

Projet tutoré :
LE ROLE DE LA SYLLABE EN LECTURE

ANGELONI Marine - DOBMANN Elodie

30 mai 2016

Encadrants : Manuel Rebuschi¹ Daniel Zagar²

M1 Sciences Cognitives Appliquées : 2015-2016

1. Responsable Master M1 SCA

2. Professeur en Psychologie expérimentale, responsable de l'axe Cognition

Organisme d'accueil :
Université de Lorraine



Interpsy



LE ROLE DE LA SYLLABE EN LECTURE

ANGELONI Marine - DOBMANN Elodie

2015 - 2016

REMERCIEMENTS

En premier lieu, nous remercions notre tuteur Daniel Zagar qui nous a accompagnées tout au long du développement de notre projet, pour ses conseils, ses explications et sa bienveillance.

Nous remercions également H  l  ne Maire pour son aide lors de la passation de l'exp  rience ainsi que tous les   tudiants volontaires qui ont accept   d'y participer.

Enfin, un grand merci aux relecteurs anonymes pour leur patience et leurs remarques pertinentes.

Table des matières

I Article Scientifique	1
1 Introduction	1
1.1 La syllabe dans le processus de lecture	1
1.2 Perception de la syllabe à l'écrit chez les lecteurs français . .	2
1.3 L'effet syllabique, la fréquence syllabique et traitement visuel des mots	4
1.4 Objectif de l'étude	8
2 Experience	8
2.1 But de l'expérience	8
2.2 Méthode	9
2.2.1 Participants	9
2.2.2 Matériel	9
2.2.3 Procédure	10
3 Résultats et discussions	11
3.1 Analyse de la structure syllabique	12
3.2 Analyse de la fréquence syllabique	12
3.3 Analyse du SOA	13
3.4 Analyse du SOA et de la fréquence	15
3.4.1 structure CV ambigu	15
3.4.2 Structure CV non ambigu	16
3.4.3 Structure CVC	16
3.4.4 Conclusion	17

II	Compléments d'informations	18
1	Mise en place de l'expérience	18
2	Programmation de l'expérience	20
3	Passation de l'expérience	21
4	Problèmes rencontrés lors de la passation	21
5	Interprétation des résultats	22
6	Conclusion sur l'ensemble du projet tutoré	23
III	Annexes	23
A	Exemple traitement du lexique	23
B	Carte de Kohonen	24
	Références	26

Première partie

Article Scientifique

Effet de la fréquence syllabique dans la reconnaissance visuelle des mots : représentation visuelle du découpage syllabique.

Résumé

Connaissant les effets de la fréquence syllabique sur le processus de reconnaissance visuelle des mots, nous avons étudié l'influence de cet effet lorsqu'on met en place une représentation visuelle du découpage syllabique d'un mot à l'écrit. Pour les mots de type CV ambigus, nous avons constaté un effet facilitateur de la fréquence lorsque le mot possède peu de voisins. Cependant, plus le SOA augmente, plus ce même effet devient inhibiteur car les sujets répondent plus lentement. Concernant les structures CV non ambigus et CVCs, les résultats obtenus sont à l'opposé. Ils ne correspondent pas à nos attentes.

Mots clés : Syllabe; Fréquence syllabique; Reconnaissance visuelle des mots; Redondance orthographique

1 Introduction

1.1 La syllabe dans le processus de lecture

Le processus de lecture est un phénomène très complexe mettant en relation la représentation écrite du mot ainsi que sa prononciation. Ce phénomène est facilement identifiable en lecture silencieuse, puisque nous entendons notre voix dans notre tête qui « prononce » les mots que nous lisons. Selon Rubenstein et Garfield (cités par Rossi 1978)[9], l'accès à la dimension

sémantique d'un mot est précédé par sa représentation mentale, une étape nécessaire pour permettre sa prononciation. Dès notre plus jeune âge, là où commence l'apprentissage de la lecture, on décompose naturellement un mot en comptant les syllabes qui le composent avant même de savoir lire, ce qui explique l'intérêt que l'on porte depuis plusieurs années à étudier l'influence de la syllabe à l'écrit. Cette dernière constitue une « unité phonologique qui se définit par la présence d'une voyelle combinée ou non à des consonnes la précédent et/ou la succédant » (Doignon et Zagar, 2006)[8] et représente l'élément de segmentation de la parole (Ecalte et Magnan, 2002)[3].

1.2 Perception de la syllabe à l'écrit chez les lecteurs français

Ces dernières années, plusieurs procédés expérimentaux ont été mis en place afin d'étudier l'impact de l'unité syllabique à l'écrit chez les lecteurs français. En effet, la frontière syllabique n'est pas toujours facile à identifier. Prenons par exemple MANÈGE et MANGER : bien que les trois premières lettres de ces deux mots soient identiques, on n'obtient pas le même découpage syllabique. Dans le premier cas, MANÈGE suit le découpage MA\NÈGE, tandis que MANGER se décompose MAN\GER.

Par ailleurs, il existe un procédé qui a permis de mettre en évidence l'influence de la première syllabe à l'écrit : le paradigme des conjonctions illusives (Prinzmetal et al, 1986)[4]. Il se déroule comme suit : on présente au sujet un mot bisyllabique (deux syllabes) bicolore sur un écran. On lui demande ensuite de nous donner la couleur de la lettre cible qui correspond à la frontière du changement de couleur dans le mot. Cette frontière peut correspondre ou non à la frontière syllabique du mot. On lui demande par exemple de rapporter la couleur du R dans CARTE ou CARTE. On repère

ensuite les erreurs commises par le sujet, c'est-à-dire lorsqu'il se trompe de couleur parmi celles affichées à l'écran, ce que l'on appelle les conjonctions illusoires. Puis, on analyse le type d'erreur commis par ces individus. Il en existe deux sortes: les erreurs de préservations et les erreurs de violation. Une erreur de préservation est avérée si celle-ci est en adéquation avec la frontière syllabique du mot (le sujet dit que le R est rouge alors qu'on lui a présenté le mot CARTE). En revanche, si le sujet identifie le R comme étant de couleur bleue alors qu'on lui a présenté CARTE, il commet alors une erreur de violation. Cette expérience a permis de montrer que la plupart des erreurs commises sont des erreurs de préservation, ce qui indique que le sujet a perçu la syllabe. Ce procédé est réutilisé par Doignon et Zagar (2005)[7]. Ils ont mis en place deux conditions expérimentales: une congruente et une conflictuelle. Concernant la condition congruente, la frontière syllabique ou phonologique (/) du mot correspond à sa frontière orthographique(*), repérable par le bigramme³ qui ne forme pas le début ou la fin d'une syllabe. Exemple: MARTEAU donne la segmentation suivante en situation congruente: MAR/*TEAU Dans le cas conflictuel, la frontière phonologique et orthographique du mot diffèrent. On obtient par exemple MAR /T*EAU. En situation congruente, si la majorité des erreurs recueillies correspondent à des erreurs de préservation, alors cela indique que le lecteur reconnaît la syllabe dans le processus de traitement du mot présenté. En revanche, en condition conflictuelle (frontière orthographique et phonologique distinctes), une erreur de préservation est caractéristique d'un recours aux informations phonologiques, tandis qu'une erreur de violation signifie que le sujet a recours à la représentation orthographique. D'après les résultats

3. Sous unité lexicale d'un mot qui correspond à un ensemble de deux lettres successives. Exemple, le premier bigramme de SOLEIL est SO

de cette expérience, lorsque la frontière syllabique et la frontière orthographique coïncident, on constate d'avantage d'erreurs de préservation ce qui permet d'affirmer que l'on accède à la syllabe à l'écrit. En condition conflictuelle, le nombre d'erreurs de préservation est quasi similaire au nombre d'erreurs de violation, donc la perception de la syllabe dépend à la fois de ses caractéristiques orthographiques et phonologiques. Le repérage de ces caractéristiques orthographiques s'explique par la redondance orthographique (Seidenberg 1987)[10]. Selon ce principe, le lecteur est capable de reconnaître la fréquence des bigrammes qui délimitent une syllabe. En effet, il sait grâce à son expérience de lecteur que RL ne constitue pas le début ou la fin d'une syllabe. Cela s'explique par la co-occurrence de lettres: Il sait déterminer si un ensemble de lettres est rencontré fréquemment. Si c'est le cas, alors cet ensemble délimitera une syllabe.

La syllabe en français s'inscrit comme un élément indispensable au traitement visuel des mots tant par ses propriétés phonologiques que par ses propriétés orthographiques. Des études ont permis d'identifier les raisons pour lesquelles le traitement syllabique va influencer la rapidité de reconnaissance des mots.

1.3 L'effet syllabique, la fréquence syllabique et traitement visuel des mots

En 1999, Colé, Magnan et Grainger[2] ont mis en place une tâche d'amorçage dont le processus expérimental est le suivant : on présente à des lecteurs experts ou débutants une séquence de deux ou trois lettres suivie d'un mot, et ils doivent dire si la séquence présentée correspond ou non au mot présenté par la suite. Par exemple, on présente PO ou POR pour PORTE. Cette

expérience a permis de mettre en évidence un effet syllabique : si l’amorce présentée correspond exactement à la première syllabe du mot lui succédant, alors la réponse est plus rapide.

L’effet de fréquence syllabique⁴ a quant à lui été démontré à travers de nombreuses tâches de décision lexicale dans diverses langues européennes, notamment la langue française et espagnole. Le principe est le même dans tous les cas : Il s’agit de présenter à des sujets, des items (mots ou pseudo-mots) à l’écran pendant quelques millisecondes et ces derniers doivent identifier s’ils appartiennent à leur lexique, à l’aide des touches réponses du clavier. Chez les espagnols, on observe un effet inhibiteur de la fréquence de la première syllabe lorsque celle-ci est élevée : les sujets mettent plus de temps à répondre qu’en cas de fréquence syllabique faible (Carreras, Alvarez et Vega, 1993)[6]. Des résultats similaires ont été obtenus chez les français. Ce phénomène provient de la compétition syllabique : lorsque la fréquence de la première syllabe d’un mot est élevée, alors cela va engendrer l’activation de tous les voisins partageant la même première syllabe qui vont entrer en compétition, ce qui explique que l’on mettra plus de temps à répondre (Mathey, Zagar, Doignon, et Seigneuric, 2006)[5]. L’effet de fréquence syllabique est une conséquence de la redondance orthographique : Si le lecteur reconnaît un ensemble de lettres fréquemment rencontrées et que cet ensemble correspond en tous points à une frontière syllabique, alors la syllabe s’active très rapidement selon un processus automatique, entraînant les compétiteurs partageant la même première syllabe. Bien que les Espagnols soient sensibles à la fréquence syllabique, ils ne seraient pas sensibles aux caractéristiques orthographiques (redondance orthographique). En effet, lorsque le premier bigramme a une fréquence élevée, cela n’a pas d’influence sur l’effet inhi-

4. La fréquence syllabique correspond toujours à la fréquence de la première syllabe

biteur de la fréquence syllabique. En revanche, chez les Français, l'effet de fréquence syllabique est liée à la fréquence du premier bigramme. Lorsque la fréquence syllabique, est élevée et que celle du premier bigramme voisin activé l'est également, alors elle aura un effet inhibiteur sur le traitement des mots (on les reconnaîtra plus lentement)A l'inverse, si la fréquence de ce bigame est faible, cette fréquence,toujours élevée, aura un effet facilitateur donc le mot sera reconnu plus rapidement (Mathey, Zagar, Doignon, Seigneuric, 2006)[5]. Par exemple, CLA/VIER et CAR/NAGE ont tous les deux une fréquence syllabique élevée, ils ont donc beaucoup de voisins partageant la même première syllabe. Cependant, CLAVIER sera reconnu plus rapidement car CL est un bigramme beaucoup moins fréquent que CA. Ce principe s'implique indépendamment du nombre de syllabes contenues dans le mot (monosyllabique, bisyllabique, ...). Ce test confirme l'importance des caractéristiques orthographiques et phonologiques, impliquées dans le traitement visuel des mots, précédemment relevées dans le paradigme des conjonctions illusoire chez les Français. Cette différence observée chez les Espagnols peut s'expliquer entre autre par un système d'apprentissage de la lecture différent entre les deux pays. De plus, en espagnol, l'orthographe des mots est très superficiel ce qui explique une insensibilité à la redondance orthographique et donc un traitement centré sur la phonologie de la syllabe (Carreras, Alvarez et Vega, 1993)[6].

Le processus de reconnaissance d'un mot peut différer de manière très significative d'une langue à une autre. La langue coréenne a par exemple aucun effet de fréquence syllabique impliqué dans le traitement des mots à l'écrit. En effet, leur système alphabétique diffère de celui des français, puisqu'il s'appuie une représentation visuelle des frontières syllabiques à l'écrit. Les lettres coréennes sont regroupées en bloc syllabiques. De ce fait,

les lecteurs coréens peuvent associer directement les composantes orthographiques et phonologiques des syllabes et ce, dès leurs plus jeune âge. Le système alphabétique coréen comparé au système européen permet de comprendre l'influence phonologique de la première syllabe dans l'identification des mots, car ils ne sont pas dépendants ou soumis au découpage syllabique (Kwon, Nam et Lee, 2015, cités par Zagar 2015)[1], comme le montre la figure ci-après.

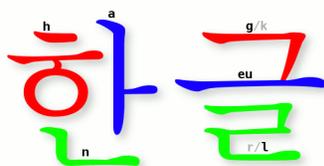


FIG. 1 – Exemple de mot à deux syllabes dans la langue coréenne : Hangeul

Toutes ces expériences montrent que la syllabe a une fonction indiscutable dans la reconnaissance visuelle des mots. Cependant, il semblerait qu'il y ait des différences plus ou moins majeures entre les différentes langues. Ces dernières s'expliquent notamment par les caractéristiques du système alphabétique utilisé, la complexité orthographique, et par des méthodes appliquées lors de l'apprentissage de la lecture. Puisque les systèmes d'écriture basés sur une segmentation visuelle des frontières syllabiques facilitent le processus de traitement des mots à l'écrit, il serait intéressant de tester l'influence d'un tel mode de représentation, en français, à l'aide d'un procédé expérimental reproduisant ce type de système.

1.4 Objectif de l'étude

Le système d'écriture Français n'étant pas basé sur un découpage visuel des syllabes, la processus de traitement des mots s'effectue selon les propriétés phonologiques et orthographiques de la première syllabe (Doignon et Zagar 2005)[7]. La présente étude vise à déterminer si un tel découpage engendrerait une variation de l'effet de fréquence syllabique qui est très influant dans le processus d'identification des mots dans notre système d'écriture actuel (Mathey Zagar, Doignon ,Seigneuric,2006)[5]. Afin de simuler un tel système, nous avons mis en place une tâche de décision lexicale permettant de faire ressortir visuellement le découpage syllabique d'un mot, en faisant varier le SOA (stimulus onset asynchrony) qui va déterminer le temps de séparation des syllabes.

2 Experience

2.1 But de l'expérience

Nous analyserons l'effet du SOA conjointement à la fréquence syllabique afin d'étudier l'effet du SOA sur l'inhibition de la fréquence syllabique. Nous supposons que plus nous augmenterons le SOA, plus l'effet inhibiteur de la fréquence syllabique sera important peu importe la structure syllabique. En effet, l'augmentation du SOA va accentuer le découpage syllabique, et donc rendre saillante la première syllabe qui apparaîtra plus longtemps. Cela devrait accentuer la compétition syllabique et par conséquent révéler un effet inhibiteur : les sujets répondraient plus lentement. Nous étudierons cette influence sur différents types de structure syllabique.

2.2 Méthode

2.2.1 Participants

Cinquante-quatre étudiants de l’université de Lorraine ont participé à l’expérience. Tous étaient de langue maternelle française avec une vision normale ou une vision corrigée se rapportant à une vision normale. Ils ont été équitablement répartis dans 3 groupes, chaque groupe se différenciant par la valeur du SOA appliqué lors de la passation.

2.2.2 Matériel

180 items ont été sélectionnés dans Lexique 3, une base de données qui répertorie 135 000 mots de la langue française. Ces items sont divisés en 2 catégories comprenant chacune 90 items : des mots et des pseudo-mots.

Les mots sont divisés en six groupes représentés dans le tableau ci-dessous Tab 1

Structure \ Fréquence	Peu de voisins (45)	Beaucoup de voisin (45)
CV ambigus (30)	Bonus	Balai
CV non-ambigus (30)	Latin	Vigile
CVCs (30)	Furtif	Formel

TAB. 1 – *Représentation du matériel utilisé*

Concernant la structure syllabique, nous avons trois groupes distincts : les CVCs qui sont des mots dont la première syllabe est composée par une consonne suivie d’une voyelle et d’une autre consonne (marteau), les CVs dont la première syllabe est composée par une consonne suivie d’une voyelle(soleil), et en les CV ambigus, qui correspondent aux mots dont la troisième lettre peut prêter à confusion (l, n, r). En effet, en fonction de la

lettre qui suit, un CV ambigu peut devenir un CVC (pa-rent, par-tage).

Dans le cas de la fréquence syllabique, « Peu de voisins » signifie que ce mot possède peu de mot concurrents c'est à dire peu de mots partageant la même première syllabe. Il a donc un rang⁵ inférieur à 5. Par exemple, BONUS est identifié plus vite que BALAI car, lorsqu'on lit BALAI, notre cerveau va activer tous les mots dont la première syllabe est BA. Or, ils sont plus nombreux que des mots commençant par BO.

2.2.3 Procédure

L'expérience est divisée en trois conditions; . La condition 1 possède un SOA de 0 ms (milliseconde), la condition 2 un SOA de 64 ms et la condition 3 un SOA de 128 ms. L'expérience repose sur une tâche de décision lexicale. On va présenter au sujet l'ensemble des items (mot ou pseudo-mot) expérimentaux de manière aléatoire en deux parties séparées par une pause. Chaque item est tronqué suivant sa première syllabe puis présenté de manière asynchrone (en deux parties), où la première syllabe correspond à l'amorce, et la deuxième correspond à la cible, avec un SOA séparant les deux parties. La personne doit décider si l'item qui apparaît à l'écran forme un mot de la langue française à l'aide des touches CTRL du clavier. Si le sujet est droitier, alors la touche réponse pour un mot existant sera la touche CTRL de droite et la touche pour un pseudo-mot sera CTRL de gauche. Les paramètres sont évidemment inversés pour un gaucher. Par ailleurs, la passation comporte une phase de familiarisation durant laquelle les temps de réaction de ne sont pas pris en compte lors de l'interprétation des résultats, puisque celle-ci permet aux sujets de découvrir la tâche qu'il

5. Un rang est déterminé au préalable par le tri de sa fréquence d'apparition dans les films.

doivent réaliser.

3 Résultats et discussion

Durant notre expérience, nous avons mesuré les temps de réactions qui correspondent à notre variable dépendante. Les tableaux ci-dessous représentent la moyenne des temps de réaction (en ms) en fonction des variables indépendantes : les SOA (0 ms, 64 ms, 128 ms), les types de fréquences syllabiques (peu de voisin, beaucoup de voisin) et enfin la structure syllabique (CV ambigu, CV non-ambigu, CVCs).

Structure \ Fréquence	Peu de voisins	Beaucoup de voisins
CV ambigu	680,30	695.75
CV non-ambigu	708.5918	692.42
CVCs	693.14	690.49

TAB. 2 – Moyenne des temps de réaction avec un SOA de 0 ms

Structure \ Fréquence	Peu de voisins	Beaucoup de voisins
CV ambigu	672.45	724.25
CV non-ambigu	688.56	643.34
CVCs	679.26	661.54

TAB. 3 – Moyenne des temps de réaction avec un SOA de 64 ms

Structure \ Fréquence	Peu de voisins	Beaucoup de voisins
CV ambigu	632.05	677.08
CV non-ambigu	658.17	639.25
CVCs	645.45	610.30

TAB. 4 – Moyenne des temps de réaction avec un SOA de 128 ms

3.1 Analyse de la structure syllabique

Nous avons étudié dans un premier temps l'effet de la structure syllabique sur le temps de réaction. On remarque que les CVCs sont perçus plus rapidement que les CVs (8 ms en moyenne). De plus, un CV ambigu est perçu plus lentement qu'un CV non ambigu (5 ms). Ceci peut s'expliquer par le fait que notre cerveau a plus de difficulté à traiter les ensembles de lettres qui peuvent présenter des ambiguïtés.

L'analyse de la variance n'est pas concluante en ce qui concerne l'effet de la structure sur le temps de réaction.

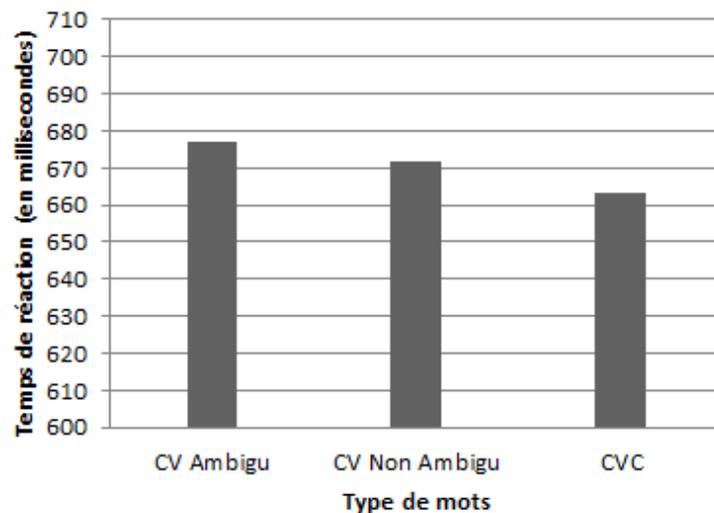


FIG. 2 – Étude du temps de réaction en fonction de la structure syllabique

3.2 Analyse de la fréquence syllabique

L'étude de la fréquence syllabique montre que lorsqu'un mot possède peu de voisins, il sera plus difficile à identifier qu'un mot possédant beaucoup de voisins, c'est à dire une fréquence élevée de sa première syllabe (6 ms).

L'analyse de la variance a permis de mettre en évidence l'influence de la fréquence syllabique dans le cas où la variable aléatoire est le sujet. Lorsque l'on fixe la structure syllabique, on obtient $F1(1,45)_{CV\text{ambigu}} = 12.64$ et $p = 0.009$ et $F1(1,45)_{CV\text{nonambigu}} = 8.65$ et $p = 0.0051$.

Elle a un donc un effet inhibiteur sur le temps de réaction. Plus un mot aura peu de voisins, plus il sera difficile à identifier. Or nous devrions apercevoir l'effet inverse. Plus un mot possède de voisins, plus l'effet inhibiteur devrait être important.

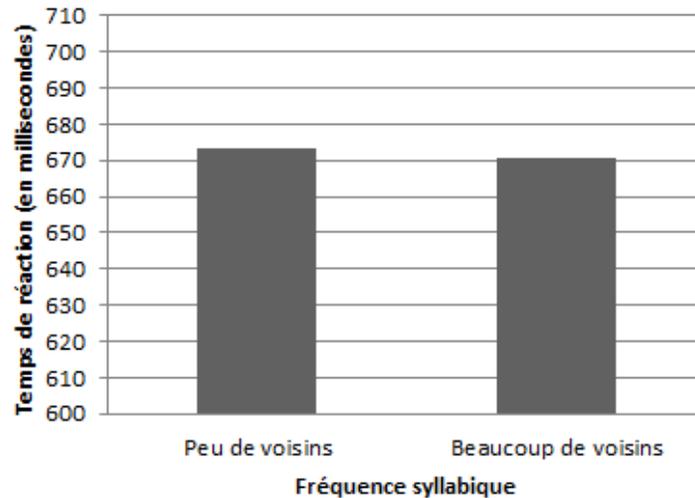


FIG. 3 – Étude du temps de réaction en fonction de la fréquence syllabique

3.3 Analyse du SOA

Afin de comprendre les résultats obtenus précédemment concernant l'effet de la fréquence syllabique, nous avons également analysé l'influence du SOA sur le temps de réaction. Les résultats montrent que, plus ce dernier augmente, plus le temps de réaction va diminuer. Le temps de réaction

moyen lors d'un SOA de 64 ms sera plus rapide (14 ms) que celui d'un SOA de 0 ms, tandis qu'un SOA de 128 ms aura des temps de réaction moyens plus courts (35 ms) qu'un SOA de 64 ms.

Cette influence a été révélée par l'analyse de variance. En prenant l'item comme variable, nous avons obtenu $F2(2,168) = 19.65$ et $p \simeq 0$. De plus, pour la même variable aléatoire, si l'on fixe la structure (CV ambigu, CV non ambigu, CVC), l'analyse de variance montre une dépendance entre le SOA et le temps de réaction $F2(2,56)_{CV\text{ ambigu}} = 5.379$ et $p = 0.0073$, $F2(2,56)_{CV\text{ non ambigu}} = 5.961$ et $p = 0.0045$ et enfin $F2(2,56)_{CVC} = 12.71$ et $p = 0.000028$

Cette dépendance s'explique par le fait que le sujet aura plus le temps d'analyser la première syllabe, avec d'un SOA élevé, avant l'apparition de la seconde partie du mot. Lors d'un SOA faible, voire nul, les deux parties du mot apparaissent de manière quasiment simultanée donc elles laissent moins de temps au cerveau pour analyser la première partie du mot (première syllabe), ce qui n'est pas le cas avec un SOA important. Le sujet devra alors analyser la syllabe plus rapidement.

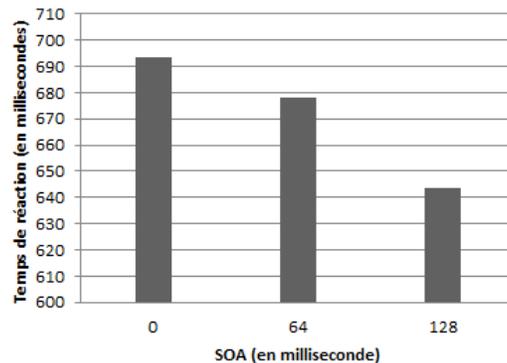


FIG. 4 – Étude du temps de réaction en fonction du SOA

3.4 Analyse du SOA et de la fréquence

Nous avons ensuite étudié l'effet conjoint du SOA et de la fréquence syllabique sur le temps de réaction, en analysant leur influence sur les différents types de structure.

3.4.1 structure CV ambigu

La courbe ci-dessous suggère que les mots ayant beaucoup de voisins sont identifiés plus lentement que les mots ayant peu de voisins, ce qui met en évidence l'effet inhibiteur de la fréquence syllabique. Avec un SOA de 0 ms, on remarque une différence de 15 ms entre les mots ayant peu de voisins et ceux en ayant beaucoup. Cette différence augmente avec le SOA. Pour un SOA de 64 ms, la différence est de 51 ms. Celle-ci est identique pour un SOA de 128 ms. Elle est due à l'effet inhibiteur de la première syllabe. On peut en conclure que pour les CVs ambigus, le SOA influe sur l'effet de la fréquence syllabique. Plus celui-ci augmente, plus l'inhibition augmente.

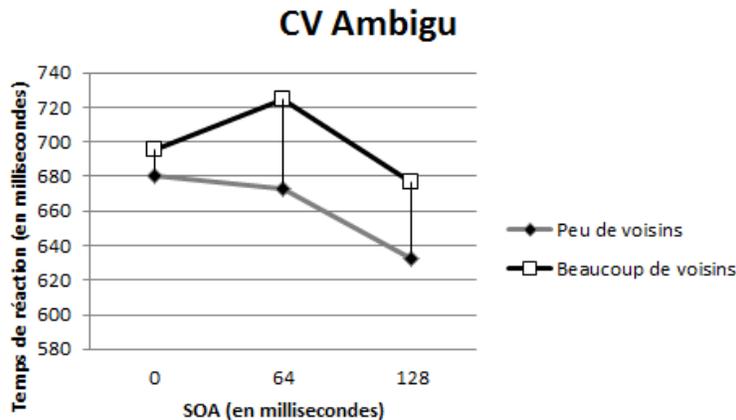


FIG. 5 – Étude du temps de réaction en fonction du SOA et de la fréquence syllabique pour les CVs ambigus

3.4.2 Structure CV non ambigu

Contrairement à ce que l'on a constaté pour les CVs ambigus, l'effet de la fréquence syllabique est inversé. En effet, plus un mot possède de voisins, plus vite il sera identifié. On a donc un effet facilitateur et non un effet inhibiteur de la première syllabe. Cet effet croît également quand le SOA augmente. Il passe de 16 ms pour un SOA de 0 ms à 45 pour un SOA de 64 ms.

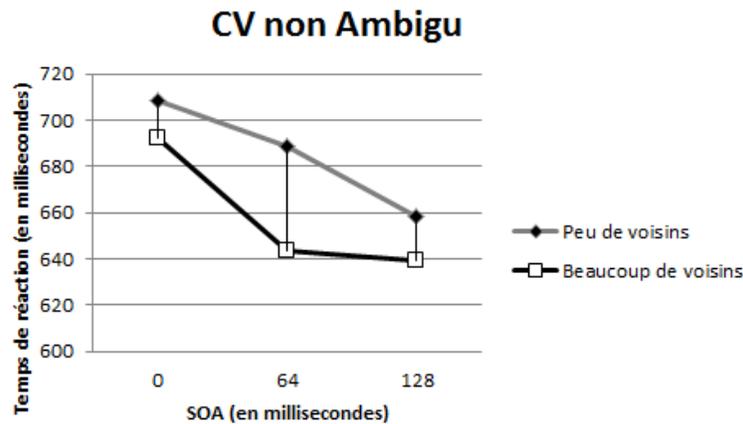


FIG. 6 – Étude du temps de réaction en fonction du SOA et de la fréquence syllabique pour les CVs non ambigus

3.4.3 Structure CVC

L'effet de la fréquence syllabique et du SOA sur des mots dont la structure est de type CVC indiquent que l'effet de la fréquence est inversé comme pour les CVs non ambigus, c'est à dire facilitateur et non inhibiteur. Cependant, contrairement aux CVs non ambigus, l'effet est moins marqué (35 ms maximum).

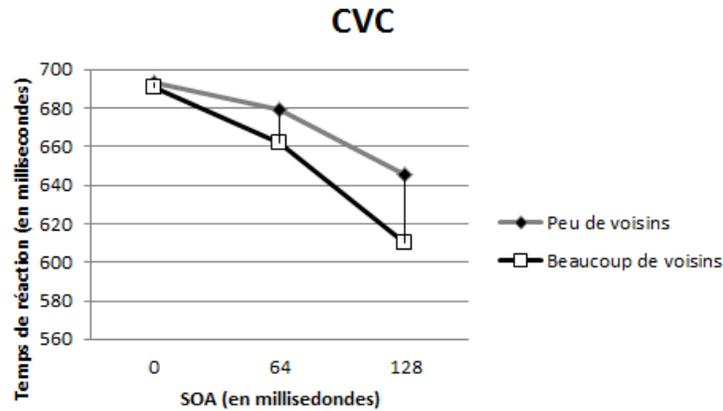


FIG. 7 – Étude du temps de réaction en fonction du SOA et de la fréquence syllabique pour les CVCs

3.4.4 Conclusion

De manière générale, on constate que l'effet de la fréquence syllabique est d'autant plus important que le SOA est élevé. Il est inhibiteur pour les CVs ambigus donc les mots sont reconnus plus lentement, et facilitateur avec les CVs non ambigus et CVCs qui seront reconnus plus vite. On pense que les incohérences que nous avons constatées sur l'effet de la fréquence syllabique peuvent être dues à la structure de la première syllabe.

En effet, on remarque que le temps de réaction nécessaire identifier un mot de ce type CVC, est le plus court parmi les trois types de structure (Voir figure2).

De plus, on peut constater que l'effet de fréquence syllabique (la différence des temps de réaction entre les mots possédant peu et beaucoup de voisins), (cf. TAB 5), est le plus faible avec des CVCs. On peut donc en conclure que les mots de type CVCs sont les plus faciles à identifier. Ce qui explique que l'effet de la fréquence syllabique est la plus faible.

Ces observations réfutent notre hypothèse initiale. En effet nous avons supposé que le SOA influencerait l'effet de fréquence syllabique indépendamment du type de structure de la première syllabe. Or, on constate une influence significative du type de structure sur l'effet de fréquence syllabique.

CV ambigu	CV non ambigu	CVCs
38	27	18

TAB. 5 – *Différence des temps de réaction (ms) pour les mots avec peu et beaucoup de voisins en fonction de la structure.*

Le raisonnement est analogue pour les CVs non ambigu. Nous pouvons cependant apercevoir un effet de la fréquence syllabique. La différence entre les temps de réaction pour les mots avec peu et beaucoup de voisins est d'autant plus grande que le SOA est élevé.

On ne peut néanmoins pas expliquer l'inversement de l'effet inhibiteur pour les classes CV non ambigu et CVC. Des expériences complémentaires permettraient d'infirmer ou confirmer ces résultats. Il faudrait notamment apporter des modifications au matériel utilisé, de façon à contrôler différemment certains paramètres (fréquences, composition, ...) qui auraient pu influencer l'effet inhibiteur sur les CVs non ambiguës et les CVCs.

Deuxième partie

Compléments d'informations

1 Mise en place de l'expérience

L'expérience repose sur le principe de décision lexicale. Une série de mot est présentée au sujet, et celui-ci a pour objectif de déterminer l'existence (ou

non) des mots proposés. Chaque mot est présenté syllabe par syllabe, séparé par un SOA (les mots possèdent chacun deux syllabes). Détermination du lexique

La première étape de l'expérience consiste à sélectionner les mots susceptibles d'être présentés au sujet. De ce fait, nous avons eu recours à la base de données Lexique 3, qui regroupe 135 000 mots de la langue française. Ensuite, nous avons filtré cette base de données selon les critères suivants :

1. Le mot doit contenir 2 syllabes
2. Le mot doit contenir entre 5 et 7 lettres
3. Le mot doit être un lemme⁶
4. Le mot ne doit pas contenir de x, de y et de double consonne
5. Le mot doit commencer par une consonne
6. Le mot ne doit pas être un mot composé
7. Les mots ayant une fréquence films comprise entre 2 et 18

Nous avons ensuite séparé les mots qui respectent ces critères en deux groupes distincts :

1. Un groupe pour les CVs
2. Un autre pour les CVCs

De plus, chaque groupe est divisé en sous-groupe en fonction de leur première syllabe (, ba, bi, bo, bu, da, di, ...). Exemple Annexe A

Dans le but d'étudier l'influence de la fréquence, nous devons apparier les mots. Nous devons pour cela diviser chaque liste en deux sous-listes :

1. Une qui contient les mots avec un haut rang
2. Une autre qui contient les mots avec un bas rang.

6. Unité lexicale, forme canonique d'un mot.

Le rang d'un mot est déterminé en fonction de sa fréquence d'apparition dans les films. Ensuite, leur position dans la cohorte de mots représente leur rang. Le mot avec la plus haute fréquence aura le rang de 1.

Ensuite, pour effectuer l'appariement, nous avons utilisé le principe des cartes de Kohonen vu dans le cadre du cours "Mémoire et Apprentissage numérique". Le programme consiste à trouver cinq mots de rang plus faible pour chaque mot de rang élevé. Puis, il faut étudier les propositions émises par celui-ci et déterminer lesquels des cinq mots sont les plus pertinents.

2 Programmation de l'expérience

L'expérience a été programmée sur le logiciel E-Prime, un outil un utilisé pour la conception d'expérience en psychologie. Il permet de présenter des éléments à l'écran (texte, image) ainsi que des sons en recueillant des temps de réponses. Nous avons commencé à programmer l'expérience, cependant le temps d'assimiler le fonctionnement du logiciel, le délai avant le début de l'expérience étant court, nous avons utilisé une version proposée par notre tuteur M. Zagar. De plus, nous avons consacré énormément de temps dans le traitement et le filtrage des mots retenus pour l'expérience. Au premier abord, un tri des mots semble simple, mais plus on avance, plus les critères de tri s'affinent et l'on a mis un certain temps à bien comprendre toutes les notions et spécificités en jeu sans lesquelles nous n'aurions pu réaliser ce travail.

3 Passation de l'expérience

L'expérience s'est déroulée sur deux semaines du 18 avril au 29 avril au CLSH. Nous avons à notre disposition une salle informatique. Les machines sur lesquelles E-prime était installé permettaient de faire passer l'expérience moyennant la condition un, deux ou trois. Nous avons mis en place un système de gommettes de couleurs différentes pour permettre au sujet d'identifier les touches, les touches CTRL du clavier.

Lors de la passation, nous demandions à chaque personne de remplir le cahier de passation pour lui attribuer un identifiant utilisé pour récupérer ses résultats. Ensuite, nous demandions à chacun sa latéralité manuelle (droitier ou gaucher) afin de sélectionner les paramètres correspondants pour les touches de réponse : pour un individu droitier la touche « c'est un mot » se situait de son côté droit et la touche « ce n'est pas un mot » se trouvait à sa gauche. Les paramètres étant inversés pour les gauchers. Ensuite, le déroulement de l'expérience puis les instructions qui étaient affichées en parallèle à l'écran ont été donnés oralement. Cette expérience est en collaboration avec Hélène Maire, une poste doctorante de Monsieur Zagar. Cette dernière a participé à la passation la première semaine.

4 Problèmes rencontrés lors de la passation

Nous avons été confrontées à des perturbations environnementales indépendantes de notre volonté (alarme incendie, travaux extérieurs, . . .), ce qui a pu perturber la concentration des étudiants. De plus, certains semblaient oublier les consignes en cours d'expérience. Par exemple, ils appuyaient uniquement sur la touche « c'est un mot » lorsque l'item présenté correspondait pour eux

à un mot mais ne faisaient rien quand il s'agissait d'un pseudo mot, alors qu'ils auraient dû presser la touche « ce n'est pas un mot ». De plus, les étudiants qui ont accepté de participer à notre expérience étaient en période de révision la dernière semaine de l'expérience, ce qui augmentait sensiblement leur niveau de stress, et certains d'entre eux se désistaient à la dernière minute, entraînant une perte assez considérable de sujets potentiels. Chaque problème rencontré faisait l'objet d'une notification dans le cahier de passation, ce qui nous permettrait en cas de résultats suspects de potentiellement en expliquer la provenance.

5 Interprétation des résultats

Lors de l'extraction des données, les personnes ayant programmé l'expérience sur les machines du campus lettres se sont aperçus d'une erreur sur les touches de réponse. Dans le protocole expérimental, les deux touches réponse étaient les touches CTRL droite et CTRL gauche et du clavier. Or, si pour nous ces deux touches sont différentes, pour un ordinateur elles correspondent à une seule et même touche. Par conséquent, nous n'avons pas pu tenir compte de la partie pseudo mots qui est donc exclue de l'analyse. Cette conséquence est fâcheuse puisque ce sont les pseudo mots qui nous sont indispensables pour comprendre comment un individu traite une séquence de lettres afin de déterminer s'il s'agit ou non d'un mot. De plus, ces items nous auraient permis de mettre en place un seuil d'erreur dans le filtrage des résultats : si un sujet a commis énormément d'erreurs, cela pouvait signifier qu'il n'était pas concentré lors de la passation, et fausser les résultats..

Par ailleurs, ces données ne comptabilisaient pas le nombre total de sujets ayant passés l'expérience. Ainsi, nous avons 54 individus répartis de

manière équitable dans les trois groupes mais le dénombrement informatique indique un nombre inférieur pour le groupe 3. De ce fait, nos analyses de variance peuvent être faussées et peuvent donc engendrer une erreur dans l'interprétation des données. Pour avoir des résultats fiables, il aurait été souhaitable de comptabiliser au minimum 30 individus par groupe.

6 Conclusion sur l'ensemble du projet tutoré

Ce projet nous a appris à adopter une démarche scientifique rigoureuse. Nous n'avons jamais réalisé un projet orienté recherche auparavant et nous nous sommes aperçues que mettre en place une expérience n'est pas si aisé et répondait à des critères très spécifiques en fonction du domaine concerné. Par ailleurs, nous avons eu l'opportunité d'associer des compétences ciblées Sciences Cognitives à ce travail telle que la programmation (E-Prime), l'analyse de données et les algorithmes d'apprentissage (carte de Kohlen) Nous aurions souhaité pour refaire cette expérience

Troisième partie

Annexes

A Exemple traitement du lexique

Le tableau ci dessous illustre un exemple de mot qui nous a été fourni par Lexique 3. Dans ce tableau, chaque mot possède une couleur qui correspond à la règle (répertoriée après le tableau) appliquée permettant. Les mots de couleur noire sont ceux qui ont été conservés pour être utilisés dans

l'expérience.

dessaisir	varier	abaissaient	périples
fortifiés	consonne	ayans	virage
oxyde	emprunt	à tue-tête	malchance
perplexe	grader	tanière	fieffés

TAB. 6 – Exemple de filtrage du lexique

1. Le nombre de syllabe fixé à 2
2. Le nombre de lettre compris entre 5 et 7
3. Les mots doivent être des lemmes
4. Les mots ne doivent pas contenir de x, de y et de double consonne
5. Le mot devra commencer par une consonne
6. Les mots de doivent pas être composés
7. Les mots ayant une fréquence films compris entre 2 et 18

B Carte de Kohonen

Les cartes de Kohonen appartiennent à la catégorie des algorithmes d'apprentissage. Une entrée est nécessaire au fonctionnement de ces derniers. Elles font partie des apprentissages non supervisés. Dans ce cas, l'algorithme peut apprendre avec uniquement l'entrée, sans connaître la sortie. Elles apprennent suivant un modèle compétitif: les neurones à l'intérieur de l'algorithme vont se « battre » pour être activés par l'entrée (neurone gagnant). Concernant notre expérience, nous donnons en entrée à l'algorithme un mot appartenant à la catégorie des mots avec peu de voisins. Dans cet algorithme, les neurones vont correspondre à tous les mots de l'autre catégorie (beaucoup de voisins). Lorsqu'un mot est donné en entrée, les données qui

lui correspondent (fréquence des bigrammes, fréquence d'apparition dans les films, ...) sont comparées avec les données des neurones (mots ayant beaucoup de compétiteurs). Les cinq neurones qui les plus pertinents par rapport au mot donné en entrée sont alors sélectionnés et affichés dans la console.

Références

- [1] Zagar D. Hangul: A fascinating writing system. a comment on kwon, nam and lee(2015). *Perceptual and Motor Skills: Learning and Memory*, 2015.
- [2] Colé P. Magnan A. et Grainger J. Syllable-sized units in visual word recognition: evidence from skilled and beginning readers of french. *Applied Psycholinguistics*, 1999.
- [3] Ecalle J. et Magnan A. L'apprentissage de la lecture : fonctionnement et développement cognitifs. *Paris : Armand Colin*, 2002.
- [4] Prinzmetal W. Treiman R. et Rho S. How to see a reading unit? *Journal of memory and Language*, 1986.
- [5] Mathey S. Zagar D. Doignon N. et Seigneure A. The nature of syllabic neighbourhood in french. *Acta Psychologica*, 2006.
- [6] Carreiras M. Carlos J. Alvares et Vega. Syllable frequency and visual word recognition in spanish. *Journal of memory and language*, 1993.
- [7] Doignon N. et Zagar D. Illusory conjunctions in french: The nature of sublexical units in visual word recognition. *Language and cognitive processes*, 2005.
- [8] Doignon N. et Zagar D. Les enfants en cours d'apprentissage perçoivent-ils la syllabe à l'écrit? *Revue canadienne de psychologie expérimentale*, 2006.
- [9] Rossi Jean-Pierre. Effet syllabique et structure de la syllabe dans l'identification des mots. *L'année psychologique*, 1978.
- [10] Seidenberg M.S. Sublexical structures in visual word recognition : Access units of orthographic redundancy. *Attention and Performance XII: the psychologic of reading*, 1987.