



HAL
open science

Constitution et traitement d'un corpus de communication médiée par ordinateur pour observer l'alternance des langues

Anitaa Paramanantham

► **To cite this version:**

Anitaa Paramanantham. Constitution et traitement d'un corpus de communication médiée par ordinateur pour observer l'alternance des langues. Sciences de l'Homme et Société. 2020. dumas-02972888

HAL Id: dumas-02972888

<https://dumas.ccsd.cnrs.fr/dumas-02972888>

Submitted on 20 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Constitution et traitement d'un corpus de communication médiée par ordinateur pour observer l'alternance des langues

PARAMANANTHAM Anita

Sous la direction de Marinette Matthey et Olivier Kraif

Laboratoire : LIDILEM

UFR LLASIC
Département Sciences du langage

Mémoire de master 2 mention Sciences du langage - 20 crédits

Parcours : Industries de la langue

Année universitaire 2019-2020

Constitution et traitement d'un corpus de communication médiée par ordinateur pour observer l'alternance

PARAMANANTHAM Anita

Sous la direction de Marinette Matthey et Olivier Kraif

Laboratoire : LIDILEM

UFR LLASIC

Département Sciences du langage

Mémoire de master 2 mention Sciences du langage - 20 crédits

Parcours : Industries de la langue

Année universitaire 2019-2020

REMERCIEMENTS

Je souhaiterais tout d'abord remercier toutes les personnes qui m'ont aidée lors de la rédaction de ce mémoire.

Je tiens à remercier Madame Marinette Matthey qui a volontiers accepté de m'encadrer et de m'aider dans ma recherche. Elle était aussi enthousiaste que moi en discutant du sujet et a proposé de nombreuses idées et suggestions. Elle a su me guider lorsque j'avais des questions et des doutes et a partagé avec moi ses connaissances dans le domaine de la sociolinguistique.

Je remercie également Monsieur Olivier Kraif pour avoir accepté de m'encadrer et m'avoir guidé pour la partie informatique. Il a su faire preuve de pédagogie et de patience me permettant de développer une démarche cohérente.

Je souhaite également remercier mon amie Laurence Pottier qui m'a beaucoup encouragée pendant les moments difficiles et pour avoir relu et corrigé mon mémoire. Ses conseils de rédaction ont été très précieux.

Enfin, je tiens à remercier ma famille pour ses encouragements.

DÉCLARATION ANTI-PLAGIAT

1. Ce travail est le fruit d'un travail personnel et constitue un document original.
2. Je sais que prétendre être l'auteur d'un travail écrit par une autre personne est une pratique sévèrement sanctionnée par la loi.
3. Personne d'autre que moi n'a le droit de faire valoir ce travail, en totalité ou en partie, comme le sien.
4. Les propos repris mot à mot à d'autres auteurs figurent entre guillemets (citations).
5. Les écrits sur lesquels je m'appuie dans ce mémoire sont systématiquement référencés selon un système de renvoi bibliographique clair et précis.

PRENOM : Anitaa

NOM : PARAMANANTHAM

DATE : 6 septembre 2020

SOMMAIRE

REMERCIEMENTS.....	4
SOMMAIRE.....	6
INTRODUCTION.....	8
CHAPITRE 1 - ÉTAT DE L'ART.....	11
CODE-SWITCHING EN TAL ET EN SOCIOLINGUISTIQUE.....	12
1. POINTS DE VUE DE DIVERS CHERCHEURS.....	13
2. CODE-SWITCHING EN SOCIOLINGUISTIQUE.....	14
2.1.Linguistique Socioculturelle.....	15
2.2.Trois processus pour rendre compte du code-switching.....	17
2.3.Code-switching dans un environnement mixte.....	19
3. CODE-SWITCHING EN TAL.....	20
3.1.Identification de langue.....	20
3.2.Prise en compte du code-switching dans les systèmes de reconnaissance vocale.....	21
3.3.Conclusion.....	22
CHAPITRE 2 - MÉTHODOLOGIE.....	24
LES RESSOURCES.....	25
1. LES DONNÉES.....	25
1.1.Choix des données.....	25
1.2.Méthodes de collecte des données.....	26
STRUCTURE DES DONNÉES.....	28
1. FICHER .NUMBERS.....	28
1.1.Feuille 1 : v ori.....	28
1.2.Feuille 2 : V1.....	29
1.3.Feuille 3 : V2.....	30
1.4.Feuille 4 : V3.....	30
1.5.Feuille 5 : V4.....	30
1.6.Feuille 6 : Locuteurs.....	31
2. SCRIPT PYTHON.....	31
CHAPITRE 3 - BIOGRAPHIE LINGUISTIQUE DES LOCUTEURS.....	33
1. ORIGINE ETHNIQUE.....	34
1.1.Locuteur 1.....	36
1.2.Locuteur 2.....	37
1.3.Locuteur 3.....	38
1.4.Locuteur 4.....	40
1.5.Locuteur 5.....	41
2. PRATIQUES LANGAGIÈRES PARMY DIFFÉRENTS GROUPES DE PERSONNES.....	42
2.1.À la maison en famille.....	44

2.2. Entre amis et connaissances.....	44
3. PRATIQUES LANGAGIÈRES À L'ÉCOLE.....	45
3.1. École maternelle	45
3.2. École primaire.....	46
3.3. Lycée.....	46
3.4. Université.....	47
CHAPITRE 4 - ANALYSES ET RÉSULTATS STATISTIQUES	49
1. LE TABLEAU DE DONNÉES.....	50
1.1. Colonne C : langues	50
1.2. Colonne E : formes normalisées	53
1.3. Colonne F : traduction	53
2. LES STATISTIQUES	54
2.1. Nombre de messages par langue.....	54
2.2. Nombre de messages par locuteur	55
2.3. Les messages mixtes	59
3. LES INTERPRÉTATIONS	61
3.1. L'environnement influence-t-il l'utilisation d'une langue?	62
3.2. Code-switching ou code-mixing?	62
Conclusion	63
Bibliographie.....	65
Sitographie	66
Table des illustrations.....	67
Table des annexes.....	68
Table des matières.....	121

INTRODUCTION

Mon pays d'origine est la Malaisie. C'est un pays composé de diverses cultures, ethnies, et surtout de nombreuses langues. Malgré nos différences, nous arrivons à nous exprimer et à communiquer avec aisance. Pour donner un aperçu, en Malaisie, lorsqu'un groupe de locuteurs aux antécédents linguistiques différents conversent, ils intègrent différentes langues dans leur conversation. Ce qui est intéressant, c'est que même quand le groupe possède les mêmes antécédents linguistiques, les locuteurs intègrent souvent différentes langues dans leur conversation, cela se manifeste notamment dans le "*code-mixing*" et également dans le "*code-switching*".

Venant d'un pays multilingue et possédant une culture nourrie de nombreuses influences, j'ai toujours eu un intérêt pour les langues, l'évolution des langues et la façon dont les gens conversent en utilisant différentes langues, en particulier dans mon pays où chacun parle au moins deux langues. C'est la raison pour laquelle au cours de ma deuxième année de master j'ai choisi de rechercher et d'étudier cette forme de communication. J'ai proposé ce sujet à mes professeurs qui ont volontiers accepté de m'aider et de me superviser.

À ma connaissance, en particulier dans mon pays, la plupart des gens ne savent pas ce qu'est l'alternance codique ou le mélange des codes même s'ils l'utilisent très souvent en parlant ou en écrivant.

Dans le cadre de la rédaction de mon mémoire, je vais étudier le phénomène de *code-switching* en utilisant un corpus de textes de locuteurs Malaisiens que j'ai constitué. Mon but est de réunir un corpus de pratiques langagières "mêlées" en exploitant les réseaux sociaux et les messageries d'un réseau de locuteurs plurilingues malais, tamoul, chinois et anglais, pour évaluer les fréquences d'utilisation de ces différentes langues. À partir de cette constitution et du traitement de ce corpus de communication, ce travail permettra de mieux décrire les phénomènes de l'alternance des codes et le mélange des codes entre locuteurs plurilingues.

Ce mémoire est rédigé parallèlement à un stage de recherche effectué dans le cadre du Master 2 Sciences du Langage, parcours Industries de la Langue proposé par l'Université de Grenoble Alpes. Mon mémoire de recherche a été encadré par Mme Marinette Matthey, professeure en sciences du langage à l'Université Grenoble Alpes ainsi

que par M. Olivier Kraif, professeur et le responsable du parcours IDL à l'Université Grenoble Alpes. Mme Matthey supervise l'aspect sociolinguistique de cette recherche et M. Kraif supervise l'aspect informatique.

Ce mémoire se décomposera en quatre chapitres consistant en l'état de l'art, la méthodologie du travail, la présentation de la biographie des locuteurs et enfin les analyses et les résultats de recherche récoltés.

CHAPITRE 1

-

ÉTAT DE L'ART

Code-switching en TAL et en Sociolinguistique

Dans l'usage courant, le *code-switching* est utilisé pour faire référence à des mélanges de deux langues. En sociolinguistique, la notion de code est considérée par certains chercheurs comme périmée, en raison de son ancrage structuraliste du siècle passé, dès lors la notion même de *code-switching* est remise en question car les langues ne sont plus considérées avant tout comme des codes mais plutôt comme des ensembles disparates de ressources multimodales. Beaucoup préfèrent ainsi parler de « crossing languages » (Ben Rampton) ou de *translanguaging* (par exemple Ofelia Garcia) plutôt que de *code-switching*. De plus, le terme ne concerne pas que les « langues », mais également les dialectes, et les variétés de langue, etc. Nous conserverons néanmoins le terme de *code-switching*, tout en étant consciente des discussions sur la notion de code.

Le *code-switching* a été l'objet de nombreuses études depuis le milieu du siècle passé. Avec l'augmentation des flux migratoires et la diffusion des grandes langues sur toutes la planète, et le développement des outils numériques de communication, le phénomène devient massif. Il est étudié en sociolinguistique, en anthropologie linguistique et dans le traitement du langage naturel.

Le *code-switching* est une caractéristique des communautés bilingues dans le monde entier et n'a commencé à attirer l'attention des chercheurs qu'au cours des dernières décennies. Afin de lutter contre la vision négative du *code-switching* et du bilinguisme en général, les chercheurs ont cherché à faire reconnaître qu'il s'agit d'une manifestation majeure du contact linguistique pour la plupart des communautés bilingues et multilingues. L'identification des diverses contraintes du *code-switching* a inspiré de nombreux travaux en linguistique ainsi qu'en traitement naturel du langage, et particulièrement en reconnaissance vocale.

Les frontières entre ce que l'on nomme *code-switching*, *code-mixing*, transfert linguistique et de nombreux autres phénomènes linguistiques sont floues. La distinction est parfois difficile à faire, et la plupart des non-experts confondent ces termes.

Le bilinguisme est étudié depuis longtemps, mais sous un angle négatif. Cependant,

depuis les années 1980, la plupart des chercheurs sont venus à le considérer comme une production courante de locuteurs bilingues et multilingues (Wikipedia contributors, *Code-switching* consulté le 15 janvier 2020 à l'adresse <https://en.wikipedia.org/wiki/Code-switching>, nous traduisons).

La communication par téléphone portable est devenue un instrument central d'interaction et de socialisation à travers le monde. Cet outil numérique permet de pratiquer un mélange des langues où les codes linguistiques fusionnent. Les locuteurs bilingues ou plurilingues utilisent régulièrement deux ou plusieurs langues dans des contextes culturels multilingues divers (Morel, E. & Pekarek Doehler, S., 2013).

Les sections suivantes présentent le *code-switching* dans les domaines de recherche de la sociolinguistique et du TAL.

1. Points de vue de divers chercheurs

On constate à travers de nombreuses études que les linguistes ne définissent pas le *code-switching* de la même manière. Nous allons, dans cette partie, développer et comparer les différents points de vue des linguistes face à ce phénomène. Certains linguistes analysent le *code-switching* comme une forme de communication erronée due à des lacunes éducatives, d'autres études définissent le *code-switching* comme un phénomène psychologique. Les nombreuses études menées sur le *code-switching* montrent que ces interactions écrites ou orales se produisent de manière très courante et naturelle, c'est-à-dire que les locuteurs qui utilisent le *code-switching* écrivent ou parlent de manière spontanée en recourant au *code-switching*. D'après Gumperz (1977), linguiste socioculturel pionnier dans l'étude du *code-switching*, la relation entre les locuteurs affecte le choix de la variété linguistique.

Romaine (cité dans Nilep, 2016) dans ses études sur l'acquisition des langues maternelles et secondes, utilise le terme *code-switching* pour décrire les capacités linguistiques cognitives des locuteurs bilingues et apprenants, ou pour décrire les pratiques en classe chez les apprenants impliquant l'utilisation de plusieurs langues.

Alvarez-Cáccamo (cité dans Nilep, 2016) constate que pratiquement tous les

travaux sur le *code-switching* reposent sur une identification stricte entre les notions de « code » et de « variété linguistique », qu'il s'agisse d'une langue, d'un dialecte, d'un style ou d'un registre prosodique. « Cependant, cette orientation structurelle ne parvient pas à expliquer de manière convaincante certains phénomènes conversationnels relatifs à la pertinence ou à la signification (ou au manque de pertinence) des alternances entre les variétés contrastées (nous traduisons) ».

Weinreich (*Languages in Contact*, 1953) est un des précurseurs dans la recherche sur le *code-switching* dans le domaine de la linguistique et de la sociolinguistique. Il souhaitait décrire l'effet du contact linguistique sur les langues, en plus de décrire les activités des communautés d'expression bilingue et multilingues. Il soutenait que « l'anthropologie devrait se tourner vers la linguistique afin de décrire correctement la pratique de la parole bilingue et le processus d'acquisition de la langue qui a lieu dans les communautés bilingues (nous traduisons) ».

2. *Code-switching en sociolinguistique*

L'approche sociolinguistique de l'interaction aborde le *code-switching* sous l'angle des significations attachées à l'usage d'une langue plutôt qu'à une autre. Le *code-switching*, autrement dit l'alternance codique, a une fonction localement construite et négociée par les participants dans le cours même de leur interaction communicative (Auer, 1984, 1999, Mondada et Doehler, P., 2003). Mondada (2007) considère que la fonction d'une alternance codique n'est pas pré-déterminée par des 'valeurs' pré-existantes des langues, mais émerge de l'interaction en train de se faire, de l'orientation mutuelle des interlocuteurs, et de son positionnement précis dans le déroulement séquentiel de l'échange communicatif.

Selon Weinreich (1953), les « switcheurs » doivent avoir au moins deux variétés linguistiques afin d'utiliser ces codes dans deux situations différentes. Il a suggéré que l'alternance fréquente était le produit de lacunes éducatives. Vogt dans Nilep (2006) affirme que le *code-switching* en soi n'est peut-être pas un phénomène linguistique, mais plutôt un phénomène psychologique, et ses causes sont évidemment extra-linguistiques.

Mais le bilinguisme présente un grand intérêt pour le linguiste car il permet d'étudier les modalités de contact entre les langues.

Différentes études suggèrent qu'il existe un ensemble de codes entre lesquels les locuteurs alternent de manière significative, la question est de se demander en quoi ces codes consistent.

2.1.Linguistique Socioculturelle

La linguistique socioculturelle développe une approche des fonctions sociales et culturelles des langues. L'approche socioculturelle comprend de nombreux aspects tels que la sociolinguistique, la linguistique conversationnelle, l'anthropologie, etc. Le "markedness model" de Myers-Scotton (1983, 1993, 1998) s'inscrit dans cette approche. Nilep (2006) discute l'approche psychosociale de Myers-Scotton, qui cherche à analyser les choix de code en lien avec l'identité et les effets du *code-switching* sur la conversation.

Le "markedness model" de Myers-Scotton (1993) considère que chaque langue dans une communauté multilingue est associée à des rôles sociaux particuliers. « En parlant une langue particulière, un participant signale sa compréhension de la situation actuelle, et en particulier son rôle pertinent dans le contexte. En utilisant plusieurs langues, les locuteurs peuvent initier des négociations sur les rôles sociaux pertinents. Myers-Scotton suppose que les locuteurs doivent partager, au moins dans une certaine mesure, une compréhension des significations sociales de chaque code disponible. S'il n'existait pas de telles normes, les interlocuteurs n'auraient aucune base pour comprendre la signification de choix de code particuliers (nous traduisons) ». Le "markedness model" est un des modèles les plus influents et les plus développés en la matière. Myers-Scotton (née en 1934) a continué d'affiner le modèle de manière cohérente tout au long de sa carrière.

De nombreux travaux en anthropologie linguistique, sociolinguistique et dans d'autres domaines de la linguistique socioculturelle fournissent des interactions documentant le *code-switching* et l'interprétant dans des contextes particuliers. Nilep (2006) étudie les travaux de Monica Heller au Québec et en Ontario qui l'ont amenée à

considérer l'économie du bilinguisme et le changement de code comme une stratégie politique. Étant donné que les langues tendent à s'associer à des situations idéalisées et à des groupes de locuteurs, l'utilisation de plusieurs langues « permet aux gens de dire et de faire, en fait d'être deux ou plusieurs choses où normalement un choix est attendu » (Heller 1988b: 93). Selon cette auteure, « le recours à plusieurs langues permet à un locuteur de revendiquer ses langues là où il devrait se décider entre l'une d'entre elles. Cette ambiguïté stratégique permet aux anglophones du Québec, par exemple, de se positionner dans une culture d'entreprise sous contrôle francophone, tout en revendiquant une identité anglophone, avec sa valeur associée sur le marché international (nous traduisons) ».

Des études sur le changement d'identité et de code montrent qu'une observation attentive du discours peut permettre une compréhension à la fois empirique et théorique des fonctions de la variation du langage dans l'interaction sociale. En liant les observations à des locuteurs et à des acteurs sociaux particuliers, plutôt que de passer trop facilement aux discussions sur les normes culturelles ou linguistiques, les chercheurs peuvent parvenir à une compréhension détaillée et fiable de la place de la langue dans la construction et la transmission des traditions sociales (Nilep, 2006, nous traduisons).

D'après Nilep, les études sur la linguistique interactionnelle sont fortement motivées par l'analyse de la conversation, ainsi que la linguistique fonctionnelle et l'anthropologie linguistique qui cherche à comprendre « la manière dont le langage figure dans l'interaction et la cognition quotidiennes » (Ochs, Schegloff & Thompson 1996: 2, nous traduisons). Cependant l'auteur précise également qu'il ne faut pas supposer que tous les éléments pertinents pour le discours et l'interaction sociale sont visibles pour l'analyste, en particulier lorsque l'analyste n'est pas ancré dans les structures sociales particulières qu'il étudie.

La représentation mentale de ces codes ne peut être directement observée, ni par les analystes, ni par les parties en interaction. L'analyste doit plutôt observer le discours lui-même et récupérer la saillance d'une forme linguistique comme indice de son effet sur l'interaction du discours. L'approche décrite ici comprend le *code-switching* comme la

pratique des individus dans des contextes de discours particuliers. Par conséquent, les constats effectués ne peuvent être étendus à une théorie générale des fonctions de l'alternance linguistique, ni définir la nature exacte de ce qu'est un code (Nilep, C., 2006, nous traduisons).

2.2. Trois processus pour rendre compte du code-switching

Les trois processus discutés par Muysken (1997) sont l'alternance, l'insertion et la lexicalisation congruente. Ce linguiste a été le premier à proposer un modèle général pour rendre compte de la manière dont les éléments de deux variétés linguistiques différentes peuvent se combiner.

L'alternance signifie que le locuteur passe d'une langue à l'autre, à la fois en grammaire et en lexique. L'alternance a lieu entre les énoncés d'un tour ou entre les tours.

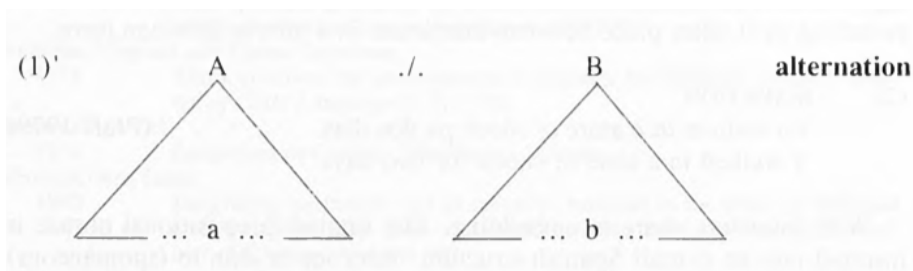


Figure 1. L'alternance dans Muysken, P.C. (1997). Code-switching processes : alternation, insertion, congruent lexicalization.

Dans la figure 1, *a* constituant de la langue *A* est suivi de *a* constituant de la langue *B*. La langue du constituant dominant *A* et *B* n'est pas spécifiée. *A*, *B* sont des étiquettes de langage pour les nœuds non terminaux et *a*, *b* sont des étiquettes pour les nœuds terminaux.

En insertion, il y a incorporation. L'insertion s'apparente à l'emprunt lexical, qui implique également une unité lexicale.

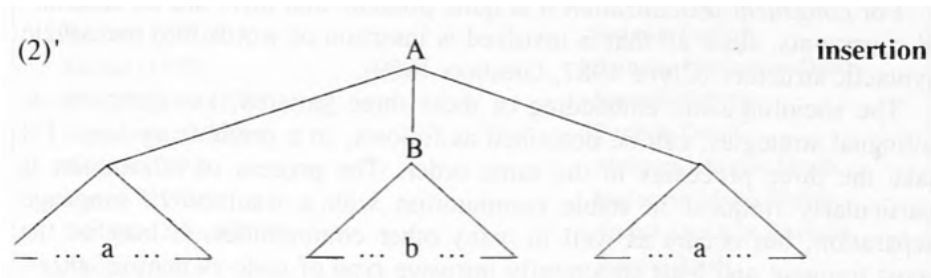


Figure 2. l'insertion dans Muysken, P.C. (1997). Code-switching PROCESSES: alternation, insertion, congruent lexicalization.

Sur la figure 2, un seul constituant *B* est inséré dans une structure définie par la langue *A*, avec les mots *a* de cette langue. *A*, *B* sont des étiquettes de langage pour les nœuds non terminaux et *a*, *b* sont des étiquettes pour les nœuds terminaux.

La lexicalisation se réfère à une situation où les deux langues partagent une structure grammaticale. Le mélange de deux langues pourrait être interprété comme une combinaison d'alternances et d'insertions, mais il peut y avoir plus, les éléments des deux langues sont insérés, comme constituants ou comme mots dans une structure partagée. Dans cette perspective, la lexicalisation congruente est similaire au style et à la variation linguistique monolingue.

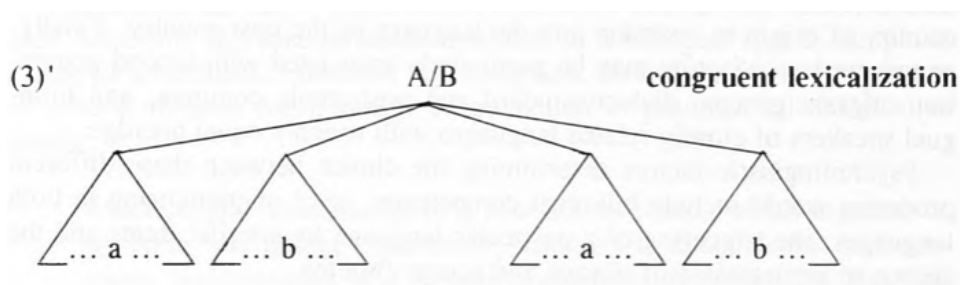


Figure 3. Lexicalisation congruente dans Muysken P.C. (1997). Code-switching PROCESSES: alternation, insertion, congruent lexicalization.

Dans la figure 3, la structure grammaticale est partagée par les langues *A* et *B*, et les mots des deux langues *a* et *b* sont insérés de manière aléatoire. *A*, *B* sont des étiquettes de langage pour les nœuds non terminaux et *a*, *b* sont des étiquettes pour les nœuds terminaux.

2.3. Code-switching dans un environnement mixte

Dans de nombreux pays ou environnements de culture mixte, nous l'avons déjà mentionné, le *code-switching* est une pratique très courante. Les locuteurs de plusieurs variétés linguistiques communiquent avec d'autres locuteurs de même variétés linguistiques ou parfois de différentes variétés linguistiques.

L'article de Stapa & Sahabudin Khan (2016) se concentre sur les types et les fonctions de *code-switching* dans une famille mixte malais-chinois à la maison. Le changement de code est la norme en Malaisie. Les Malaisiens changent de code entre le malais et l'anglais dans leurs conversations quotidiennes et cela se produit dans les interactions formelles et informelles. Cet article se concentre sur les types et les fonctions de *code-switching* dans une famille mixte malais-chinois dans le cercle familial. Le but de l'étude est d'examiner le comportement de l'alternance codique qui se produit dans leurs conversations. Cette étude montre que les locuteurs changent de code dans leurs conversations, de manière consciente ou inconsciente. Il y a plusieurs raisons pour lesquelles la famille mixte malaise-chinoise de cette étude a recours au changement de code, mais aucune n'est déterminante. L'une des raisons est que le sujet abordé dans les conversations peut être un déclencheur. La famille de cette étude passe à une certaine langue en fonction des sujets d'actualité qui sont discutés. Ils utilisent le malais ou l'anglais selon leur lexique pour ce sujet. Les autres raisons sont les citations du discours de quelqu'un d'autre, l'insistance sur quelque chose (répétition du propos dans une autre langue), les interjections et les besoins lexicaux (emprunts). Les résultats et les données de cette étude fournissent de nouvelles idées sur les fonctions de *code-switching* dans une famille mixte, ainsi que la prise de conscience qu'il existe différents types de modèles de *code-switching* qui se produisent dans certaines situations. Cette étude a sensibilisé au phénomène linguistique de l'alternance codique dans les familles interethniques, en particulier dans les familles malaises-chinoises.

3. *Code-switching en TAL*

Afin d'étudier en détail le *code-switching* dans le corpus constitué, je vais mentionner des outils informatique pour calculer la fréquence des mots dans telles langues, et pour détecter la langue utilisée à quel endroit à quel moment, etc.

3.1. *Identification de langue*

L'identification de la langue est importante pour une grande variété d'applications telles que les systèmes d'extraction d'informations, les interfaces de l'assistant vocal, la traduction automatique, ainsi que pour les outils d'aide à l'évaluation de la langue chez les enfants bilingues (Gupta et al., 2014; Chandu et al., 2017; Roy et al., 2013, nous traduisons). Selon Solorio, Maharjan, & Mave (2018) l'identification de la langue est la première étape de la modélisation informatique du *code-switching*.

Dans l'article *Language Identification and Analysis of Code-Switched Social Media Text* (Solorio, Maharjan, & Mave, 2018), les auteurs détaillent leurs travaux sur la comparaison de différents systèmes d'identification de la langue au niveau des mots pour les données Hindi-Anglais à *code-switching* et un ensemble de données standard Espagnol-Anglais. À cette fin, ils construisent un nouvel ensemble de données de *code-switching* pour l'Hindi-Anglais. Pour comprendre les modèles de *code-switching* dans ces paires de langues, ils étudient différentes mesures de l'alternance codique. Cependant, l'identification de la langue dans un texte contenant du *code-switching* n'est pas évidente. Premièrement, la nature de l'élément commuté peut être un segment de phrase, un mot et voire un morphème. Deuxièmement, les types de langues impliquées dans la commutation et leur relation augmente encore la complexité de la tâche (Solorio, Maharjan, & Mave, 2018).

Pour cette étude, les modèles d'identification de la langue impliqués sont: les CRF (un modèle d'étiquetage de séquence pour capturer la structure dans les données), les

LSTM¹ bidirectionnels et Word-Character LSTM². La complexité de l'identification de la langue des données à *code-switching* dépend de la source de données, du comportement de *code-switching*, de la typologie et de la relation entre les langues impliquées. Les chercheurs dans cet article utilisent avec succès deux architectures d'apprentissage profond différentes, sans impliquer une ingénierie sophistiquée pour la tâche, et obtiennent des résultats compétitifs.

Au fil du temps, des chercheurs ont exploré l'identification des langues pour différentes paires de langues et variétés de dialectes. La série de campagnes (shared tasks) FIRE (Roy et al. 2013; Choudhury et al., 2014; Sequiera et al., 2015b) se concentre sur l'identification de la langue pour des requêtes de recherche mixtes en anglais et en langues indiennes (p.e. Hindi) pour la recherche d'informations. Les première et deuxième tâches partagées sur l'identification de la langue dans les données de *code-switching* (Solorio et al. 2014; Molina et al. 2016) montrent la nécessité d'un traitement automatique du texte comportant des *code-switching* et rendent compte de la comparaison des différents systèmes d'identification de la langue.

3.2.Prise en compte du code-switching dans les systèmes de reconnaissance vocale

Il existe déjà des études sur la prise en compte du *code-switching* dans les systèmes de reconnaissance vocale. Par exemple, dans l'article de Xu, H., Pham, V. T., Kyaw, Z. T., Li, H., Chng, E. S., & Lim, Z. H. (2018), les auteurs visent à présenter le *code-switching* de Mandarin-Anglais dans un système de reconnaissance vocale. Les chercheurs constatent qu'il est important d'étudier et implanter ce phénomène dans le système, puisqu'il est fréquent et se produit de plus en plus dans les sociétés multilingues. Ils ont créé une application de bureau où les utilisateurs peuvent enregistrer leur discours contenant du *code-switching* et obtenir automatiquement le texte retranscrit. Ce faisant, ils ont

¹ Long Short Term Memory networks, (Hochreiter et Schmidhuber, 1997) sont une forme de réseaux de neurones récurrents, qui permettent de traiter le problème du gradient de fuite (Hochreiter, 1998) en étendant les RNN avec des cellules de mémorisation.

² Ce modèle est une réplique du modèle proposé par Samih et al. (2016). La couche d'entrée de ce modèle a un mot et l'intégration de caractères. Ces derniers sont utilisés pour capturer les caractéristiques morphologiques d'un mot.

développé plus de 1000 heures de données d'entraînement pour la modélisation acoustique. Ils ont construit les modèles acoustiques en utilisant un Time-delay Neural Network (TDNN). Avec ces modèles, ils ont pu créer un système qui génère des transcriptions en temps réel avec des performances satisfaisantes pour les applications visées.

Dans l'article *Automatic Detection of Code-switching Style from Acoustics* (Black, Rallabandi, & Sitaram, 2018) il est expliqué que pour intégrer ce type de mélange dans les systèmes de reconnaissance vocale, nous devons entraîner les systèmes et détecter d'abord le style de "switching" d'un énoncé à partir de l'acoustique, puis utiliser les modèles de langage spécialisés pour décoder la parole. Pour ce faire, nous avons besoin des deux métriques qui indiquent le niveau de mélange dans l'énoncé qui sont le "CodeMixing Index (CMI)" et le "CodeMixing Span Index". « Le CodeMixing Index est utilisé pour quantifier le niveau de mélange entre les langues participantes. Bien que le CodeMixing Index capture le niveau de mélange, il ne prend pas en compte la fréquence du mélange. Par conséquent, ils utilisent le CodeMixing Span Index pour classer les énoncés en différents styles. Dans l'article, les chercheurs montrent leur tentative de catégoriser le style de code-switching en utilisant des données acoustiques, qui peuvent être utilisées par un système de reconnaissance vocale. Les études sur le modèle de langage indiquent les résultats prometteurs, et ces chercheurs prévoient d'améliorer leurs modèles d'extraction et de classification et de tester leurs modèles de langage sur la reconnaissance vocale (nous traduisons) ».

3.3. Conclusion

De nombreuses études doivent être menées pour parvenir à une compréhension plus complète du code-switching. Dans une communauté mixte, diverses langues sont généralement utilisées dans la communication quotidienne. Il est courant que les personnes de même origine linguistique ou d'origine linguistique différente mélangent leurs échanges. Il existe de nombreuses raisons pour le code-switching et de nombreux différents types de modèles de code-switching qui se produisent de manière attendue et inattendue dans certaines situations. La représentation mentale de ces mélanges ne peut pas être directement observée, par conséquent, on ne peut pas spécifier les fonctions générales de

l'alternance linguistique, ni définir la nature exacte d'une langue avant l'interaction. De ce fait, il est difficile de développer un système de reconnaissance vocale de *code-switching*. Les systèmes de reconnaissance de la parole sont capables de ne reconnaître qu'une seule langue, ce sont donc des systèmes de type monolingue. Il manque actuellement des corpus de *code-switching* pour améliorer ces systèmes. Néanmoins, les tests sur les modèles de langage indiquent des résultats prometteurs pour améliorer les modèles d'extraction et de classification de *code-switching* sur la reconnaissance vocale.

CHAPITRE 2

-

MÉTHODOLOGIE

Les ressources

1. Les données

Dans le cadre de mon travail, je vais étudier un corpus de textes de locuteurs Malaisiens que j'ai constitué à partir de l'application mobile WhatsApp, j'ai pu réunir et exploiter les messages de locuteurs Malaisiens. Le corpus est enregistré sous la forme de fichiers .numbers qui contiennent six feuilles :

v ori	ce fichier contient toutes les métadonnées, les notifications et tous les tokens de messageries y compris les ponctuations et les émoticônes.
V1	ce fichier contient les tokens et les notifications
V2	ce fichier ne contient que les tokens
V3	ce fichier est ordonné par fréquence
V4	ce fichier contient le tableau de v3 avec leurs étiquettes
Locuteurs	ce fichier contient les métadonnées de locuteurs

Figure 4. Feuilles de données

1.1. Choix des données

Pour réaliser ce travail, j'ai discuté avec mes directeurs à propos de la plate-forme à utiliser pour étudier une utilisation mixte des langues comme forme de communication. Nous avons la possibilité d'extraire du texte de "Twitter", "Facebook" et d'autres applications de réseaux sociaux, mais nous avons décidé de ne pas le faire car ce type de communication dans ces plate-formes n'était pas celui que nous voulions étudier.

Notre choix s'est porté sur l'application mobile WhatsApp qui est une application mobile qui permet aux gens de communiquer en envoyant des messages, des notes vocales et il est également possible de passer des appels audio et vidéo. Le téléphone doit être connecté à Internet lors de l'utilisation de l'application. Cette application offre également la possibilité d'envoyer des photos, des documents et d'autres fichiers multimédias.

L'interaction entre les locuteurs sur WhatsApp était plus naturelle. Beaucoup de gens utilisent WhatsApp et j'ai vu à quel point les conversations étaient fluides en raison

du cadre informel puisque généralement les gens se connaissent, voire même sont liés par des liens d'amitié. Agréable, simple, et surtout, convivial avec un chiffrement de bout en bout pour chaque message, les gens se sentent plus en sécurité en utilisant WhatsApp qu'en utilisant d'autres plates-formes pour communiquer.

J'ai parcouru mes propres messages de groupes et j'ai constaté que dans tous mes groupes, il y avait plusieurs locuteurs plurilingues. La plupart du temps, les locuteurs mélangent une ou deux langues mais il y a aussi beaucoup de passage d'une langue à une autre. L'étude de ce corpus a été proposée à mes directeurs qui ont convenu que ce serait une bonne source pour étudier le phénomène de l'alternance codique, le nombre de langues utilisées et le rôle du contexte sur ce phénomène.

Dans un premier temps, nous pensions collecter des données orales et écrites, mais pour pouvoir collecter des données orales, il aurait fallu organiser des entretiens. De plus, en raison de contraintes de temps et par manque de ressources, nous avons décidé de travailler uniquement sur des données écrites. J'ai collecté les données de quelques groupes mais dans le but de faciliter l'exploration des données disponibles, nous n'étudierons qu'un seul groupe. Le groupe auquel je me suis intéressée était constitué d'étudiants qui étaient des colocataires faisant leurs études en France, issus de différents contextes linguistiques et originaires de différents États de Malaisie.

1.2.Méthodes de collecte des données

Comme indiqué ci-dessus, après avoir réfléchi aux différentes options, nous avons fait le choix d'utiliser WhatsApp Messenger parce que cela permet de télécharger les messages avec les images ou sans les images au format texte. C'était la première étape de la collecte de mes données. J'ai donc téléchargé les messages de quatre groupes, comme je l'ai déjà mentionné, j'ai finalement utilisé les messages d'un seul groupe. La taille de ce groupe étant conséquente, les données sont suffisantes pour la recherche.

J'ai demandé à ma famille et à mes amis la permission de télécharger et d'utiliser les messages du groupe comme données. Ils étaient tous d'accord verbalement, il n'y avait pas de document officiel signé mais je me suis engagée à utiliser ces données uniquement dans le cadre de mon mémoire, à ne pas les diffuser et à conserver l'anonymat de ces personnes.

Après avoir terminé la première étape qui consistait à télécharger les messages, j'ai ensuite tokenisé le texte. La tokenisation consiste à découper le texte en tokens qui sont des composants graphiques de base (mots, ponctuations, nombres, émoticônes, etc.). J'ai converti le texte tokenisé dans un format .numbers pour travailler sur chaque occurrence. Le groupe de chat se compose de cinq locuteurs avec 4137 messages de chat, 26 083 tokens. La date de début des échanges entre les cinq locuteurs est le 25 Août 2015 et la date de fin est le 30 Novembre 2019.

Structure des données

1. Fichier .numbers

J'ai utilisé .numbers, l'équivalent d'Excel sur Mac pour travailler sur mes données. Au cours de ce travail, pour suivre les modifications apportées aux données, j'ai créé un nouveau fichier à chaque modification de celles-ci. Les feuilles .numbers ont été faites à l'aide de formules trouvées dans Numbers et dans Excel afin d'obtenir les résultats nécessaires à une analyse plus approfondie.

1.1. Feuille 1 : v ori

La première feuille contient les messages en entier et leurs étiquettes. Chaque ligne contient un token. Cette feuille contient toutes les données sans aucune modification, aucune information n'a été ajoutée à la colonne A et rien n'a été supprimé. La colonne A comporte les tokens, les métadonnées concernant les locuteurs qui sont regroupées au début de chaque tour de paroles. Chaque locuteur se voit attribuer un identifiant unique dans la colonne B, qui sert de référence par la suite. Également chaque langue se voit attribuer un identifiant unique dans la colonne C. Cette colonne C contient non seulement les langues utilisées mais aussi les termes linguistiques associés aux noms propres et aux connecteurs des énoncés.

Il y a 6 colonnes dans le v ori contenant des informations qui contribueront à l'analyse du sujet :

Tokens	Locuteurs	Langues	Erreurs Typographique	Formes Normalisées	Traduction
l'ensemble des textes tokenisés	les membres de ce chat en groupe représentés par un identifiant unique	les langues utilisées pour communiquer représentées par un identifiant unique	les mots avec erreurs grammaticales	les mots sont restructurés selon l'orthographe correcte trouvée dans le dictionnaire	la traduction en français selon la conversation

Figure 5. Les colonnes de v ori

Tokens	Locuteurs	Langues	Erreurs typographique	Formes Normalisées	Traduction
[25/08/2015 17:04:55]	▲				
Vous	▲	F		*	*
avez	▲				
créé	▲				
le	▲				
groupe	▲				
“	▲				
Barang2	▲				
dapur	▲	M		*	la cuisine
”	▲				
[25/08/2015 17:05:11]	▲				
Jeremy:	L1				
Hsemates	▲				
ku	▲	M		aku	moi(je)
yg	▲	M		yang	ça/qui(que)
cemerlang	▲	M		cemerlang	brillant(intelligent)
[25/08/2015 17:05:29]	▲				
Jeremy:	L1				
Jom	▲	M		Jom	aller/on y vas
planning	▲	A		*	planner
brg2	▲	M		barang-barang	choses/articles/produits
nk	▲	M		hendak	vouloir
bwk	▲	M		bawa	apporter/amener
[25/08/2015 17:05:42]	▲				
Jeremy:	L1				
Blih	▲	M		boleh	pouvoir
bahgi2	▲	M		bahagi-bahagi	diviser/partager
sikit	▲	M		sikit	un peu
[25/08/2015 17:03:35]	▲				
Nazmi	▲	EN		*	anthroponyme

Figure 6 . Feuille v ori

1.2.Feuille 2 : V1

La feuille V1 contient l'ensemble des textes mais avec certaines modifications apportées au texte. Les noms, les dates et l'heure ont été supprimés. Ceci est dû au fait qu'en supprimant les métadonnées, je dois organiser le reste de texte en fonction du nombre d'occurrences afin de pouvoir analyser l'utilisation des mots dans le chat.

1.3.Feuille 3 : V2

Tout comme la feuille ci-dessus, les métadonnées ont été supprimées dans cette feuille et les notifications ont également été supprimées. Comme expliqué ci-dessus, cela devait être fait afin d'exécuter des fonctions pour obtenir le nombre d'occurrences des mots.

1.4.Feuille 4 : V3

La feuille V3 contient les tokens ordonnés par fréquence. Il n'était pas nécessaire de créer un programme python pour obtenir ce résultat car il y a une fonction dans .numbers lui-même qui fonctionne selon les besoins pour obtenir le nombre d'occurrences de chaque mot. Chaque mot est compté et ordonné dans une liste de fréquence dans un ordre décroissante. Le nombre d'occurrences de ces mots est obtenu en utilisant la fonction de tri dans .numbers. L'inconvénient de cette fonction est que je ne pouvais pas obtenir le nombre d'occurrences, mais une liste de tokens répétés comme celle-ci :

Hsemates	
	Hsemates
	hsemates
ku	
	ku
	ku

Figure 7. Feuille 4

1.5.Feuille 5 : V4

Cette feuille contient le tableau de la feuille v3, avec uniquement la version modifiée du texte. Autrement dit, les locuteurs, les dates, l'heure et les notifications ne sont pas incluses ici, seuls figurent les mots sans leurs répétitions (type). La plupart du travail a été effectué dans ce tableau sur cette feuille. Le tableau a été complété ici et par la suite en utilisant la fonction de VLookup sur .numbers, j'ai synchronisé les données de V4 et relié toutes les informations dans v ori. Dans cette feuille, il y a une colonne de moins que dans le tableau dans v ori :

Tokens	Langues	Erreurs Typographique	Formes Normalisées	Traduction
l'ensemble des textes tokenisés	les langues utilisées pour communiquer représentées par un identifiant unique	les mots avec erreurs grammaticales	les mots sont restructurés selon l'orthographe correcte trouvée dans le dictionnaire	la traduction en français selon la conversation

Figure 8. Les colonnes de V 4

1.6. Feuille 6 : Locuteurs

Enfin la dernière feuille contient l'ensemble des locuteurs représentés par un identifiant et les informations sur les locuteurs :

Locuteurs	ID	Âge	Langues	Origine ethnique
les noms de locuteurs	identifiant pour chaque locuteur	les âges des locuteurs	les langues parlées par les locuteurs	les origines des locuteurs

Figure 9. Les colonnes de feuille de locuteurs

Parmi les membres de ce groupe, il n'y a qu'une seule personne d'origine indienne et les autres sont malaises. J'ai également utilisé la fonction de VLookup pour relier l'identifiant de locuteurs dans le fichier v ori. Cette feuille est présentée dans l'Annexe 1.

2. Script python

Pour obtenir la statistique des langues, et pour analyser l'utilisation de l'alternance codique, il faut d'abord coder un script python pour créer un tableau de messages comportant les métadonnées et la liste des tokens avec leurs étiquettes. Le script python *extractStats.py* a été créé et pour l'entrée j'ai utilisé le fichier CSV. Le script *extractStats.py* a permis de créer un array d'array de chaînes de caractères, puis en parcourant les lignes nous pouvons détecter si la ligne est une ligne de la métadonnée. Cela va aussi détecter la fin d'un message et le début d'un autre. Chaque morceau de ligne entre le début de message et la fin de message est associé à un nom :

form	Tokens
loc	Locuteurs
lang	Langues
error	Erreurs Typographiques
norm	Formes Normalisées
glose	Traduction

Figure 10. Noms des dictionnaires dans le script avec leur valeurs.

Et un tableau de messages en sortie contenant les propriétés suivantes :

language
date
time
name
loc
tokens
withEN
withOno

Figure 11. Tableau de messages.

Ensuite, des boucles sont créées pour calculer le champ des langues et elles sont nommés comme ceci :

T	tamoul
F	français
A	anglais
M	malay
mixed	mélange de plusieurs langues

Figure 12. Identifiants des langues dans le script python.

CHAPITRE 3

-

BIOGRAPHIE LINGUISTIQUE DES LOCUTEURS

La manière de communiquer en Malaisie est différente par rapport aux pays avec officiellement monolingues, cela peut être très déroutant. Ceci est dû au fait qu'il y a une population multiraciale, multiculturelle et multi-religieuse en Malaisie. Il est donc important que dans ce chapitre, je présente les origines ethniques des locuteurs dans le groupe de WhatsApp que j'ai récolté et leur biographie linguistique. Ceci est important car, pour mieux comprendre le fonctionnement du *code-switching* chez les Malaisiens, nous aurons besoin de savoir comment ils acquièrent leurs compétences linguistiques, quelle est leur maîtrise des langues apprises et de manière plus générale quelle a été leur éducation linguistique scolaire. Par conséquent, je présenterai brièvement le système éducatif en Malaisie et les pratiques langagières dans les écoles et les universités. Je présenterai également les principales attractions de chaque état des locuteurs car les attractions ont augmenté le nombre de touristes au fil des années, ce qui a également augmenté l'utilisation de la langue anglaise.

1. Origine Ethnique

La Malaisie géographiquement séparée en deux régions : la Malaisie péninsulaire également connue sous le nom de la Malaisie occidentale et la Malaisie orientale, située sur l'île de Bornéo.



Figure 13. Malaisie péninsulaire et la Malaisie orientale à l'adresse <https://www.1clit1planet.com/malaisie.htm>

La population de la Malaisie comprend de nombreux groupes ethniques. Traditionnellement, les Malais formaient la majorité du pays, ce qui est encore le cas de nos jours. Ils sont également le groupe le plus puissant politiquement parmi les autres groupes ethniques. Les autres communautés, à l'exception des aborigènes, sont pour la plupart des immigrants et leurs descendants. Le diagramme ci-dessous distingue les différents groupes ethniques qui constituent la population malaisienne (Population Quick Info à l'adresse <http://pqi.stats.gov.my/>), la copie de cette ressource est présentée dans l'Annexe 2:

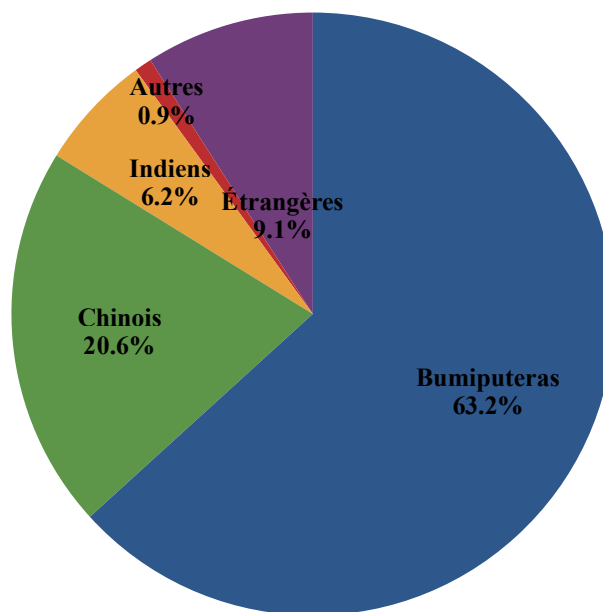


Figure 14. Population par groupe ethnique, MALAISIE , 2020.

La population de Malaisie est d'environ 32 657 300 habitants, dont 20 655 500 de Bumiputeras, 6 715 200 de Chinois, 2 020 800 d'Indiens, 305 300 d'autres groupes ethniques et 2960,400 d'étrangers.

La population malaisienne est traditionnellement divisée en deux types, les "Bumiputeras"³ et les "Non-bumiputeras". Les "Bumiputeras" sont les Malais, les aborigènes, les Sarawakiens et les Sabahans. Les Sarawakiens et les Sabahans sont les habitants de l'île de Bornéo. Les "Non-bumiputeras" sont notamment les Malaisiens d'origine chinoise, indienne et les asiatiques du sud.

³ Bumiputera est un terme politique réservé aux malais, aux sarawakiens, aux sabahans et aux aborigènes de la Malaisie.

Naturellement, les aborigènes sont les plus anciens habitants de la Malaisie. Au sein de la communauté aborigène, il existe de nombreux petits groupes ethniques en Malaisie péninsulaire et en Malaisie orientale. Les Chinois de la Malaisie sont des migrants chinois du sud-est de la Chine. Ils sont le deuxième groupe ethnique le plus important.

Ils ont émigré en raison des opportunités d'emploi dans les colonies britanniques. Les Chinois dominent le commerce depuis les temps anciens. Les Indiens ont été amenés par les Britanniques pendant leur colonisation, comme main-d'œuvre bon marché. Les migrants indiens ont été mis au travail dans les plantations. Ils forment le plus petit groupe ethnique de Malaisie. Il y a aussi un petit nombre d'étrangers : les Européens, les Eurasiens, les Thaïlandais et d'autres qui sont en Malaisie depuis longtemps et leurs descendants sont bien intégrés.

Les trois communautés ethniques constituées par les Malais, les Chinois et les Indiens étaient les principaux groupes responsables de l'indépendance du pays en 1957.

1.1.Locuteur 1

Le premier locuteur représenté par l'identifiant L1 est de Kuantan qui est la capitale de Pahang, un état situé sur la côte est de la Malaisie. Pahang est le plus grand état de la Malaisie péninsulaire. Kuantan est une grande ville réputée pour ses plages et aussi ses chutes d'eau, qui sont très fréquentées par les touristes. La ville est également connue pour ses craquelins de poisson, ses poissons salés et ses poissons grillés.

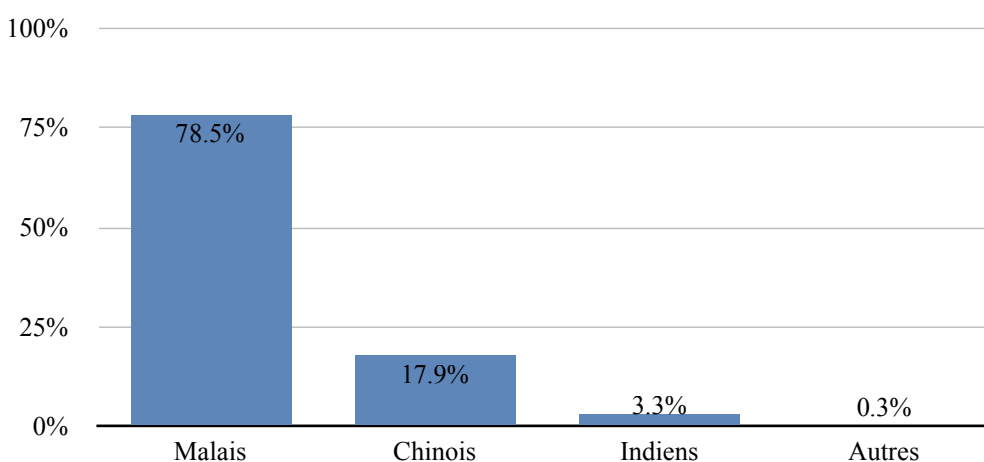


Figure 15. Population par groupe ethnique à Kuantan.

Selon le recensement de 2010 du Département des statistiques de Malaisie illustré dans la figure 15 (Wikipedia contributors, 2020a) la population de Kuantan est d'environ 427 515 habitants, dont 78,5% de Malais, 17,9% de Chinois, 3,3% d'Indiens et 0,3% d'autres groupes ethniques. La culture malaise est majoritaire, elle a un fort impact dans cette ville.

Le locuteur L1 de 27 ans, un homme d'origine indienne et pratique le christianisme. Ses parents sont des Malaisiens de la minorité indienne et ses grands-parents étaient des immigrants de l'Inde. La raison pour laquelle je mets l'accent sur sa religion, est qu'elle nous permet de déduire quelle langue il parle le plus couramment. La plupart des Indiens chrétiens en Malaisie parlent plus couramment l'anglais que le tamoul. De même, les Chinois qui sont chrétiens parlent plus couramment l'anglais, langue de leur église, que le mandarin.

Par cette observation, nous pouvons supposer que L1 a été exposé à des l'enfance principalement à trois langues: l'anglais, le tamoul et le malais. Sans oublier que chaque état a son patois ou sa manière différente de parler le malais et d'utiliser certaines expressions malaises, L1 aurait donc pu aussi acquérir la manière locale de communiquer.

1.2.Locuteur 2

Le deuxième locuteur représenté par l'identifiant L2 est originaire de l'état de Terengganu dans la ville de Kuala Terengganu : la capitale de cet état. Terengganu est située dans la péninsule orientale de la Malaisie et il est bordé par Kelantan et Pahang. À l'est, l'état est bordé par la mer de Chine méridionale. Cet état comme Pahang est également connu pour ses belles plages et ses îles. Les îles Perhentian, Redang et Kapas sont bien connues chez les Malaisiens et les touristes. Ces îles sont un véritable incontournable en Malaisie. Une autre attraction particulière de Terengganu repose sur le fait que de nombreuses espèces de tortues viennent nicher sur ses plages chaque année. La capitale elle-même est célèbre pour ses sites patrimoniaux.

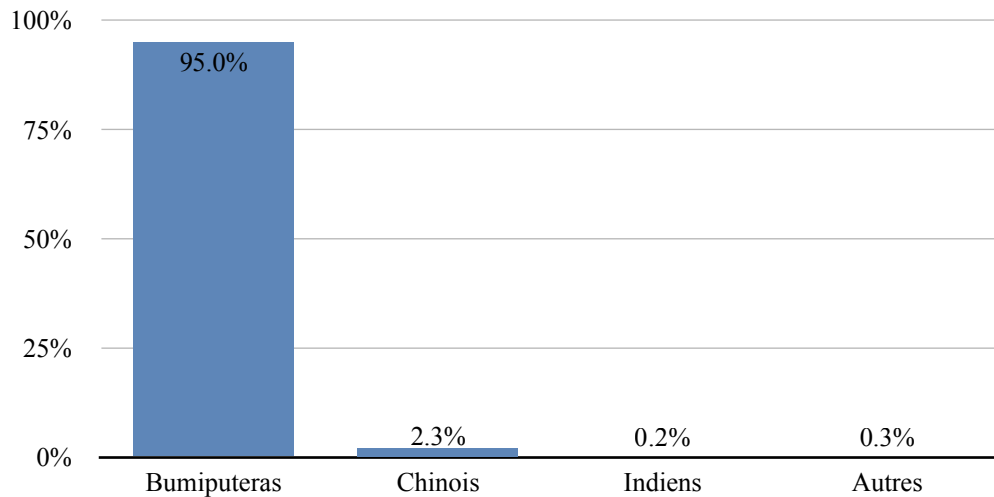


Figure 16. Population par groupe ethnique à Terengganu, 2020.

Cet état a une forte population malaise, l'une des plus nombreuses parmi les autres états de Malaisie. La figure 16 est extraite de Population Quick Info (à l'adresse <http://pqi.stats.gov.my/>) montre que la majorité de la population fait partie de l'ethnie des Bumiputeras et que les Indiens sont très minoritaires. Les Chinois sont la plus forte minorité. La population de Terengganu est d'environ 1 259 200 habitants, dont 1 196 500 Bumiputeras, 28 500 Chinois, 25 500 Indiens, 3 400 d'autres groupes ethniques et 28 300 d'étrangers. La copie de cette ressource est présentée dans l'Annexe 3.

Le locuteur L2 est malais et comme tous les Malais, il est musulman. A noter que tous les Bumiputeras ne sont pas des musulmans. Certains Bumiputeras ont une autre religion. Ceci n'est applicable qu'à certains autochtones. Ses parents sont malais ainsi que ses grands-parents. L'histoire retient que la religion islamique a été accueillie pour la première fois à Terengganu. Avant l'arrivée de l'islam, Terengganu pratiquait les religions hindoue et bouddhiste. Étant Malais⁴, sa langue maternelle est le malais⁵. Généralement les Malais en Malaisie ont besoin d'apprendre l'arabe, en raison de leur religion.

1.3.Locuteur 3

Le troisième locuteur est représenté par l'identifiant L3 et comme L2, il est malais. Ses parents aussi sont malais ainsi que ses grands-parents. Il est de Gemas, Negeri Sembilan. L'état est situé sur la côte ouest de la Malaisie péninsulaire. Le nom de l'état est

⁴ Malais renvoie ici au peuple musulman de Malaisie.

⁵ Malais renvoie ici à la langue parlée en Malaisie (Bahasa Malaysia) peu importe la religion.

dérivé des neuf villages de la langue Minangkabau, également connue sous le nom de Luak de nos jours. Minangkabau est un peuple originaire d'Indonésie et ses caractéristiques sont encore visibles aujourd'hui dans l'architecture et le dialecte parlé dans cet état (Wikipedia contributors, 2020b).

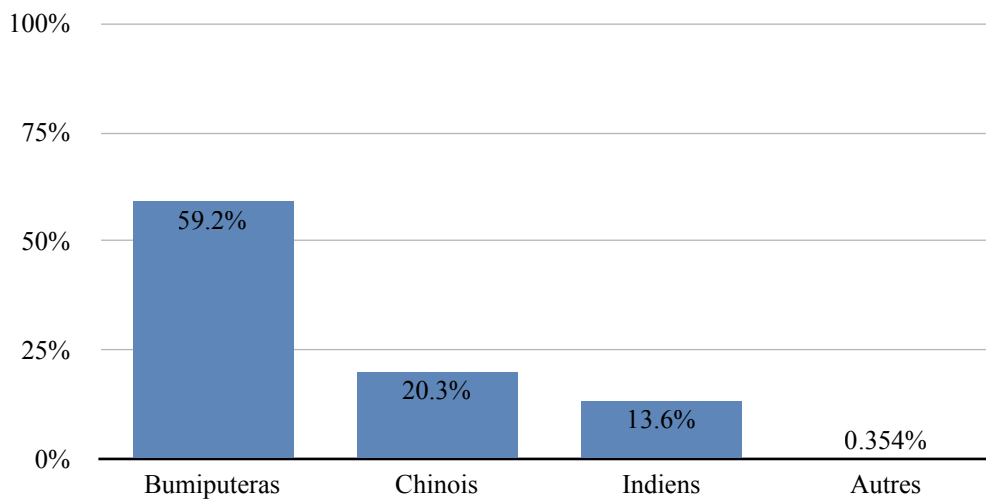


Figure 17. Population par groupe ethnique à Negeri Sembilan, 2020.

Selon Population Quick Info (à l'adresse <http://pqi.stats.gov.my/>), la population de Negeri Sembilan est d'environ 1 128 800 habitants, dont 668 400 Bumiputeras, 229 700 Chinois, 154 000 Indiens, 4 000 d'autres groupes ethniques et 72 700 d'étrangers. La copie de cette ressource est présentée dans l'Annexe 4.

Gemas est une petite ville avec peu d'habitants mais c'est une ville animée et un nœud ferroviaire. La ville est connue comme la ville des transports, les trains de toutes les directions s'arrêtent à la gare de Gemas. Cette ville est bien connue chez des Malaisiens car pendant la seconde guerre mondiale, le pont de Gemencheh qui est près de Gemas était un lieu de bataille entre l'armée japonaise et l'armée australienne. Ces armées devaient traverser ce pont afin de s'infiltrer à Gemas. Cela a eu lieu lors de l'invasion japonaise de Malaya⁶.

⁶ La Malaisie était appelée le Malaya avant d'accéder à l'indépendance le 31 août 1957.

1.4.Locuteur 4

Le quatrième locuteur représenté par l'identifiant L4 est de Kangar, Perlis. Kangar est la capitale de Perlis. Perlis est le plus petit état de Malaisie. Perlis possède de belles rizières qui sont l'une de ses principales attractions. Cet état était autrefois sous l'influence des Siamois⁷. L'état est situé sur le côté nord de la côte ouest de la Malaisie péninsulaire.

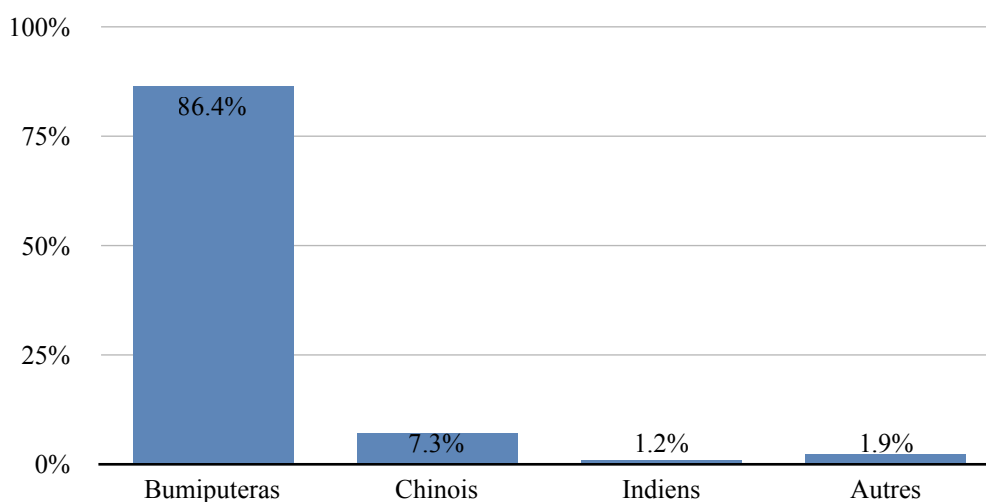


Figure 18. Population par groupe ethnique à Perlis, 2020.

Les Malais ou bien les Bumiputeras représentent la grande majorité dans cet état comme la plupart des états en Malaisie. Comme à Terengganu, les chiffres dans la figure 18 montrent une grande différence entre le groupe ethnique de malais et les trois autres groupes ethniques. La minorité indienne étant le plus petit groupe ethnique. Le nombre d'étrangers dépasse celui de la population indienne. Le nombre de la population de Perlis tiré du site Population Quick Info (à l'adresse <http://pqi.stats.gov.my/>) est d'environ 254 900 habitants, dont 220 400 Bumiputeras, 18 800 Chinois, 3 100 Indiens, 4 900 d'autres groupe ethniques et 7 600 d'étrangers. La copie de cette ressource est présentée dans l'Annexe 5.

L4 est malais ainsi que ses parents et ses grands-parents. Toutefois, ce locuteur a certainement des influences diverses qui peuvent s'expliquer notamment par l'histoire ainsi que par la situation géographique de l'état de Perlis. Autrefois, il était gouverné par le Siam, il y a donc une forte influence thaïlandaise, pays frontalier.

⁷ Les Siamois sont les citoyens du Siam. Siam est l'ancien nom du royaume de Thaïlande.

1.5.Locuteur 5

Le cinquième et le dernier locuteur est représenté par l'identifiant L5 est de Jerantut, à Pahang. L5 est malais ainsi que ses parents et ses grands-parents. Il vient du même état que le premier locuteur L1. Jerantut est une petite ville de Pahang et la principale porte d'entrée du plus ancien Parc National de Malaisie où se situe cette ville.

La figure ci-dessous, tirée du site Population Quick Info (à l'adresse <http://pqi.stats.gov.my/>), montre le nombre total des habitants de Pahang qui n'a pas été mentionné ci-dessus:

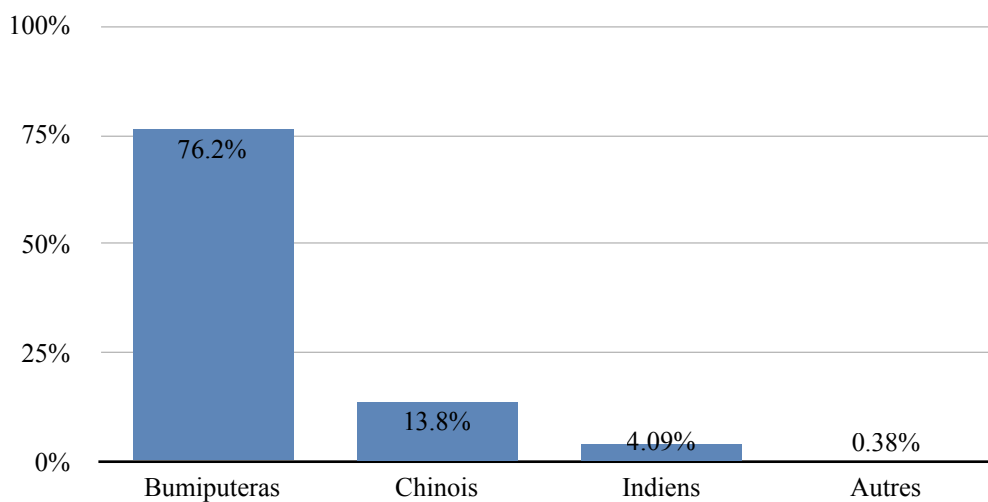


Figure 19. Population par groupe ethnique à Pahang, 2020.

Les chiffres ci-dessus montrent que la majorité est malaise et que les Chinois sont le deuxième plus grand groupe ethnique. La population de Pahang est d'environ 1 678 600 habitants, dont 1 279 700 Bumiputeras, 232 600 Chinois, 68 800 Indiens, 6,500 d'autres groupes ethniques et 91 100 d'étrangers. La copie de cette ressource est présentée dans l'Annexe 6. Nous pouvons constater que dans cet état aussi, le nombre d'étrangers dépasse celui de la population indienne et chinoise.

Locuteurs	Sexe	Ethnie	Influence linguistique de l'état	Influence linguistique dans la famille
L1	H	Indien	malais, mandarin, tamoul, anglais	anglais, malais, tamoul
L2	H	Malais	malais, mandarin, tamoul, anglais	malais, arabe, un peu d'anglais
L3	H	Malais	malais, mandarin, tamoul, anglais	malais, arabe, un peu d'anglais
L4	H	Malais	malais, mandarin, tamoul, anglais	malais, arabe, un peu d'anglais
L5	H	Malais	malais, mandarin, tamoul, anglais	malais, arabe, un peu d'anglais

Figure 20 . Récapitulatif d'influences linguistiques de cinq locuteurs

En raison du développement dans les grandes villes et les villes métropoles, le nombre de touristes et d'étrangers est élevé, ce qui a augmenté surtout l'utilisation de l'anglais et cela a aussi augmenté la pratique de l'alternance codique et du *code-mixing*.

2. *Pratiques langagières parmi différents groupes de personnes*

Nous allons maintenant nous intéresser aux pratiques langagières des Malaisiens. Les principales langues parlées en Malaisie sont le malais, l'anglais, le chinois et le tamoul. Le malais et l'anglais sont les langues officielles du pays et chaque citoyen doit apprendre ces deux langues. Outre ces quatre langues principales, il existe de nombreuses autres langues et dialectes, il y a environ 137 langues parlées en Malaisie. Presque tous les Malaisiens, surtout les jeunes générations sont multilingues, ils sont capables de parler plusieurs langues avec une fluidité variable.

La langue malaise serait originaire de Sumatra, elle est maintenant parlée non seulement en Malaisie, mais aussi à Singapour et dans d'autres pays. Le malais moderne est dérivé de l'ancien malais et une grande partie de cette langue est influencée par d'autres langues telles que l'arabe, le sanskrit, le portugais, etc.

L'anglais utilisé en Malaisie est l'anglais britannique car le pays était gouverné par les Britanniques avant l'indépendance. Au fil des années, les Malaisiens se sont adaptés aux nouvelles tendances et certains changements ont été apportés à l'utilisation de la langue anglaise. Ceci est notable chez les jeunes. Pour cette raison, nous avons aujourd'hui un nouveau dialecte appelé le manglish qui n'est parlé que par les Malaisiens.

Les Chinois en Malaisie parlent non seulement le mandarin mais ils communiquent également dans d'autres langues chinoises telles que le cantonais, le hokkien (min), le hakka, le hainanais, etc. Le mandarin est la langue standard parlée principalement par la jeune génération tandis que les plus âgés parlent d'autres langues et dialectes chinois.

D'autre part, la majorité des Indiens et des Asiatiques du sud en Malaisie parlent le tamoul. Mais il existe également d'autres langues indiennes parlées, telles que l'hindi, le malayalam, le punjabi, le telugu, etc. Il existe deux sortes de tamouls, l'un parlé par les Tamouls de l'Inde et l'autre par les Tamouls du Sri Lanka. Le tamoul parlé au Sri Lanka est une variété ancienne de cette langue. Les Indiens de Malaisie dont les ancêtres viennent de l'Inde parlent le tamoul de l'Inde, tandis que les Indiens de Malaisie avec des ancêtres sri-lankais parlent le tamoul du Sri Lanka.

Toutes les langues ci-dessus sont utilisées au quotidien en fonction de la situation et de l'environnement. Les langues utilisées pour communiquer dépendent de différents facteurs, tels que l'appartenance ethnique, la localisation, la religion, etc. Par exemple, le malais parlé dans la capitale de la Malaisie est le malais standard et l'anglais est très souvent utilisé pour communiquer, ce qui pourrait ne pas être le cas dans les petits états où l'anglais n'est enseigné que dans les écoles et utilisé pour communiquer dans certaines familles. Des états comme Kelantan, Terengganu, Kedah et Negeri Sembilan ont leur propre dialecte malais et ces dialectes ne sont utilisés que par les habitants.

Généralement en Malaisie, des mots de différentes langues sont adoptés et utilisés partout au quotidien. Par exemple, 'annai' est un mot tamoul qui signifie grand frère que toutes les ethnies en Malaisie utilisent pour désigner un homme indien qu'on ne connaît pas forcément. C'est un mot qui est fréquemment utilisé lorsqu'on se rend dans un magasin ou dans un restaurant avant de commencer la discussion. De même, tous les Malaisiens utilisent "abang" pour s'adresser un homme malais. La phrase ci-dessous est un exemple d'une phrase courante utilisée pour commander de la nourriture par tout le monde, et la phrase comprend les trois principales langues parlées en Malaisie :

“ Annai, tapau satu nasi goreng, please. ”

Traduction directe : Grand frère, un riz frit à emporter s'il vous plaît.

Annai	tamoul
tapau	mandarin, cantonais
satu	malais
nasi	malais
goreng	malais
please	anglais

Figure 21. Langue de chaque mot

Les deux sections suivantes présentent les langues parlées dans les environs de chaque locuteur et le système linguistique dans les écoles.

2.1.À la maison en famille

Dans le cas de L1, lui et sa famille communiquent principalement en anglais et parfois en tamoul. Ils communiquent souvent en malais aussi, ce qui n'est généralement pas le cas dans d'autres familles indiennes. Ils utilisent l'anglais et le malais pour parler et envoyer des messages au téléphone contrairement au tamoul qui est peu utilisé dans ces situations.

Dans les quatre autres cas, les locuteurs et leur famille conversent le plus souvent en malais. Ils conversent rarement dans d'autres langues. Les états d'où proviennent les locuteurs ont leur dialecte malais local. Et ces dialectes sont utilisés pour converser avec la population locale. Ce qui est également intéressant, c'est que la plupart des Malais aiment regarder des films en hindi et c'est une activité de passe-temps courante parmi les membres de la famille.

2.2.Entre amis et connaissances

Converser avec la famille est différent de converser avec des amis et des connaissances. Selon mon expérience, avec des amis, les conversations sont simples et amusantes, non seulement une langue est utilisée pour communiquer, mais parfois deux langues ou plus sont utilisées. De nombreux mots d'argot sont également utilisés, tout le monde mélange les langues et adopte une façon différente de parler quand ils sont avec leurs amis. Habituellement, dans un groupe mixte d'amis, ils utilisent à la fois le malais et l'anglais pour converser ou l'anglais et le tamoul ou l'anglais et le chinois.

Dans le cas des locuteurs de cette étude, les cinq locuteurs étaient principalement entourés d'amis d'ethnie malaise. Après avoir terminé leurs études au lycée, ils sont partis en France pour poursuivre leurs études. Pendant ce temps, ils conversaient en malais entre eux, car tous les cinq étaient en colocation en France et avec leurs amis français, ils conversaient en français et parfois ils utilisaient des mots en anglais s'ils ne pouvaient pas s'expliquer en français.

3. Pratiques langagières à l'école

Sous le règne des Britanniques en Malaisie, le principal medium d'enseignement dans les écoles était l'anglais. Après l'indépendance en 1957, cette pratique était encore en vigueur jusqu'à une date ultérieure (1970), lorsque la langue d'enseignement pour les écoles nationales a été changé au malais. Pendant un certain temps, un medium d'enseignement mixte a été introduit dans les écoles afin que les élèves apprennent les sciences et les mathématiques en anglais. Cependant, en raison de certains problèmes politiques, le gouvernement a changé le medium d'enseignement pour le malais en 2011.

Il existe trois types d'écoles en Malaisie: les écoles nationales, les écoles vernaculaires et les écoles internationales. Dans les écoles nationales, le medium d'enseignement est le malais et l'anglais est une matière obligatoire. Le tamoul et le mandarin sont proposés comme matières optionnelles. Le mandarin est la langue principale dans les écoles vernaculaires chinoises et le tamoul est la langue principale dans les écoles vernaculaires tamoules. Ces deux écoles enseignent également l'anglais et le malais comme matières obligatoires. Les écoles internationales sont un peu différentes car elles suivent un système éducatif international. On enseigne aussi d'autres langues internationales dans ces écoles.

3.1. École maternelle

À l'âge de 5 ou 6 ans, les parents peuvent inscrire leurs enfants dans les écoles maternelles (jardins d'enfants). L'éducation précoce n'est pas obligatoire mais les parents peuvent choisir d'envoyer leurs enfants à l'école maternelle à l'âge de 3 ans (WENR. consulté le 3 septembre 2020, à l'adresse <https://wenr.wes.org/2014/12/education-in-malaysia#:~:text=School%20Structure,pre%2Duniversity%20senior%20secondary%20study>). L'éducation dans l'école maternelle dure généralement 2 ans, puis ils passent à l'école

primaire. Les habitants dans les zones urbaines envoient généralement leurs enfants dans ces écoles, mais très peu le font dans les zones rurales car il n'y a pas de règles fixes. L'anglais et le malais sont des langues obligatoires qui doivent être enseignées à ces enfants dans les écoles maternelles et dans certaines écoles, le tamoul et le chinois sont également enseignés. Cela est devenu plus courant maintenant.

Les cinq locuteurs sont tous allés à l'école maternelle et ont appris l'anglais et le malais, par contre ils n'ont pas appris d'autres langues.

3.2.École primaire

L'enseignement primaire commence à l'âge de 7 ans et se termine à 12 ans. À la fin de la sixième année, tous les élèves doivent se présenter à un examen très important s'appelant 'UPSR' (Test de réussite à l'école primaire). Les étudiants sont testés à la fois en anglais et en malais et dans les écoles vernaculaires, la langue tamoule et la langue chinoise sont également incluses.

Les cinq locuteurs sont allés aux écoles publiques nationales. À l'exception du premier locuteur, les autres ont appris la langue arabe car, à partir de l'école primaire, elle est indispensable pour les étudiants musulmans.

3.3.Lycée

Tout comme les écoles primaires, les parents peuvent choisir d'envoyer leurs enfants soit aux écoles secondaires (lycées) nationales, soit dans des écoles secondaires (lycées) vernaculaires. Et s'ils en ont les moyens, ils peuvent envoyer leurs enfants dans des écoles privées. Le lycée est divisé en deux sections: secondaire inférieur et secondaire supérieur. Les trois premières années sont l'enseignement secondaire inférieur où les élèves ont un minimum de huit matières qui comprennent le malais, l'anglais, les sciences, l'histoire, la géographie, les mathématiques et l'éducation morale. Les matières facultatives comprennent d'autres langues, des études islamiques, des compétences de vie. Selon le type d'école, les cours sont dirigés dans leurs langues respectives. Dans certains cas, les élèves des écoles vernaculaires doivent d'abord faire une année de transition avant de commencer la première année du secondaire inférieur. À la fin de la troisième année, ils doivent se présenter à une autre évaluation importante qui est le 'PTS' pour pouvoir

continuer vers le secondaire supérieur. Les résultats de cette évaluation détermineront dans quelles classes ils seront intégrés.

Le secondaire supérieur dure deux ans et il est divisé en trois filières principales: la filière académique qui comprend la filière scientifique et la filière artistique, la filière technique et professionnelle et enfin la filière religieuse. À la fin de ces deux années, il existe une autre évaluation appelée 'SPM' qui permet aux étudiants d'obtenir leur diplôme. Cette évaluation aura un impact considérable sur la vie d'un étudiant car elle détermine s'il sera en mesure de poursuivre ses études. Les étudiants qui ne réussissent toujours pas doivent reprendre les matières et s'ils ne réussissent pas bien, leurs choix d'orientation dans la formation supérieure seront limités.

De même, les cinq locuteurs sont allés aux lycées publiques nationales, et l'un d'eux est allé à un internat public. Comme à l'école primaire, les élèves musulmans doivent entreprendre des études religieuses et continuer à apprendre la langue arabe. Jusque-là, tous les cinq ont acquis des compétences linguistiques dans au moins deux langues. À l'exception de L1, les autres ont acquis une troisième langue qui est l'arabe.

3.4. Université

Après avoir terminé leurs études au lycée, les étudiants peuvent passer à l'enseignement supérieur. Certaines universités exigent que les étudiants fassent d'abord une préparation qui peut durer un à deux ans, où ils suivent différents cours avant d'accéder aux études supérieures. Il existe un certain nombre d'universités publiques et d'universités privées établies en Malaisie. Les universités privées acquièrent aujourd'hui une bonne réputation pour une éducation de qualité internationale et des étudiants du monde entier y participent. Beaucoup de ces universités proposent des cours et des programmes d'échange avec un institut ou une université étrangère, en particulier au Royaume-Uni et en Australie. L'anglais est le principal médium utilisé dans tous les instituts et universités. Dans certaines universités publiques, la langue malaise est le principal médium d'enseignement.

Les cinq locuteurs qui ont obtenu de bons résultats pour leur évaluation de 'SPM' ont reçu une bourse parrainée par le gouvernement pour étudier en France. Ils ont quitté la Malaisie et ont passé un an à apprendre la langue française avant de commencer leurs

études en licence et en master. Ils ont passé sept ans en France et aujourd'hui ils parlent couramment le français.

CHAPITRE 4

-

ANALYSES ET RÉSULTATS STATISTIQUES

Je vais dans ce chapitre présenter les données récoltées, les colonnes contenant les informations supplémentaires dans le tableau de données, les analyses basées sur les sorties de données, puis je présenterai les résultats statistiques obtenus. Comme expliqué précédemment, les origines des locuteurs sont différentes, donc leur manière et leur propension à mélanger les langues et à s'exprimer peuvent varier.

1. Le tableau de données

Pour cette partie je vais mettre l'accent sur la feuille v ori qui contient les messages en entier et leurs étiquettes. La colonne A dans la feuille contient toutes les données sans aucune modification. A part la segmentation en tokens, aucune information n'a été ajoutée dans cette colonne et rien n'a été supprimé. Par contre les colonnes qui suivent contiennent les informations supplémentaires : la colonne C contenant les différentes langues utilisées pour communiquer entre les locuteurs (chaque langue se voyant attribuer un identifiant unique), la colonne E contenant les formes normalisées des tokens et la colonne F contenant les traductions en français de chaque token. Nous verrons pourquoi ces trois colonnes ont été créées et quelle est leur finalité. L'ensemble de ces données en format PDF est consultable dans un dossier joint avec le mémoire.

1.1.Colonne C : langues

La colonne C qui sert à l'extraction des statistiques de ce travail contient non seulement les langues utilisées par les locuteurs mais aussi les termes linguistiques associés aux noms propres et aux connecteurs des énoncés, pour lesquels il ne nous semble pas pertinent d'associer une langue en particulier. Cette colonne a été créée pour pouvoir compter les langues utilisées et à leur fréquence d'utilisation afin qu'un nombre précis puisse être obtenu après l'exécution du script python.

Entité Nommée	EN
Marqueur Discursif	MD
Interjection	I
Onomatopée	O

Figure 22. Identifiants de termes linguistiques.

L'entité nommée ici sert à désigner les noms propres tels que les anthroponymes, les toponymes, les ethnonymes, les noms de marques, les noms de films, etc.

Les marqueurs discursifs qui apparaissent assez souvent dans les discussions de groupe sont : “*la*”, “*ar*” et “*dowh/doh*”. Les Malaisiens aiment souvent utiliser “*la*”, le plus souvent dans un langage familier. “*La*” ou “*lah*” est un suffixe sans signification standard, il ne change pas le sens sémantique de la phrase, mais il assume des fonctions pragmatiques telles que l’indication du ton ou de l’émotion. Ce ne sont que des marqueurs, normalement ajoutés à la fin des phrases pour exprimer ou insister sur une émotion.

Les interjections sont des mots invariables utilisés pour exprimer un état de surprise ou d’incrédulité parmi d’autres types de sentiments. Les mots peuvent être séparés d’une phrase pour être utilisés seuls ou ils peuvent être ajoutés au début d’une phrase et généralement avec un point d’exclamation. Dans les discussions entre les cinq locuteurs, les interjections souvent utilisées sont : “*oh/ohh*”, “*weh*”, “*eh*” et “*uish*”. “*Oh*” peut introduire une phrase exclamative en exprimant, à l’oral, selon l’intonation, la surprise, la compréhension et parfois quand soudainement une idée surgit dans l’esprit. “*Weh*” est généralement utilisé pour s’adresser à un ami proche. Il peut également être utilisé pour exprimer la joie ou l’incrédulité extrême. “*Eh*” de la même manière que “*weh*” est utilisé pour exprimer la surprise et la colère, ou comme dans une question pour demander une confirmation, et parfois utilisé pour appeler quelqu’un à haute voix. Enfin “*ah*” est utilisé dans les moments pour exprimer de la compréhension et selon l’intonation, il est également utilisé pour exprimer l’ennui ainsi que lorsqu’on est en accord avec quelqu’un.

L’onomatopée peut être définie comme un : « Processus permettant la création de mots dont le signifiant est étroitement lié à la perception acoustique des sons émis par des êtres animés ou des objets. » (Larousse en ligne). J’ai choisi de classer les mots utilisés pour exprimer le rire comme l’onomatopée parce que ces mots ont été formés à partir du son du rire. Les mots onomatopées utilisés dans cet exemple sont : “*haha*”, “*hehe*”, “*hihi*” et “*kahkah*”. Ajouter des lettres à un mot est une façon de vouloir moduler un sens ou un sentiment que nous voulons exprimer. Ce jeu sur la graphie permet transmettre des nuances émotionnelles qui ne peuvent pas être faites simplement par écrit, c’est une façon de souligner un sentiment derrière un mot. Ceci est couramment utilisé lors de l’envoi de SMS ou des conversations sur les réseaux sociaux. Le nombre de lettres ajoutées et le point ajouté affecteront le sens véhiculé. Par exemple, “*haha*” exprime un rire normal ou un rire ennuyé et “*hahaha*” ou “*hahahahhhhh*” exprime une immense joie, les lettres supplémentaires jouant un rôle d’intensification. Tandis que les textos courts peuvent sembler trop neutres, ces lettres supplémentaires jouent une fonction expressive pour les participants à la conversation. Ceci est fréquemment repéré dans les données.

ANGLAIS	A
MALAIS	M
FRANCAIS	F
TAMOUL	T
CHINOIS	C
MIXTE	MX

Figure 23. Identifiants des langues.

Le tableau ci-dessus montre les langues utilisées par les cinq locuteurs avec les identifiants attribuées aux langues. Le code “mixte” désigne tous les messages contenant deux langues ou plus.

1.2.Colonne E : formes normalisées

Les cinq locuteurs ont utilisé de nombreuses formes courtes dans leur discussion de groupe. Ils ont utilisé des formes abrégées pour la plupart des mots, ce qui engendre une certaine complexité dans l'analyse des données, car ils ont chacun leur propre manière de raccourcir les mots, et celle-ci change en fonction du contexte de chaque phrase. Pour pouvoir extraire un résultat statistique précis, j'ai donc besoin de normaliser la forme abrégée. Je l'ai fait manuellement une par une, car il y avait différentes formes pour les mots. Par exemple, le mot "aku" qui signifie "moi" a au moins trois formes abrégées : "ak", "aq", "ku". Pourtant il y avait beaucoup de mots que je ne connaissais pas, parce que c'étaient des formes issues de dialectes malais locaux, et la signification de ces mots était discutable. Donc, parfois, un mot pouvait être sous une forme abrégée, mais je ne le savais pas en pensant que c'était sa forme standard, mais en cherchant en ligne et en enquêtant autour de moi, j'ai pu normaliser le mot dans la forme appropriée.

Il y avait aussi des cas où deux mots étaient combinés pour faire un mot composé, dont la forme n'existe pas dans le dictionnaire. Par exemple, "kolumpo" qui signifie Kuala Lumpur, qui est la capitale de la Malaisie ou encore "takde", une fois décomposé en "tidak ada" qui signifie "il n'y pas". D'autres exemples sont : "bdk2", "xsmar", "takyah", etc. Le caractère latin *x* est utilisée pour exprimer le déni ou le refus. Il remplace le mot "non". Par exemple : "xde" => "tidak ada" qui signifie "il n'y a pas", "xfree" => combinaison d'un mot en malais et d'un mot en anglais => "tidak free" qui signifie "n'est pas disponible".

Dans la langue malaise, la règle pour former les pluriels d'une forme singulière d'un mot consiste à le dupliquer mot en insérant un trait d'union. Par exemple, "kawan" signifie "ami" et "kawan-kawan" signifie "amis". Mais dans ce cas, les locuteurs utilisent le chiffre 2 pour effectuer cette transformation : "kawan2", "mantap2", "awl2", etc. Le plus souvent, les Malaisiens utilisent cette façon de doubler les mots pour d'autres langues, telles que: "okay-okay"/"okok". De tels exemples sont fréquemment trouvés dans ces données.

1.3.Colonne F : traduction

Cette colonne n'allait en aucun cas aider l'extraction statistique mais afin de mieux comprendre les données et de pouvoir analyser chaque phénomène, la traduction nous a semblé importante. C'est la raison pour laquelle j'ai choisi de traduire chaque mot en

fonction de son contexte dans la phrase ainsi qu'en fonction de la signification originale du mot trouvé dans le dictionnaire. Dans quelques cas, après avoir tokenisé un mot composé, il n'a pas été possible ou il était difficile de traduire ces mots. Comme dans les cas de toponymes : “*Bukit Kluang*”. Le mot “*kluang*” n'a plus de sens une fois séparé de “*bukit*”. Cependant il est facile de traduire “*bukit*”, ce mot signifiant “colline”.

2. Les statistiques

J'ai eu recours à un script python pour extraire les fréquences d'utilisation de différentes langues parmi les locuteurs dans le but de réaliser les analyses statistiques. Pour ce faire donc le script python *extractStats.py* été créé. On obtient en sortie le nombre de messages par langue et le nombre de messages produits par les locuteurs dans telle ou telle langue. Il y a d'autres langues qui sont apparues dans les données avec très peu d'occurrences, mais nous ne les analyserons pas car nous allons nous concentrer uniquement sur les quatre langues principales parlées par les cinq locuteurs : le malais, l'anglais, le français et le tamoul. La sortie .txt de ces résultats est consultable dans l'annexe 7.

2.1. Nombre de messages par langue

T	3
F	421
A	108
M	1125
mixed (mixte)	1488

Figure 24. Nombre de messages par langue.

Le tableau ci-dessus montre le nombre de messages dans les langues utilisées par les cinq locuteurs, et l'on constate que la majorité des messages produits est considéré comme mixte. Le malais est la deuxième langue la plus utilisée, suivie du français, ensuite de l'anglais et enfin du tamoul. Nous pouvons constater qu'il y a de nombreuses occurrences de *code-switching*, selon les chiffres récoltés.

2.2. Nombre de messages par locuteur

Dans cette section, il y a des tableaux présentant les résultats du nombre de messages produits par les locuteurs.

T	1
F	85
A	50
M	268
mixed (mixte)	408

Figure 25. Nombre de messages produits par L1.

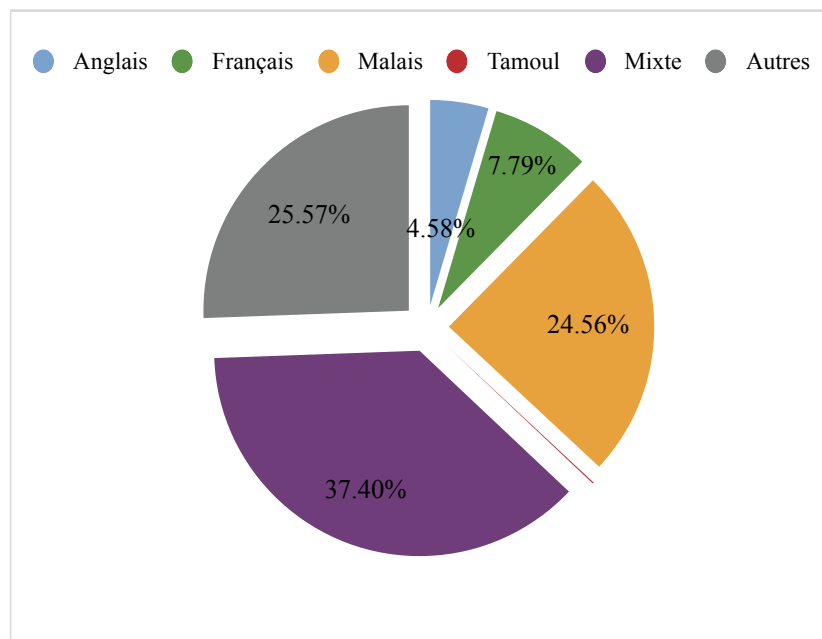


Figure 26. Répartition des langues utilisées par L1

Dans le tableau ci-dessus, nous avons le nombre de messages produits par L1 en quatre langues ainsi que le nombre de messages en mélangeant les langues. Après observation, nous pouvons constater que ce locuteur mélange son usage des langues. La langue qu'il utilise le plus fréquemment sans mélanger avec d'autre langue est le malais.

T	0
F	177
A	23
M	379
mixed (mixte)	472

Figure 27. Nombre de messages produits par L2.

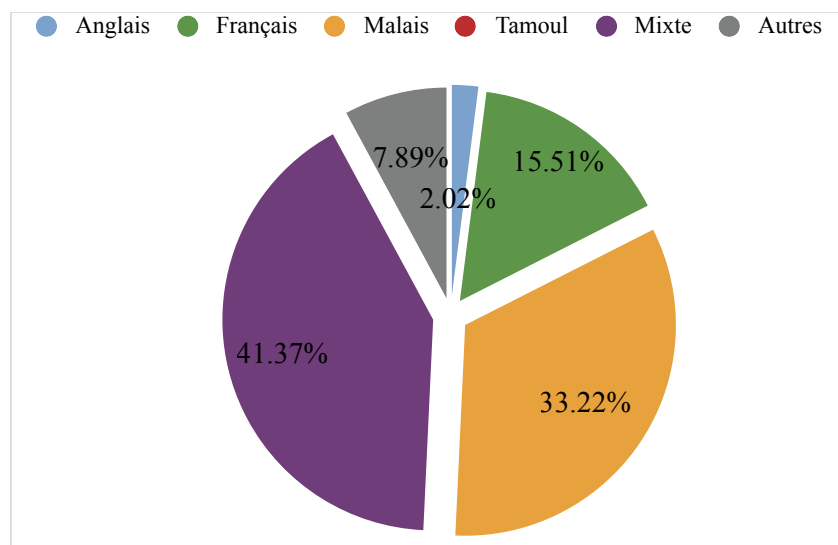


Figure 28. Répartition des langues utilisées par L2

L2 comme L1 a mélangé ses langues en communiquant avec ses collègues. Selon les chiffres dans le tableau ci-dessus, nous pouvons observer les langues les plus souvent utilisées toutes seules sans mélanger par L2 : le malais avec 472 messages est utilisé le plus fréquemment suivi du français avec 177 messages et ensuite de l'anglais avec 23 messages. La majorité des messages produits par ce locuteur est mélangée dans deux langues.

T	2
F	60
A	11
M	226
mixed (mixte)	251

Figure 29. Nombre de messages produits par L3.

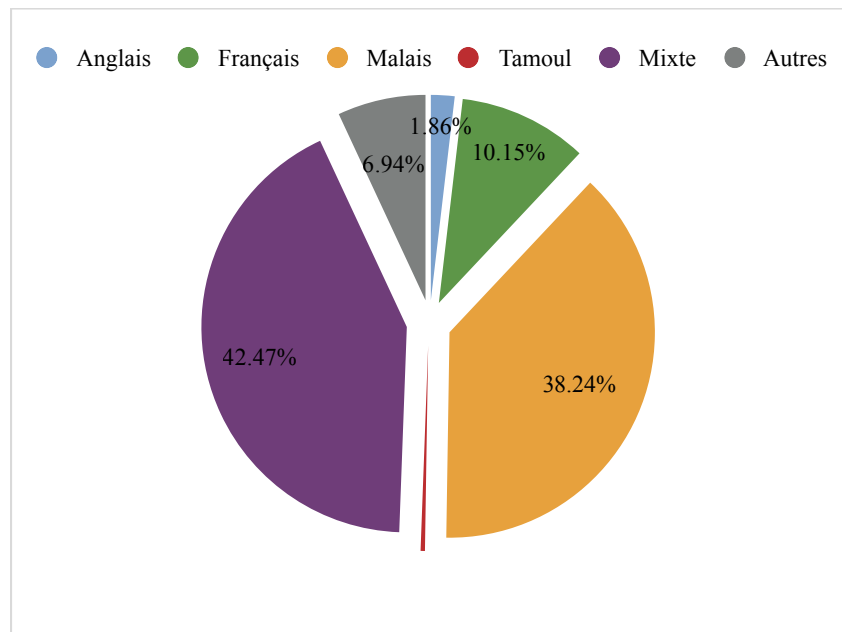


Figure 30. Répartition des langues utilisées par L3

Selon les chiffres dans le tableau ci-dessus, il est évident que, le profil du locuteur L3 est très similaire à celui de L2 car il mélange aussi les langues. Les langues les plus fréquemment utilisées par L3 sans les mélanger sont le malais et le français. Le nombre total de messages dans ces quatre langues est de 550 qui est composé de 251 messages mixtes, de 226 messages en malais, de 11 messages en anglais, de 60 messages en français et de 2 messages en tamoul. Il est intéressant de noter qu'il a également utilisé le tamoul, mais seulement deux fois, par exemple, "*macha*" et "*laddu*". Cela est intéressant car il ne parle pas le tamoul et n'a jamais appris le tamoul auparavant.

T	0
F	44
A	16
M	70
mixed (mixte)	138

Figure 31. Nombre de messages produits par L4.

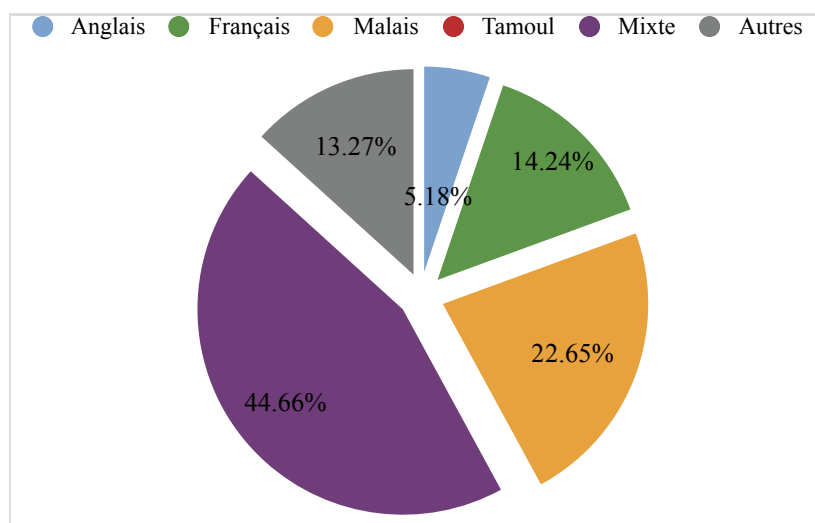


Figure 32. Répartition des langues utilisées par L4

Parmi les autres membres du groupe, ce locuteur a produit le moins de messages pour des raisons inconnues. Tout comme le montrent les résultats ci-dessus, le locuteur L4 a également mélangé son utilisation de deux différentes langues, avec 138 messages mixtes. Les langues les plus utilisées sans mixage sont le malais avec 70 messages et le français avec 44 messages. L'anglais est composé de 16 messages.

T	0
F	55
A	8
M	182
mixed (mixte)	219

Figure 33. Nombre de messages produits par L5.

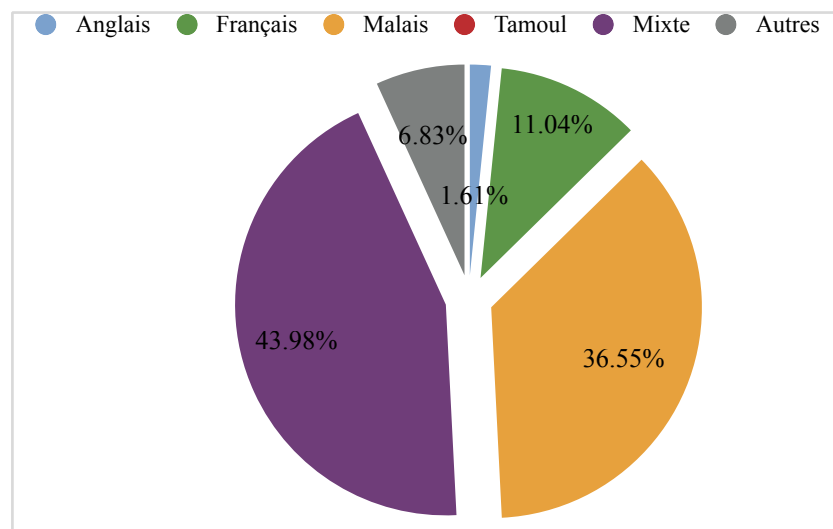


Figure 34. Répartition des langues utilisées par L5

Dans le tableau ci-dessus, nous avons le nombre de messages produits par L5 dans les quatre langues utilisées pour communiquer entre les cinq locuteurs. Nous pouvons constater que ce locuteur mélange son usage des langues aussi. La langue qu'il utilise le plus fréquemment sans mélanger avec d'autre langue est le malais. La deuxième langue plus souvent utilisée après le malais est le français. Le nombre total de messages dans ces quatre langues est de 464 qui est composé de 219 messages mixtes, de 182 messages en malais, de 8 messages en anglais, de 55 messages en français et aucun en tamoul.

2.3. Les messages mixtes

Avec le script python, on obtient en sortie également un fichier "mixed.csv" contenant les messages mixtes produits par les locuteurs. La langue de base des messages

produits par les locuteurs est le malais. Dans les exemples ci-dessous, nous pouvons observer les messages mixtes :

*“L2 : Ak dah pegi **shopping** aritu brg dapur”*

Dans ce message produit par L2, il a utilisé shopping dans une phrase de base syntaxique malaise. Le mot “shopping” n’est pas un emprunt ici. Ce que nous observons ici est un mélange d’un mot anglais dans une phrase malaise. Le locuteur a remplacé “beli-belah” qui signifie “faire les course” ou “faire du shopping”.

*“L3 : Bwk **laddu** skit”*

*“L3 : Pagi2 lepas smbhyg boleh mkn **laddu**”*

*“L1 : **Laddu** mne leh tahan lame”*

Dans le deuxième exemple, ce sont des messages produits par L3 et L1, ils ont mentionné “laddu” un mot en tamoul qui est un type de dessert. L3 connait ce dessert même s’il ne parle pas le tamoul et il est origine d’ethnie malais.

*“L1 : Dok aq ambik **flight** pgi dr kuantan”*

*“L1 : Smpt aq dh thun blk camtu **bro**”*

Dans les deux messages ci-dessus, L1 comme L2 ont mélangé un mot anglais “flight” qui signifie “vol” et “bro” qui signifie “mec” (usage familier) dans une phrase produite en malais. Ils auraient pu utiliser le mot “penerbangan” au lieu de “flight”.

*“L1 : **Thanks** ar doakn aq”*

*“L2 : Dah **check in** blum nk pilih **seat** ni”*

*“L2 : Ntah boleh **survive** ke tak sbulan ni”*

Dans les trois exemples ci-dessus, les locuteurs ont utilisé la structure syntaxique malaise et remplacé quelques mots malais en anglais. Souvent les gens remplacent les mots qui sont longs par des mots en anglais plus courts. Dans ces exemples, “thanks”, la forme abrégée pour le mot “thank you” est normalement utilisée à la place de “terima kasih” qui signifie “merci” en malais par les Malaisiens. Cela est dû au fait que, la forme courte est plus pratique à utiliser.

“L1 : *Formule ape lak*”

“L2 : *Ak dah virer kat ko*”

“L1 : *Aq virer blk no worries*”

C'est le même principe pour d'autres langues à part l'anglais. Dans les exemples ci-dessus, le mot “virer” est utilisé probablement parce qu'ils utilisent les banques locales, d'où le fait qu'ils préfèrent utiliser des termes français relatifs aux banques.

En regardant les exemples ci-dessus, déjà nous pouvons constater que les locuteurs utilisent la structure syntaxique de base malaise en remplaçant les mots qu'ils veulent dans d'autres langues en plus de l'alternance entre deux langues qui se fait principalement avec le malais et l'anglais. L'ensemble de ces messages mélangés en format PDF est consultable dans un dossier joint avec le mémoire.

3. *Les interprétations*

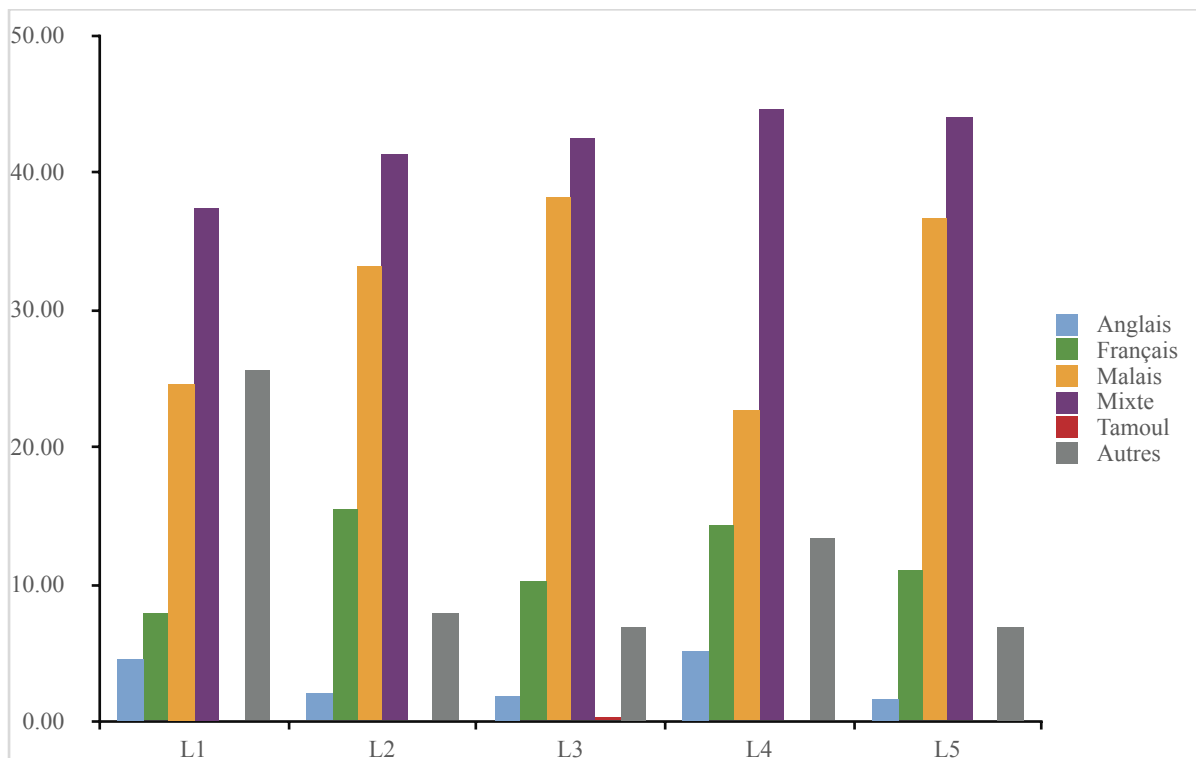


Figure 35. Pourcentage des langues produits par les cinq locuteurs.

Selon les résultats statistiques, nous pouvons interpréter que dans les données récoltés, il y a un grand nombre de messages où les cinq locuteurs mélangent leur utilisation des langues. La sortie de messages mixte est obtenue en comptant l'apparition

d'au moins deux langues dans le message. Comment pouvons-nous interpréter cela et pourquoi est-ce très couramment pratiqué par les cinq locuteurs?

3.1.L'environnement influence-t-il l'utilisation d'une langue?

Les données obtenues pour cette recherche étaient des messages envoyés dans un chat de groupe de cinq étudiants qui à l'époque étudiaient en France et qui étaient également des colocataires.

En regardant la figure 30 ci-dessus, le pourcentage pour la langue malaise et la langue française est le plus élevé sans compter le pourcentage de mélange de deux langues. Cela peut être dû au fait qu'ils vivaient en France, ils auraient adopté des mots ou des expressions françaises dans leur vie quotidienne. La raison pour laquelle le malais domine est que parmi les cinq locuteurs, un seul d'entre eux est indien et les autres sont malais, donc ils parleraient sans aucun doute la langue parlée par la majorité d'entre eux. Avec ce raisonnement, nous pouvons probablement dire que l'environnement affecte l'utilisation d'une langue particulière. Nous pouvons également constater que le fait d'avoir un groupe d'amis mixtes favorise l'alternance des langues ainsi que le mélange des langues. Des études pragmatiques supplémentaires seront nécessaires pour analyser ce phénomène.

3.2.Code-switching ou code-mixing?

Selon Waris (2012), le *code-switching* désigne l'utilisation de plus d'une langue ou une variété de langues dans la conversation. Parfois, des locuteurs bilingues peuvent rencontrer des problèmes dans une conversation. Pour faire passer ses idées le locuteur change de langue ce qui est possible uniquement lorsque l'interlocuteur possède le même contexte linguistique. Pour l'auteur, le *code-mixing* se produit généralement lorsque les locuteurs maîtrisent les deux langues ensemble, basculent entre deux langues dans la mesure où ils changent quelques mots d'une langue à l'autre au cours d'un seul énoncé.

Lors de la rédaction de ce mémoire, j'ai surtout utilisé le mot "mélange" au lieu d'"alternance". La raison vient du fait que, lors de l'observation, la plupart des messages ont été écrits en malais avec l'incorporation des unités d'autres langues. Il y a une ligne très fine entre *code-mixing* et *code-switching* qui s'est avérée être un problème au cours de cette étude.

Conclusion

A travers ce travail de recherche, j'ai essayé de décrire le phénomène de *code-switching* chez les Malaisiens. Pour être capable d'établir une description de travail, cette recherche m'a demandé de collecter les données.

Pendant la période de travail sur mon mémoire, mon objectif était d'étudier les conversations entre les Malaisiens pour voir à quelle fréquence ils changent ou mélangent les codes. Le but de cette recherche, après avoir réussi à réunir le corpus de pratiques langagières, est d'observer quelles langues sont les plus utilisées par les locuteurs lors du *code-switching*.

La partie qui a pris le plus de temps a été de normaliser et de traduire les messages, mais en faisant cela, j'ai beaucoup appris, j'ai appris que même avec la même nationalité, les locuteurs de différents états avaient leur propre style de communication en utilisant les langues qu'ils maîtrisent. Les habitants d'une ville plus urbanisée étaient plus à l'aise dans l'utilisation des autres langues que leur langue maternelle ce qui a augmenté la pratique de l'alternance codique.

Concernant les résultats sur le *code-switching* entre les locuteurs, je pense qu'une étude plus détaillée est nécessaire. J'ai besoin d'étudier le contexte des phrases en utilisant des outils de traitement du langage. D'autres études doivent être menées pour analyser la structure lexicale des messages.

Les résultats mènent à plusieurs interprétations, pouvons-nous dire que les locuteurs alternent les langues ou ont-ils simplement mélangé des mots d'autres langues dans leurs conversations? Du point de vue de mes observations, nous pouvons constater que même s'il n'y a pas beaucoup d'occurrences de l'alternance codique, il en reste encore quelques traces. Selon les données récoltées, il y a un mélange de *code-switching* et de *code-mixing*.

Pour conclure, je suis toujours très intéressée par ce phénomène et pour pouvoir continuer de l'étudier, je devrai effectuer une étude statistique et une étude lexicale plus détaillée sur un nombre de données plus important. Ce faisant, je pourrai déterminer si les informations supplémentaires comme les traductions ou les structures lexicales, permettraient-elles d'améliorer mes analyses sur ce phénomène et dans quelles mesures le feraient-elles?

Bibliographie

1. Auer, P. (1999). From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech*. *International Journal of Bilingualism* 1999 3: 309, 3(4).
2. Black, A. W., Rallabandi, S., & Sitaram, S. (2018). Automatic Detection of Code-switching Style from Acoustics. *Proceedings of The Third Workshop on Computational Approaches to Code-Switching*.
3. Morel, E. & Pekarek Doehler, S. (2013). Les ‘textos’ plurilingues : l’alternance codique comme ressource d’affiliation à une communauté globalisée. *Revue française de linguistique appliquée*, vol. xviii(2), 29-43. doi:10.3917/rfla.182.0029.
4. Muysken, P. (s. d.). CODE-SWITCHING AND GRAMMATICAL THEORY. 23.
5. Nilep, C. (2006). “Code Switching” in Sociocultural Linguistics. *Colorado Research in Linguistics*, 19.
6. Stapa, S. H., & Sahabudin Khan, N. N. B. (2016). Functions of Code-Switching: A Case Study of a Mixed Malay- Chinese Family in the Home Domain. *Pertanika*.
7. Solorio, T., Maharjan, S., & Mave, D. (2018). Language Identification and Analysis of Code-Switched Social Media Text. *Proceedings of The Third Workshop on Computational Approaches to Code-Switching*.
8. Waris, A. M. (2012). Code switching and mixing. *Jurnal Dakwah Tabligh*, 13(1), 13.
9. Weinreich, U. (1953). *Languages in contact, findings and problems*. New York, États unis : Generic.
10. Xu, H., Pham, V. T., Kyaw, Z. T., Li, H., Chng, E. S., & Lim, Z. H. (2018). Mandarin-English Code-switching Speech Recognition. *Interspeech*.

Sitographie

1. 1clic1planet. Consulté le 17 août 2020, à l'adresse 1clic1planet. Consulté le 17 août 2020, à l'adresse <https://www.1clic1planet.com/malaisie.htm>
2. Department of Statistics Malaysia Official Portal. (s. d.). Consulté à l'adresse https://www.dosm.gov.my/_v1/
3. Larousse. (s. d.). Réussite. Dans Le Dictionnaire Larousse en ligne. Consulté le 24 août 2020 sur <https://www.larousse.fr/dictionnaires/francais/onomatop%C3%A9e/56061>
4. Population par groupe ethnique, MALAISIE. (2020). Consulté le 17 août 2020, à l'adresse <http://pqi.stats.gov.my/result.php?token=dc8cf2810e9328de7c597e7edbcdbc70>
5. Population Quick Info. Consulté le 17 août 2020, à l'adresse à l'adresse <http://pqi.stats.gov.my/>
6. StudyMalaysia. Consulté le 17 août 2020, à l'adresse à l'adresse <https://www.studymalaysia.com/education/higher-education-in-malaysia/a-glance-at-the-malaysian-education-system>
7. WENR. Consulté le 3 septembre 2020, à l'adresse <https://wenr.wes.org/2014/12/education-in-malaysia#:~:text=School%20Structure,pre%2Duniversity%20senior%20secondary%20study>.
8. Wikipedia contributors. (2020, janvier 7). Code-switching. Consulté le 15 janvier 2020 à l'adresse <https://en.wikipedia.org/wiki/Code-switching>
9. Wikipedia contributors. (2020a, juillet 6). Kuantan. Consulté à l'adresse https://en.wikipedia.org/wiki/Kuantan#cite_note-2010census-5
10. Wikipedia contributors. (2020b, août 16). Negeri Sembilan. Consulté à l'adresse https://en.wikipedia.org/wiki/Negeri_Sembilan
11. Zakaria Bin Ahmad, Craig A. Lockard, & others. (2020, 27 août). Malaysia. Consulté à l'adresse <https://www.britannica.com/place/Malaysia>

Table des illustrations

Figure 1. L'alternation dans Muysken, P.C. (1997). Code-switching processes : alternation, insertion, congruent lexicalization.....	17
Figure 2. l'insertion dans Muysken, P.C. (1997). Code-switching PROCESSES: alternation, insertion, congruent lexicalization.....	18
Figure 3. lexicalisation congruente dans Muysken, P.C. (1997). Code-switching PROCESSES: alternation, insertion, congruent lexicalization.	18
Figure 4. Feuilles de données.....	25
Figure 5. Les colonnes de v ori	29
Figure 6. Feuille v ori.....	29
Figure 7. Feuille 4	30
Figure 8. Les colonnes de V 4.....	31
Figure 9. Les colonnes de feuille de locuteurs	31
Figure 10. les noms des dictionnaires dans le script avec leur valeurs	32
Figure 11. Tableau de messages.	32
Figure 12. Identifiants des langues dans le script python.	32
Figure 13. Malaisie péninsulaire et la Malaisie orientale consulté à l'adresse https://www.lcliclplanet.com/malaisie.htm	34
Figure 14. Population par groupe ethnique, MALAISIE , 2020	35
Figure 15. Population par groupe ethnique à Kuantan.....	36
Figure 16. Population par groupe ethnique à Terengganu, 2020.	35
Figure 17. Population par groupe ethnique à Negeri Sembilan, 2020.	36
Figure 18. Population par groupe ethnique à Perlis, 2020.	37
Figure 19. Population par groupe ethnique à Pahang, 2020.	38
Figure 20 . Récapitulatif d'influences linguistiques de cinq locuteurs	42
Figure 21. Langue de chaque mot.....	44
Figure 22. Identifiants de termes linguistiques	50
Figure 23. Identifiants des langues.....	51
Figure 24. Nombre de messages par langue.....	53
Figure 25. Nombre de messages produit par L1	54
Figure 26. Répartition des langues utilisées par L1	54
Figure 27. Nombre de messages produit par L2	55
Figure 28. Répartition des langues utilisées par L2	55
Figure 29. Nombre de messages produit par L3	56
Figure 30. Répartition des langues utilisées par L3	56
Figure 31. Nombre de messages produit par L4	57
Figure 32. Répartition des langues utilisées par L4	57
Figure 33. Nombre de messages produit par L5	58
Figure 34. Répartition des langues utilisées par L5	58
Figure 35. Pourcentage des langues produits par les cinq locuteurs.....	60

Table des annexes

Annexe 1 Les informations de locuteurs de groupe de chat	69
Annexe 2 Population par groupe ethnique en Malaisie	70
Annexe 3 Population à Terengganu	71
Annexe 4 Population à Negeri Sembilan	72
Annexe 5 Population à Perlis	73
Annexe 6 Population à Pahang	74
Annexe 7 Les résultats de statistique	75
Annexe 8 Les messages mixtes.....	76

Annexe 1
Les informations de locuteurs de groupe de chat

	ID	Age	Langues	Origine Ethnique	Localisation
	L1	27	Anglais	Indien	Kuantan, Pahang
			Malais		
			Tamoul		
			Français		
	L2	27	Malais	Malais	Kampung Losong, Kuala Terengganu, Terengganu
			Anglais		
			Français		
	L3	27	Malais	Malais	Gemas, Negeri Sembilan
			Anglais		
			Français		
	L4	27	Malais	Malais	Kangar, Perlis
			Anglais		
			Français		
	L5	27	Malais	Malais	Jerantut, Pahang
			Anglais		
		Français			

Annexe 2

Population par groupe ethnique en Malaisie

17/08/2020

POPULATION QUICK INFO

POPULATION QUICK INFO

Penduduk / Population ('000) : 32657.3

Penduduk mengikut kumpulan etnik, MALAYSIA, 2020

Population by ethnic group, MALAYSIA, 2020

Kumpulan etnik <i>Ethnic group</i>	MALAYSIA
Jumlah <i>Total</i>	32657.3
Bumiputera	20655.5
Cina <i>Chinese</i>	6715.2
India <i>Indians</i>	2020.8
Lain-lain <i>Others</i>	305.3
Bukan warganegara Malaysia <i>Non-Malaysian Citizens</i>	2960.4

Sumber: Jabatan Perangkaan Malaysia

Nota: 1. Hasil tambah mungkin berbeza kerana pembundaran.

2. Anggaran.

Source: Department of Statistics, Malaysia

Notes 1. The added total may differ due to rounding.

2. Estimate.

Annexe 3 Population à Terengganu

19/08/2020

POPULATION QUICK INFO

POPULATION QUICK INFO

Penduduk / Population ('000) : 1259.2

Penduduk mengikut kumpulan etnik, TERENGGANU, 2020

Population by ethnic group, TERENGGANU, 2020

Kumpulan etnik <i>Ethnic group</i>	TERENGGANU
Jumlah <i>Total</i>	1259.2
Bumiputera	1196.5
Cina <i>Chinese</i>	28.5
India <i>Indians</i>	2.5
Lain-lain <i>Others</i>	3.4
Bukan warganegara Malaysia <i>Non-Malaysian Citizens</i>	28.3

Sumber: Jabatan Perangkaan Malaysia

Nota: 1. Hasil tambah mungkin berbeza kerana pembundaran.

2. Anggaran.

Source: Department of Statistics, Malaysia

Notes 1. The added total may differ due to rounding.

2. Estimate.

Annexe 4 Population à Negeri Sembilan

20/08/2020

POPULATION QUICK INFO

POPULATION QUICK INFO

Penduduk / Population ('000) : 1128.8

Penduduk mengikut kumpulan etnik, NEGERI SEMBILAN, 2020

Population by ethnic group, NEGERI SEMBILAN, 2020

Kumpulan etnik <i>Ethnic group</i>	NEGERI SEMBILAN
Jumlah <i>Total</i>	1128.8
Bumiputera	668.4
Cina <i>Chinese</i>	229.7
India <i>Indians</i>	154
Lain-lain <i>Others</i>	4
Bukan warganegara Malaysia <i>Non-Malaysian Citizens</i>	72.7

Sumber: Jabatan Perangkaan Malaysia

Nota: 1. Hasil tambah mungkin berbeza kerana pembundaran.
2. Anggaran.

Source: Department of Statistics, Malaysia

Notes 1. The added total may differ due to rounding.
2. Estimate.

Annexe 5 Population à Perlis

20/08/2020

POPULATION QUICK INFO

POPULATION QUICK INFO

Penduduk / Population ('000) : 254.9

Penduduk mengikut kumpulan etnik, PERLIS, 2020

Population by ethnic group, PERLIS, 2020

Kumpulan etnik <i>Ethnic group</i>	PERLIS
Jumlah <i>Total</i>	254.9
Bumiputera	220.4
Cina <i>Chinese</i>	18.8
India <i>Indians</i>	3.1
Lain-lain <i>Others</i>	4.9
Bukan warganegara Malaysia <i>Non-Malaysian Citizens</i>	7.6

Sumber: Jabatan Perangkaan Malaysia

Nota: 1. Hasil tambah mungkin berbeza kerana pembundaran.

2. Anggaran.

Source: Department of Statistics, Malaysia

Notes 1. The added total may differ due to rounding.

2. Estimate.

Annexe 6 Population à Pahang

20/08/2020

POPULATION QUICK INFO

POPULATION QUICK INFO

Penduduk / Population ('000) : 1678.6

Penduduk mengikut kumpulan etnik, PAHANG, 2020

Population by ethnic group, PAHANG, 2020

Kumpulan etnik <i>Ethnic group</i>	PAHANG
Jumlah <i>Total</i>	1678.6
Bumiputera	1279.7
Cina <i>Chinese</i>	232.6
India <i>Indians</i>	68.8
Lain-lain <i>Others</i>	6.5
Bukan warganegara Malaysia <i>Non-Malaysian Citizens</i>	91.1

Sumber: Jabatan Perangkaan Malaysia

Nota: 1. Hasil tambah mungkin berbeza kerana pembundaran.

2. Anggaran.

Source: Department of Statistics, Malaysia

Notes 1. The added total may differ due to rounding.

2. Estimate.

Annexe 7

Les résultats de statistique

3630 messages pour 5 locuteurs dont 762 avec EN

Nombre de messages par langue :

T => 3

F => 421

A => 108

M => 1125

mixed => 1488

Nombre de messages par locuteurs :

L1 => 1091 messages

A => 50

F => 85

M => 268

T => 1

mixed => 408

other: => 145

other: A/F => 1

other: Arabe => 1

other: EN => 7

other: EN EN EN => 1

other: I => 1

other: MD => 2

other: O => 121

L2 => 1141 messages

A => 23

F => 177

M => 379

mixed => 472

other: => 64

other: A/F => 1

other: A/F/M => 1

other: Arabe => 2

other: Arabe Arabe => 1

other: C => 1

other: EN => 7

other: EN EN => 1

other: I EN => 1

other: Indonesian => 1

other: MD => 1

other: O => 9

L3 => 591 messages

A => 11

F => 60

M => 226

T => 2

mixed => 251

other: => 15

Annexe 8
Les messages mixtes

L1	Jom planning brg2 nk bwk
L1	Aq bwk kicap sebotol milo speket n sos cili
L1	Yg lain aq bwk skit2 ar perch nasi grg tu smua
L1	Enth la
L3	Xpelaa kau peluh2 nanti tadah
L2	Ak dah pegi shopping aritu brg dapur
L2	Serbuk2 tu ak beli semua jenis tp sikit2 je la
L3	Bwak barang rare2 sikit
L2	Bwk laddu skit
L2	Pagi2 lepas smbhyg boleh mkn laddu
L1	Laddu mne leh tahan lame
L2	Selempang xknl
L3	Buat2 x tau Cmne nak jd padiyappa
L2	Bwk laddu ayam
L3	Mu duo tido kolumpo ke kelik pagi
L1	Dok aq ambik flight pgi dr kuantan
L2	Tido dulu la
L3	Flite kau pkul brp
L3	Tinggal la kau
L1	Smpt aq dh thun blk camtu bro
L1	Flight kul smpi klia kul camtu
L3	Aku doakan eh Aminn
L1	Thanks ar doakn aq
L1	Caring btul aq nie
L2	Weh mcm mane nk buat ayt yg ditawarkan kepada saya dlm french
L3	Geng klau jumpa adapter audio tv tuh belikan dulu
L2	Ak takot silap beli je boleh nk try caru
L2	Dah check in blum nk pilih seat ni
L3	Dh bleh cek in ke Ak nak business laa
L1	Blh check in dh
L1	Aq punye ade satu je seat
L1	Dubai-lyon takde seat lgi
L2	Dubai Lyon nnti ko dah check in dia akn keluar

L1	Klia-dubai 35g
L2	Time pilih seat nnti
L2	Tp ak nk try pilih td da penuh
L1	Aq dubai-lyon 24g
L2	Babi ak dah kat aepot dah ni
L1	Aq grak sok pgi ar
L1	Ko stay hotel yg biase tu ke
L2	Weh tanya shahir bila duit nk masuk ak da kering ni
L2	Ntah boleh survive ke tak sbulan ni
L1	Sme la situasi kite smua
L2	Ak ada je dlm bank
L1	Inscription lum buat
L2	Ak nk jual jam kat sale grenoble
L1	Nie tuk tket blk grenoble sok bkn harge
L1	Sbb aq pnye dh expired
L1	Beli sok ar bile smpi aeroport
L2	Ko smpai kol brape agak2 kat klia
L2	Ak xtau nk gerak pkul brape dr hotel
L1	Flight smpi kul
L1	Dlm kul ke ade kt departure hall dh aq rase
L3	Klia ad subway x
L3	Babi laa teriyaki x rasa lg
L1	Aq dh ade kt departure hall
L3	Jay ckp kt mak aku xyah que dulu
L3	Lama pulak nak strap ni
L1	Ayah ko dh call
L1	Takyh ar nk strap tu
L3	Haha strap la jay
L3	Dpt bisnes x
L2	Korg kat mana ni ak otw nk ke klia
L1	Kitrng kat foodcourt
L2	Naik la ak kat check in ni
L1	Check ar dulu
L1	Kitrng dh check in

L2	Tax d'habitation yg kita byar haritu dpt remburse total dlm mggu lgi
L1	Amboi virer sorng skit
L4	Jay nk nंबर tacis de stade
L3	Numericable ni kome nk trus pkai compte aku ke nak tukar
L1	Tkar ar compte bank ke acct nabil
L1	Compte numericable kekal cume compte paiement ko tkar
L3	Nah acc dia
L4	Pic Kuew teow goreng mamak satu
L2	Jay nk order ke tak Taqip
L2	Awkward sbb kenal
L1	Ucapan good morning plg ceria aq pernh dpt
L3	Kau dah gud morning kat tulus minggu pon xpuas lg
L3	Dasarr black dick hahaha
L3	Bil nanti pas no BIC kau kt aku
L3	Salam gaisss Aku nk pkai duit urgent Nk mintak pinjam duit malu pul
L2	Gdf virer kat ak euro so chacun virer
L2	Jay virer kat ak ak tolak duit mkn flex aritu
L3	Salah colour tu jay Ni kau punya
L1	Agak2 ar
L1	Mcm nigga btul dah ti
L2	Tgh nangis dlm bilik ke tu jay ptut la abis game laju je msuk bilik pint
L3	Hahaha xpe jay Next year ok
L3	Ride along dah keluar bil Shalom
L4	Kau dh tngok ke Ride Along tu mi
L1	Sweater tu dh arwah ar bro haha no more
L1	Sewa fevrier Jay Nabil Matmin Takip
L1	Nabil motif Itak ABDUL MUNIR rue alexander fleming SMDH
L1	Matmin virer kt agence takip virer kt matmin part ko
L1	Nazmi duit tax habitation dh dpt blk nk virer kt ko ke cam mane
L4	Ada org main bola x
L1	Nk rootbeer ke tak aq kt casino
L2	Sewa rumah bulan dan elektrik bln Jay virer kat oralia kat matmin Na
L4	Main bola x
L1	Bgn la dri tido tu kalian

L2	Kena bayar lebih sewa yg tertunggak tu plus denda lmbt bayar plus
L2	Rib oralia lepas ni bayar kat oralia
L3	Kt forum ke cha
L1	Tak bro
L1	Abis kul pm pastu aq g
L3	Sekali lah
L1	Orait jumpe kt gab faure nnti
L1	Bro mane ko
L1	Whatsapp official appart
L1	Aq tak bwk kunci otw blk
L1	Otw balik
L4	Xde la
L2	Chicken alps que dites vous
L2	Ak kat caf ni
L2	Sejuk + lapar
L3	Mana Aku oder dlu lah
L1	Mai umah dulu lah
L1	Mai ar cpat
L3	Aku kt sini la gila
L1	Away g awal sgt
L3	Nak balik dah Nk tapau ke x ni
L1	Shit sial ko
L1	Order dulu ar tuk kite
L3	Mcm ckp dgn india hmmm
L4	Bak pic menu skit
L2	Jay memang india la igt xnk bgtau sape2 tapi dah keliru ni
L1	Alps master algerien tacos red tikka nuggets bbq alg
L3	Brp total ni
L2	Panggilan bola kul pdg futsal
L1	Panggilan mkn malam kul pm di rumah hafiz pada
L2	Labu ajk pergi umah dia kalau nk mkn daging panggang sbb xjd bbq
L1	Aq kt casino ade nk pesan ape2
L3	Nak battry AA dgn poivre
L1	Battery aa yg kecil kan

L3	Dgn tomato skit
L3	Aa je Azwan ali
L1	Ko nk brape battery
L3	Nak french kopi vanilla latte creme leger yoplait satu
L1	Dh Imbt ar bro
L1	Skrng tgh shopping
L1	Rugby ade kt tf1 tak
L3	Ada klau ada vanilla latte pumpkin spice
L1	Syafiq sudi msk mlm nie kene deliver ayam je nnti
L3	The de coco
L1	Tak g bubble tea arr
L3	Aku letak siap2 formule
L1	Formule ape lak
L2	The perles sauf pearl noir
L2	Ak dah virer kat ko
L1	Sorry bros kami dh otw blk
L2	Tapi ak dah virer
L1	Aq virer blk no worries
L4	Dh siap masak ke J
L1	Code entree
L1	Mai ar cpat
L1	Syamim n takip dh dtg
L1	Tacos nugets chili
L1	Tacos red nugets chilli
L3	Sapa berminat nak berlakon filem tamil ngan superstar ranjini SALA
L2	Dalam kedai buku lg eh
L4	Datang la bilik
L4	Ke mana Gare
L4	Kau kt ne
L1	Loyer fr n fr Nabil byr sorang
L4	Teh noir mangue froid sans tapioca benda lain letup2
L4	Nk main
L1	Register name kt fb tu
L2	Bayaran gdf Chacun virer euro kat ak

L5	Apa pucat burger kau ni
L1	Asyik gbr mkn je sharing ar gbr janda wicet tu haha
L1	N no secu sociale ko tu jgn lupe
L2	Biasa ak tggk melayu je rempiwxxxt ini rare
L1	Nami yg eja camtu aq follow je
L3	Magreb ni cha Bawak2 ambik wudhu
L1	Match mohamed kul h
L5	Ada sapa leh ganti aku masak x mlm ni Aku tgh wat projet lambat ko
L5	X yah la aku balik masak Td ni gaya mcm nak balik lmbat
L3	Bereh Sata skali la
L2	Pegi tgganu mkn kepok celop tepung ibarat bdk junior tours dtg gren
L3	Opening je min
L3	Min kenduri tganu mmg smpi ptg ke doh
L3	X leh blah sial
L3	Utk sek2 yg pergi paris taon lepas yg x tahan panas dunia
L2	Asal kena byr fidyah Xleh ganti biasa ke Ahhaha
L1	Siap bgi exemples
L2	Blik blik mlysia je la ko tu xpyah nk sibuk hal org kat france
L1	Aq tgh caring psl mmember2 ku yg tersayang kot
L2	Ak dah virer kat agence Kepada bro bro yg lain anda tahu anda siapa
L4	Dh virer
L1	Next ko bil
L2	Mcm da rosak je
L1	Bkn aq punye ar
L4	Bil internet jay taqip matmin virer kat aku
L1	Guna la
L1	Ehh btul ke guna anta anta tu aq ke ko Lupe plk
L4	Ana means saya
L4	Anta means awak
L1	Ko bile nk dtg Bju ski ko smua ade kt aq
L1	Aq virer kt ko dh
L1	Aq virer kt matmin
L1	Min virer kat nabil nnti
L2	Virer apa

L1	Bkk compte air so nami jay n matmin je
L1	virer kat aq
L1	Nami numericable brape
L1	Best fren ko menari piring
L2	Salam jay thé rouge coco + manque tout sauf tapioca ak dah virer eu
L1	Sorry ar takg bubble tea
L1	Tgh lyan concert jazz
L4	Happy train
L1	Sewa bulan Virer kat agence matmin nami kat matmin jay nami
L5	Ada casing alcatel
L4	Movie kul
L2	Mcm xsempat je td igt nk pegi tp dah dgr tram lalu mls ar rakam je la
L1	Kip gune ar blik aq nnti teman abg ko tido
L1	Blih buat bro's pillow talk mlm2
L4	Min kat rumah ada telur n ayam x
L2	Sorry bil ak tidur td
L2	Nasib baik dapur lawa xde la malu sgt msuk fb
L5	Tram kol min skali
L1	Qip pinjam pc jap nk main game gta
L1	Birthday ko bila
L1	Boleh main dari awl xni
L5	Main je
L5	Aku dah abis main
L1	Ok ak buat nouvelle mission lah
L1	Story mode eh
L2	Ak kat caf ngan nabil ni
L2	Kitorg tunggu kat cesar je la
L2	Sedih nya jgn nangis2 lak kat airport malu kat org
L2	Pegi la eiffel
L4	Sory doe taqip
L4	Nnti ak blnje tacos
L3	Thanks guys Halalkan makan minum Mintak maaf klau gurau trkasar
L1	Good luck bro nnti jumpe kat gemas nk gk g konsert merdeka gemas
L1	Jgn lupe passing no MY ko

L2	Ko virer sndiri la kat nabil yg ko dah virer kat ak tu kira burn je la
L1	Aq takde rib nabil
L2	Salah ko la xmntak
L1	N klau aq add die anta msg kt no france
L2	Ko virer baru je la kat nabil
L5	Aku virer kat kau gak min kau virer kat nabil
L2	Ak dah virer kat nabil
L1	Bkl millionaire tu
L2	Bil piscine jom
L2	K jumpa kol kat tamu
L1	A macha drinking cha-time
L5	Org kat umah aku trtinggal kertas algo kat atas meja jay Leh ambik g
L5	Ambik gambar pages dia bkn ambik gamba kertas tu atas meja
L2	Nk ak scan hntar ke
L5	Scan pun boleh
L5	Smua page min
L2	Dah scan
L2	Ak hntar kat email
L4	Jay aku dh tukar RIB tau kat numericable
L1	Kite bestfriends zaim
L2	Ko follow x jay
L2	Tkot ak sorg je yg xfollow
L1	Aq xfollow zaim legacy
L1	Aq follow insta perso die je miazocool
L3	Dah la guna aq
L2	Wat camnerxx matpit taman tas
L1	Bhse sms standard ar
L4	Ada org main bola x
L2	Ak bru baca kat fb
L2	Rasanya baru start main xpon xmain lg tu
L1	Tito cant touch tis
L1	Skg matmin's touch
L1	Anta video entry yeng skli biar terubat rindu die
L2	Susu kat bhgian mana kat casino

L1	Tis is hw a macha eats
L5	Uish kuat nye anta jay
L1	Bile nk turun kuantan nami Blh banana leave skli
L3	Dtg kuantan kne la readykan kumalasari
L1	Tu ko ejas sndri ar dri gemas
L1	Msg ar awek2 wechat ko
L4	Kemon min kau lak next
L4	RDV kat OU la
L2	Misi div xtercapai lg ni
L4	Dh sampai dh kau tngok game warriors x DAmN curry
L2	Ak dah cuba x tido smlm last2 kol lebih tido gak bgun2 da 4th qtr
L2	Kalau nk tgk curry pergi kat channel youtube ak
L2	Cha mana cha time spurs bantai manU laju je watsapp ak ni lah bala
L2	Babi best gila game
L1	Bangang freekick plk
L1	Video utube ko tak kene remove lgi min
L1	Brape view highest
L2	Sakho power siak mlm ni
L1	Nk fan liverpool min Skg mu mcm kesian je ak tgk haha
L3	Ajak la syamim Dia nk jgk dating kt geant
L1	Tu ar smpi ati korng lupekn ex hsemate sedih syamim
L1	Cam biase ar tacos nugget chili thai dgn chicken burger
L5	Call la jay
L1	Slalu ak call die tak jwb
L1	Klau ko dgn min call bru die jwb
L1	Die tak suke macha call die nk yg melayu gk call
L1	Min tak g main frisbee ke
L2	Main basket xcukup hipster nk main frisbee
L2	Jay nami hntar attestation de bourse kat ak ak nk hntar declaration r
L2	Hntar kat email ak
L2	Passport skali
L2	Montant bourse ko eh
L3	Lama dh xdpt duit euro min
L2	Bg numero securite sociale

L2	Ak tgk process declaration revenu en ligne ni xtau asal ko punya lain
L2	Dah siap suma dpt montant impot kosong mcm thun lps tp ak xtau a
L3	Doc mne plk tu Xpela jnji
L4	Min kau landing KLiA bila
L2	Kalau ko nk ambik brg time tu boleh la sbb ak stay airpot je mlm tu
L4	Macam lagi senang amik kt airport Bleh aku lepak sekali kt hotel kau
L2	ak xtau duk hotel ke tak
L2	Mcm duk kat klia je mlm tu
L2	Bajet kurg skrg
L2	Dulu boleh la arab money
L4	Hahaha orait2 Nnt aku usha dulu macam mana n gitau lain kali
L2	Update saint martin Husni dah tido sini
L3	Update skit peti ais dgn sofa
L3	Tgok civil war weh Baik
L4	Jay cuba check bank kau Ada prelevement x dari numericable
L1	Smlm takde lgi nnti ak cek
L4	Yg numericable tu aku baru check Nnt hb dia buat prelevement kat a
L4	Sesiapa nnt lepas hb ke try la modify balik account kat website dia
L2	Surat elektrik da smpai
L2	Husni xmsuk lg time tu
L4	Kira la aku skali bahagi
L2	Ko boleh buat virement lg ke
L4	Aku rasa aku boleh lagi virer ke account sesiapa yg tlg bayar lebih
L4	Sebab aku se conecter kat website dia ptg tadi ok je lagi
L2	bahagi so nabil taqip jay virer kat ak
L2	Yg ni mntk tlg virer cpt skit lah kalu xmati ak dia tolak nnti
L2	Fuck mon gouvernement malaisien
L2	Lps ni msk guna kayu api je la
L1	Ak dh virer
L1	Bil gaz yg mahal
L1	Dh asal ko kabo msk gune kayu api stove gune gaz
L3	Keje la part time sana Jaga bubel tea
L1	Die dh nk blk MY week lgi bro tak gune g keje skrg
L2	skrg bubble tea da boikot mlyisian sejak2 tumpah air luar kedai aritu

L2	Nk tea rouge pon da susah skrg tea noir je
L1	Insiden bubble tea disepak kt dpn kedai die
L3	kau xde kad fidelite de jeremy
L1	Nabil trauma lpas tu
L4	Dh virer tau min
L1	Ape mksd name tu
L1	Asal mcm name kelantan
L2	Abe uji ko xknl alahai susah la mcm ni
L2	Tu la dulu ajk pegi perhentian xnak razin ajak pegi redang ikut pulak
L1	Sial la
L1	Anep ajk ak okay bkn best fren ko
L2	dah heppy2 gurau2 mnja kat redang xnk mengaku pulak da
L2	Ko punya virement ak da dpt jay par contre nabil punya xde à venir p
L2	Ak rasa account ko da bloqué kot
L1	Weyh mskkn la husni tu dlm group ni gk
L4	Sebab aku contact agence by mail dossier aku dia traitez may
L1	Bah parce que parce que
L2	Bil Nba playoffs ada kat astro weh astro supersport
L1	Pagi kul am biasenye
L4	xde sape nk tngok kat rumah so stop subscription
L1	Btw welcome home min
L1	Ko buat ape doe kt umah skrg
L4	jay cuba try contact admin pasal inscription kau
L4	matmin cakap dia ada terima confirmation email tapi kau belum
L2	Aah xigt ak cek td ada live lak dah la english hd pulak tu dtg umah ak
L2	Tp ak xde internet kat umah skrg pakai celcom E je
L2	Siot la
L1	Ohh tu dh settle kot
L1	Sbb bile nk try deposer bru die tak bgi
L1	Klau takpun aq buat reinscription je la bln nnti
L1	Bile ko grak g aussie
L4	xsempat ar kelas start jun ni
L1	Ko blja mane pasnie Mmu Ke tukar course economy
L1	Min ak bru dpt emel confirmation masters tu

L2	Oh ye ke ak tgh drive smlm xigt nk reply alhamdulillah jay
L2	Duk umah baca buku papertown je
L1	Hobby bru skrg
L2	Weh europa league final smlm ke ?????? fuckkk
L1	Tgk highlight je ko
L2	Baru tgk result kat livescore hheheeeeeee
L1	Senyap arrr
L2	Tp liverpool 1st yg score smlm so bg ak liverpool gak menang sevilla
L1	Takyh la nk cover2 klau nk kutuk tu
L2	saje je nk usik2 fan liverpool jgn sedih kita sokong manU msuk final f
L1	Yg die forward emel akmal kt die tu
L4	nk mintak tolong trnya mdme Baras or mr Cholet tu pasal syllabus RT
L4	RT sekali untuk nami
L4	untuk transfer kredit
L5	Aku gi masa daftar nnti la kot Segan lak nak jumpa dia dia hantar em
L5	Tengok la kalau jumpa
L3	Please qip please Tayang tetat tetat
L5	Ntah ada prof x sbb skrg bkn iut buat stage
L4	kalau boleh nk gak setelkan cepat2 aku pun xtau sangat camne proc
L5	Kau hantar tu pakai email biasa ke email ujf
L5	Biasa prof x baca sbb email baisa akan masuk trash x pun spam
L4	aku hntar pakai email biasa
L3	Ujf email hidup lg ke Ak nak login pon xdpt dah
L4	aku hntar email biasa reply je vedel
L5	Try hantar pakai email ujf Tgok prof gak mcm baras nak urus ramai st
L5	Forward la mail kau ke aku nami lut skrg dah kosong ntah camna na
L4	bak email qip jaggi aku forwardkn
L4	dh send qip
L2	Selfie dgn husni la satu
L2	Ak ada byk gmbar time gemok pon ada harga boleh nego
L5	jam bergayut nnti aku selfie dgn hasil dia
L1	Siot crop gmbr
L1	Nie bkn gmbr full
L5	Nak highlight exhousemate

L5	Apa ni crop
L5	Nak full
L2	Hensem sgt dgn misai pornstar di
L1	Mmbe ak ambik mase nk send off
L4	qip baras or chollet ada reply x email kau
L5	X reply
L4	western Conference champion
L2	Power weh comeback
L4	xdapat nk tngok la tapi comeback 2nd half gempak gila comeback d
L2	Tetiba dia bgtau hebat sungguh nama Curry ni
L3	Nak dtg ktn la gitu
L1	Ko trun ar dulu
L3	Ayuh lah ayuh
L1	Ko fulus msk sini dua org fulus tk msk bro
L1	Ko nk g weekend nie ???
L3	Wiken ni lah Mnggu depan puasa
L1	Ak tak smpt ganti ar bro
L1	Byr fidyah je la nnti
L2	Cibai bayar fidyah
L3	Dgn macha Jap tgh ajak nabil
L3	Helmi xdpt la Dia busy jaga anak
L1	Macha mane ko bwk nie
L2	Baru duk sbulan kuantan dah xngaku cha ke jay
L3	Baru bayar fidyah xnk jd macha dah
L3	Pergi buat roti naan sana
L1	Nami kabo macha je mane tau die jumpe macha bru kt mmu tu
L1	Btw macha nie tak kabo pun blh join ke tak weekend nie
L3	Gedik r chaa Dh hbis intern kot
L2	Tp korg kena stay hotel la kalau dtg sini xleh ajak dtg umah ak umah
L2	Kalau boleh stay primula ke boleh ak tumpang mandi kolam hahaha
L3	Bawak kemah laa
L1	Weekend nie ak dh ade plan arr
L3	Hmm mcm tu lah diaa
L1	Bkn ar

L1	Nie bdk2 smart
L3	Bdak smart racist doh jay
L2	Xde la nk sgt ak nk bg baju ko ni
L2	Kalau pegi bukit kluang mcm best gak randonné
L3	Kemon la jayy
L1	Ak takleh join ar bro weekend nie
L1	Dh plan lame dgn bdk2 nie
L2	Dah la jay pegi la dgn bdk2 smart ak tahu ko malu nk lepak dgn bdk
L1	Hahaha takyh nk emo sgt
L2	Jgn sesekali berpaling setelah kau pergi membawa diri bersama sma
L2	Igt nk ajk ko pegi kl pegi atas pasar tp xjd la
L3	Dan isa dgn capiq balik x
L3	G perlis la jom
L1	Tpi tgh intern kt pekan tu
L1	Klau g mase cuti maybe bdk tu blh join
L3	Klau kau xde jay malas la nk dtg ktn esok
L2	Tapi ahad ak xfree sgt kena pegi ambik adik ak blik kat airport
L2	Ak buat byk benda dah jay balik ni hmmm
L2	Kau da kena tapis awl2 dgn abg dia
L2	Lps ni boleh buat rapport stage sesama
L2	Screenshot* post kat fb
L1	Tpi klau razin jdi adik ipar ko lgi byk possibilies dripd jdi abg ipar ak
L1	Nami kene compete dgn razin tuk spot bertuah sbgi adik ipar ko
L2	Jilaka mkn escargot
L2	Dan Retweet dia punya tweet
L1	Escargot tu bdk laki blkg yg bgtau kt die
L1	Nk profil fb die ke min
L1	Bale numericable tu ko potong ke transfer ke rib ak
L4	cuba try check kat website dia
L1	Sbb takde pun die ptg kt acct ak
L1	Cek bgi username n password
L1	Nnti sok ak try tgkkn
L1	Yg numericable tu die tak tkur lgi rib
L1	Still rib ko

L1	Tgh try buat skrg
L4	sebab tu kot bank aku ni xsettle2 lagi
L4	takip prof2 tu still xreply eh email kau
L5	Tak diorg cuti kot Aku x pegi cari lg bizi boh projek Ni nak ratttrapage
L5	Minggu dpn la aku pegi free skit
L3	Selamat berbuka gais
L4	qip ada pape update x
L4	try gi jumpa M Fayolle ke
L5	Aku gi iut td dah kosong abis nabil Semua dah start cuti
L4	ye ke xpe la thanx qip
L5	Minggu lpas aku rattprg malas lak nak gi Sbb hb tu last prof ada kot
L5	Aku pun x tau asal diorg x baca mail Try call la pulak
L3	Qip ko xde simpan syllabus lama ke
L5	Aku ada simpan tp rt je
L5	Yg aku sempat download la
L5	ada scan semua x silap
L3	Zizi ada eh
L3	Dia letak dlm web dia kan
L5	X tau wujud lg x web dia Mintak kat dia ada lg kot
L2	Salems freros selamat hari raya passez tres une tres tres bonne soire
L5	Jay min nami bg email no fon alamat rumah parents Nak isi borang t
L5	Smua no fon ws ni kan
L2	Aik borang tax Rumah sapa Ak dah buat declaration aritu org dpt ko
L5	Isi rumah pieces smua
L3	Fakof taman tas
L1	Buat cam still under bourse je
L5	So tu biar kosong la
L1	Alamat google street takleh g
L5	X yah la alamat msia tu
L5	Alamat x yah email ko je nami
L2	Dah alamat rumah parents parent ak kat ganu ko st etienne lain la
L5	No fon ko aku isi yg jay
L5	Yg mohon crous tu boleh hantar lg kan
L1	Anta ape weyh

L2	Mcm xduk crous je tahun dpn dah kita dah org ok la tu kn
L5	Kita x hantar lg kan surat uni
L5	Hantar je mana tau ada prob aku kena balik ke before bln ni
L2	Nnti ak cek kena hntar kat alamat mana
L2	Ada dlm email
L3	Xyah balik la qip
L3	Esaimen Tapi awek ngamm
L5	Ko dah start kelas ke Tu la masalahnye x de myvi
L5	Ada saga aerobek je
L1	Nie alamat crous tu qib
L3	Saga pon ngam janji ada turbo skit Aku sem dah qip
L5	Kahkah Ada brapa sem lg Ko blaja bnde sama ke lg expert
L1	nie numero INE aku gune tuk reference
L5	Forward ke aku email tu Sbb dia kena antar skli fichier notif
L1	Fichier notif
L2	Nnti ak forwardkan yg ak punya skali
L5	Last email ko dpt kan dia ada bg pdf satu tu notif
L5	Antar la bila free
L2	rindu ke weh
L4	rindu gak la masakan nasi ayam kau tu
L2	Busy x
L2	Aah ko xde fb
L4	bleh tahan la busy dia
L4	xde transport je yg susah tu
L2	Mcm mana ni labu diam2 dah nk kawen da tu
L2	ak lek lu baru fasa keluar mkn kedai hipster
L4	kau jay balik sana bile lepak sekali lagi sebelum korang chow
L4	kat eh majlis tu
L4	nnt dapat alamat roger aku lagi nnt aku usha jadual
L4	hahaha adiktiri money maywheater ni
L5	Apa password dgn username utk configurer wifi
L1	Haha bru nk type tnye soaln ko
L1	Ce try yg nie kip
L5	X leh reset je la

L1	reset je
L1	Haha ak dh tgg bro
L3	Mmg la
L1	Ko bile nk ambik brg2 ski ko
L4	kim salam wei n cakap tahniah kat labu
L1	Nami ko dtg tak weyh
L3	Selpie sikit
L1	Ye la bro
L1	Dlm kul smpi doh
L3	Sassenage dlu la gais
L3	Isa samad dgn celan x join
L3	Send location sikit
L3	Gambar atas tv tu yg mahal
L3	Penat putcal
L3	Rice cooker aku kne jual ehh siot jay
L2	Awek nk beli abg jay hntar dpn rumah siap tolong pasang bg tips ma
L3	Dey chaa sunat dulu laa
L1	Siot min Die yg buat promo tu siap ltak condition mcam uqasha
L1	Salam penghuni penghuni alexfleming dan alumni alumni penghuni s
L1	Mane prnh bayar bro
L1	Nie first time dpt bil air
L3	gadai la brg2 ak yg tinggal buat byr
L3	Klau xdo ermm xdpek den nak nolong ehh
L1	ak dah buat kira2 sorg kena bayar matmin nami nabil Kalau korang b
L4	nk bank acc cimb qip
L2	gadai apa weh
L2	Baju ski dah letak warga grenoble
L1	Thanks bil
L3	Baru lepas tggok koh lanta ke guana abey semua
L1	Koh lanta kt boh
L1	Epreuve same je
L3	Kirim salam laa
L1	Hahaha Takde ar philip tu legend
L3	X main wechat dah jay

L3	Nanti laa ak carik yg ada serbuk2 kari sikit
L2	Xkari pon takpe dpt yg pakai sweater hijau dah mntap dah
L3	Sweater hijau xde Tuala carlsberg je available
L1	Ko prnh güne tuale tiger gak
L1	Rindu casino
L3	Bosan ahh chamechaude
L3	Damm jay org sarawak jgk rupanya
L1	Baru discover origin ak bro
L1	Setan Semangat buat research nmpk
L3	Bau weed special min
L1	Rakam match mane tau kitrng muncul
L5	Buat la spanduk tulis sack
L1	Tak kenal doh ke org yg snggup ambik ko kul pagi kt kuantan tu
L1	Bukak full brightness phone tu
L1	Haha Lupe nk buat sign tuk omar
L2	Gentel pedih laa nnti tertanggal daki2 kat bwh tu
L4	kalau kumpul daki boleh buat pemadam stabilo kaler hitam tu
L4	bahaya min kang buat kawan baru pulak kat swimming pool tu
L4	ce quoi medicine
L3	Ni mandi kat mana doh
L3	Haha adehh terobek betul laa
L1	Kat nice
L2	MasyaAllah brader ni chapter forthi thu verse twentieth eight
L2	Veda chapter four verse three
L3	Aik dinner sanmatandeh je jay X balik amien ke
L1	Amiens dh takde org boh
L1	Stay kt smh je la
L3	Berehh selamat hari raya boh
L1	Mekasih bro
L3	Hahaha taik r jay
L2	Mkn daging black pepper laa
L3	Masak la min
L1	Haha Tak hairan ar kfc Tacos nugget chili thai lgi mantap
L3	Fakk tacos rindu sial

L2	Flex laa mlm nib
L3	Hahaha jahat kau jay Merangkak la kt mahsyar nk mtk maaf nnti
L1	Ak tkg ambik gmbo je idea dgn tulisan org lain boh
L3	Aishh tv besorr haha
L2	Fuuuuuuuuckkkkkk suppooooooort laaa bodoh fuck fuck fuckkk la
L4	team rohana lawan team shahir
L2	Cuti datang shopping casino beli zakia
L3	Uishh jay Mentang2 BBC haha
L2	Big Black Cha
L4	ladder match
L2	Alliance kena betul
L1	Lagu wwe lame dh tak main
L4	apa citer wei kat sana rindu sial rumah alexandre fleming
L1	Tv still selamat
L4	main bola lagi x kt sana pepagi weekend
L1	Pagi susah skit ar mlm jumaat biase main kat tailles
L4	dh xmain badminton ke malam jumaat
L1	Tak ak dgn min ambik sport noté badmintn so main ptg isnin
L4	jay tolong gi beli shit kt notre dame
L1	Kedai kt notre dame dh ttup boh
L1	Notre dame ade lgi kot
L4	damn ada kat situ eh kasut futsal aku hahaha xprasan pun
L3	Serius laa
L3	Haha apo isi eh tu
L4	serius la
L3	Eh asal ak rply kt sini haha keliru sial
L3	Wtf hijau Kene bgtau bukit aman ni
L2	Beli sayur pon on9
L1	Sayur special min
L1	Ko ar tak prnh
L1	Bila nk buat match rohaini vs akmal ni
L2	Nice letak dlm frame
L2	Resit courir ada lg ke tak tu
L3	Fak teringin pain bagnat siak

L2	Bagnat apa do
L3	Shiall ah jay
L3	Dap uhh
L1	Sok ak ambik gmbur pain bagnat yg sebenar khas tuk ko
L3	Sos piquant jgn lupa
L2	Mcm nipples jay je
L1	Aku punye tajam lgi agak2 arr
L3	Cepat la jay belanjaaa
L3	Perhh ni mahal siak Yg hensem je dapat kakkah
L4	tu macam daun sup je xpun salad
L3	Ni kalau pekena x siuman terus bai
L4	hati2 min kalau shoot ni sampai esok pagi kentot bau weed
L2	Power barang ni taqip tido tersenyum smlm
L2	Potong2 dulu qip nnti kluar beli feuille roulant jap
L4	ni barang bara ni min kick dia xpower sngt letak rokok skit je time ch
L2	Ada lg botol caraf d'eau
L2	Lain skit sebut Allah imam tours Allahu
L4	malam ni lepak Pathé la jom xtngok lagi John Wick
L2	Bil tgganu dah ada ciné
L4	gi travel jepun ke mi
L3	Jepun boh carik sasuke
L5	Cari miyabi lama x jumpa
L4	tolong bawak balik skateboard min kalau boleh la
L4	pastu bong kaca cm satu pls
L3	Kedai hipster dpn bubble tea xde kasut baru ke
L2	Nnti nk blik ak try sumbat dlm luggage
L2	Kedai hispter destockage hari tu mcm nk tutup je
L2	Seat keras mcm batu siakk
L2	Selesa lg bas sklh smart
L2	Td couple india dpn ak boleh nyanyi mcm dlm movie
L2	Tp teruk sgt seat ni seat kancil ak selesa lg
L2	Eh pfc kat stalingard kn
L5	aritu aku pegi kol h x de biryani
L2	Kedai cina poulet caramel arret cansier desbeton kn

L5	Aku turun gare austerlitz jz
L4	bawak awek jalan2 paris ke min
L2	Lejen ni mana gmba pegang ksut ahaha
L4	kasut tu pun dh masuk kotak kaca buat display je min
L2	Nice gmba skit
L4	gantung display camtu je
L4	koh lanta malam ni jngn lupa goreng kropok lekor bnyak skit min
L2	Dah lama tinggal kohlanta
L4	ye ke apa gaya hidup kat rumah time weekend skang ni xkan artist y
L2	Bola pon tak main dah
L2	start dgn kurma dulu
L2	Baru plat principal
L4	hahaha azan kat youtube kena pasang depan tv
L1	Pas azan layan boboiboy skrg tak tgk doraemon dgn shin chan dah
L3	Orang nak rayo dah ni poit
L3	Bawak ole2 bong satu dgn the rouge mangue
L2	Ada thé rouge manque tous sauf tapioca the noir satu lg
L3	Thé noir curry balik sekali kaa
L2	Skali laaa
L2	Dah hbis stock kicap jalen
L2	Plan la satu trip
L3	Tadah peluh dia dpat ahh kicap botol hahah
L3	Trip tioman laa ngidam nak poie sano
L2	Cari date korg cuti laa
L2	Ada yg smbut ak tak ni kat airport tune hotel dah booking
L3	Hahaha kau esok smpi baru nk roger
L3	Guano boh
L3	Aku kijo xdpt ahh maaf
L3	Jual baju rayo boh
L3	Woi chaaa xyah la acah sweet
L1	Dengki ar tu
L2	Cha sedar diri xmcm ken
L2	Cari chienne
L3	Carik la melayu jay

L1	Ko dh bahan aku kat insta kat whatsapp bahan lgi
L3	Congrats ahh
L2	awal lg nnti skali la semua jd dulang boys abé uji duk depan
L1	Amboiii double standard
L1	Aku kene kutuk min ko kabo congrats
L1	Nmpk racist di situ
L3	Ntah nye kau mic jayy
L3	Ktorng umno
L5	m nasir skang Mat ri ye smua
L2	Nk guna ambik ar
L5	Aku guna eh beg kcik aku dah antar msia
L2	Guna lah
L5	Buat haji iAllah
L5	Dah lama da aku prasan
L5	Dia suruh contact agence antar plumber
L2	Ko boleh tlg pergi agence bgtau
L5	Antar email je la x leh
L5	Msti nak kena ada assurance d'hab
L1	Assurance kite ade takpe
L5	Call ke
L2	Ko ada no jay
L2	Cuba cek kat surat loyer
L1	Takde name
L2	Qip mohamed nk dtg umah ambik set bbq
L2	Qip tlg cari no bank ak utk compte online ada dlm file bnp kat rak pu
L4	Min vs Mcgregor August ni
L2	Press conference pon dah mcm apa
L4	tu la mcgregor trash talk bapak mayweather pun lawak gak
L4	pastu bapak dia start tunjuk boxing skill dia
L2	you are a fan floyd ahaha
L4	he's in a tracksuit hahaha
L2	Yg hari ni punya press da bomb jgak tu
L2	Tgk bil dance for me boy dance for me weh ak mgkin pegi kl hjung m
L4	weekend aku ada la kt KL tu

L2	Nnti ak roger balik kalau jd
L3	ni first year ke qip Salji atas kepala ada laa
L2	Ni hari last tu kn
L2	Lejen ni
L2	Kalau ada vdeo hari first yg sesat tu lg lejen
L3	Kalau ad video aku excident dkt curam baek gak
L2	Ambil la bas dtg tgganu
L2	Zmn usrah ni
L5	bnyk gamba zmn ni aku jumpa lam pc umah aku
L5	dtg la ambik aku
L2	Dtg ambik pulak da jom pegi cherating laa
L3	Dtg la jalan sini
L5	wat mende pegi taman negara org bandar ye la pegi sana
L3	je qip Dtg rogerr
L3	Acah ah minn
L2	Ni zmn baru lps beli controller ni mkn maggi sampai xkluar bilik
L5	aku tgok quran tu berabuk teghok lam almari
L2	Ak letak dekat ngan katil ko qip sbnarnya nk bg hint nk suh ko baca
L2	Tgk tafsir ko pon byk note kaler kaler tu
L2	Bila last bukak tu
L5	alhamdulillah mnggu lepas baca yasin kat surau
L2	Tp xleh lwn laa blakon time rdh
L2	J'en ai marre punchline power
L5	punchline mnde tuh
L2	First year dah klako
L3	Cibai berlakon rdh
L3	Xde video ke
L2	Ada dlm harddisk ak tinggal kat saint martin dheres
L2	Dulu time summer pon dia berubah elok je
L4	xnak plan gi pulau mana2 ke sementara korang ada kt sini ni
L2	Ak On
L5	x de pitih
L1	Aku tak sure lgi Sbb ade plan nk g tioman gk dgn mmber2 lain
L4	fine jay fine sampai sekarang lebih kan keluarga Amien dari keluarga

L2	Korg free bila
L4	try la Jumaat sampai ahad week August ke
L5	player issue ke muroh da tu
L2	Asal anti sgt putih jay
L2	Racist sial
L1	Bengong Jersi home malaysia ah
L2	Gmbar please
L2	Tp byar euro la
L2	Murah la tu
L2	Prrrr cha pakai gshock putih skrg mcm mat rempit kuantan jeee
L1	Tu kogu punye die bg kat aku
L2	Tu mcm ak punya je jay kalau ak tahu ko nk yg tu baik ko beli yg digit
L2	K tak kacau
L2	Bgtau bonjour kat bdk2 ni nnti
L1	Haha Ko lek ar kat masjid kristal tu Aku g lepak kat venice
L1	Aikk bkn cinque terre ke
L1	Oh yeke husni ajk cinque terre
L2	last2 duk umah sahaja
L3	Cuti lamo bwk keluar laa myvi hijau
L1	Aku dh kt dubai nie nami
L3	Adehh rindu pulek
L1	Banjaran tak nmpk ar nie Awan ttup
L3	Aku bkn nk gebang laa
L1	Haha Amboi dh training byk sgt nmpk
L3	Xdehal la jay
L3	Xkan nak bg shahir champion pulak
L3	Rohana kata bak datangg pakai liverpool je Mane score nanti
L2	Pakai liverpool xde coutinho laaa
L1	Min dh join geng mac skrg bil
L4	waa welcome the club la min edit la video skit guna iMovies
L2	Ni legend ni qip xigt dah nama dia
L2	Dah edit dah
L5	x kenal ah
L1	Die yg diornng g interlaken tu qip Canyoning

L1	Beli ar se
L1	Tknk curry
L2	Ak xnak la
L2	Beli la
L2	Tlg beli nata de coco
L1	Air lychee se
L1	Fuck u min
L1	Hotel de ville
L1	Sblh tour perret
L4	nnt makan tacos snap pic lu rindu la pulak tacos tepi stade tu
L2	Dah lama xmkn situ xsedap dah skrg tacos stade cesar jugak yg po
L3	Faris mcm jaafar on cili padi
L3	Xdo kojo ko chaa
L5	Ni mesti bangun tahajjud
L3	Cilaka ada je bahan ehh
L2	Qip jumpa trungggg kat ciné echirolles siakkkk
L2	rindu nya trung kat ko
L2	Nigga je boleh buat camni
L4	hahaha video nigga mans not hot kau dah tngok
L1	Haha Tu la lagu fav min
L4	sumpah fav aku gak
L4	nigga tu sumpah crack gila babi
L4	tapi video baru yg ada dj Khalid tu legend ar
L2	Tu legend
L2	Kena la tgk
L1	Thanks frerot
L3	Nanti dapat hadiah chas
L2	Stoking xde boxer ngan spender je byk kat heater
L1	Blk la Back for good bak kata
L1	Ade je Tpi kitrng tak daftar undi so takleh undi bro
L1	Krg ajo Tgh mode camouflage tu
L3	Nasib laa gigi kaler cerah hmmm kalau takk
L5	thé rouge à la mangue stp frère
L3	Balik la balik

L3	Kita pergi bazaar pulak
L1	Tk smpt ar bazaar
L2	Xsabar nk dgr abang akip takbir mlm ni dulu azan sedap sampai dpt
L3	Raya Grenoble ke embassy
L3	Jangan nangis ² gais pagi raya
L3	Sedap doh leman rendang
L3	Tapi tu semua dunia gaisss kahkah
L3	kopiah kene hensem
L2	Org lain ak tgk okay je cuma jay laa sedih skit ramadan dah hbis
L3	Haaa apa lagi rosakkan lah kami
L2	Haritu dia ada tnya ak air mustakmal ni kalau berhadassah eh
L3	Series laa
L2	Hilmi dah tinggal buku panduan senang laa
L3	apaa eiffel tower yg ituu
L3	Yg ori adoo poie
L3	Cam taik cicak ah jeremsss Selamat hari raya
L4	Selamat Hari Raya semua harap sihat sentiasa mon freres
L1	Siot aku dh improve kot Come on arr
L2	Tmpat lain ni kat cannes aritu kita kat nice
L5	Junior dtg bila
L2	Exact time xtau la
L3	Pehh memang bro habis jaga junior sekarang
L5	Haha Aku naqib usrah skrg
L5	Tgh tggu promotion naik president arif
L2	Taqip yg print table tu bg kat junior
L2	Smlm ak tgk tambah column baru tasbih 300x sblum weekend
L3	Baekk eh francee giroud ngam terox
L2	Yo freros tv nk jual brapa euro
L1	Haha Blake main phone je tak hormat senior
L2	Ak suruh diorg zikir
L1	Amboii Bonding beno dgn
L2	Ni peringkat pertama dlm usrah tgk diorg main game dulu utk menjali
L3	jay kau boleh masuk doh
L1	Tak lpas nk tgk video kene block

L5	Jay brapa lama kena rendam kueue teow
L1	Entah Smpi teow tu lembik ar
L2	nami bg ko punya no account ak nk tranfer duit jual brg brg rumah
L2	masyukkk laaa euro ni
L4	min flight bila kalau timing ok nk gi sambut kt klia
L2	nami no account jgn lupa euro en attente ni xtermasuk tv + xbox dua
L4	min sbtu ni xleh jumpa la ada event lak
L2	event shuffle ke bil
L1	Lawan ravers dpan klcc tu
L2	xpe la nnti ak pegi KL
L2	dia tnya possible ke tak jumpa korg
L2	kalau korg free boleh lpak mlm
L3	Apa hari weekday
L3	Tv xbox aduhh kenangann
L3	Gambar nice jangan jual
L3	Masuk muzium
L2	sebata bai
L2	nnti dah jual tv boleh angkat sebata lg
L3	Aduihh view
L2	bata ke dua yeayyyhhhh
L3	Lelong ah tepi rumah
L2	yard sale laaa
L3	Betul lah macam amerika
L3	Yg depan tu sean kingston kan
L2	mi ak dah virer duit tv euro RM
L3	Orait Thanks min Subuh jangan tinggal
L1	Haha Subuh last kat france
L5	ah tepat mlm letrik x de
L5	Sampai dah Faris letak kunci rumah dia lam boîte aux lettres
L5	Lamappa pegi nice
L2	tp mlm ni ko tido je la dulu kot kat umah kita
L2	esok la cungkil
L5	X de letrik boh gila babi bosan
L5	Batterie phone ni pun nak mati

L1	G ar umah zamry tu
L5	Bak no
L2	gi umah dia laa
L2	sleeping bag ada kn
L5	Aku dah call dia td mlm ni aku tido sana la kot
L5	Tp aku nak try ambik kunci tu dulu
L2	contact dia je laa esok kalau camtu
L2	tp mesti chiant duk serumah dgn dia
L5	Ckp faut la frapper fois puis ça marchera bien
L4	frapper bien sur la tête pour entrant la mode 3D
L2	tgk ar dia ckp apa
L2	sbb sblum ni kita xpernah ada problem pon télécommande
L5	Ada baru2 ni mmg jd camtu batterie nak abis
L2	sbb bateri la kan
L2	sbb hari yg kita jual tu ak bukak ok je
L2	ak on tv ok je
L5	tu jual kat junior x
L2	brg dapur pinggan periuk kalau ko nk ambik he
L2	yg xnak je bg junior
L5	aku ambik ah
L1	Haha Hsemate dgn lamappa siakkkk
L1	Gntng ar kt bilik ko nnti
L2	haram tinggal kat junior
L5	Kunci rumah aku x bg la agence
L2	bg la bila2 pon
L2	ko tau xagence kat mana
L5	Rumah faris tingkat brp ek
L5	Fuck aku punya headphone x jumpa
L2	ko ada headphone ke
L2	kotak beats ak je ak buang tu pon dah cek xde pape kat dlm
L2	tp server asia lah
L1	Last bile ko gune
L2	last ak nmpk dlm peti sebelah ayam
L5	Last aku letak lam kotak beats pastu aku letak atas sofa

L2	mais non en vrai ak rasa ada dlm kotak kat umah faris
L1	Tu la ade kt situ je qip cari lok2
L2	gaya mak mak *
L5	X de dah kali cek kotak ni
L2	dlm bilik husni xde gak eh ko xbwk pegi kerja kn kot2 tertinggal dlm
L1	Last aku nmpk kt hall atas meja depan sofa lpas tu aku igtkn ko dh si
L5	Nak gi cari kat rumah ah
L1	Gi ar cari
L2	ak rasa dlm luggage ko
L2	dia yg still nk pegi
L2	lain kali ar
L3	Kau ada buat transfer lagi ke
L3	Dapat msg ni
L2	ak transfer duit dua kali je
L2	duit tv duit brg rumah
L1	Asal tutup muke last tu
L3	Ajak membe kau pergi ah jom
L1	Kitrng tak lpas ar bro plan dh penuh
L5	Dtg la jerantut skrg tahun melawat jerantut
L5	Bwk la makan ikan patin
L5	X pakai dah kasut adidas bling2
L2	pedas boh
L2	mkn burger pon xhbis sbb pdas
L5	Haha agak ah Mohamed x pnah mkn mcdo daging
L5	Mkn burger ramly tepi jln ah sedak
L5	Aku balik ni x de org nak sambut ke Hmm
L5	Balik terus métropole jerantut
L3	tak mampu qip adidas Asadi je lah sekarang
L3	Balik ahh boh
L2	letak note laa
L2	xpon bgtau agence
L2	the fuck smgg
L5	Aku dah nganjing dah x jd apa Dia lak keje kat agence tu
L2	babi yg kerja agence lak camtu

L5	Aku ingat agence rumah bukak hari ni Aku x hantar lg kunci rumah
L1	Flight ko bkn sok ke
L1	Tnggl je la kunci tu kt zamry
L1	Nnti kitrng bgtau junior
L1	Agence pun tak call dgn anta mail tanye
L5	Bg zamry ah gitu
L5	Tiket flight aritu korg check in on9 x
L5	X kan kat guichet tu je
L1	Check in kt sane je
L5	Junior balik bila
L1	Entah la Dlm kot
L3	Jay membe kau pergi jugak ke bali
L3	Sana gegor lagi boh
L1	Entoh la mi
L1	Drng nie kitrng dh ckp tapi drng cam tak respons pun
L2	seb gagap2 skit td kat ilaha”
L2	mi postcard natalia aritu dah hntar ke
L3	Jangan khuatir dong
L3	Kirim bisou satu eh kat dia
L1	Mai bgi sndri arr tak main ar kirim2 skrg
L3	Doce au pegang owang puan ltew akut
L3	Pakofffff esok ada test wakluuu bisous padiaa
L1	Good luck test sok
L1	Tu ar Dkt klia pulak
L1	Ape rahsia kejayaan ko qip dri kire stock ke owner mall
L1	Amboi pnggl ko eh
L3	Laki ke tino rumet ni
L5	Tu msti dak cina yg msg
L2	aah dak cina tu laa
L5	Kunci cave diorg x jumpa Aku dah ckp kat box putih tu
L5	Putain sweater aku smua junior buang
L2	sweater mana ni asal diorg buang bodoh tul
L5	Semua baju winter ada kotak
L2	bgtau la dia ambik blik

L5	Tp bodoh gak la diorg semua pakai buang je
L5	Nasib beg laptop aku diorg x buang
L1	Konfem ade kt bwh tu lgi
L1	At least tanye la sblum buang
L1	Byk hal btul junior2 nie
L1	Dh la kotak pulak tu
L5	Diorg buang smua kecuali beg laptop dgn ski
L1	Manteau husni skli dlm tu
L2	dia mmg ada tnya ak haritu ak ckp tu sume brg korg laa ak xtau la ad
L5	Diorg ckp camtu la tp salah aku gak
L5	redho je la
L1	Dhtu ko kene beli blk ar manteau smua
L5	Ig laa duit
L5	nice tu aku tggal kat rumah ntah diorg buang skali
L3	Aku ada qip baju sejuk kaler rainbow kalau kau nak
L5	Haha thnx Simpan la buat naik kinabalu
L5	Nnti aku cari la kat pasar notre dame
L3	cerita pasal Kinabalu hahah aku baru bkk doh video chamechaude
L3	Nasib cha bgtau benda ada kat yt
L5	Baru tau ke lol Lama dah aku upload
L5	Pkai camera alcatel je
L2	hat tu ada lg
L4	wahahaa kalau la aku mmpu elaun skang bukan cam dulu qip makan
L3	Kau ada klang eh
L3	Mari minummmm braderr
L5	Aku baru smpai rumah tenet x dok Ko msg bila ni
L3	Baru lg boh
L3	Mlm ni ah kott
L5	Allahu udah pukul dong
L5	Aku ni dah la dduk pulau indah x tau jln
L5	Pulau indah kat hulu ni coverage pun x dak
L3	Pulau indah ehh
L3	Minul laa suruh aku roger
L3	Dokpe ahh

L3	Lain kali ork
L3	Uishhh rohana kena ada grup jugakk
L1	Bincang strategy fifa
L2	masih freedom
L2	masih freedom
L3	Serius la ni beliau
L2	kena tag taqip ni
L3	Apa doh
L2	semua sudah kufur
L5	Nasib matmin x ajet french Tuko joseph
L3	Tak sekarang apesal shirt less
L2	ak nk tuko nama kat fb la gnie
L2	shirtless dah lama dah
L2	zmn workout lg
L4	ni siapa wei
L2	ak dgr cerita dia ni pernah share rokok ngan taqip kat klia
L2	ko xckup legend lg ni bil
L2	kena open table kalau nk bincang psl dia ni
L3	tetiba keluar kat suggestion aku dohh
L3	Sorang pergi follow
L5	Dia follow aku kot tp aku dah lama unfollow dia
L3	Kau dah share rokok sama ²
L3	Woi mari lepak oihh
L2	open la skali
L2	ak ambik flight dtg xde hal hahaha
L2	tp kena jmput kat airport laa
L5	hb fly
L2	ah shit kwn ak kawin lak da
L5	min mak aku x bg drive jauh2
L2	xnnti bncang blik nie dlm flight nk blik kl ni
L1	Blh la adjust biasiswa blk
L2	rajin2 laa g embassy lps ni
L5	Segan la aku
L1	Konfem die blh tlg

L2	kalau jln2 paris pegi laa mlwat dia xpon time dia dtg grenoble ke
L2	jd org penting grenoble la qip
L5	Malas ah
L3	Kau kenal ² qip Mana tauu tahun lagi kau pulak orang embassy
L3	Konfem free kalau aku dtg sana
L1	Power tak Ke tak setaraf ko
L5	Yg pkai spek tu power Fifa x main pun mana dia blaja ntah
L2	saje je tu xnak ngaku pernah main fifa nnti kalau kalah malu
L1	Happy birthday Kiting
L1	Dh tua doh
L1	Haha Nie je chance nk buat cam nie min nnti dh keje takleh dh
L2	ada jwpn pape x ko apply ak baru je buat account linkedin
L1	Haha Dpt invitation ko
L1	Apply dh byk tpi satu pun tak reply
L2	ko apply kat mana job street
L2	jobstreet priority xsilap ak
L2	ak pon baru buat account jobstreet smlm
L2	buat laa
L1	Still byk je offer ko dpt
L2	ko search google
L2	offer mmg byk tp kalau ko byar nama ko akn kat atas dlm list yg appl
L1	Mmg ade yg priority tu klau tak silap
L2	junior apa entah
L1	Okay2 ko dpt buat tu
L3	Happy birthday min Dah dahh Boleh kahwin dah boh
L2	nk kahwinnnn boh h xthn nii
L4	kahwin nnt min aku bagi satu bonghit sebelum naik pelamin haha Ha
L2	tenang is good
L1	Chicken burger tuk aku min menu cesar
L2	apa ni sauce barbeq
L2	tu m'as decu frere
L1	Tu duit garantie yg kite bayar dulu mase masuk rumah
L3	Banyak nor duit nye Mari lah boleh ikan bakor lagi sekali
L4	nk pinjam K gais bil elektrik kena potong bulan tertunggak

L4	wahaha dah hobi min kerja collect mesin basuh je
L2	dah ckup satu album belum ahaha
L1	Vous ce
L2	Ce a été
L5	Tu lagu amalina versi tamil la jay
L5	Nita pe cita jay Aku x join dak2 grenoble ni
L5	X tau dia join x mkn2 before ni
L1	Haha Die pun tak join bro
L1	Die klau harap ko join bru die join
L1	Since die kenal ko je
L1	Ko ar admin group
L5	Aku admin aktiviti bola
L5	Kasut bola ko aku pkai ni ha
L1	Hahaha Asal lina pulak admin
L1	Come on ar kip Ko super senior
L5	Diorg active kat wsapp je fb x hidup
L5	Esok diorg nak wat potluck X gi lg aku Busy
L5	Stage x jumpa lg boh
L5	Guane interview
L1	Potluck tuk ape Deepavali
L1	Awal lgi kip Stage start bln kan
L1	Okay je Kene tnggu ari isnin tuk result
L1	Tak sure g ke tak kot
L5	Dpt tu inchallah frère
L1	Harap2 ar long dpt first job
L2	wei jay mntk laa yg wd yg capik mntk tu
L2	dia nk background engineer dah ko dr grenoble satu univ ngan bdk2
L1	Tapi sok la
L1	Tgh layan dgn kogu n mber2 aku
L5	Santesh join x
L1	Mimpi la ko qi
L5	amalina versi tamil satu jay
L1	Tak yah nya nk report ar bro
L2	so cepat2 la kalau nk

L5	place utk aku next year
L2	kael kerja Osram hati hati kalau nk mntk intern hahah
L1	Kael program lain bro
L2	ada lobang kita bgtau dont worry
L5	pinggan pun ok ada pengalaman
L2	blik intern kat sini la
L2	ada chance nk smbung lps dah hbis
L5	Cari sini je la mantap skit cv
L2	enjoy dulu
L2	ajk mohamed pegi dotty dulu
L2	dah final year ni
L2	all out laa
L3	Jangan nak berchivas sangat
L3	Bawak ² laa minum air kelapa je
L1	Takde la Bdk baik skrg Arak
L3	Minum barbican jelah
L3	Nak rogere
L1	Hahaha Takpe ar dutch lady dh ar
L2	Ce a été
L2	Ce a été
L4	cuka tu jangan pedas sangat kena manis2 skit ni ayat isa kan
L2	isa atau capik ni
L2	isa kot
L2	mcm mana ko boleh igt ayt ni siakk legend ni
L2	Ce a été
L4	rindu sial kat isa capik
L4	rusDan lagi la
L4	kau sape min kat penang tu
L4	sampai sekarang aku still xpaham apa maksud ayat ni
L2	dgn bdk2 yg ko rindu la
L2	xde choice weh
L2	so status menumpang la
L2	kael kerja penang gak osram
L4	housemate aku kt cyber pun kenal Kael tu aku trus

L4	sabo min nnt aku jak nami gi lawat penang
L2	yg ak tau dia kerja osram penang haritu dia pergi training kot kat san
L2	ahhaha saka femes ni
L3	Ni matmin suka ah ni
L2	mana jumpa vdeo legend nii
L2	sbb tu kepala ringan je main periuk
L1	Main periuk
L2	baru perasan kedai cina tu xbg bju fut
L5	Baru beli ke Aku x main da fifa main game pc skrg
L2	tp klau tenet xlaju xleh main online mmg sia sia la beli
L2	weh bdk2 junior buat mkn2 kat rumah kita
L1	Hahaha Rumah kite terlalu high class tkleh buat mkn2 smua
L2	skrg junior jd mangsa
L1	Tak buat openhse
L2	weh qip surat ak dah post eh
L3	Marilah rohana vs shahir
L2	dtg penang laa
L1	Natang Main kasar skrg bdk insurans nie
L3	Kahkah tu pujian kot jay Kau ingat kau level alex hunter ke broo
L2	ak dgr crite jay dah start main gym skrg
L1	Hahaha Aku level rajinikanth la bro alex hunter ape brg
L3	SERIOUS AHHH
L2	ni taqip lps kls ali freedom
L5	Mari join group usrah ana insyaallah lurus jalannya
L2	sweater baru siakk
L2	sweater lama dh buang
L1	Haha Biasa la bro
L1	Takip dh upgrade skrg nie min
L3	Macam black m ah qip
L5	Haha Aku pun confius mula2 camna ada org muka mcm nita kat lab
L5	Sweater aku yg kena buang tu tau x sapa pakai
L5	Pakcik mystère bwh rumah
L4	baru nk start intern next sem pstu ada year lagi
L4	gossip mengatakan bahawa TM dah full so dia x dapat nk bagi kerja

L5	Keje sini la
L4	haha bagi habis belaja lu nnt dtg grenoble la balik mintak kerja SFR
L4	kau tng M2
L5	Web développeur blambak offer sini
L5	X de guna aku buat info classique slama ni
L5	Last2 stage aku buat web
L4	hahaha kerja reseaux xde ke
L4	Java ok lagi dari PHP
L5	Jay kasi no nita stp
L5	Dengar citer husni jd consultant
L1	Aku keje kt kuantan je kip kt pahang cement
L5	Best kot dpt mkn resto zaman tiap pagi
L1	Jauh arr resto zaman
L2	junior xbgtau ke diorg buang kat bwh gmbar skit
L5	buang kat bawah tu la
L5	Aku nmpak pakcik tu pkai veste hijau aku kat casino
L5	X pe la dia pakai dr buang camtu je
L1	Haha Kire sedekah la tu kip
L1	Bdk tu MIA sial nnti bile ko jumpe mintak no die pastu passing kt aku
L2	dia smbung PhD tp skrg PhD dia convert part time sbb baru dpt kerj
L1	Missing in action
L2	mmg missing laaa
L2	aritu meeting dgn ak
L2	asyik pegi gym skrg
L2	td blik pon ak nmpk dia baru blik gym
L1	Die mmg gym body dri dulu
L2	aritu dlm lift dah laa ramai org tetiba dia tnya ak dr dulu pendek
L1	Haha lgtkn dh masuk hospital sakit jiwa sbb tu hilang
L5	Jay nak join company la pasni
L1	Hahaha Sbb raj ade kt situ aku dgn rela hati stay kat kuantan
L2	boleh thn sakit hati meeting dgn raj ni
L1	Ko meeting dgn did je ke ade org lain gk
L2	ada org lain gak laa
L2	otw ni

L1	Amboiii main kasar nmpk abg retirement plan kite
L1	Screen brightness tak ckup tu
L3	betul siaa jayy Aku full brightness baru nampak
L5	Pkai alcatel
L3	Kau mesti baru lepas makan pain bagnat kan kip
L1	Hahaha Aku tau teknik die broo
L1	Dgn sauce pizza tu
L5	Letak sauce piquante keping
L1	Nie resipi blaja dri sifu
L3	Moh la lepak anep kawin
L3	Atuk kawin dah la aku tak pergi
L3	Atuk dgn bebe ke orang lain akhirnya
L1	Aku keje ar sabtu bro Sabtu yg aku keje la majlis bdk2 nie
L1	Dgn bebe la Live happily ever after
L3	Kau dgn tanuj bila jayy
L1	Wtf Sejak bile pulak aku dgn tanuja sial
L1	Aku Imbt lgi ar bro Tnggu anitaa abis blaja dulu
L3	sampai sial keluar ehh
L3	Kau jgn main ² jay Aku dgr cerita tanuj sexy sekarang Lawo dahh
L3	Tahun depan qip inshaAllah
L1	Uishh Nie wajib pegi Gtau date awal2 blh apply cuti
L3	Takyah balik ahh Kau hadi kat sana
L5	Haha Masaf game ni aku team dgn amin
L5	Wat phd kat lyon
L3	Kau patutnya dah ada team dgn anak ² usrah ni
L1	Haha Fuyohh geng super senior skrg Masaf game kt mane tahun ni k
L5	Sbb tu aku join haha
L5	Team ni hanafi yg recrute Nak emas katanya
L5	Dia letak aku jd RB Kuangja Mbappé ni
L5	kau dah hilang jay sorry
L2	ni mode silhouette
L2	tomyam pomme de terre dgn salad ke rindu siakkkkk
L2	ak pegi dgn bdk2 ni lepak laa

L2	dah hbis balik laa
L1	Complete meal kot Ade protein carbohydrate minerals
L5	Razin msg aku weh Haha
L2	ss skit
L2	asal dia nk smbung phd ke
L4	nk pinjam GoPro la tu
L3	Esok boleh laa masukkan razin dalam group
L2	ak dah tau qip psl apa dia contact ko
L2	dia nk mntk tlg amikkan diplome
L1	Die tak post ke MY
L2	dia xbuat la tu
L2	ko dah hntar email ke jay
L1	Email add die pun aku dh lupa
L1	Nnti die akn anta email ke dok utk bgtau diplome dh ready
L1	Takpun aku srh anitaa ambikkn je la tuk
L2	gerak bgtau laa
L4	boleh xde hal aku memang ade RDV delnondedieu
L3	Pinjam kad tram min
L3	Controlleur banyak sekarang
L1	Bubble tea jom
L3	Min kalau aku pegi penang
L3	Boleh musafir rumah kau tak
L3	Ke kena bawak sleeping bag
L2	dtg je laa
L2	boxer boleh pnjm capik punya
L5	Kad tram min dah expired
L5	Haha Coloc rembat
L3	Anep punya jemputan aku tgok petang eh
L3	Ke tu membe ² je
L2	kitorg kat OU ni mai laaa pekena lagu karaoke lu spukul
L4	send pic skit
L5	Syamim jean skinny skrg
L2	qip tahun nie xnak tinggal2 lah kalau boleh
L2	babi RT

L4	malas ar
L2	mohon taqip xpegi paris time puasa
L1	Selamat berpuase bros
L4	xde nk jalan2 paris ke apa
L5	Kena gi paris sekali lg ni nak cari air gedegang nami haritu
L5	Aku bukak puasa skrg commande livraison je
L1	Order flex jom
L2	buat la qip
L2	jay kalau final dia keluar qip
L1	Keluar pun still tgk bruh
L1	Tapi tahun nie tak keluar ar Jauh punye
L5	Jay tgh warming up ke
L1	Warm up versi tamil
L5	Join ah jay
L1	Baru trun boh Dh letih
L1	Salam YNWA
L1	Plant tgh buat stoppage skrg So kene keje bro
L1	Haha Bile plant stop ni la masa kene buat major maintenance jobs s
L2	kije apa bro ak cuti smgg
L2	msjk shift la jay terok sgt tu
L2	subuh dulu jay dekat dah tu lps tu tido la
L1	Selamat hari raya bros
L5	Maaf zahir batin
L1	Ada buat potluck ke
L5	Ada week-end nnti
L5	X de rasa raya pun Mcm hari biasa haha
L5	Aku legend x yah bwk apa2
L1	Mmg ar Ko layan maitre gims dgn jules skrg nie
L1	Mana dgr lagu m nasir dh skrg
L2	nnti ak virer kat ko qip
L2	aritu pegi bank susah sgt
L1	Kip bile blk boh
L1	October nie ada event besar
L5	X balik boh

L5	Event mnde
L2	dgr cite kapel dgn lina skrg qup
L2	bros save the date
L1	Tapi krng ajo buat mase deepavali
L1	Aku lmbt lgi bro
L5	Cibai apa bekas isa
L2	sah time tu cuti jay nk buat mcm mana
L2	ak xsmbut deepavali xpe kot
L2	ok laa
L5	Awal ngat min Mai sini lu en boite
L2	zaman allez danser la boite de nuit dah hbis bro
L1	Haha Post2 fb die skrg mcm post2 ustad dh
L1	Smpt lgi g love lane dgn leboh chulia tu
L2	mlm ni ak try virer kat ko qip duit elektrik nnti dah virer ak bgtau
L2	apa siak jgn bg lagu dj santesh jay ak xtau
L1	Hahaha Ko ddk penang takkan taktau tmpt2 happening mase malam
L1	Come on arr
L2	dey ko tgk ak duk penang dgn sape
L2	nk release stress kerja
L2	nnti dtg penang laa jay bwk amoi2
L5	sini aku ada kwn party girl lesbienne lgbt
L1	Aku bru turun penang mase wesak aritu dgn kogu blake bro
L2	ak pon ada kwn lgbt dulu
L5	la x kan nak post family smua ada réseaux soc
L2	time tu ak blik tgganu nmpk je story ko
L2	dah get laid belum
L1	Share kt grp whatsapp sdh
L5	Puasa yak yak ye la sorg2
L2	mcbook lama
L2	aidil fiti nal va tikam
L4	SHR MZB semua ada rezeki jumpa kat rue Alexandre Fleming lagi nn
L3	Pe cerita budak ² curry
L2	kawhi power siakkk tahun ni
L2	Siakam power gak

L2	kena warriors
L3	Abes la dynasty eh Takde dah min
L4	jngn la camni
L4	fighting kot game tadi
L4	warrior bodo kalah first games dah la Durant injured
L1	Amboii bosan sgt hang nie smpi buat editing skrg
L5	X de org main bola boring2
L2	sorry jay terjumpa gmbar ni
L2	gemok laaa
L2	kena cari blik vdeo jay mkn kayu manis ni
L5	Video pakoera jay ada nak
L1	Capoeira kip truk sgt spelling ko tu Haha
L2	capoeira dgn husni legend tu
L1	Bkn video tu ko yg rakam ke min
L2	ak la rakam
L2	ada dlm harddisk kot
L2	tggg jemputan laa kbo ke dia
L2	tp utk dlm group ni kat taman tamadun islam
L2	save the date
L2	cancel jay satu mjlis je
L5	Aku x confirm lg
L5	Tp nak balik la kot lpas stage
L1	Hahaha Buat satu dh la min jimat duit utk
L2	xyah bgtau laa bdk2 grenoble qip ak nk invite yg rapat2 je
L5	Orait Aku pun bkn kenal pun dal lain
L3	Lek luuu
L2	bln tu smpat lg kalau nk pegi songkran tembak air
L2	awek ko lina jgn acah xknl
L5	Kawin la nami cegah maksiat
L1	Blk grenoble smula september
L2	weh jay diplome dh ambik
L2	weh qip tlong ak utk recuperer diplome
L1	Haha Dulu die kluar sbb aku bwk Skrg nie mmg payah la die nk kluar
L1	Kat accueil phitem

L2	xpon tlg hntar smpul surat dgn timbre kat phitem
L2	ak xtnya pejabat post boleh hntar ke tak timbre tu
L2	baby jay mntak tlg ambikkan confirm laaa anita ambik
L1	Baik sruh takip ambikkn je min lgi senang bgi lettre procuration dgn
L2	taqip stage busy kot
L5	Aku kalau boleh ambik masa soutenance je la
L1	Accueil bukak pagi je
L1	Ambik bln la
L1	Confirm dh bukak mase tu
L2	taqip dh habis bro
L1	Soutenance takkn dh takde kt situ
L5	Sbb aku stage jauh x de kat smh
L5	Tp lpas stage boleh je
L5	Dah la Cam haram
L5	Mmber lain semua ada congé
L5	Entah Aku sign je contract x tgok pun haha
L5	Last2 boss aku cek aku x de congé
L5	Redho je la
L2	bdk2 ni tahun lps semua ada congé
L5	Bil engie tu dah bayar setel
L5	Aku nak delete rib bank aku lam compte tu
L2	delete lah
L2	francais dh skrg jay aritu p grand place
L5	Aku tgh recrute produk mlm
L1	Hahaha Thank you for ur service sir bagai
L2	bru sedap skit ayat ada idiom
L2	ce mec a tué des innocents en iraque et toi tu lui dit
L5	Okay min kita sini melayu je
L1	Dah ar poyo balas dlm english bkn francais
L2	post gmba tram la qip kat insta mana tau TAG repost
L3	Merci jean michel
L5	Babi x leh hidup aku dgn org toksik camni
L2	ahh tu m'as tué mdr
L3	Thaqif Une inspiration pour malaisien

L5	Haha Nak buat record melayu paling lama kat grenoble
L5	bos stage dah ajak bincang salaire
L2	ba la la la laaaaaa
L2	stay sana laa gaji besar brg2 murah time kerja ikut law
L1	Blh ar kip mintak pr jdi francais
L3	Bisnes sini
L2	kerja luar qip dtg sini terus naik level manager
L3	Carik bini French Nanti bawak balik jerantut baru semua melopong
L1	tak anak en hassim kawin dgn org mat salleh
L2	mari laaa penang
L2	ramai dah kat penang
L3	Jumpa sana laa
L3	Bila eh
L3	Aku kalau takde carpool pon tak dapat jugak ni
L5	X pegi la baru sampai
L2	bkn dia kat group warga grenoble
L2	ak xde dlm group
L2	sbb semua org dia jmut dlm group je qip xde personal
L2	dtg penang je laa
L2	tgk husni main game
L2	Ce a été
L2	Ce a été
L1	Haha Mmg die tgh semua kt fb kip
L1	Aku pun dpt invitation tu
L5	Mai dtg jerantut naik train pastu gerak skali haha
L2	selamat pulang kpd frero Thaqif abdulla hashim félicitation pour ton
L5	Otw ke pekan ni

Table des matières

REMERCIEMENTS.....	4
SOMMAIRE.....	6
INTRODUCTION	8
CHAPITRE 1 - ÉTAT DE L'ART.....	11
Code-switching en TAL et en Sociolinguistique	12
1. Points de vue de divers chercheurs.....	13
2. Code-switching en sociolinguistique	14
2.1.Linguistique Socioculturelle	15
2.2.Trois processus pour rendre compte du code-switching.....	17
2.3.Code-switching dans un environnement mixte.....	19
3. Code-switching en TAL	20
3.1.Identification de langue	20
3.2.Prise en compte du code-switching dans les systèmes de reconnaissance vocale	21
3.3.Conclusion	22
CHAPITRE 2 - MÉTHODOLOGIE	24
Les ressources.....	25
1. Les données.....	25
1.1.Choix des données	25
1.2.Méthodes de collecte des données	26
Structure des données	28
1. Fichier .numbers.....	28
1.1.Feuille 1 : v ori	28
1.2.Feuille 2 : V1	29
1.3.Feuille 3 : V2	30
1.4.Feuille 4 : V3	30
1.5.Feuille 5 : V4	30

1.6.Feuille 6 : Locuteurs	31
2. Script python.....	31
CHAPITRE 3 - BIOGRAPHIE LINGUISTIQUE DES LOCUTEURS ...	33
1. Origine Ethnique.....	34
1.1.Locuteur 1	36
1.2.Locuteur 2	37
1.3.Locuteur 3	38
1.4.Locuteur 4	40
1.5.Locuteur 5	41
2. Pratiques langagières parmi différents groupes de personnes.....	42
2.1.À la maison en famille	44
2.2.Entre amis et connaissances.....	44
3. Pratiques langagières à l'école	45
3.1.École maternelle.....	45
3.2.École primaire.....	46
3.3.Lycée	46
3.4.Université.....	47
CHAPITRE 4 - ANALYSES ET RÉSULTATS STATISTIQUES	49
1. Le tableau de données.....	50
1.1.Colonne C : langues	50
1.2.Colonne E : formes normalisées	53
1.3.Colonne F : traduction.....	53
2. Les statistiques	54
2.1.Nombre de messages par langue	54
2.2.Nombre de messages par locuteur.....	55
2.3.Les messages mixtes	59
3. Les interprétations.....	61
3.1.L'environnement influence-t-il l'utilisation d'une langue?	62

3.2.Code-switching ou code-mixing?	62
Conclusion	63
Bibliographie	65
Sitographie.....	66
Table des illustrations	67
Table des annexes	68
Table des matières	121

MOTS-CLÉS : alternance codique, corpus, applications de messagerie, Malaisie

RÉSUMÉ

La diversité colorée des ethnies et des cultures en Malaisie se compose de nombreuses langues parlées, la plupart des Malaisiens étant multilingues. Au fil des années, cette communauté mixte a développé une façon de communiquer qui n'est pas considérée comme conventionnelle ailleurs. L'alternance codique chez les Malaisiens est devenue la norme, en raison d'un manque de connaissances linguistiques, ils ne sont pas conscients de ce phénomène, plus connu d'eux comme une habitude de conversation. Le manque de données et le défi de la compilation des données ont entravé les études dans ce domaine. Dans ce mémoire, j'ai étudié le phénomène de l'alternance codique à l'aide d'un ensemble de données obtenues en rassemblant des conversations de textes entre des étudiants Malaisiens qui ont étudié en France. L'objectif de cette étude était de créer un corpus de pratiques langagières « mixtes » en évaluant l'utilisation des langues dans les applications de messagerie, souvent utilisées par les Malaisiens.

KEYWORDS : code-switching, corpora, messaging applications, Malaysia

ABSTRACT

The colourful diversity of ethnics and cultures in Malaysia bears many spoken languages, with most Malaysians being multilingual. Over the years this mixed community has developed a way of communicating that is not the convention elsewhere. Code-switching amongst Malaysians has become the norm, due to lack of linguistic knowledge they are unaware of this phenomenon, more commonly known to them as a conversational habit. Lack of data and the challenge of compiling data has hindered studies in this field. In this dissertation, I studied the phenomenon of code-switching using a set of data obtained by assembling texts conversations amongst Malaysian students who studied in France. My goal of this study was to create a corpora of “mixed” language practice by assessing the use of languages in messaging applications, often used by Malaysians.