résultats de Britt et al. (2016)

Fig. 7 : Résultats des corrélations entre les pools français (Sirois et al., 2006)

Fig. 8 : Résultats de la comparaison entre deux clés vocales et analyse manuelle (Rastle & Davis, 2002)

Fig. 9 : Histogramme (Rastle & Davis)

Fig. 10 : Résultats des cinq expériences sur la comparaison de deux clés (NEVK et CTVK)
(Duyck & al., 2008)

Fig. 11 : Histogramme des résultats de Jansen et Watter (2002)

Fig. 12 : Résultats de la comparaison entre plusieurs dispositifs (Schillingmann, 2018)

Annexe : Démarches méthodologiques

Mes recherches se sont étendues sur plusieurs sujets qui étaient en majorité du type *background questions*, pour citer quelques exemples de questions : Quelles sont les hypothèses formulées pour expliquer les causes du ralentissement observé ? Quels sont les facteurs qui peuvent influencer le manque du mot ou la vitesse en dénomination ? A quel âge situe-t-on l'apparition de l'augmentation de la latence en dénomination ? Quel moyen est disponible pour mesurer la latence en dénomination ? Toutefois, le sujet principal de mes recherches, le temps de latence, est, lui, du type *foreground questions*. C'est pourquoi, je le prendrai comme exemple de canevas de recherche.

Avant d'aborder le sujet, j'avais l'avantage de posséder deux travaux de référence : celui de Verhaegen et Poncelet (2013) et celui de Truffert (2014).

J'ai commencé par établir une question de recherche initiale selon le modèle PICO (en parallèle des problématiques) :

(P) personnes d'âge différent (entre 30–80 ans), tout-venant ; mais surtout dans le contexte d'une tâche dénomination d'images

(I) techniques, programmes pour mesurer le temps de latence

(C) la technique ou le programme devra être autre que la technique manuelle et devra y être comparé pour examiner si c'est une bonne alternative

(O) fiabilité des résultats (mesure très précise en millisecondes)

De là, j'ai délimité des mots-clés : « latence dénomination », « vitesse en dénomination », « mesure latence dénomination », « naming latency », « naming speed », « naming measure », « naming assessment/evaluation », ... mais j'ai pu également élargir les mots-clés grâce à ceux proposés par la base de données comme, par exemple, « time response », « reaction time », « picture naming ». Des troncatures ont été employées pour élargir les résultats (dans le cas de singulier-pluriel ou de verbe-nom), ainsi que les opérateurs booléens (AND principalement pour chaque fois associer la tâche de dénomination). J'ai également utilisé la fonction « Similar (*chercher des articles similaires*) » ou « cité par » qui s'est montré très efficace pour trouver les derniers articles en date avec des titres s'éloignant des premiers

mots-clés initialement pensés (par exemple, *speech onset detection* ou *detecting speech onset*).

Les bases de données utilisées étaient celle de l'Université de Liège (notamment pour trouver les mémoires qui se fait aussi par Matheo) et PsycINFO/Ovid.

TABLE DES MATIERES

Introduction générale	p.3
------------------------------------	------------

I. Introduction théorique	p. 4-23
--	----------------

1. Le manque du mot et son évaluation	p.4-7
--	--------------

a) Le manque du mot : définition et caractéristiques	p.4-5
--	-------

b) Autres phénomènes : le « mot sur le bout de la langue » et le « Feeling of Knowing »	p. 5
--	------

c) Méthodes d'observation et d'évaluation dans la littérature	p. 5-7
---	--------

2. La tâche de dénomination : la problématique de l'âge	p.7-12
--	---------------

3. Pourquoi l'âge influence-t-il le manque du mot ?	p. 12-22
--	-----------------

a) Accès lexical : mécanismes et modèles cognitifs	p.12-16
--	---------

b) Hypothèses de l'effet de l'âge sur le manque du mot	p. 17-18
--	----------

c) Quelques études examinant les différentes hypothèses	p.18-22
---	---------

4. Problématiques	p. 22-23
--------------------------------	-----------------

II. Dénomination orale d'images	p.24-42
--	----------------

1. Tâche de dénomination orale d'images : préambule	p. 24-26
--	-----------------

2. Facteurs influençant la tâche de dénomination orale d'images	p. 27-38
--	-----------------

a) Facteurs visuels	p. 27-30
---------------------------	----------

b) Facteurs sémantiques	p. 30-31
-------------------------------	----------

c) Facteurs lexicaux	p. 31-35
----------------------------	----------

d) Facteurs phonologiques	p. 36
e) Autres variables	p. 36-38
3. Ensembles d'images normés	p. 38-41
a) Au préalable	p. 38-39
b) Pools disponibles en français	p.40-41
c) Différences entre les formats	p. 41
4. Conclusions	p. 42
III. Moyens pour mesurer la latence	p. 43-63
1. Revue des moyens proposés dans la littérature	p. 43-53
a) Analyse manuelle	p. 44
b) Clé vocale	p. 44-48
c) DMDX, CheckVocal et Visual DMDX	p. 48-50
d) SayWhen	p. 51
e) Chronset	p. 52
f) AlignTool	p. 52-53
2. Mise à l'épreuve des moyens et techniques dans la littérature	p. 54-63
Discussion	p. 64-72
I. Préambule	p. 64
II. L'effet de l'âge	p. 65-67
III. La problématique de la latence	p. 67-70
IV. Les dispositifs	p. 70-71

V. Conclusions	p. 71-72
Bibliographie	p.73-77
Listes des abréviations et des figures	p. 78
Annexe méthodologique	p. 79-80