N° 126

UNIVERSITE D'ABOMEY-CALAVI (UAC)

****

Ecole Doctorale des Sciences De l'Ingénieur (ED-SDI)

****

Master de Recherche en Télécommunication et Réseaux Informatiques

****

Rapport de stage

Thème :

Revue des Techniques d'apprentissage profond pour la synthèse de Langue Locale du Bénin

Présenté par **:**

Ing. BABADJIDE Marie-Charbel

Encadré par **:**

Dr. AOGA John

Enseignant Chercheur à l'UAC

Sous la direction de **:**

Pr. EZIN C. Eugène

Professeur Titulaire des Universités du CAMES

Enseignant Chercheur à l'IFRI

Laboratoire d'Électrotechnique, de Télécommunication et d'Informatique Appliquée (LETIA)

Sommaire

Remerciements

L'aboutissement de ce projet de fin de formation a été possible grâce aux différentes contributions techniques, organisationnelles et financières de plusieurs personnes. Nous tenons à témoigner nos sincères gratitudes :

- ✓ au Pr Ezin Eugène pour avoir accepté superviser ce travail, nous lui témoignons notre profonde gratitude ;
- ✓ au Directeur de l'École Doctorale des Sciences de l'Ingenieur (ED-SDI) Pr. Mohamed GIBIGAYE et son équipe, pour vos conseils ;
- ✓ au Dr. John AOGA, pour avoir accepté encadrer ce travail, vos apports, suggestions et critiques ont été d'un apport très important pour la réalisation de ce mémoire ;
- ✓ à tous les enseignants du master TRI, pour la formation reçue ;
- ✓ à toutes les personnes qui ont contribué de proche ou de loin à la réalisation de ce travail notamment Dr Charles LIGAN, je vous dis merci.

Liste des figures

Liste des tableaux

Liste des sigles et abréviations

ACM DL :	Association for Computing Machinery Digital Library
API :	Alphabet Phonétique International
ASR :	Automatic Speech Recognition
BERT :	Bidirectional Encoder Representations from Transformers
BLEU :	BiLingual Evaluation Understudy
BTEC :	Basic Travel Expression Corpus)
CNN :	Convolutional Neural Network
GAN :	Generative Adversarial Networks
GRU :	Gated Recurrent Unit
HMM :	Hidden Markov Model
HTER :	Human-targeted Translation Edit Rate
IA :	Intelligence Artificielle
IEEE :	Institute of Electrical and Electronics Engineers
INSAE :	Institut National de la Statistique et de l'Analyse Economique
LSTM :	Long Short-Term Memory
METEOR :	Metric for Evaluation of Translation with Explicit Ordering)
MFCC :	Mel Frequency Cepstral Co-efficients
MOS :	Mean Opinion Score
MT :	Machine Translation
NER :	Name Entity Recognition
NIST :	National Institute of Standards and Technology
NLP :	Na- tural Language Processing
PER :	Position-independent Word Error Rate
PICO :	Population Intervention Comparator and Outcome
POS :	Part Of Speech
PRISMA :	Preferred Reporting Item for Systematic Reviews and Meta-Aalyses
RNN :	Recurrent Neural Network
SNA :	Social Network Analysis

SVO : Sujet Verbe Objet
STT : Speech To Text
TTS : Text To Speech
TALN : Traitement Automatique du Langage Nature
TER : Translation Edit Rate
TIC : Technologies de l'information et de la communication
VAE : Variational AutoEnconders
VQ-VAE : Variational AutoEncoder that uses Vector Quantisation
WER : Word Error Rate

Résumé. La révolution de l'intelligence artificielle et de l'apprentissage profond a permis de grandes avancées dans le domaine du traitement automatique du langage naturel en général et de la traduction automatique en particulier. Cette révolution technologique présente des avantages pour bon nombre de pays, surtout développés, sur le plan, de l'inter-compréhension et de l'intégration socio-culturelle et des échanges économiques. Pour les pays en voie de développement comme le Bénin, dont les langues sont issues d'une longue histoire d'oralité et très peu numérisées, elle pourrait représenter un atout majeur. Cependant, il n'existe pas à notre connaissance des travaux sur les méthodes de traduction automatique en Speech To Speech, identifiant leurs forces et faiblesses et faisant un diagnostic de l'éclosion de nos langues locales. Ainsi, nous avons effectué la synthèse des méthodes de traduction automatique speech to speech, en apprentissage profond, vu le caractère dominant de l'oralité dans nos langues. En suite, nous avons répertorié les problèmes liés à l'éclosion de nos langues locales du Bénin, en Afrique en général. Enfin, nous avons proposé des directives futures pour une mise en œuvre plus efficace et effective, pour créer des applications en traduction automatique d'utilité sociale et de développement. En nous servant de moteur de recherche tel que Google Scholar, de la base de donnée indexée Scopus, nous avons dans un premier temps utilisé la bibliométrie pour faire une revue globale. Cette revue nous a permis d'avoir une indication de l'importance des papiers les uns par rapport aux autres dans le domaine, les communautés existantes, la collaboration entre auteurs. Dans un deuxième temps, fort de cette bibliométrie, nous avons sélectionné un certain nombre de papiers que nous avons analysés (description, exigences, forces, faiblesse). Grâce ces exigences, nous avons pu identifier les défis avec nos langues locales (les difficultés) et, nous proposons les réseaux de neurones Transformer pour la traduction de nos langues locales.

Mots clés : *NLP*, *speech to speech*, *langue peu dotée*, *bibliométrie*, *PRISMA*

Abstract. The revolution in artificial intelligence and deep learning has led to great advances in the field of natural language processing in general and machine translation in particular. This technological revolution presents advantages for many countries, especially developed ones, in terms of inter-comprehension, socio-cultural integration and economic exchanges. For developing countries like Benin, whose languages have a long history of orality and are not very digitized, it could represent a major asset. However, to the best of our knowledge, there is no work on automatic translation methods in Speech To Speech, identifying their strengths and weaknesses and making a diagnosis of the emergence of our local languages. Thus, we have made a synthesis of speech to speech machine translation methods, in deep learning, given the dominant character of orality in our languages. Then, we have listed the problems related to the development of our local languages in Benin, in Africa in general. Finally, we proposed future guidelines for a more efficient and effective implementation, to create machine translation applications of social utility and development. Using search engines such as Google Scholar, the Scopus indexed database, we first used bibliometry to make a global review. This review allowed us to have an indication of the importance of the papers in relation to each other in the field, existing communities, and collaboration between authors. In a second step, based on this bibliometry we selected a number of papers that we analyzed (description, requirements, strengths, weaknesses). Thanks to these requirements, we were able to identify the challenges with our local languages (the difficulties) and we propose Transformer neural networks for the translation of our local languages.

Keywords: *NLP*, *speech to speech*, *Poorly endowed language*, *Bibliometry*, *PRISMA*

Introduction

La communication à notre ère est un véritable garant de la cohésion, et de développement social du fait qu'elle fait évoluer la société à travers des échanges d'information. Avec la révolution des technologies de l'information et de la communication, le besoin d'interaction (communication) homme-machine (ordinateur, téléphone) a augmenté. Il est donc devenu primordial de trouver des moyens de permettre à l'homme de s'exprimer dans un langage naturel et à la machine de pouvoir comprendre ce langage et l'interpréter convenablement. La discipline, Traitement automatique du langage naturel (TALN), en anglais Natural Language Processing (NLP), s'est dédiée à la résolution de ce problème. Depuis les années 1950, cette discipline travaillait déjà sur la traduction de phrases simples, et une expérience de 1954 à Georgetown a fait la traduction automatique d'une soixantaine de phrases russes en anglais [1], puis évolution a ensuite permis l'émergence des chatsbot dont ELIZA en 1964 [2]. Elle a connu une première révolution avec l'augmentation de la puissance informatique et surtout l'introduction du Machine Learning dans les années 1980. Ainsi, la TALN a permis donc l'essor des tâches telles que :

- la segmentation de texte, déterminant le début et la fin des phrases ;
- la classification des documents ;
- le résumé de document ;
- le balisage morphosyntaxique qui permettent de déterminer la classe morphosyntaxique de chaque mot à partir de connaissances lexicales et du contexte dans lequel il est utilisé ;
- l'identification d'entités nommées qui permettent de reconnaître dans un texte un certain type de concepts catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, quantités, distances, valeurs, dates, etc.

Mais aussi une des tâches qui est des plus difficiles du TALN est la traduction automatique qui repose sur des algorithmes prédictifs qui apprennent à partir d'un ensemble de textes d'une langue avec leur correspondant dans d'autres langues. Aujourd'hui, à mesure que les machines deviennent plus puissantes et moins chères, la quantité de données open source toujours plus importante et l'utilisation du Deep Learning, les résultats obtenus pour les tâches de la TALN ont connu une amélioration significative de leur performance. Ainsi, le TALN se trouve donc être en pleine expansion. Ces améliorations et cette expansion sont de plus en plus visibles au travers des chatbots et assistants personnels tels que Google Now, Cortana ou Siri, et des outils de traductions automatiques, Google Translate, Microsoft Translator et

Skype Translator qui ont évolué de la traduction des textes à la traduction de la parole prononcée (Speech to Speesch). Tous ces outils améliorent au jour le jour les échanges entre les peuples de tous les coins du monde, ce qui représente un atout majeur pour le développement des nations et favorise l'accès à une infinité d'informations, de connaissances et de compétences dans diverses langues et révolutionnant ainsi la vie des êtres humains.

Cependant, nous constatons que pour les langues africaines, surtout dans les pays en voie de développement, que ces prouesses technologiques pourront avantager de toute évidence, sont restés en marge de ce toutes ces avancées. C'est le cas de notre pays, le Bénin où selon l'INSAE, en 2018, seulement 41,7% de la population d'adulte de 15 ans et plus, était alphabétisée et où malgré le fait que le français joue un rôle important dans la gestion des affaires, elle n'est pas la langue la plus parlée [3] mais plutôt les langues locales dont tradition très orale. Malgré les efforts du gouvernement pour la promotion des langues locale, marquée surtout par l'introduction de l'apprentissage de dix langues locales dans les écoles qui sont encore très peu numérisées, beaucoup d'effort reste encore à consentir. La question qui se pose donc est : Comment faire bénéficier nos langues locales de toute cette révolution numérique ?

1- Problématique

Nos langues locales ne bénéficient pas encore des avantages de la révolution de la traduction automatique, car il existe très peu, à notre connaissance, de travaux s'effectuent dans ce sens. La principale raison à cela selon nous est le fait qu'il existe de nombreuse technique de ce genre pour créer des systèmes de traduction Speech to Speech cependant il n'existe pas un travail de référence pour faire la synthèse de toutes ces techniques et donnant des directives sur comment créer un système speech to speech pour nos langues locales.

2- Contexte, justification

Il est primordial de faire bénéficier nos langues locales de toute cette révolution, car de nos jours la traduction automatique joue d'important rôle sur le plan économique, socio-culturel et logistique. Sur le plan économique, le Speech to Speech en permettant de communiquer, de faire des affaires avec les interlocuteurs dans leurs langues et en permettant de traduire son partenaire dans plusieurs langues pour une augmentation du taux de visibilité sur les marchés, permet d'être compétitif dans un marché globalisé. Sur le plan socio-culturel, elle facilite les brassages interculturels, aide à une meilleure compréhension de leur valeur. Elle contribue et valorise le patrimoine culturel et linguistique du pays, mais aussi améliore l'alphabétisation des langues locales.
Au niveau de la logistique, elle augmente la rapidité de la traduction entre les hommes et réduit les coûts de traduction qui sont considérables quand on recrute des traducteurs.

Un travail de référence pour faire la synthèse de toutes les techniques donnant des directives sur comment créer un système speech to speech pour nos langues locales permettrait justement l'éclosion de travaux scientifique pour du Speech to Speech dans nos langues locales, orienterait vers des lignes directives de leur développé, de diagnostiquer l'état de santé de ce domaine et d'armés pour faire face aux goulots d'étranglements, de regrouper et de faire la revue des travaux qui ont dans le temps constitué une référence de l'évolution dans le domaine du Speech to Speech. Cela a été le cas pour l'application des techniques de neuronale pour la traduction automatique suite à la sortie de l'article «A survey of current paradigms in machine translation» en 1999. Avant cette sortie en 1999, les recherches concernant les techniques neuronales en machine translate pour l'intervalle 1900 à 2000 présentait 12 000 résultats environs pour une augmentation entre 500 et 1500 résultats par décennie. Mais après, nous constatons une augmentation de 8 000 résultats entre 2000 et 2010.

3-Objectifs

Ainsi, dans le cadre de notre étude, nous nous sommes fixés comme objectif d'effectuer l'état de l'art des méthodes de traduction automatique speech to speech en Deep Learning afin de répertorier les problèmes à leur éclosion pour les langues locales du Bénin, en Afrique en général et de proposer des directives

futures pour une mise en œuvre plus efficace et effective.

4-Méthode et contribution

En nous servant des outils de rechercher tels que Google Scholar, Scopus, pour atteindre nos objectifs, nous avons fait une revue systématique des méthodes de Deep Learning, fait une revue systématique de nos langues locales (une description globale des langues FONGBE YORUBA, BARIBA), fait la synthèse des papiers de speech to speech en Deep Learning (essentiellement) depuis 2010, identifier les forces et limitation de ces approches (inventaire des goulots d'étranglement globaux), catégoriser ces approches selon un certain nombre de critères, fait une étude comparative des approches et des comparaisons inter groupe, fait une étude de compatibilité avec nos langues locales (identifier les goulots d'étranglement spécifique et les forces) et recommander de nouvelles directives pour une mise en œuvre effective et efficaces des méthodes de traduction Speech to Speech pour nos langues locales.

5-Organisation du document

Ce mémoire est subdivisé en trois (03) grandes parties :
Le premier chapitre présente la revue des techniques de traduction de langue et identifie leurs forces et leurs faiblesses.
Le deuxième chapitre présente l'ensemble des outils utilisés, les choix techniques effectués et les étapes de la mise en œuvre de nos solutions constituant la méthodologie
Le troisième chapitre donne les détails sur les résultats obtenus au vu de la méthodologie adoptée. Elle fait aussi l'analyse de ces résultats et les critiques.

Chapitre 1

Revue de litterature

Introduction Le domaine du traitement automatique du langage naturel de façon générale et la traduction automatique de façon spécifique demande une certaine précision sur les facteurs qu'ils font intervenir. Dans ce chapitre, nous avons ressorti les caractéristiques de certains de nos langues, identifié et apprécié les approches et les méthodes de traduction automatique et déterminé les critères d'évaluation des systèmes de traduction automatique.

1.1 Langues et caractéristiques

Selon linguiste suisse Ferdinand de Saussure, la langue est un système de signes vocaux propres aux membres d'une même communauté. C'est un outil de communication au sein d'une même communauté et, d'un point de vue sociolinguistique, un symbole d'identité et d'appartenance culturelle. En tant que code, la langue demeure une convention sociale, à priori indépendante des variations individuelles [68].

1.1.1 Phylums et familles de langues

Pour l'Afrique, la plupart des linguistes acceptent aujourd'hui les grandes lignes de la classification des langues africaines en ce qui concerne les familles des langues (Phylum) proposée par J. GREENBERG (1963) [4] qui sont au nombre de quatre et qui sont la famille Afro-Asiatique, la famille Nilo-Saharienne, la famille Nigéro-Congolaise et la famille Khoisan.

- Le phylum Nigér-Congo, regroupe la grande majorité des langues ouest-africaines avec la grande famille Bantu qui couvre la quasi-totalité de l'Afrique au sud de l'équateur.
- Le phylum nilo-saharien comprend les langues parlées essentiellement en Afrique centrale, orientale et nord-orientale.
- Le phylum afro-asiatique comprend celles sémitiques (dont l'arabe), l'égyptien ancien et le berbère, ainsi qu'un grand nombre de langues du Nigeria et du Cameroun (famille tchadique), d'Éthiopie et des régions avoisinantes (familles couchitique et omotique).

- Le phylum Khoisan, en fin, regroupe des langues parlées principalement en Afrique australe (Afrique du Sud, Namibie et Botswana). Mais sa zone d'extension allait autrefois beaucoup plus au nord.

Selon le recensement général de la population et de l'habitation (2002), sur les 48 langues répertoriées, toutes les langues du Bénin appartiennent à la famille des langues nigéro-congolaises, à l'exception du dendi qui est une famille des langues nilo-sahariennes [69].

1.1.1.1 Les groupes du phylum Nigér-Congo

Dans la famille Niger-Congo, on a dénombré sept groupes à savoir : ouest-atlantique, mandingue (mandé), voltaïque (gour ou gur), adamawa-oubangien, ijoïdo-défaka, kwa, méridional [70].

Le groupe Ouest-Atlantique Le groupe Ouest-Atlantique compte environs 30 millions de locuteurs répartis en plus de 60 langues. Ceci dans les pays tels que le Sénégal, la Gambie, la Guinée, la Guinée-Bissau, la Sierra Leone et le Liberia. Les langues célèbres telles que le Fulani (1 million), le Peul (1 million), le Wolof (3,2 millions) et le Sérère (1 million) [70] sont considérés comme des langues à forte importance numérique.

Le groupe Voltaïque (Gur) Le groupe Voltaïque (Gur), compte 20 millions de locuteurs et peut comprendre une bonne centaine de langues parlées surtout au Mali, en Côte d'Ivoire, au Nigeria, au Burkina, au Ghana, au Togo et au Bénin. Parmi ces langues, l'Ibo (18 millions) se distingue de toutes les autres, ainsi que le Yoruba (465 000), le Sénoufo (260000) et le Bariba (100000) locuteurs [70].

Le Mandingue (Mendé) Ce groupe compte plus d'une soixantaine de langues ouest Africaines et parmi ces langues figure le Bambara (2,7 millions), le Dioula (1,3 million), le Mandingue (1,4 million), le Kpellé (800000), le Bisa (430000), le Malinké (340 000), le Soninké (150000) et le Bozo (100000) [70].

Le groupe Adamawa-oubangien Pour le groupe Adamawa-oubangien, environs 12 millions de locuteurs répartis entre 160 langues du Nigeria, du Cameroun, du sud du Tchad, de la République Centrafricaine et du nord du Congo-Kinshasa. À l'exception du Sango (360000) et du Mumuye (400000) [70], toutes ces langues ne sont parlées par un nombre relativement restreint de personnes.

Le groupe Kwa Le groupe Kwa, composé d'environ 80 langues avec 20 millions de locuteurs, comprenant une partie de la Côte d'Ivoire (sud-est), du Ghana, du Togo, du Bénin et du sud-ouest du Nigeria. L'Akan (7 millions au Ghana) et le Baoulé (2,3 millions en Côte d'Ivoire) sont les Thèmes de la Sierra Leone (1,2 million)[70] ;

Le groupe Ijoïdo-Défaka (Ijoïdo) Le groupe Ijoïdo-Défaka (Ijoïdo) regroupe une dizaine de langues parlée par environs 1,6 million de personnes. L'Izon compte à lui seul un million de locuteurs, les autres

langues numériquement importantes étant le Kalabari du Nigeria (258000) et le Irike (248000)[70].

Le groupe Méridional Ce groupe compte une vingtaine de parlées pour environs deux millions de locuteurs, surtout en Guinée, en Sierra Leone et au Liberia. Seule le themne de la Sierra Leone atteint le million de locuteurs (1,2 milion) ; le kissi de la Guinée (286 500) et le gola du Liberia (99000) [70].

Au Bénin, il existe plus de 73 langues parlées, qui ont été réduites à 25 sur la base de définitions sociolinguistiques et de statistiques lexicales, de données morphosyntaxiques, de tests d'intelligibilité et de tentatives de reconstitution d'une ascendance commune, et sont toutes reconnues par la Constitution de 1990. Dans sa politique linguistique, le Bénin a introduit 10 langues dans son système éducatif, réparties par région du pays : pour le nord, le bariba, le fulfulde, le dendi, le yom et le ditamari, pour le centre le Fongbé, le gingbé et le Adjagbé et enfin pour le sud le gungbé et le yoruba [3]. En outre, la situation sociolinguistique des radios communautaires indique pour le sud 4 langues communes à 11 radios pour 23 communes, à savoir le Fongbé le Goungbé le Yoruba et le Nagot. Pour le centre, quatre langues sont répertoriées à savoir le Maxi, le Fongbé, le Ajagbé et le Saxwegbé pour 10 radios dans 27 communes. Pour le nord, quatre langues, à savoir le Baatonou, le Filfulde le Lokpa et le Dendi.

Nous constatons que des langues introduites dans le système éducatif, le Fongbé et le Yoruba couvrent un plus grand espace, du fait de la présence du Fongbé langue de la famille des Gbé comme lange commune pour les radios au sud et au centre du Bénin et du fait de la présence des langues de la famille Idé telle que le Nagot au sud et au centre de la carte des ethnies du Bénin.

De cette analyse, pour le cadre de notre étude, nous nous intéresserons à ces trois langues qui sont le Fongbé, le Yoruba et Bariba

1.1.2 Langues et structure vocaliques

Le Fongbé, le Yoruba et Bariba dispose chacune de sa propre structure interne.

1.1.2.1 Langue yorùbá et sa structure vocalique

Le yorùbá est une langue d'origine africaine, et l'une des langues du groupe voltaïque (gour ou gur), de la grande famille des langues Niger-Congo. Il est parlé au sud-ouest du Nigeria (2□ plus important groupe ethnique en nombre), au Bénin et au Togo par plus de trente millions de personnes [5]. Son système vocalique comporte trois(03) ensembles de sons à partir desquelles les mots sont formés, à savoir : les voyelles, les consonnes et les tons.

Le yorùbá a douze (12) voyelles dont sept (07) voyelles ouvertes et cinq (05) voyelles nasales. Il s'agit de : a, e, ẹ, i, o, ọ, u et an,ẹn, in, oṇ, un. Les sept (07) voyelles ouvertes sont obtenues par la sortie de l'air de la bouche (éventuellement du nez) et elles sont appelées des phonèmes. Toutes ces voyelles sont nasalisées si elles sont précédées d'une consonne nasale (n ou m). C'est le cas des mots «mu» (boire), «nà» (chicoter), «mọ» (connaître), ⋯

Le tableau 1.1 présente les voyelles ouvertes avec leur représentation en alphabet phonétique international (API)

Tab. 1.1 – Les voyelles ouvertes avec leur représentation en Alphabet Phonétique International (API).

API	Orthographe	Exemple	Signification
[i]	i	i̲lé	Maison
[e]	e	ire̲	bienfait, faveur
[ϵ]	e	ilè̲	Terre, terrain
[a]	a	bà̲ta̲	Chaussures
[⊃]	ọ	lọ̲	Partir, quitter
[o]	o	gbo̲gbo̲	Tout
[u]	u	ìlú̲	ville, cité, pays

Les cinq (05) voyelles nasales du yorùbá ne sont pas obtenues à cause d'une précédence comme le cas des voyelles nasalisées, elles sont des voyelles natives du yorùbá. Elles sont produites après une consonne orale. Le tableau 1.2 présente quelques mots avec l'utilisation de voyelles nasales et donne la signification pour chacun en français ainsi que leur représentation dans l'API.

Tab. 1.2 – Les voyelles nasales avec leur représentation en alphabet phonétique international (API)

API	Orthographe	Exemple	Signification
[ĩ]	in	ikì̲n	noix de palme
[ϵ̃]	ẹn	iyẹ̲n	celle-ci, celui-là
[ã]	an	ika̲n	Fourmi blanche
[⊃̃]	ọn	ibọ̲n	arme, fusil
[ũ]	un	iku̲n	Écureuil

Il est important de remarquer qu'en yorùbá le changement d'une voyelle par une autre change (dans la majorité des cas) le sens du mot [«mu» (boire) et «mo»(connaître)].

Le yorùbá a dix-huit (18) consonnes. Elles sont présentées dans le tableau 1.3 avec leur représentation en API, des exemples de mots et leurs significations.

Tab. 1.3 – les consonnes avec leur représentation en alphabet phonétique international (API)

API	Orthographe	Exemple	Signification
[b]	b	b̲á	Rencontrer
[m]	m	m̲ò	Connaître
[t]	t	t̲à	Vendre
[d]	d	àd̲à	Coupe-coupe

[s]	s	s̲ò	Dire
[n]	n	n̲à	Frapper
[l]	l	àl̲á	Rêve
[r]	r	r̲à	Acheter
[f]	j	j̲à	Se battre
[ʃ]	ṣ	ṣ̲á	couper
[j]	y	ay̲a	Femme
[k]	k	k̲à	Lire
[g]	g	àg̲a	Chaise
[kp]	p	p̲a	Tuer
[gb]	gb	gb̲a	balayage
[w]	w	w̲á	Chercher
[h]	h	h̲a	gratter

Les sons yorùbá produits s'obtiennent par une obstruction partielle ou complète de la voix.

Le yorùbá possède également trois (03) niveaux de tons, à savoir : le ton haut - H (accent aigu), le ton moyen - M (absence d'accent) et le ton bas −L (accent grave). Ils jouent un rôle déterminant pour distinguer les unités lexicales et donne différents sens selon le fait qu'un mot soit prononcé avec un ton haut, moyen ou bas. Par exemple, on a :

—kò. (H) = construire ;

—ko. (M) = chanter ;

—kó. (L) = Refuser.

Pour les mots monosyllabiques, on a trois(03) possibilités et pour les mots dissyllabiques, on peut avoir jusqu'à neuf(09) possibilités à cause des trois tons.

Les mots yorùbá résultent d'une structure syllabique très simple. Cependant, l'obtention d'un mot ne se fait par une combinaison quelconque de voyelles, de consonnes ou de tons, les combinaisons se font en se basant sur des règles bien précises. L'ensemble de ces règles est désignée par structures syllabiques. Les structures syllabiques sont décrites en se basant sur les notations suivantes : « C » pour les consonnes et « V » pour les voyelles. Le yorùbá dispose de deux (02) types de syllabes qui sont : —« V » par exemple dans le mot àlá [à - lá] (rêve) ; —« CV » dans le mot wá [wá] (viens). À part les pronoms qui peuvent être de simples voyelles et donc représentés par V, on retrouve les syllabes de type V dans les noms commençant par une voyelle (à - lá), dans tous les mots où les consonnes nasales (m et n) font office de syllabe (ò-ro-m-bó = orange ; gé-n-dé=robuste jeune homme). Quatre(04) combinaisons des deux(02) type de syllabes, précédemment citées, sont possibles[6]. On a :

1. V-V → à-á-nu (pitié) ;
2. V-CV → é-tí (oreille) ;
3. CV-V → dí-è. (peu) ;

4. CV-CV → bà-tà (chaussure).
Remarques :
1. les noms dans le yorùbá standard sont au moins de la forme V-CV et tous les verbes commencent par une consonne ;
2. la plupart des noms yorùbá commencent par une voyelle et tous se termine par une voyelle.

1.1.2.2 Langue Fongbé et sa structure vocalique

Le Fongbé est une langue nationale africaine, voire la langue la plus parlée au Bénin. Le Bénin compte 55 langues nationales (dont 50 langues autochtones et 5 langues non autochtones), selon ”Ethnologue : Languages of the world” [5] (19 ième Edition, 2016).

Aujourd'hui, plus de la moitié de la population béninoise parle le Fongbé à cause de son caractère véhiculaire. Par conséquent, il est très courant dans les médias et est également utilisé pour l'éducation et l'alphabétisation des adultes. Il est aussi utilisé au Togo et au Nigeria. Il est classé dans le groupe des langues Kwa de la famille nigero-congolaise [7] et l'écriture de son alphabet est connu officiellement depuis 1975, se fait à partir du latin avec les caractères de l'API.

Le Fongbé est une langue tonale avec un système complexe de deux tons lexicaux, haut et bas, qui peuvent être modifiés pour produire trois autres tonalités : montée basse-haute, descente haute-basse et moyen [8]. Des signes diacritiques sont utilisés pour retranscrire ces différentes tonalités. Les tons haut, bas, moyen, Modulation, sont les quatre tons qui sont retenus par la Commission Nationale de la langue Fongbé [9] Le système vocalique du Fongbé, bien dessiné par les premiers phonéticiens, compte douze (12) timbres :
-sept (7) voyelles orales avec quatre degrés d'ouverture comme indiqué dans le tableau 1.4
-cinq voyelles nasales avec trois degrés d'ouverture comme indiqué dans le tableau 1.5

Tab. 1.4 – Les sept (7) voyelles orales avec quatre degrés d'ouverture avec leur représentation en alphabet phonétique international (API)[10]

API	Orthographe	Exemple	Signification
[a]	a	a̲fo	Pied
[i]	i	ali̲	Voie
[e]	e	se̲	Entendre
[u]	u	wu̲	Peau
[o]	o	to̲	Oreille
[ɛ]	e	gbɛ̲]	Monde
[ɔ]	ɔ	bɔ̲	Alors

Tab. 1.5 – Les cinq voyelles nasales avec trois degrés d'ouverture et leur représentation en alphabet phonétique international (API)[10]

API	Orthographe	Exemple	Signification
[a]	a̱	xa̱o	Avec
[u]	u̱	u̱n	J'ai
[i]	i̱	nyi̱	Lancer
[ɛ]	e̱	mɛ̱	dans
[ɔ]	ɔ̱	zɔ̱n	Commander

Le système consonantique se compose 22 phonèmes [10] comme indiqué dans le tableau 1.6 [10].

Tab. 1.6 – Les consonnes avec leur représentation en alphabet phonétique international (API)

API	Orthographe	Exemple	Signification
[f]	f	f̱o	Finir
[t]	t	ṯa	Tête
[c]	c	c̱e	À moi
[s]	s	s̱o	Montagne
[kp]	kp	kp̱a	Barrière
[k]	k	ḵwe	Argent
[x]	x	x̱o	Parole
[v]	v	v̱i	Enfant
[d]	d	ḏo	Porter
[ʤ]	j	ʤ̱ i	Haut
[z]	z	ẕa	Balailler
[gb]	gb	gḇe	voix
[h]	ɣ	hu̱n	Sang
[b]	b	ḇa	Chercher
[y]	y	y̱e	Eux
[l]	l	aḻɔ	Main
[w]	w	w̱a	Faire
[m]	m	m̱i	Vous
[n]	n	ṉu	Chose
[ɖ]	d	ɖo	Avoir
[ɲ]	ny	ɲikɔ	Nom

Sauf les voyelles /i̱ / et /u̱ / et les consonnes / kp/, / c/ et / x/ qui lui sont propres, le Fongbé partage les mêmes sons avec le français. Son écriture est basée sur un ensemble de conventions dont les règles pratiques suivantes en sont dérivées :

✓ toutes les voyelles après les consonnes nasales sont systématiquement nasalisées et la marque de

nasalisation n'est plus écrite [10] ;
Exemple : [nu̱] ("chose"en français) s'écrit |nŭ| ;

- ✓ les voyelles nasales s'écrivent en remplaçant le tilde ( ) par la consonne /n/ [10].
  Exemple : [tá̱] ("marigot"en français) s'écrit |tán| ;
- ✓ la seule syllabique nasale dans le système phonétique Fongbé ñ s'est formée en combinant la voyelle /u/ avec la consonne /n/ [10].
  Exemple : [ñ min] ("j'ai braisé" en français) s'écrit |un min| ;
- ✓ toute voyelle sans ton est prononcée avec un ton moyen [10] ;
  Exemple : [m̱] ("vous" en français) s'écrit |mi| ;
- ✓ la voyelle /a/ venant devant un mot est toujours prononcée bas /à/ [10] ;
  Exemple : [àkpa ("côté" en français) s'écrit | akpa |.

Notons aussi que, lorsque les tons sont utilisés dans des phrases, ils modifient l'orthographe des mots. Ainsi, pour maîtriser l'écriture du Fongbé, il faut regarder le texte dans le sens de la structure interne des mots. Tous les mots peuvent être divisés en trois structures syllabiques différentes[10] :
monosyllabique
V → à ?(est-ce que ?) ,é (il, elle)

CV → tɔ (père), kpan (porter sur le dos)

dissyllabique
VCV → àwà (bras), àtàn (vin de palme)
CVCV → finlin (se rappeler), gali (gari : farine de manioc)
CVV → lɛ̌è (la manière dont...), wĭn (miel)
VV → éõ (non)
et trisyllabique.
VCVCV → àvivɔ (froid), àwewè (parcimonie)
CVCVCV → còkòtò (culotte)
VCVV → azĭì (arachide)
CVCVV → Béléú (rapidement)
CVVCV → cáunká (culotte)

La publication du premier dictionnaire Fongbé-Français a marqué le debut de la recherche scientifique en linguistique en 1963 [11]. Depuis 1976, plusieurs chercheurs linguistes ont travaillé sur le Fongbé et de nombreux articles sur les aspects linguistiques de la langue ont été publiés. Mais à l'opposé de la plupart des langues occidentales (français, anglais, espagnol, etc.), asiatiques (Chinois, Japonnais, etc.) et africaines (Wolof, Swahili, Haussa), le Fongbé manque cruellement de données linguistiques sous forme numérique (corpus audio et de textes) malgré le grand nombre d'ouvrages linguistiques (phonologie, lexique et syntaxe).

1.1.2.3 Le Bariba et sa structure vocalique

Le bariba est une langue nationale africaine et est aussi l'une des langues du groupe voltaïque (gour ou gur), de la grande famille des langues Niger-Congo, parlée majoritairement dans les régions septentrionales du Bénin par environ 10% de la population selon le recensement général de la population et de l'habitation (2002). Son écriture se base sur le latin avec les caractères de l'API. Il dispose d'un système tonal de quatre ton ponctuel et de deux tons modulés : un descendant et un montant [12]. Le système vocalique du bariba comprend donc sept (7) et quinze (15) consonnes et les syllabes phonétiques sont essentiellement de type CV avec une seule consonne en position post vocalique. En bariba, les seules consonnes géminées sont les nn et ll. Il existe aussi en bariba une nasalisation conditionnée de la voyelle par les consonnes nasales et un sous-système de phonèmes vocalique nasalles dialectes du songhay : contribution des changements linguistique pg 260.

Les types syllabiques suivants sont enregistrés : CV, CVV et CVC. Parmi les morphèmes verbaux, seuls les radicaux peuvent avoir la structure CVC. Les exemples suivants illustrent les types syllabiques du bariba :

(1) a. CV ye 'cuir'

b. CVV díí.rì 'trembler'

c. CVC.V seb.e 'vêtir'[13]

1.1.2.4 Comparaison de ces trois langues

Langue	FONGBE	YOURUBA	BARIBA (BAATONUM)
Voyelles	12	12	7
Consonnes	22	17	15
Tons	Haut, Moyen, Bas, Modulé	Haut, Moyen, Bas	Haut, Moyen, Bas, Modulé, moyen-descendant, descendant et montant
Type syllabique	V, CV, VCV, CVCV, CVCVCV	VV, VCV, CVV, CVCV	CV, CVV, CVC.V
Typologie syntaxique	SVO	SVO	SOV
Expression des Temps	Passe : ko, futur : na [71]	Passe : se, futur : yio [72]	Passe : gui, futur : do
Caractères particuliers	ɔ, GB, KP, DJ, ξ, SH, NY	ɔ,ξ, SH, GB, KP, DJ	ɔ,ξ, SH, GB, KP

Tab. 1.7 – Tableau de comparaison des langues

De l'analyse de ce tableau, nous pouvons constater que malgré le fait que les types syllabiques et les caractères particuliers respectifs ne sont pas les mêmes, ces trois langues ont en commun grand nombre d'éléments. En outre, une chose intéressante est le fait que pour ces langues, l'expression des temps se

fait par l'ajout de certains mots sans tenir compte du sujet. Aussi, alors que le Fongbé et le Yoruba sont de type SVO le Bariba est de type SOV

1.1.2.5 Notion de langue proche

La notion de « distance » en linguistique est une image qui présente un rapport entre différences et ressemblances. Cette image suggère et implique une possibilité de quantification. On y fait souvent recours dans les méthodes lexicologiques en faisant ressortir par exemple des variétés « proches » par l'importance du vocabulaire que des langues ont en commun, et de fonder des hypothèses historiques et génétiques [73]. Elle permet aussi de mettre en rapport des données linguistiques avec des données de génétique des populations, en y relevant des correspondances significatives [73]. on y fait aussi recours en dialectométrie qui, en plus du lexique, se base également sur la phonétique, et à partir de décomptes des différences géolectales, elle fait ressortir des structurations sous-jacentes de l'espace linguistique, auxquelles peuvent être confrontées des données qualitatives et historiques [73]. Cependant, lexique et la phonétique ne sont suffisants pour exprimer toute la complexité lier aux langues et a la distance entre elles. Alors, la grille de classification suivante, qui repose sur trois critères : a) la « parenté », c'est-à-dire, l'existence d'une même variété ascendante, dont on pourra préciser la « distance génétique », b) la « distinction », c'est-à-dire l'existence d'un consensus socio-politique (objectivable en discours et objectivé par des institutions) sur l'existence de deux langues distinctes, et c) l'intercompréhension, ont été fixer provisoirement pour mieux exprimer cette notion de « distance »[73].

Notion d'Énonciation (en anglais Utterance) en NLP Ce sont les entrées des utilisateurs qui doivent être traduites. Les énoncés sont comme des clauses. Tout ce que l'utilisateur dit est un énoncé. Nous utilisons l'énoncé pour entraîner la NLP afin qu'elle puisse identifier correctement l'intention de l'utilisateur[74].

Notion MFCC —Mel Frequency Cepstral Co-efficients Dans l'analyse conventionnelle des signaux temporels, toute composante périodique (par exemple, les échos) apparaît sous forme de pics aigus dans le spectre de fréquence correspondant (c'est-à-dire le spectre de Fourier. Ceci est obtenu en appliquant une transformée de Fourier sur le signal temporel). Cela peut être vu dans l'image suivante[75].

En prenant le log de l'amplitude de ce spectre de Fourier, puis en reprenant le spectre de ce log par une transformation en cosinus, nous observons un pic partout où il y a un élément périodique dans le signal temporel d'origine. Puisque nous appliquons une transformée sur le spectre de fréquence lui-même, le spectre résultant n'est ni dans le domaine fréquentiel ni dans le domaine temporel [14] ont décidé de l'appeler le domaine quefrence [1](en anglais quefrency). Et ce spectre du log du spectre du signal temporel a été nommé cepstre(en anglais cepstrum).

La figure 1.2 est un résumé des étapes expliquées ci-dessus. La hauteur tonale est l'une des caractéristiques d'un signal vocal et est mesurée comme la fréquence du signal. L'échelle Mel est une échelle qui

1. transformée sur le spectre de fréquence

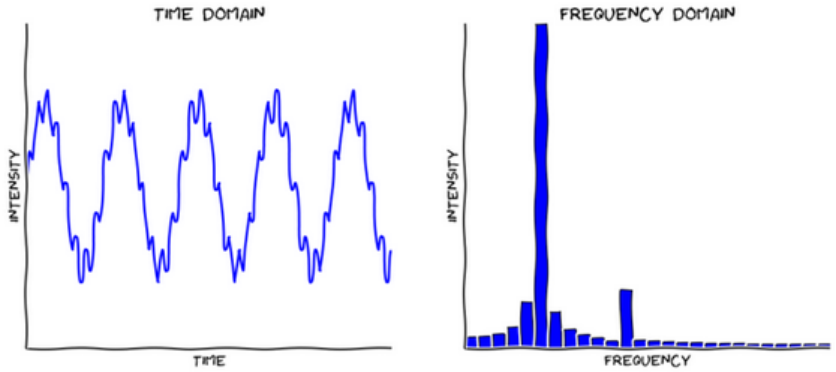

Fig. 1.1 – Transformer de Fourier du spectre de la fréquence

relie la fréquence perçue d'une tonalité à la fréquence réelle mesurée. Il l'échelonne la fréquence afin de la faire correspondre plus étroitement à ce que l'oreille humaine peut entendre (les humains sont mieux à même d'identifier les petits changements dans la parole à des fréquences plus basses), elle a été dérivée d'ensembles d'expériences sur des sujets humains[76].

1.2 Les approches de traduction automatique

1.2.1 Traitement Automatique du Langage Naturel TALN

1.2.1.1 Définition et utilité générale

Le traitement automatique du langage naturel (TALN) est une branche de l'intelligence artificielle qui aide les ordinateurs à comprendre, interpréter et manipuler le langage humain. Le TALN s'inspire de nombreuses disciplines, dont l'informatique et la linguistique computationnelle, et cherche combler le fossé entre la communication humaine et la compréhension informatique [77] au travers d'outils :

- ✓ de question réponse, comme ce que font What Siri, Alexa, et Cortana do,
- ✓ d'analyse des sentiments,
- ✓ de reconnaissance vocale,
- ✓ d'agent conversationnel,
- ✓ de classification des documents,
- ✓ de synthèse automatique de documents,
- ✓ de cartes d'images en générant des légendes pour les images d'entrée
- ✓ et la traduction automatique, sur laquelle cette étude se concentrera [78].

1.2.1.2 La traduction automatique

La traduction automatique, c'est avoir une application, un système, qui échange un texte ou contenu audio d'une langue source vers une autre langue cible, ou simplement changer son format de texte en un format

audio, et vice versa, sans aucune intervention humaine. Bien que cette définition soit simple, c'est un vrai défi, impliquant beaucoup de concepts et en créant de nouveaux lui correspondant.

Les approches de traduction automatique Il existe quatre principaux types d'approche de traduction automatique. Il s'agit de la traduction Speech to Text, qui traduit toutes les phrases prononcées dans une langue A en texte écrit correspondant dans la même langue A, de la traduction texte to texte, qui traduit tout texte d'une langue A en texte correspondant d'une langue B, de la traduction Text to Speech qui traduit tout texte d'une langue A en phrase audio correspondante dans la même langue A et de la traduction Speech to Speech qui traduit toutes phrases audio d'une langue A en Phrases correspondantes dans une autre langue B.

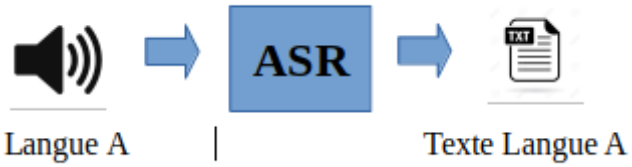

Fig. 1.2 – Système Speech to Text

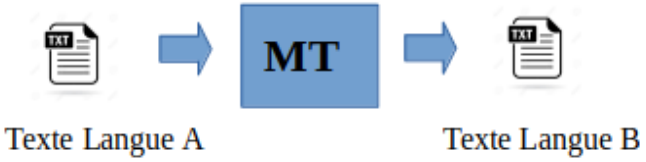

Fig. 1.3 – Système Text to Text

Fig. 1.4 – Systeme Text to Speech

Fig. 1.5 – Système Speech to Speech

1.2.2 Le Speech To Speech en Traitement Automatique du Langage Naturel

1.2.2.1 Définition

Les systèmes speech to speech translation sont des systèmes de traduction qui reçoivent à l'entrée des textes audio dans une langue source et les traduisent en texte audio correspondant dans une langue cible. Elles ont été développées au cours des dernières décennies dans le but d'aider les personnes qui parlent différentes langues à communiquer entre elles et comble le fossé linguistique dans le commerce mondial et interculturel [57].
Les systèmes de traduction speech to speech disposent déjà de plusieurs applications performantes qui sont de nos jours très employés. Nous pouvons énumérer par exemple Google Translate qui est le système de traduction le plus fréquemment utilise par les internautes et qui supporte plus de 100 langues, nous pouvons également citer Verbmobil avec les trois langues, à savoir l'allemand, l'anglais et le japonais qu'il supporte offre dans un contexte de téléphonie mobile la possibilité de traduire les échange de deux correspondants pour que chacun le reçoive dans sa langue. Nous avons aussi le Microsoft Translator qui est un système qui peut servir pour les besoins personnels ou d'entreprise dans les travaux collaboratifs que Microsoft a mise en place [80].

1.2.2.2 Les approches de méthode Speech To Speech en TALN

Deux approches de méthode se dégagent des multitudes recherche ayant développé un système Speech To Speech au cours de cette dernière décennie. Il s'agit de la méthode en cascade qui est celle-là plus traditionnelle et de la méthode directe qui est la plus récente [59].

Approche en Cascade L'approche traditionnelle en Speech To Speech consiste à séparer en trois modules le processus de traduction. Il s'agit de la reconnaissance vocale (en anglais automatic speech recognition ASR), la traduction automatique (en machine translation MT) et la synthèse du texte en signal vocal (en anglais text-to-speech TTS), qui sont toutes entraînées et réglées indépendamment. Compte tenu de l'entrée vocale, ASR traite et transforme la parole en texte dans la langue source, MT transforme la langue source texte au texte correspondant dans la langue cible, et enfin TTS génère de la parole à partir du texte dans la langue cible. Grâce aux différentes avancées du Deep Learning en NLP, des performances impressionnantes sont désormais observées dans les différents composants des taches entrant

dans la traduction automatique. D'énorme progrès sont faits dans cette approche au point ou de multiple système de traduction commercial existe déjà pour de multiples langues [15]. Cependant, cette approche rencontre un problème majeur. C'est le fait que, plus de la moitie des langues du monde n'ont pas de forme écrite, rendant leur traduction impossible du fait que cette approche telle qu'elle se présente nécessite pour chaque étape une quantité importante de ressource écrite. L'approche du Deep Learning en NLP, offre la possibilité d'apprendre une correspondance directe entre la longueur variable de la source et les séquences cibles qui ne sont souvent pas connues a priori. Plusieurs travaux ont étendu la couverture des modèles séquence à séquence du Deep Learning en NLP, l'approche directe en effectuant directement une traduction vocale de bout en bout en utilisant uniquement une architecture de réseau neuronal unique au lieu de se concentrer séparément sur ses composants (ASR, MT et TTS)[15].

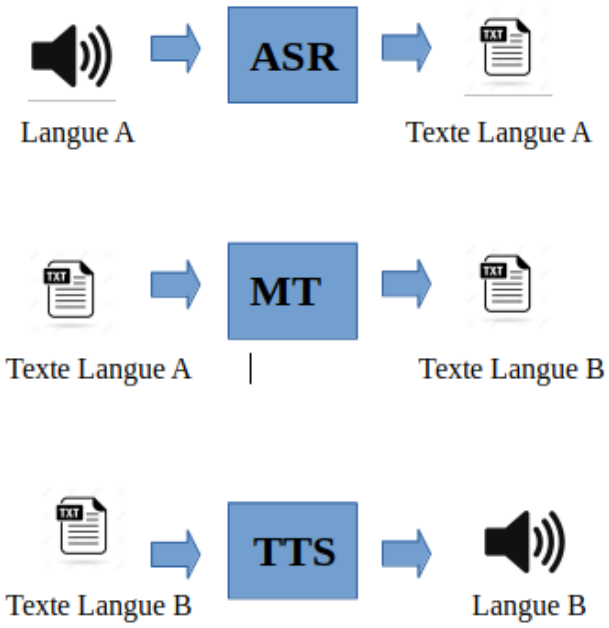

Fig. 1.6 – Les trois étapes de l'approche en cascade

Approche Directe L'approche directe permet de traduire la parole d'une langue en celle d'une autre langue sans compter sur la génération de texte intermédiaire.
Le développement rapide de l'approche direct est motivé par ses avantages théoriques et pratiques, à savoir :
- pendant la phase de traduction, il a accès aux informations présentes dans l'audio qui sont perdues dans ses transcriptions (ex. caractéristique de la voix du locuteur),
- il n'y a pas de propagation d'erreur (en cas systèmes cade les erreurs introduites par l'ASR sont propagés au MT, etc),
- la latence est plus faible (car les flux de données via un système unique au lieu de deux), - la gestion est plus facile (car il existe un modèle unique à maintenir et aucune intégration entre les différends module

est nécessaire).
Nous y distinguons trois approches. La première orientation de l'approche directe a été une traduction de donnée audio en langue source vers des données textuelles. Le premier de ce modèle réussi a été produit par Bérard [16] pour la synthèse de corpus French-English. Au vu de la nécessiter de ressource écrite dans la langue source, une deuxième orientation s'est établie et se base sur l'analyse des spectrogrammes des données audios de la langue source et de la langue cible accompagnée de réseaux de décodeur pour prédire les phonèmes nécessitant ainsi des ressources textuelle. Ainsi donc, une troisième orientation a été introduite par le projet nommé "Zero Ressource"d'une communauté ou uniquement des données audio sont entraînées en vue de la traduction automatique[15].

Fig. 1.7 – Système Speech to Speech

1.2.3 Les corpus parallèles bilingues

Un corpus parallèle est un ensemble de textes accompagnés de leurs traductions dans une autre langue. Comme le note Véronis J.[18], la systématisation de l'exploitation de ce type de corpus en NLP ne date que de la fin des années 1980. L'existence de textes parallèles remonte à la plus haute Antiquité : en attestent les inscriptions bilingues des tombes des princes d'Éléphantine en Égypte, qui datent du troisième millénaire avant J.- C., bien avant la pierre de Rosette (196 av. J.-C.) [17]. L'usage des textes parallèles pourrait donc être aussi ancien que la pratique de la traduction écrite. D'ailleurs, durant l'antiquité, certains textes sacrés, déjà, étaient présentés dans des versions bilingues parallèles, afin d'en faciliter l'accès et leurs études : c'est par exemple le cas d'une des plus anciennes versions des Évangiles, le Codex Bezae Cantabrigiensis, que l'on date vers la fin du IV^{e} siècle.

L'idée de rassembler des corpus de textes traduits dans une perspective de recyclage des traductions est apparue à la fin des années 1970, entre le Xerox Parc et la Brigham Young University. Le début des années 1980 verra également la constitution du premier corpus parallèle bilingue de grande envergure, le corpus Hansard, qui regroupe des textes issus du Sénat canadien. Le terme de corpus parallèle s'est peu à peu imposé dans les années 1990, la propriété géométrique du parallélisme désignant par analogie une propriété caractéristique de la traduction : sa compositionnalité –que Pierre Isabelle [19], définit ainsi « (...) les traductions obéissent à un principe dit de compositionnalité : la traduction d'un segment complexe est

généralement une fonction de la traduction de ses parties, et ce, jusqu'au niveau d'un ensemble d'unités élémentaires». Notons ainsi donc que le parallélisme implique que les segments de texte issus de cette décomposition se succèdent dans un même ordre. Ainsi, deux idées sous-tendent en général la notion de parallélisme :

- ✓ La compositionnalité : la relation d'équivalence traductionnelle, globalement mise en jeu entre deux textes, peut se décomposer au niveau de segments plus petits (Ex. Des chapitres, des paragraphes, des phrases, etc), également équivalents sur le plan de la traduction ;
- ✓ La séquentialité : les segments équivalents apparaissent dans le même ordre dans la cible et dans la source [20].

Notons qu'un corpus parallèle ne contient pas nécessairement le texte original en langue source. Teubert [21] donne une définition générale indiquant différents cas de figure que l'on peut rencontrer : Un corpus parallèle est un corpus bilingue ou multilingue qui contient un ensemble de textes en deux langues ou plus. Il y a plusieurs cas de figure, parmi lesquels :

- ✓ un corpus parallèle contient une quantité égale de textes originaux dans les langues A et B, et leurs traductions respectives ;
- ✓ un corpus parallèle contient seulement des traductions de textes dans des langues A, B et C, originellement écrits dans une langue Z.

Alors pour faciliter, la construction des systèmes de traduction automatique, plusieurs bases de données de corpus parallèle de langues à ressource ont été construites dont quelques-uns sont :

-le corpus BTEC (Basic Travel Expression Corpus) qui regroupe plus de 160 000 phrases employées souvent par des touristes, qui existe pour plusieurs langues dont le français, l'anglais, le japonnait et le chinois,
-le corpus grande échelle de nature « conversationnelle » de texte parallèle en audio et en texte, le corpus espagnol de conversations téléphoniques de Fisher et sa traduction anglaise correspondant,
-le corpus de la traduction de la bible qui existe en plus de 102 langues et plusieurs autres. Cependant, la majorité des bases de donne existant ne sont pas libres d'accès et très peu sont de taille conséquente pour la construction de système de Speech to Speech translation.

1.3 Les méthodes de traduction automatique

La traduction automatique est une traduction d'une langue source en une langue cible, obtenue au travers d'un système informatique (algorithme) plus ou moins sans intervention humaine. Les défis a relevé dans ce domaine de la traduction automatique ne sont pas des moindres face à la diversité, de langue, d'alphabet, de grammaire. La difficulté est encore plus grande pour les machines qui ne fonctionne qu'avec des chiffres, de travailler sur des lettres. Mais encore, à l'exemple des langues sans descendance de genre où «il» et «elle» sont désignés de la même façon, il n'existe pas une réponse correcte.

Au travers des années, trois grandes approches de traduction automatique ont vu le jour. Il s'agit de la

traduction automatique basée sur des règles qui a prédominé 1970 à 1990, de la traduction automatique statistique qui a prédominé de 1990 à 2010, de la Traduction automatique neuronale qui prédomine depuis 2014 [79].

1.3.1 La traduction automatique basée sur des règles

Le système de traduction automatique basé sur les règles est un système de traduction dans lequel des programmeurs informatique implémentent des règles linguistiques de traduction entre les deux langues préétablies par des experts en linguistique. Il est donc nécessaire pour ces experts d'avoir de bonnes connaissances de la langue source et de la langue cible pour développer des règles syntaxiques, sémantiques et morphologiques pour réaliser la traduction. Elle offre l'avantage de ne nécessiter aucun document, de texte source et de texte traduit, d'être totalement contrôlable et réutilisable pour des traductions de petite quantité donnée. Mais comme inconvénient, la nécessité d'avoir un bon dictionnaire, la nécessité d'avoir une expertise des deux langues afin d'établir manuellement les réglés et le fait qu'il devienne de plus en plus difficile de maintenir les performances du système quand les règles augmentent. Apertium et GramTrans sont deux exemples encore d'actualités de ce système [79].

1.3.2 La traduction automatique statistique

Le système de traduction automatique statistique est un système qui, au travers de l'analyse des données de corpus parallèle, défini des modèles statistiques capables de traduire des textes d'une langue vers une autre et, à partir des poids statistiques des phrases choisies, la phrase la plus probable. Google Translate et Microsoft Translator se basait sur ce système de traduction jusqu'en 2016 [79].
L'approche de la traduction automatique statistique nécessitant donc un corpus de texte bilingue fut introduite en 1955 [22]. Cette technique a pris de l'importance en 1988 suite à son utilisation par IBM Watson Research Center [23] pour la traduction par un alignement mot à mot. L'idée derrière cette approche : Étant donné une phrase T dans une langue cible, nous cherchons la phrase S à partir de laquelle le traducteur a produit T. Nous savons que notre risque d'erreur est minimisé en choisissant la phrase S la plus probable étant donné T. Ainsi, nous souhaitons choisir S donc pour maximiser $Pr(S|T)$ 1.1 [24]. Le théorème de Bayes, nous permet de transformer ce problème de maximisation en produit de $Pr(S)$ et $Pr(T|S)$, où $Pr(S)$ est la probabilité du modèle de langage de S, S étant la bonne phrase à cet endroit et $Pr(T|S)$ est la probabilité de traduction de T étant donné S. Ainsi, nous recherchons la traduction la plus probable compte tenu de l'exactitude d'une traduction candidate et de sa pertinence dans le contexte.

$$Pr(S/T) = \frac{Pr(S)Pr(T/s)}{Pr(T)} \tag{1.1}$$

L'évolution de cette approche a permis le développement d'une stratégie de traduction de plusieurs unités de mot (phrase) [25]. Ainsi donc, elle offre l'avantage de très peu nécessiter l'intervention manuelle d'un expert en linguistique contrairement à l'approche basée sur les règles. On peut réutiliser pour plusieurs autres paires de langue un système établis. Cependant, elle a pour désavantage de ne pas être adapté, de

ne pas être convenable pour les paires de langue avec de grande différence dans leur structuration des mots, de nécessiter de texte bilingue et la difficulté de corriger les erreurs spécifiques.

1.3.2.1 L'approche de l'apprentissage automatique du modèle de Markov caché

Le modèle de Markov caché, Hidden Markov Model (HMM) en anglais, modélise un automate d'états cachés dans lequel chaque état a une certaine probabilité de transition vers chacun des autres états ; chaque transition engendre une observation, l'observation suit une loi de probabilité associée à l'état courant. Les observations peuvent être discrètes, dans ce cas à chaque état sera associé la probabilité d'effectuer l'observation de chacun des symboles discrets possibles ; ou elles peuvent être continues, dans ce cas, on associe à chaque état une fonction de densité (souvent un modèle de mélange gaussien). Généralement, on note un HMM sous la forme d'un triplet : $= (A, B,)$. A est la matrice de probabilité de transition de chaque état vers chaque état, B est l'ensemble des fonctions de probabilités des observations associées à chaque état, est le vecteur des probabilités d'émission initiales, $((i)$ est la probabilité d'être à l'état i à l'état initial)[26].

1.3.3 La traduction automatique neuronale

Un réseau de neurones est un algorithme qui, par ses résultats, reproduisent ou prédisent le comportement de tout processus aussi fidèlement que possible, en fonction des facteurs qui déterminent ce comportement. "Processus" désigne tout système, naturelles ou créées par l'homme, et « facteur » pour toute grandeur qui peut avoir une influence sur le processus [27].
Ainsi, la traduction automatique par réseau de neurones, utilise un réseau de neurone pour effectuer la traduction en mettant en réseau un ensemble de données statistiques tirées du traitement de données textuelle bilingue.

Le premier modèle neuronal pour le langage a été proposé en 2001 [28], visant à prédire le mot suivant d'un texte à partir de représentation vectorielle des n premiers mots. Ces vectorielles de mots conservent les caractéristiques fondamentales que l'ordinateur pourra donc utiliser. Cela a été utilisé dans les claviers intelligents qui proposent l'écriture finale des mots entamés, proposant les mots suivants et les suggestions de mot dans les mails [29].

Après avoir été proposé par Rich Caruana en 1993 et appliqué à la surveillance des routes et à la prévision de la pneumonie [30], en 2008, l'apprentissage multitâche, une méthode de partage de paramètres entre des modèles entraînés sur plusieurs tâches, fut utilisée en NLP en liant les poids de différentes couches [31]. L'apprentissage intuitif et multitâche se traduit donc par des modèles apprenant des représentations utiles pour de nombreuses tâches.

En 2013, Amélioration de la représentation vectorielle des mots pour faciliter l'atteinte de certains objectifs. Ainsi, deux possibilités de représentation sont apparues pour deux buts opposés. La première consiste à prédire un mot central $w(t)$ à partir des mots environnants $(w(t-2), ...W(t+2))$, ce qui est très utile, par exemple, pour garder le sens des textes et apporter des corrections orthographiques, syn-

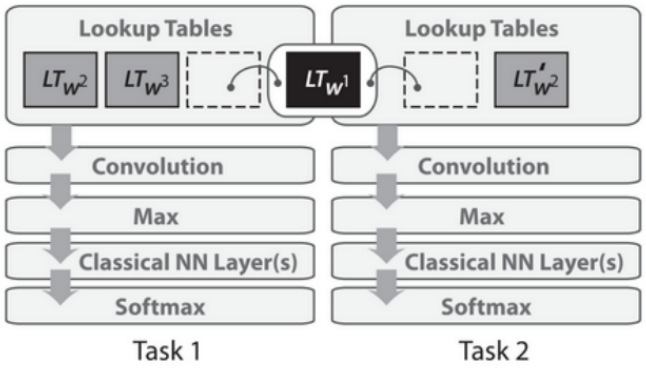

Fig. 1.8 – Partage de représentation vectorielle entre deux tâches [31]

taxiques, et la seconde consiste à prédire les mots $(w(t-2), ...W(t+2))$ à partir d'un mot central $w(t)$, pour faire par exemple des résumés de texte. Par conséquent, pour le traitement de grande donnée, elle capture les relations entre les mots tels que le genre, le temps des verbes et les relations entre les mots [32]. Il a été démontré que l'utilisation de ces représentations améliore les performances de grand nombre de tâches.

1.3.3.1 Le réseau de neurone convolutif

Cette approche est basée sur des neurones représentant des fonctions appliquées à la constitution de mots pour extraire les relations profondes des paramètres qui existent entre eux. Par conséquent, ces extraits peuvent être utilisés pour déterminer les règles de traduction qui dirige la traduction automatique, par exemple. Une des premières méthodes développées par Collobert et Weston [28] visait à la conversion du mot en une représentation vectorielle via une table de recherche donne, la méthode primitive d'incorporation de mots dans des vecteurs qui apprennent des poids pendant l'entraînement du réseau. Il s'agit donc d'appliquer des filtres de convolution à la matrice de mots de dimension d de donnée textuelle disponible. S'ensuit la réduction de la dimension de la sortie pour obtenir la représentation de la phrase finale par l'opération de regroupement maximum (Max pooling) en appliquant un filtre maximum à la dérivée de la représentation initiale [81].
Les figures 1.9 et 1.10 pressentent respectivement un modèle réseau de neurone convolutif et les étapes de conception d'une représentation de phrase en avec le réseau de neurone convolutif.

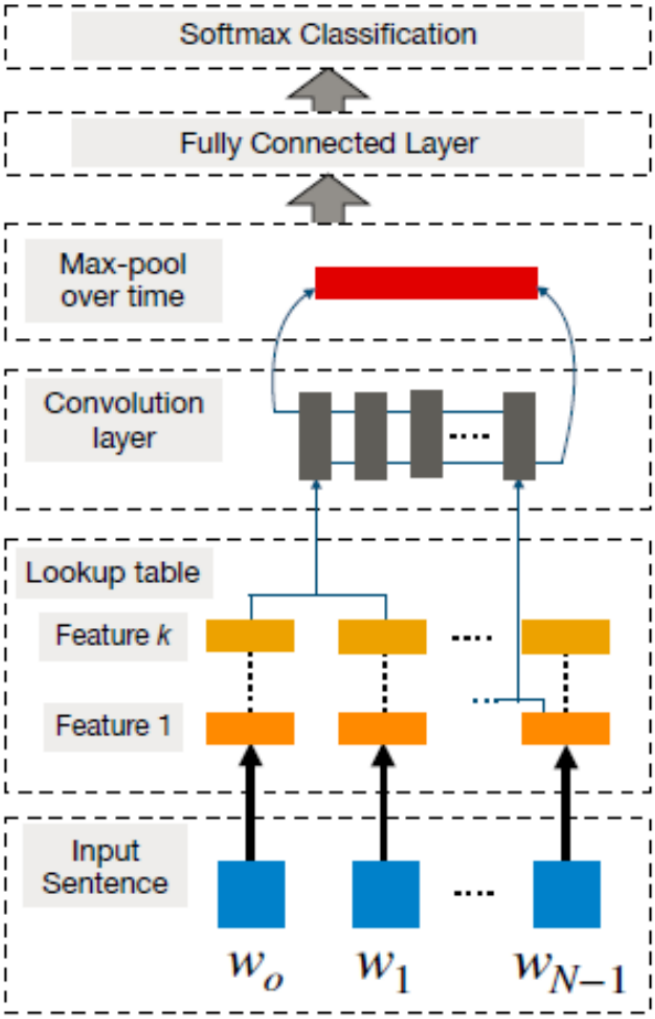

Fig. 1.9 – Modèle réseau de neurones convolutif avec table de recherche [81]

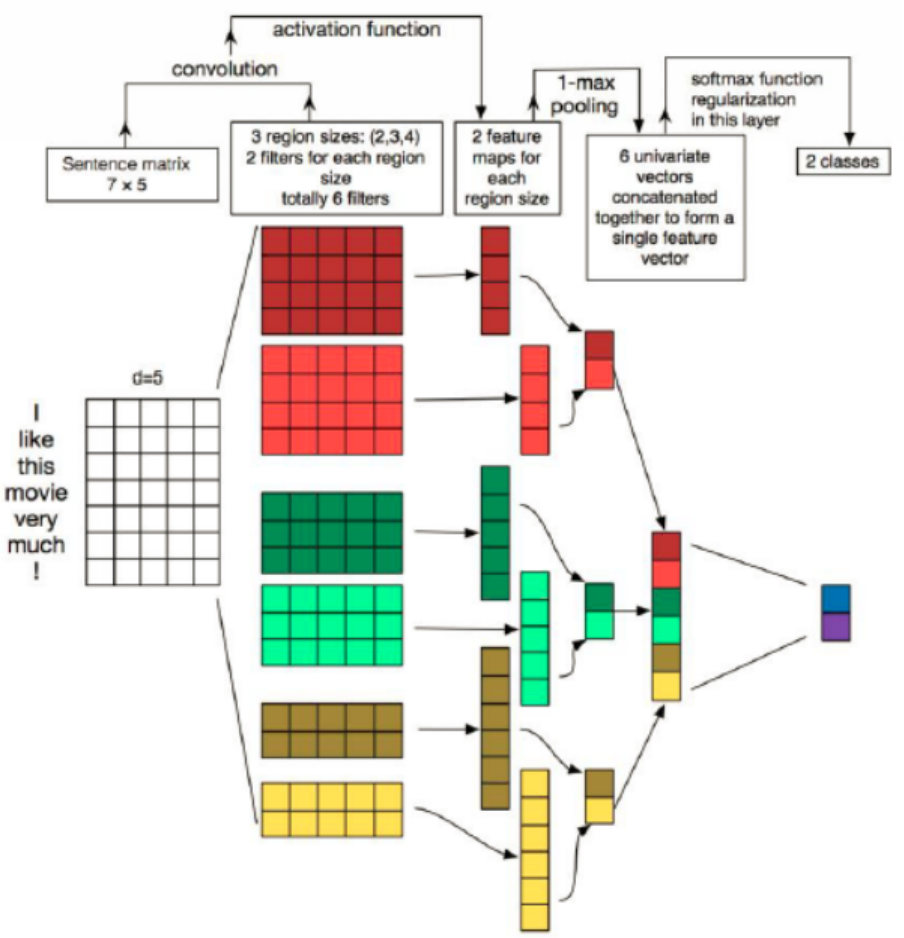

Fig. 1.10 – Étape de conception d'une représentation de phrases en avec le réseau de neurones convolutif [81]

Cette approche a obtenu un grand succès dans les tâches telles que la reconnaissance des noms d'entier ou name entity recognition (NER), le part of speech (POS), détection d'aspect, mais n'est pas beaucoup efficace sur d'autre tâche comme la traduction automatique. C'est parce qu'il ne peut représenter efficacement les dépendances distantes (les relations entre entités lointaines). De meilleurs résultats sont obtenus par des méthodes telles que les réseaux de neurones convolutifs dynamiques ou le couplage de réseaux de neurone convolutifs avec des réseaux de neurones temporisés, mais toujours sans grande amélioration pour le domaine de la traduction automatique [81].

1.3.3.2 Le réseau neuronal récurrent RNN

Un RNN est une méthode neuronale spécialisée qui peut traiter efficacement les informations séquentielles. Un RNN applique de manière récursive le calcul à chaque instance de la séquence d'entrée conditionné sur le résultat d'un calcul précédent, lui permettant de retenir les dépendances lointaines entrer les mots. Ces séquences sont généralement représentées par un vecteur de lexème de taille fixe qui sont transmis séquentiellement (un par un) à une unité récurrente. La figure 1.11 illustre un cas RNN simple.

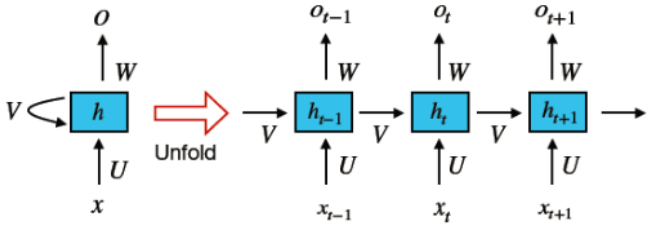

Fig. 1.11 – réseau neuronal récurrent RNN simple
[81]

Avec x l'entrée et U sa fonction d'entrée, o la sortie et W sa fonction de sortie,h et V sa fonction d'état et t le temps.

Le principal avantage d'un RNN est sa capacité à recevoir des informations, jusqu'à un certain degré, des résultats des calculs précédents, (une courte mémoire) et à utiliser ces informations dans le calcul en cours. Cela rend les modèles RNN adapté pour intégrer les dépendances de contexte dans des entrées de longueur arbitraire afin de créer une composition appropriée de l'entrée. Les RNN sont utilisés pour effectuer diverses tâches TALN telles que la traduction automatique, le sous-titrage d'images et la modélisation du langage [81].

Les RNN simples souffrent du problème de fuite de gradient, lorsque le gradient devient trop petit, il empêche le poids de changer et au pire empêcher complètement le réseau de neurone de continuer à s'entraîner, car dans cette approche, chacun des poids du réseau de neurones reçoit une mise à jour proportionnelle à la dérivée partielle de la fonction d'erreur par rapport au poids actuel dans chaque itération d'entraînement. Ce phénomène rend difficile l'apprentissage et le réglage des paramètres dans les couches précédentes. D'autres variantes, telles que les réseaux de mémoire à long terme (long short-term memory, LSTM) et les réseaux récurrents fermés (GRU) ont été introduites ultérieurement pour surmonter cette limitation.

1.3.3.3 Les réseaux neuronaux LSTM et GRU

Les réseaux de neurone LSTM Un LSTM se compose de trois portes (portes d'entrée, d'oubli et de sortie) et de deux états (état caché et l'état de la cellule) [82]. La porte de
l'oubli détermine, à partir de l'état caché précédent et de l'entrée actuelle, quelles informations doivent être conservées ou supprimées, tans disent que la porte d'entrée détermine les informations importantes pour l'entraînement des nouvelles entrées que fourni l'état caché. L'état de la cellule est simplement déterminé par la porte de l'oubli et la porte d'entrée. Les grandeurs de la sortie de l'oubli son multipliée par l'ancien état de la cellule, permettant d'oublier certaines informations de l'état précédent qui ne sont pas utilisées pour faire de nouvelles prédictions [82]. Ensuite, la sortie de la porte d'entrée lui est ajoutée, ce qui permet de conserver ce que le LSTM considère pertinent (entre l'entrée et l'état caché précédent) dans l'état de la cellule. La porte de sortie doit décider de quel est le prochain état caché et transmettre un vecteur unique contenant des informations sur les entrées précédentes du réseau et utilisé pour la prédiction [82].

Les réseaux neuronaux GRU Le GRU (Gated Recurrent Unit) est similaire au LSTM mais ne contient que deux portes (porte de réinitialisation ou reset, porte de mise à jour) et l'état de la cellule en sortie, donc c'est moins compliqué [82]. La porte de réinitialisation contrôle la quantité d'information passée que le réseau doit oublier dans l'ancien état caché, tandis que la porte de mise à jour, (comme les portes d'oubli et d'entrée dans les LSTM) décident les informations à conserver ou supprimer pour l'entraînement. En sortie, nous avons une combinaison des rendus des portes de réinitialisation et de mise à jour avec l'entre du réseau et l'ajout des données de l'état caché précédent jugé utile pour l'entraînement dans un seul vecteur.

Une étude a montré qu'il est difficile de dire lesquels de ces RNN sont le plus efficace, et il est généralement choisi en fonction de la puissance de calcul disponible. Divers modèles basés sur le LSTM ont été proposés pour le mappage séquence à séquence (via des cadres encodeur-décodeur) pour des tâches telles que la traduction automatique, à la synthèse de texte, à la modélisation de conversations humaines, à la réponse aux questions, à la génération de langage basé sur l'image, entre autres tâches[82].

1.3.3.4 Le mécanisme d'Attention

Les performances des réseaux LSTM et GRU du RNN sur de petites séquences démontrent leur efficacité. lorsque la taille de la séquence devient grande, le vecteur unique envoyé au décodeur a une grande limite, ce qui est également dû au problème de fuite de gradient. Afin de pallier, voire de résoudre cette déficience, le mécanisme d'Attention, contrairement à la pratique antérieure, conserve sous forme de matrice le vecteur de données généré après chaque étape du réseau RNN, ce qui implique également le problème d'excès de données. Pour réduire ces données, une fonction d'alignement sera utilisée pour déterminer la relation entre les mots des séquences et les données non essentielles seront supprimées de la matrice. Ainsi donc, lorsqu'une seule relation est établie, on parle de Self attention, mais si plusieurs relations sont établies (relation de conjugaison, de grammaire) on parle d'attention multi-head, multitête en français. Ceci est particulièrement utile pour les tâches qui nécessitent un certain alignement entre le texte d'entrée et de sortie [83].

1.3.3.5 Le réseau de neurone transformer

Le Transformer est un modèle de Deep Learning type sequence to sequence (donc un réseau de neurones), et sa particularité est qu'il n'utilise que des mécanismes d'attention, pas de réseaux récurrents ou convolutifs. Le réseau Transformer est également basé sur le modèle Encodeur/Décodeur, mais il est différent. Ici l'encodeur et le décodeur sont respectivement une pile d'un même nombre d'encodeurs et de décodeur dont la sortie des premiers deviennent l'entrée des suivants. De plus, chaque décodeur reçoit la dernière sortie de l'encodeur comme entrer supplémentaire, sauf le premier décodeur qui l'a comme seul entre. L'entrée du premier encodeur est une représentation vectorielle de la séquence. L'encodeur est constitué de deux blocs (tous deux des réseaux de neurones) : Une couche dite de « Self-attention » et un réseau à propagation avant (ou Feed-forward Neural Network). La couche de Self-attention est l'élément central de l'architecture du Transformer. Son rôle est de maintenir l'interdépendance des mots dans la représentation des séquences. Le décodeur est également composé d'un bloc de Self-attention et d'un Feed-forward, mais il contient en plus une couche « Encoder-Décoder Attention », conçue pour laisser le décodeur effectuer l'attention entre la séquence d'entrée (encodage) et le mécanisme de séquence de sortie (décodage en cours). L'idée du Transformer qui est de préserver l'interdépendance des mots d'une séquence en n'utilisant seulement le mécanisme d'attention au centre de son architecture au lieu d'un réseau récurrent, fait qu'il est plus rapide à entraîner et beaucoup plus parallélisable [84].C'est à la vue de ces avantages que nouvel algorithme de recherche de Google dénommé BERT (bidirectional

encoder representations from Transformers) s'est basé dessus.

1.3.3.6 Modèles génératifs profonds

Des modèles génératifs profonds tels que les auto-encodeurs variationnels (en anglais variational autoenconders, VAE) et les réseaux antagonistes génératifs (en anglais generative adversarial networks GAN), sont également appliqués en TLAN pour découvrir une structure riche en langage naturel grâce au processus de génération de phrases réalistes à partir d'un espace de code latent.

Il est bien connu que les auto-encodeurs de phrases standard n'arrivent pas à générer des phrases réalistes en raison de l'espace latent non contraint. Les VAE imposent une distribution préalable sur l'espace latent caché, permettant au modèle de générer des échantillons appropriés. Les VAE sont constitués de réseaux de codeurs et de générateurs qui codent une entrée dans un espace latent puis génèrent des échantillons à partir de l'espace latent. L'objectif de la formation est de maximiser une limite variationnelle inférieure sur la vraisemblance logarithmique des données observées dans le modèle génératif[81].

1.3.3.7 Récapitulatif

1.4 Évaluation des systèmes de traduction automatique

La nécessité et l'importance de la traduction automatique signifient aussi qu'elle doit de pouvoir l'évaluer. Mais contrairement à d'autres problèmes du NLP, ou il existe de mesures de qualité généralement acceptées par les experts, l'évaluation des systèmes de Traduction Automatique, en particulier la qualité de la traduction, reste un problème difficile, sujette à de nombreux débats [33]. Dans le cas particulier de la Traduction Automatique, deux approches sont utilisées. Il s'agit de l'évaluation humaine et l'évaluation automatisées.

1.4.1 Évaluation humaine

Pour l'évaluation humaine de la traduction automatique, un nombre donné de participants évaluent chaque traduction selon certains critères prédéfinis. Les critères de qualité peuvent inclure des critères tels que l'exactitude grammaticale et la signification textuelle.

Les travaux de HTER nous fournissent une forme de mesure couramment utilisée pour les critères humains, qui est une mesure artificielle où les humains ne classent pas directement les traductions, mais génèrent à la place de nouvelles traductions de référence plus proches de la sortie du système, en conservant la fluidité du sens de la référence [34]. Ainsi, à partir de cette référence, nous pouvons calculer les erreurs produites par le système, en la comparant avec la traduction automatique. Cette approche est une véritable mesure de la qualité du système, mais nécessite une intervention manuelle coûteuse. De plus, toute évaluation subjective souffre de problèmes de non-reproductibilité et de variabilité inter-annotateur. C'est pourquoi plusieurs mesures automatiques ont été développées au fil des années. Leur objectif est

Tab. 1.8 – Les techniques de traduction automatique (API)

Les techniques de traduction automatique	Avantages	Faiblesses
Basée sur les règles	Première approche de traduction automatique, Facile à maintenir pour la traduction de petite donnée	Besoin d'intervention humaines, Difficile à maintenir pour de grands contenus, Prend en compte une paire de langues à la fois
Statistique	Réduis l'intervention humaine, Prend en compte plusieurs langues, Prendre en compte de plus grands contenus à traduire	Peu sensible aux spécificités des langues, Faible résultat pour des langues éloigné
Neuronal Convolutif	Pas d'intervention après établissement, N'est spécifique à aucune langue	Lent à entraîner, Pas efficace pour établir les relations entre entité éloignées, Besoin de beaucoup de ressource
Neuronal Récurrent	Pas d'intervention après établissement, N'est spécifique à aucune langue, Établis les relations lointaines,	Lent a entraîné, Besoin de beaucoup de ressource, Problème de fuite du gradient
Neuronal Récurrent avec le mécanisme d'attention	Pas d'intervention après établissement, N'est spécifique à aucune langue, Établis les relations lointaines,	Lent a entraîné, Besoin de beaucoup de ressource, Léger problème de fuite du gradient
Neuronal Transformer	Pas d'intervention après établissement, N'est spécifique à aucune langue, Établis les relations lointaines, Très rapide à entrainer, Plus efficace résultat du moment	Besoin de beaucoup de ressource

d'être corrélé avec les scores que produirait une évaluation humaine, tout en étant reproductible, beaucoup moins coûteuse et rapide pour pouvoir optimiser et comparer les systèmes.

1.4.1.1 Le Mean Opinion Score (MOS)

Le Mean Opinion Score (MOS) est une mesure numérique de la qualité globale perçue par humaine d'un événement ou d'une expérience. Dans les télécommunications, le score d'opinion moyen est un classement de la qualité des sessions voix et vidéo. Plus Généralement, jugés sur une échelle de 1 (mauvais) à 5 (excellent), le score d'opinion moyen est la moyenne de nombreux autres paramètres personnels notés par une personne. Alors que les scores d'opinion moyens d'origine étaient dérivés d'enquêtes auprès d'observateurs experts, aujourd'hui, un MOS est souvent produit par une méthode de mesure objective se rapprochant d'un classement humain [85].

1.4.2 Évaluation automatique

Les évaluations qui utilisent des mesures automatisées sont appelées évaluations automatiques ou évaluations objectives. L'évaluation automatique est l'un des principaux défis dans le développement d'un système de traduction automatique. C'est un domaine de recherche actif en soi [34], alors que dans de nombreux autres domaines, il est souvent considéré comme un problème résolu et trivial. Par exemple, en reconnaissance vocale, pour toute entrée, il n'y a qu'une seule sortie correcte. Cependant, dans la traduction automatique, il existe un ensemble de traductions correctes. Diverses méthodes ont été proposées dans le but d'automatiser l'évaluation de systèmes de traduction automatique. Cette évaluation nécessitait auparavant une intervention humaine, nécessitait un temps et des ressources considérables. Une solution actuellement employée consiste à générer un ou plusieurs scores pour refléter la similarité ou la distance entre la sortie du système et une ou plusieurs références et nous pouvons énumérer les scores WER, PER, TER, BLEU, NIST, METEOR.

1.4.2.1 Mesures reposant sur des ressemblances avec des références : Le score WER

Le score WER (en anglais : Word Error Rate) était à l'origine utilisé en reconnaissance automatique de la parole. Il compare l'hypothèse e_h à la référence e_r en se fondant sur la distance de Levenshtein au niveau de mots. Il compte le nombre minimum d'opérations à effectuer sur e h pour la convertir en e_r . Moins il y a d'opérations à effectuer, meilleur est le score. Nous pouvons calculer le score WER en utilisant la formule suivante :

$$WER(e_h) = \frac{n_{ins} + n_{sup} + n_{sub}}{|e_r|} \tag{1.2}$$

où $n_{ins}, n_{sup}, n_{sub}$ sont respectivement les nombres minimums d'insertions, de suppressions et de substitutions pour transformer e_hàe_r.
Lors de l'utilisation de plusieurs références, la formule peut être modifiée pour avoir le numérateur, le nombre minimum d'opérations pour toutes les références (la référence la plus proche est considérée) et le dénominateur qui devient la moyenne des longueurs de référence. Malheureusement, cette mesure simple

n'est pas très adaptée à la traduction, car elle pénalise les hypothèses correctes où l'ordre des mots ne correspond pas à la référence. Un mot qui est traduit correctement, mais qui est mis à la mauvaise position sera pénalisé par une suppression et une insertion par exemple. On peut ainsi assigner des scores WER différents pour deux hypothèses équivalentes «je vois un chat et un chien» et «je vois un chien et un chat» avec la même référence «je vois un chat et un chien».

1.4.2.2 Le score PER

Le score PER (en anglais : Position-independent Word Error Rate) est semblable au score WER, mais il ne prend pas en compte l'ordre des mots. Il considère l'hypothèse e_h et la référence e_r comme des ensembles de mots non ordonnés plutôt que des phrases totalement ordonnées. Le score PER a été proposé par Tillmann en 1997 [35] et ne compte que le nombre de fois que des mots identiques sont produits dans les deux phrases. Les mots qui ne correspondent pas sont comptés comme des substitutions. Selon que la phrase traduite est plus longue ou plus courte que la référence, le reste des mots est compté comme insertion ou suppression. Ainsi, le score PER assigne le même score pour les deux hypothèses «je vois un chat» et «un chat vois je» lorsque la référence est «je vois un chat».

1.4.2.3 Le score TER

Le score TER (en anglais : Translation Edit Rate) compte aussi le nombre minimum d'opérations à effectuer sur e_h pour la transformer en e_r . Comme le score WER, les opérations considérées sont l'insertion, la suppression, la substitution, mais aussi le déplacement d'une suite de mots [36]. Un déplacement permet de déplacer un groupe de mots contigus vers la gauche ou la droite. Chaque déplacement est compté comme une seule opération, quel que soit le nombre de mots déplacés et l'amplitude du déplacement. Le score TER est donc formulé comme suit :

$$TER(e_h) = \frac{n_{ins} + n_{sup} + n_{sub} + n_{dep}}{|e_r|} \tag{1.3}$$

où n_{dep} est le nombre minimum de déplacements et où $n_{ins}, n_{sup}, n_{sub}$ sont respectivement les nombres minimums d'insertions, de suppressions et de substitutions pour transformer e_h à e_r.

HTER (en anglais : Human-targeted Translation Edit Rate) est une version modifiée du score TER avec l'intervention des traducteurs humains [36]. Après avoir lu les références, les traducteurs humains éditent l'hypothèse du système pour générer une nouvelle phrase qui a la même signification que les références originales. Cette phrase est considérée comme une nouvelle référence humaine de l'hypothèse. Puis, le score HTER est le score TER minimum calculé entre l'hypothèse et les références originales plus la nouvelle référence humaine.
Le score HTER est moins subjectif que les jugements humains, mais il est encore coûteux, en ce que le traducteur perd environ de 3 à 7 minutes pour éditer chaque phrase.
La métrique d'évaluation TER-Plus (TERp) est une autre extension du score TER avec des paramètres ajustables et l'incorporation avec la morphologie, la synonymie et des paraphrases [37]. TERp aligne un mot de l'hypothèse avec un mot de la référence, non seulement quand deux mots sont les correspondances

exactes, mais aussi quand ils possèdent la même racine ou ils sont les synonymes. Plus, TERp utilise aussi la substitution de groupe de mots (des paraphrases) pour aligner deux phrases. Il utilise donc toutes les opérations de TER : l'insertion, la suppression, la substitution de mots, le déplacement d'une suite de mots ; et trois nouvelles opérations : la correspondance en racine, la correspondance en synonyme et la substitution de paraphrases. Le coût de toutes les opérations est optimisé afin de maximiser la corrélation avec les jugements humains.
Le score TERp permet de mieux aligner l'hypothèse et les références, mais le calcul dépend du dictionnaire de synonymes, de la liste de mots possédant la même racine, de la liste des paraphrases, qui ne sont pas toujours disponibles pour toutes les langues.

1.4.2.4 Le score BLEU

Depuis ces dernières années, la métrique la plus souvent utilisée est le score BLEU (en anglais : BiLingual Evaluation Understudy). Le score BLEU est proposé par Papineni [38]. Il ne considère pas seulement la ressemblance au niveau des mots, mais aussi la ressemblance au niveau des n-grammes entre l'hypothèse et les références.
La tâche principale est de comparer les n-grammes de l'hypothèse avec les n-grammes de la référence et de compter le nombre d'équivalences. Les correspondances sont indépendantes de la position. Plus il y a de correspondances, meilleure est l'hypothèse. Tout d'abord, les précisions modifiées de n-gramme Pn avec l'ordre de 1 à $N(n = 1...N)$ sont calculées pour chaque paire d'hypothèses et sa référence (ou ses références lorsque plusieurs références sont utilisées).

$$p_{\text{n chaque paire}} = \frac{\sum_{n-gram \in e_h} Compte_{clip}(n-gram)}{\sum_{n-gram \in e_h} Compte_{e_h}(n-gram)} \tag{1.4}$$

Pour un n-gramme donné, soient $Compte_{e_h}(n-gram)$ le nombre de fois que ce $n-gram$ apparaît dans e_h . Si nous notons c le nombre de mots de l'hypothèse e_h , e_h contient $c-n+1$ n-grammes. Le dénominateur devient $c-n+1$.

$Compte_{clip}(n-gram)$ est le nombre d'appariements de ce $n-gram$ entre e_h et e_r , donc il est calculé par : min($Compte_{e_h}(n-gram)$, max$_{er}$ ($Compte_{e_r}(n-gram)$)) où max$_{er}$ ($Compte_{e_r}(n-gram)$) est le nombre maximal de fois que ce $n-gram$ apparaît dans une référence, parmi toutes les références disponibles.
Pour calculer la précision n-gramme modifié sur le corpus de test entier, nous accumulons simplement les comptes pour chaque paire d'hypothèses et sa référence.

$$p_{\text{n corpus}} = \frac{\sum_{e_h \in corpus} \sum_{n-gram \in e_h} Compte_{clip}(n-gram)}{\sum_{e_h \in corpus} \sum_{n-gram \in e_h} Compte_{e_h}(n-gram)} \tag{1.5}$$

Pour combiner les N précisions n-grammes modifiés, le score BLEU utilise le logarithme moyen pondéré, ce qui est équivalent à une moyenne géométrique, et pour pénaliser les hypothèses plus courtes que leurs références, une pénalité de brièveté BP est introduite. Le score BLEU est finalement calculé comme suit :

$$ScoreBLEU = BP \cdot exp(\sum_{n=1}^{N} w_n \log p_n) \tag{1.6}$$

w_n sont les poids positifs tels que $\sum_{n=1}^{N} = 1$, souvent, nous utilisons des poids uniformes $w_n = 1/N$ et $N = 4$.

$$BP = \begin{cases} 1 & c > r_p \\ e^{1-r_p/c} & c \geq r_p \end{cases} \tag{1.7}$$

pour une paire de phrases, c'est la longueur de l'hypothèse e_h , et r_p est la longueur de la référence la plus proche de e_h parmi les références. Pour le corpus entier, la somme totale de c et la somme totale de r_p de toutes les hypothèses du corpus sont calculées. Dans le domaine des logarithmes,

$$\log ScoreBLEU = \min\left(1 - \frac{r_p}{c}, 0\right) + \sum_{n=1}^{N} w_n \log p_n \tag{1.8}$$

BLEU est un score de précision, sa valeur varie de 0 à 1. **Plus le score est élevé, meilleure est la traduction**. Une hypothèse se voit attribuer un score BLEU de « 1 » lorsqu'elle est identique à une des références ; au contraire, elle aura un score BLEU de « 0 » si aucun de ses n-grammes n'est présent dans une référence.

1.4.2.5 Le score NIST

Le score NIST (dont le nom vient de National Institute of Standards and Technology) a été proposé dans [39] et reprend le principe du score BLEU, mais avec quelques adaptations légères. Il repose aussi sur la précision n-gramme, mais il utilise la moyenne arithmétique des n-grammes au lieu de la moyenne géométrique. L'expression de la pénalité de brièveté est différente de celle utilisée pour le score BLEU. Dans le score NIST, les n-grammes sont aussi pondérés selon leur fréquence d'apparition : les n-grammes rares contribuent plus au score final que les n-grammes fréquents. Par exemple, le bi-gramme anglais *« interesting calculations »* contribue plus au score que le bi-gramme *« of the »* qui apparaît souvent en anglais.

Les poids $Info()$ d'un n-gramme $n - gram = m_1....m_n$ sur un ensemble de références sont calculés par

$$Info(n-gram) = Info(m_1....m_n) = \log_2 \left(\frac{compte(m_1....m_{n-1})}{compte(m_1....m_n)} \right) \tag{1.9}$$

où $compte(m_1....m_n)$ est le nombre de fois que le n-gramme $m_1....m_n$ apparaît dans l'ensemble.
Le score NIST est alors ainsi calculé :

$$ScoreNIST = \sum_{n=1}^{N} \left\{ \frac{\sum_{n-gram \in e_h \cap n-gram \in e_r} Info(n-gram)}{\sum_{n-gram \in e_h} compte(n-gram)} \right\} \cdot exp\left\{ \beta \cdot \log^2 \left[min\left(\frac{c}{\bar{r}}, 1 \right) \right] \right\} \tag{1.10}$$

où c est le nombre de mots de l'hypothèse e_r, $\bar{r}$ est le nombre moyen de mots de toutes les références et β est un facteur pour ajuster la pénalité de brièveté.

1.4.2.6 Le score METEOR

Les scores BLEU et NIST sont des scores de précision. Banerjee a proposé le score METEOR (en anglais : Metric for Evaluation of Translation with Explicit Ordering) qui équilibre entre la précision et le rappel[40]. Ce score est calculé sur la base d'un alignement entre les uni-grammes d'une hypothèse et ceux d'une référence. Un alignement est un ensemble d'appariements d'uni-grammes. Un uni-gramme d'une phrase est mis en correspondance avec zéro ou un seul uni-gramme d'une autre phrase. Les appariements sont établis d'abord sur les formes orthographiques, puis les mots de même racine (par exemple : «joli» et «jolie») et enfin sur les synonymes (par exemple : «joli» et «beau»). Il permet de valider la fidélité de la signification de l'hypothèse avec plusieurs choix de lexiques, différents de la référence. Le meilleur alignement est celui qui contient le plus grand nombre d'appariements d'uni-grammes avec le plus petit nombre de réarrangements. Le score METEOR est déterminé à partir de ce meilleur alignement.

$$ScoreMETEOR = F_{moyenne} \cdot (1 - \text{Pénalité}) \tag{1.11}$$

où $F_{moyenne} = \frac{P \cdot R}{\alpha \cdot P + (1-\alpha) \cdot R}$ et Pénalité $= \gamma \cdot \left(\frac{Compter(segment)}{Compter(unigram.align.)} \right)^{\beta}$

La précision P est le nombre d'uni-grammes appariés de l'hypothèse divisé par la taille de cette hypothèse, et le rappel R est ce nombre d'uni-grammes appariés divisé par la taille de la référence. La Pénalité favorise l'hypothèse qui contient des segments d'uni-grammes consécutifs appariés plus longs.
$Compter(unigram.align.)$ est le nombre d'appariements d'une hypothèse. $Compter(segment)$ est le nombre de segments (le plus long) de l'hypothèse mise en correspondance avec les segments de la référence. Les poids originaux sont α=0, 9 ; β=3 ; =0, 5.

Conclusion Ce chapitre nous a permis de voir que le domaine de la traduction automatique est un domaine très actif ou des efforts sont fournis pour mettre à la disposition des langues et surtout peu fourni des approches et méthodes nécessaires pour leur traduction.

Chapitre 2

Matériel et méthode

Introduction En vue de mener à bien cette étude, nous avons adopté une approche méthodologique.

2.1 Matériel

En vue de mener à bien nos travaux, plusieurs outils nous ont été nécessaires. Il s'agit des moteurs de recherche Google, Google Scholar, IEEE, Scopus, des technologies de recherche et d'analyse Blibliometrix et VOSviewer au travers de l'environnement de développement pour le langage de programmation utilisé pour le traitement de données et l'analyse statistique nome R.

2.1.1 Les techniques Bibliométriques

La scientométrie est officieusement définie comme la discipline qui étudie la quantification des paramètres et les caractéristiques de la science et de la recherche scientifique, de la technologie et innovation. Au sein de la Scientométrie, la Bibliométrie s'occupe de l'analyse statistique de livres, d'articles ou d'autres types de publications[41]. Par des méthodes mathématiques statistiques, les données bibliographiques sont traitées et les résultats sont présentés sous forme de tableaux et de graphiques. Par exemple, Vargas-Quesada et Moya-Anegón [42] ont proposé une méthodologie pour créer des représentations visuelles des domaines scientifiques. Ils se sont concentrés sur l'illustration des interactions entre auteurs et articles par le biais de citations et de cocitations. D'autres auteurs ont ensuite concrétisé l'idée et développé des méthodes et des outils alternatifs (par exemple, [[43], [44]]) pour créer cartes des éléments liés (publications scientifiques, revues scientifiques, chercheurs, organismes de recherche, pays ou mots-clés). Différents types de liens entre paires d'items peuvent être envisagés. Par exemple, introduisons brièvement le concept de cocitation d'items. Considérant un ensemble d'items, tous les liens potentiels entre paires d'items peuvent être caractérisés par la mesure de cocitation standardisée [45] comme suit :

$$MCN_{ij} = \frac{Cc_{ij} + n_{sup} + n_{sub}}{\sqrt{c_i . c_j}} \tag{2.1}$$

où Cc signifie cocitation, c signifie citation, i et j, sont deux items différents. Les valeurs de lien (MCN_{ij}) définissent la matrice d'adjacence d'un graphe qui peut être analysé et visualisé avec des techniques d'analyse des réseaux sociaux (SNA) [46]. Ces techniques ont déjà été appliquées à de multiples domaines de recherche, tels que développement de logiciels (par exemple, débogage de systèmes multiagents [47]), scientométrie (par exemple, l'analyse de grands domaines scientifiques [48]), ou la modélisation de la logique floue (par exemple, l'analyse la logique floue basée sur les règles avec des fingrammes [49]). Il existe de nombreuses mesures conçues pour évaluer l'importance d'un nœud dans un graphe bibliographique (par exemple, degré de centralité, proximité, rang de page) [42]. De plus, il existe de nombreuses méthodes différentes pour la visualisation de graphes [50]. Parmi eux, les algorithmes "force-directed" sont les plus largement utilisés pour les sciences de l'information [51]. Leur but est de localiser les nœuds d'un graphe dans un 2D ou espace 3D, de sorte que toutes les arêtes soient approximativement de longueur égale et qu'il y ait aussi peu de bords croisés que possible, en essayant d'obtenir le plus esthétique à voir. Il existe également de nombreuses techniques de regroupement visant à découvrir la communauté (ou groupes de nœuds hautement liés) en fonction de l'importance de chaque nœud et comment il est connecté aux autres.

2.1.2 Blibliometrix

Bibliometrix est développé par Massimo Aria et Corrado Cuccurullo et est un outil open source pour la recherche quantitative en scientométrie et bibliométrie qui comprend toutes les principales méthodes d'analyse bibliométriques. Le package Bibliometrix fournit diverses routines pour importer des données bibliographiques à partir des bases de données *SCOPUS, Clarivate Analytics Web of Science, PubMed et Cochrane*, effectuer des analyses bibliométriques et créer des matrices de données pour la cocitation, le couplage et l'analyse de collaboration scientifique [86].

2.1.2.1 Choix de Scopus

Les données bibliographiques peuvent être lues à partir de différentes sources telles que Web of Science (WoS) ou Scopus. WoS semble ne pas être adéquat pour évaluer les publications et citations en Informatique. De plus, certaines autres sources telles que Google Scholar qui peuvent être trop étendues ou trop spécialisé comme Association for Computing Machinery Digital Library (ACM DL), IEEEXplore (Institute of Electrical and Electronics Engineers). Par conséquent, dans ce travail, nous nous concentrons sur Scopus qui offre également une fonctionnalité de recherche avancée utile pour sélectionner des ensembles significatifs d'éléments qui peuvent être considérés comme base pour construire notre analyse bibliométrique.

2.1.3 VOSviewer

VOSviewer est un outil logiciel de construction et de visualisation de réseaux bibliométriques. Ces réseaux peuvent par exemple inclure des revues, des chercheurs ou des publications individuelles, et ils peuvent être construits sur la base de relations de citation, de couplage bibliographique, de cocitation ou de co-auteur. VOSviewer offre également une fonctionnalité d'exploration de texte qui peut être utilisée pour construire et visualiser des réseaux de cooccurrence de termes importants extraits d'un corpus de littérature scientifique [87].

2.2 Méthodologie PRISMA

Dans le cadre de ces travaux de recherche, nous nous sommes basés sur la méthodologie PRISMA Framework (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) qui offre une approche de rechercher en 4 étapes pour une recherche efficiente.

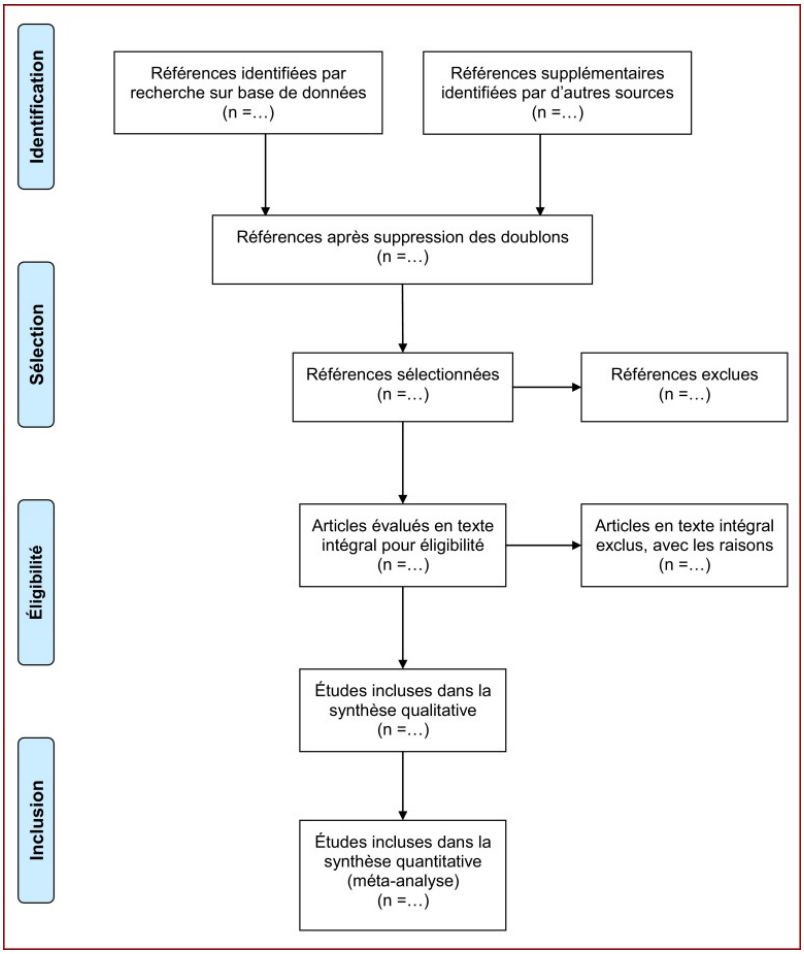

Fig. 2.1 – Les étapes de la méthodologie PRISMA

La première étape nommer identification nous permet de déterminer les mots clé spécifique à notre sujet de recherche, de déterminer les bases de donnée adéquate, de définir les critères de recherche et

d'extraire donc les données. La deuxième nommé dépistage permet donc de filtrer les données recueillies par la suppression des doublons et des données hors contexte

La troisième nommé éligibilité permet un autre filtrage à travers la vérification de la conformité des documents par rapport à des critères prédéfinis au travers de la lecture de l'abstract La quatrième nommé inclusion permet donc de recenser les différents papiers retenus.

Ainsi, il permet de démontrer la qualité de la revue de littérature, de permettre aux lecteurs d'évaluer les forces et les faiblesses, de permettre la réplication des méthodes d'examen et de structurer et de mettre en forme la revue à l'aide des méthodologies PRISMA.

2.3 Stratégie de recherche

Pour cette étude, nous avons donc développé une stratégie pour faire ressortir les littératures relatives à notre sujet de recherche en nous intéressant aux bases de données de Scopus, et d'IEEE ou les termes ont été utilisés pour extraire les données. Ces termes qui dérivent de la question de recherche que nous avons posée en nous basant l'approche PICO (population, intervention, comparator and outcome) [52] qui permet de tenir compte de la population, de l'intervention, des comparateurs et de l'objectif que cela a permis d'atteindre. Cette question de recherche est donc qu'elle est la performance des systèmes de traductions automatiques speech to speech en NLP pour les langues à forte ressource et celle a faible ressource entre 2012 et 2021 ? Toutes les données sélectionnées comprennent les articles de journal, les articles et les rapports de conférence et les chapitre de livre publié en anglais.

2.4 Critères de sélection

Basée sur la méthode PRISMA, nos recherches se sont portées sur SPEECH-TO-SPEECH TRANSLATION en NLP et couvrent les papiers parus entre 2012 et 2021. Ainsi, Tous les articles parus avant 2012 ont été exclus. Mais encore, au travers des outils que nous présente bibliometrix, l'analyse de la courbe d'évolution des publications, nous avons déterminé le point d'explosion de rechercher dans ce domaine de la traduction automatique et classement des auteurs selon le h_index, le total de citations et le nombre de publications nous a aidé à déterminer les auteurs les plus impactants. Pour découvrir les publications les plus impactantes, un classement des publications selon le nombre total de citations, et un autre selon la moyenne annuelle de citation nous a permis de constater que le top trois des deux classements comporte les mêmes publications dans des ordres différents. Nous tenons compte uniquement de ce top trois au vu de l'écart entre le nombre moyenne de citation par an du troisième et du quatrième dans le deuxième classement qui est d'environs 3 points pour ne plus autant grandir pour les documents suivent. Ainsi, nous notons ici que les deux premiers de la liste du classement des moyennes des citations annuelle, on a bordé l'approche directe de traduction automatique speech translation. De plus, l'analyse du classement des sources de publication nous a permis de déterminer que plus de publications sont faites dans les "Conférences"que dans les "Journaux". À travers le graphe des mots clés des publications, nous avons pu déterminer que les sujets les plus importants son speech recognitin et speech translation et qu'ils

sont immédiatement liés aux sujets tels que : end to end, machine translation, speech to speech translation, deep neuronal network deep learning. Alors, nous avons accentué notre recherche sur les méthodes de traduction, les approche directe, les techniques de parole à parole, les techniques de traduction automatique, surtout celles neuronales, excluant alors les comparaisons, les techniques de reconnaissance de la parole, les techniques qui ont traité seulement une partie du speech to text. Une attention particulière a été apporter aux publications de conférence. Un total de 53 articles ont été exclues et 12 ont été retenus à cette étape.

2.5 Évaluation de la qualité

Afin de nous assurer de la qualité de document collecté qui sont reparties entre articles de journal et rapports de conférence, nous avons procédé dans un premier temps à une nouvelle vérification d'existence de duplication afin de les éliminer, mais il n'y en avait plus. Puis dans un autre temps, nous avons procédé à la lecture rigoureuse et approfondis des résumés. Cela nous permit de nous assurer de la conformité de ces papiers par rapport à notre contexte d'étude et au domaine et critères prédéfini. À cette étape, 8 papiers ont été éliminés et donc 4 ont été retenus.

Conclusion L'utilisation de la méthodologie Prisma combine à la technologie bibliométrique un parcours beaucoup plus efficace des travaux publié au cours de cette dernière décennie et a facilité notre orientation vers les travaux ayant traité du speech to speech.

Chapitre 3

Résultats et discussion

Introduction Dans ce chapitre, nous avons mené une étude bibliométrique afin de déterminer les papiers ayant traité du speech to speech. L'analyse de ces papiers a permis en suite l'étude de faisabilité de la traduction automatique pour nos langues locales.

3.1 Présentation de l'état de la recherche dans le domaine de la traduction automatique

3.1.1 Évolution de la recherche

Dans l'optique d'évaluer l'état de la recherche dans le domaine de la traduction automatique, nous nous sommes servis des différentes statistiques qu'offre Bibliometrix.

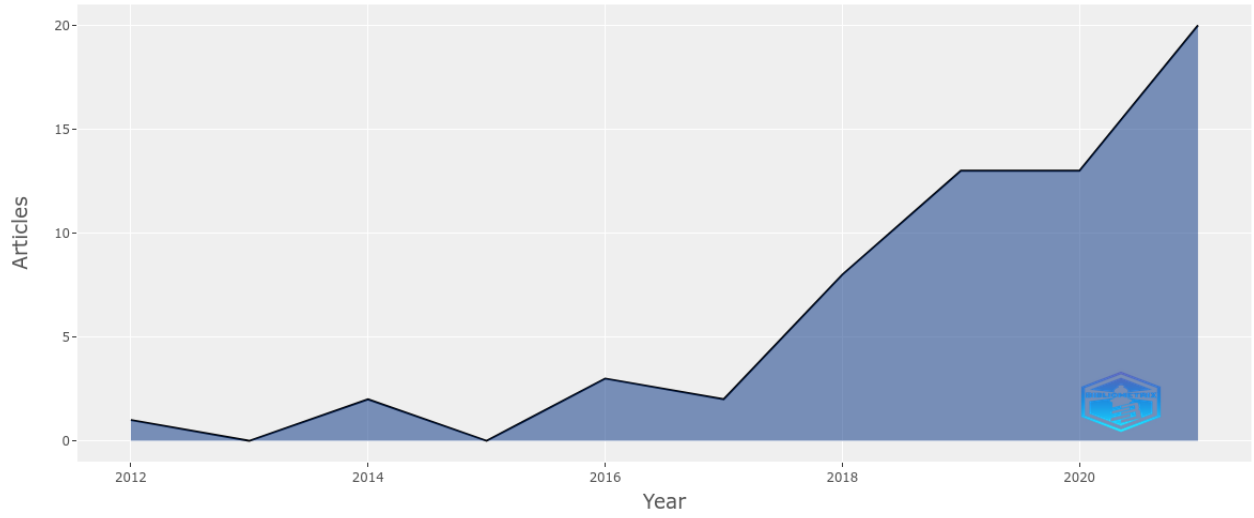

Fig. 3.1 – Évolution de la production scientifique annuelle
[81]

La figure 3.1 est la représentation graphique de l'évolution de la parution des publications dans ce

domaine. Jusqu'en 2015, l'on observe un taux extrêmement faible de publication annuelle et qui peu même être de zéro, indiquant donc l'existence d'une difficulté en cette période qui empêchait son essor. Mais à partir de 2016, on constate une montée fulgurante des publications dans ce domaine, démontrant un engouement de l'intérêt des communautés scientifiques dans ce domaine de recherche. Cette période est marquée par le début des prouesses observé grâce aux LSTM des réseaux neuronaux récurrent et des réseaux d'attention dans le domaine de la traduction automatique.

3.1.2 Auteurs et communauté d'auteurs

Tab. 3.1 – Top 5 du classement des auteurs respectivement avec leurs h-index, le total de leur nombre de citations et de publication obtenu par Bibliometrix

Element	h_index	TC	NP
SPERBER M	2	22	3
NEGRI M	2	21	3
TURCHI M	2	21	3
DI GANGI MA	2	19	2
SAKTI S	15	4	Tête

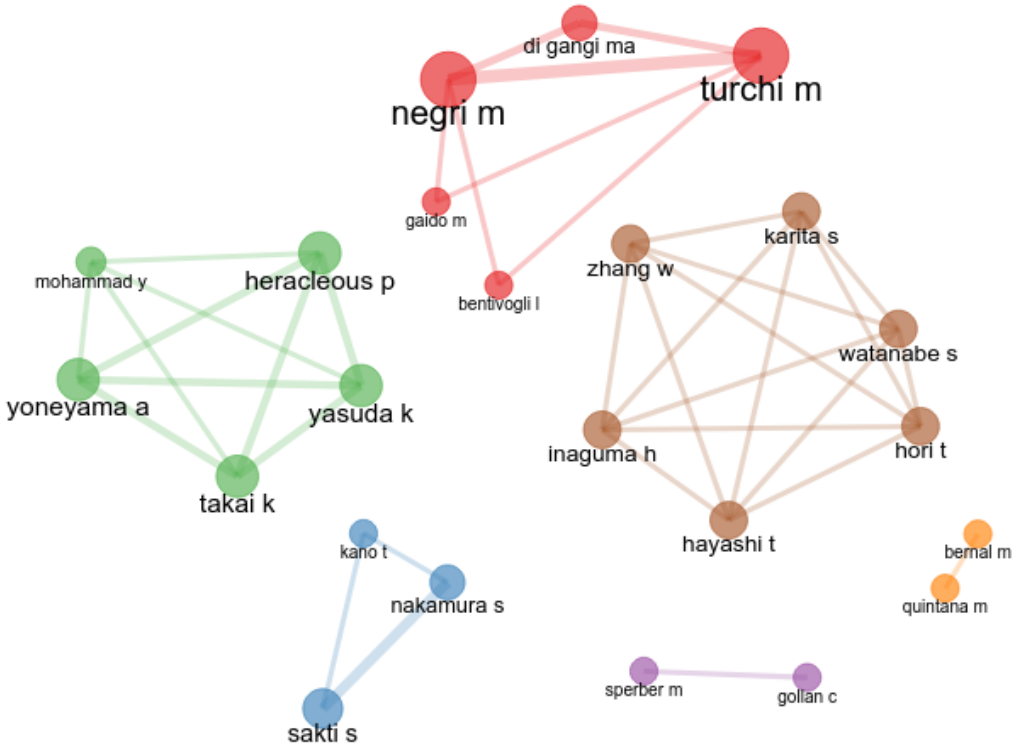

Fig. 3.2 – Communauté d'auteurs obtenue par Bibliometrix

Fig. 3.3 – Les pays les plus actifs obtenus par Bibliometrix

Le tableau 3.1 présente le top 5 du cassement des auteurs respectivement avec leurs h-index, le total de leur nombre de citations (TC) et de publication (NP). SPERBER M se trouve en tête et est suivi de NEGRI MS. Également, la figure 3.2 nous présentant les communautés de travail les plus actifs dans la publication des résultats de leur recherche dans le domaine de la traduction automatique et nous permet d'observer la présence des cinq auteurs du top 5 du classement précédant. De plus, trois des auteurs, à savoir NEGRI M TURCHI M et DI GANGI MA du classement, sont d'une des plus grandes communautés. Mais également, la figure 3.3 nous montre que les pays les plus actifs sont la Chine et les USA qui ont un certain lien de collaboration entre eux.

3.1.3 Les plus impactantes des publications

Tab. 3.2 – Top 3 du classement des publications selon le nombre total de citation obtenu par Bibliometrix

Papier	Total de citations	TC par an
Karita S, 2019, IEEE Autom speech recognit underst workshop, ASRU - PROC	136	34
Duong L, 2016, Conf north am chapter assoc comput linguist : hum lang technol, NAACL HLT - PROC CONF	52	7.429
Jia Y, 2019, Icassep IEEE int conf acoust speech signal process PROC	34	8.500

En nous intéressant aux publications ayant le plus impacté dans le domaine de la traduction automatique, nous avons étudié le classement des publications selon le nombre total de citations et le classement des publications selon une moyenne de citation annuelle. Les tableaux 3.2 et 3.3 nous présente donc, respectivement le top 3 des plus impactantes publications listées selon le nombre total de citations et selon la moyenne de citation annuelle. KARITA S[53] présente l'utilisation du modèle Séquence To Séquence dans les approches directes de la traduction automatique. Cette publication est la plus citée, mais également celle ayant la plus forte moyenne de citation annuelle. DUONG L [54] présente l'impacte de

Tab. 3.3 – Top 3 du classement des publications selon la moyenne de citation annuelle obtenu par Bibliometrix

Papier	Total Citations	TC par Année
Karita S, 2019, IEEE Autom speech recognit underst workshop, ASRU - PROC	136	34
Jia Y, 2019, Icassep IEEE int conf acoust speech signal process PROC	34	8.500
Duong L, 2016, Conf north am chapter assoc comput linguist : hum lang technol, NAACL HLT - PROC CONF	52	7.429

l'utilisation des données audio pour l'exécution des algorithmes de traduction automatique pour les langues a faible ressource.

Celle-ci es la seconde plus cite, mais troisième pour la moyenne de citation annuelle. JIA Y [55] présente les avantages de l'approche directe par rapport à l'approche en cascade. Cette publication a la troisième place en ce qui concerne le nombre total de citations, mais la seconde place selon la moyenne de citation annuelle.

Notons ici que les deux premiers de la liste du classement des moyennes des citations annuelle, on a bordé l'approche directe de traduction automatique speech translation, et que le même article, premier des deux listes, préconise les méthodes séquence à séquence des techniques neuronales, tant dise que l'article ayant la deuxième place dans la liste du classement des nombres totale de citation et troisième dans celle des moyennes de citation annuelle préconise les ressources audio pour les langues à faible ressource.

3.1.4 Les sources en tête de liste

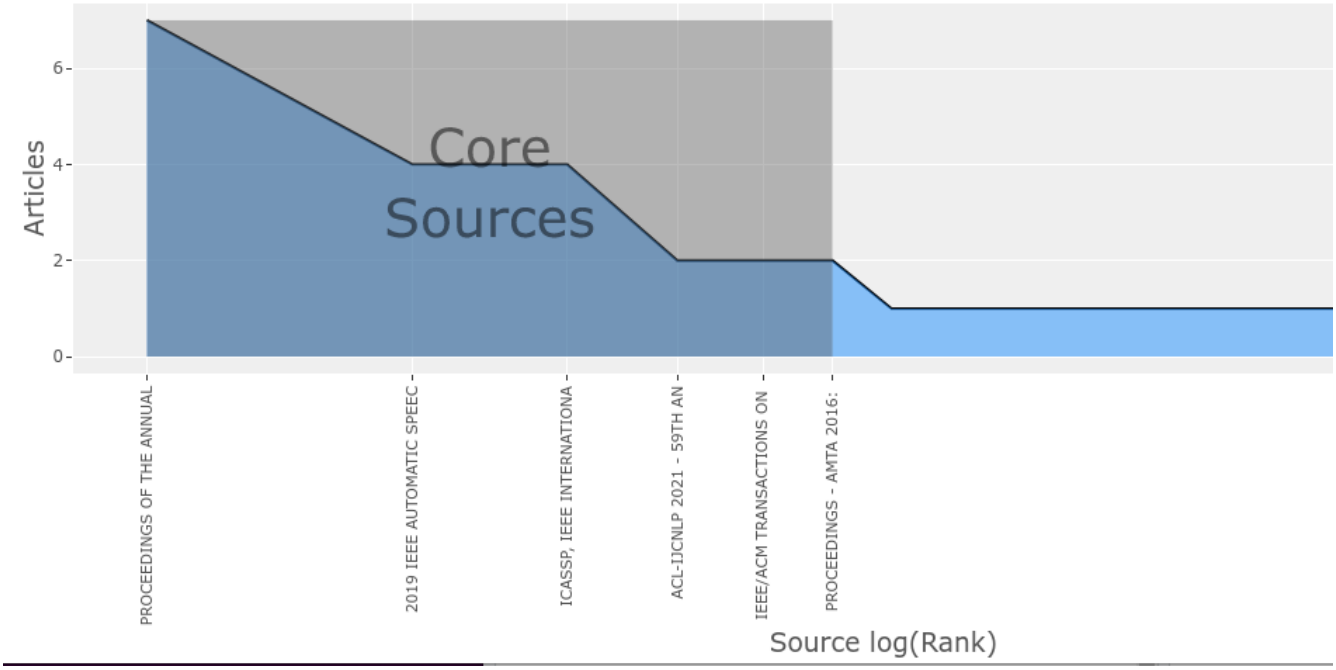

Fig. 3.4 – Les sources les plus importantes obtenues par Bibliometrix

La figure 3.4 présente les sources importantes selon la loi de Bradford par la répartition des publications au niveau des sources ayant les plus grands nombres de publications. Nous constatons que la majorité des publications sont publiées dans des « conference proceedings » et que 'PROCEEDINGS OF THE ANNUAL CONFERENCE OF THE INTERNATIONAL SPEECH COMMUNICATION ASSOCIATION INTERSPEECH"est la source en tête de liste. De plus, a l'exception d'une seule publication, toutes les autres publications précédemment présentées sont faites dans le top 3 des sources les plus en vue et deux des publications sont faites dans des « conference proceedings ».

3.1.5 Les principaux centres d'intérêts de la rechercher dans le domaine de la traduction automatique

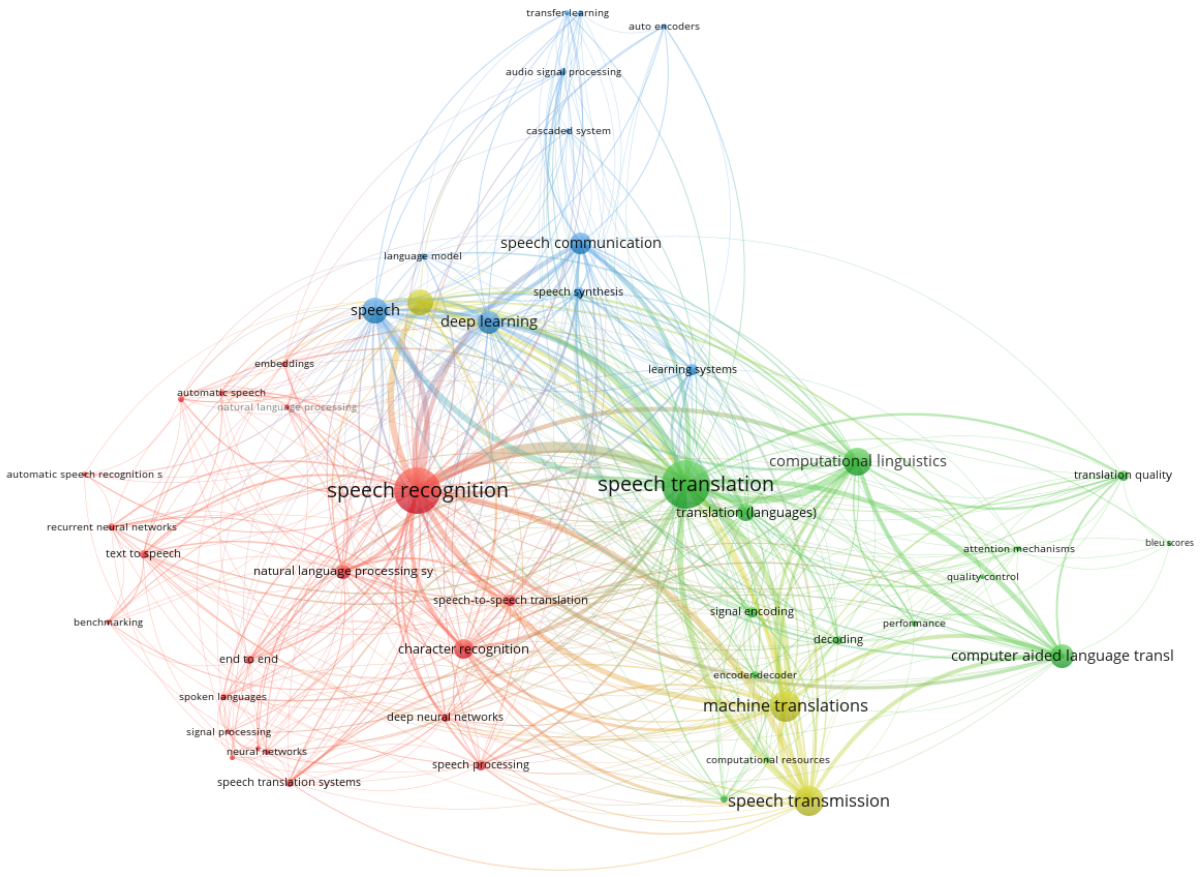

Fig. 3.5 – Les importants mots clés obtenus par Bibliometrix

La figure 3.5 montre un graphique avec les mots-clés d'auteur les plus populaires dans les publications étudié. Chaque nœud du graphe est associé à un mot-clé et sa taille est proportionnelle au nombre de documents où le mot-clé apparaît. Speech recognition et speech translation sont les mots clés principales de ces groupes de mot. En second plan, ils sont suivis des mots tels que, machine translation, computer, aided langage translation, speech transmission, deep learning, speech communication, speech, computational lingustics. Les liens entre les nœuds relient des mots-clés qui apparaissent généralement ensemble dans les mêmes documents. Ce graphique donne un aperçu global des principaux sujets d'intérêt dans le Champ de recherche de la traduction automatique, avec des groupes de nœuds étroitement liés peints dans la même couleur. D'une part, speech recognition et speech translation sont les plus proches des sujets tels que le end to end, machine translation, speech to speech translation, deep neuronal network deep learning. Les principaux mots-clés sont également fortement reliés.

De l'analyse de toutes ces données dans le cadre de notre étude qui vise à déterminer les approche et technique de traduction automatique pour nos langues locales, nous pouvons dire au vu du top trois des

tableaux 3.2 et 3.3 que nous devons nous intéresser aux travaux employant le ”Sequence to Sequence”, ceux utilisant l'approche directe et ceux utilisant les donnes audio pour le traitement en ce qui concerne les langues peu dotées. De plus, au vu des mots clé principaux, les mots clé tels que 'speech translation', 'end to end', 'deep neuronal network'et 'deep learning' pourront nous aider.

3.2 Présentation des travaux analysés

3.2.1 Hindi-English Système de traduction automatique Speech-to-Speech pour les expressions des voyageurs [56]

Ce travail présente la traduction « speech to speech » du couple de langue Hindi-English pour permettre au touriste de communiquer avec les riverains.

Pour ce faire, un corpus de phrase parallèle Hindi-English a été constituer. Ce corpus contient 306 phrase Ensuite, ses phrases sont enregistrées pour former le corpus parallèle de parole Hindi-English.

Ce travail a utilisé l'approche en cascade à trois niveaux Pour le premier niveau, un modèle ”Hidden Markov Model ”(HMM), est utilisée pour effectuer la reconnaissance de mot à (ASR). Le deuxième niveau qui consiste en la traduction text to texte (Hindi-English) permettant de convertir les phrase d'une langue à l'autre à travers une approche statistique. Finalement, le dernier niveau représentant un text-to-speech permettant de convertir les textes dans une langue en parole de la même langue au travers aussi d'un modèle HMM.

La perfection du système résultant dépend de la précision de chaque niveau. Ainsi, le premier niveau a utilisé comme métrique est WER et abouti à une précision de 100%. Le deuxième niveau est évalué grâce au TER et donne 49,46 pour la traduction Hindi −English et 59,19 pour le couple English −Hindi. Le troisième niveau a été évalué en utilisant le MOS et a obtenu un score de 3.21 et puisque le système est en cascade, c'est la performance de ce niveau qui est celui du système.

Ainsi, nous pouvons retenir de cette méthode en cascade que malgré l'acceptabilité du résultat, au niveau de la reconnaissance de la parole, les différences au niveau des langues et au niveau de la disponibilité des ressources entre ces langues affectent de façon notoire la traduction machine dont la réduction d'erreur nécessite de recourir à d'autre méthode, qui induise l'augmentation du temps et des ressources pour le travail. En fin, ces erreurs induisent des erreurs au niveau de la synthèse de parole.

3.2.2 Traduction directe speech-to-speech avec un modèle sequence-to-sequence [57]

Ce travail présente la traduction « speech to speech » du couple de langue Espagnol-English pour faciliter les échanges conversationnelle et l'intercompréhension entre des personnes parlant pour les uns l'espagnole et pour les autres l'anglais.

Pour ce fait, deux corpus de phrase parallèles Espagnol-English ont été utilisées. il s'agit d'un grand corpus « conversationnel » de texte parallèle et des enregistrements de lecture, et le corpus espagnol Fisher

de conversations téléphoniques et la traduction anglaises correspondantes, qui est plus petit et plus difficile en raison du caractère spontané et informel et qui permet d'apprécier la qualité de cette approche sur les bases de données de petite taille.

Ce travail a utilisé l'approche directe à travers un système de deux composantes d'entraînement séparé en trois parties. Le premier composant qui est un réseau neuronal récurent séquence à séquence LSTM avec un mécanisme d'attention, qui est entraînée avec le spectrogramme des données audios des langues source et cible pour la traduction. Le second étant le couplage d'un vocodeur Griffin-Lim et d'un décodeur de spectrogramme afin de convertir respectivement les données audios en spectrogramme a l'entrée et les spectrogrammes en voix audible a la sortie.

La perfection du système résultant a été évalué par comparaison de ses résultats aux résultats d'un système basée sur l'approche en cascade et ceci avec comme métrique les scores BLEU. Ce système obtient un score de 42.7 inférieurs de 6 point au score BLEU de la méthode en cascade avec le corpus de texte conversationnel et avec la base de donnée téléphonique de Fisher, ils obtiennent un score 30.1 inférieur de 9.3 points au score BLEU de la méthode en cascade. De plus, il est aussi observé avec la base de donnée téléphonique de Fisher que tant dise que la méthode en cascade traduit le prénom Guillermo (espagnol) en William (anglais), la méthode directe tante par contre de conserver le nom d'origine, mais donne « of the ermo ».

Nous pouvons donc en conclure que cette méthode directe, malgré le fait d'être nouvelle, elle présente des résultats vraiment satisfaisants, permet d'éviter le cumul des erreurs comme celui observé au niveau de la méthode en cascade et promet de meilleurs résultats même avec des données de petite taille, offrant donc une plus grande fiabilité dans ces traductions.

3.2.3 Traduction automatique speech-speech entre des langues non connues et non écrite [58]

Ce travail présente, la traduction « speech to speech » pour les paires de langues Francais-Anglais et Japonais-Anglais pour permettre à des touristes de communiquer avec les riverains.

Pour ce fait, un corpus de phrase parallele pour les paires de langues Francais-Anglais et Japonais-Anglais a été constituer. Ce corpus contient 162318 phrases dont les audio de ces phrases ont été générés pour chaque paire de langue avec API Google text-to-speech.

Ce travail a utilisé une approche directe basée sur trois composent Pour le premier composent, est constitué d'un auto-encodeur variationnel à quantification vectorielle VQ-VAE qui permet d'extraire dans les MFCC des données audios des paires de langue, le contenu des messages qui nous intéresse,
qu'il organise ensuite en livre de code. Le second composant est un réseau de neurone récurent basée sur un modèle d'attention séquence à séquence entraînée pour déterminer le chemin entre les MFCC des audio et leur correspondant dans le livre de code. troisième composant dénommer codebook inverter, qui du chemin déterminer, recéper dans le livre de code le message dans la langue cible qu'il convertit en représentation audio. La perfection du système résultant a été évalué par comparaison de ses résultats aux résultats d'un système basée sur l'approche en cascade et ceci avec comme métrique les scores BLEU.

Pour la paire français-anglais, un score BLEU de 25.0 et un score METEOR de 23.2 ont été obtenus face à un score BLEU de 47.4 et un score METEOR de 41.2 obtenus pour une approche base sur une méthode en cascade. Pour la paire japonais-anglais, un score BLEU de 15.3 et un score METEOR de 15.3 ont été obtenus face à un score BLEU de 37.4 et un score METEOR de 32.8 obtenus pour une approche base sur une méthode en cascade.

Après analyse de ces résultats, nous constatons que malgré la supériorité des scores de l'approche en cascade qui est l'approche de référence, face à celui de l'approche directe propose dans ce travail, que les scores obtenus sont assez pertinents vu la nature pionnière de ces travaux. Respectivement, pour la paire français-anglais deux langues proches, les scores BLEU et METEOR de la méthode directe proposée par ce travail sont supérieurs à la moyenne de celle de la méthode de référence, tant dise que ceux de la paire japonnais-anglais, des langues assez éloigné, ils s'en rapprochent.

3.2.4 Traduction directe Speech to Speech to Speech basé sur un réseau Transformeur avec un Transcodeur [59]

source Ce travail présente, la traduction « speech to speech » pour les paires de langues Anglais-Espagnole et Japonais-Coreen comme langue syntaxiquement proche et Anglais-Japonais comme langue syntaxiquement éloigné pour permettre à des touristes de communiquer avec les riverains.

Pour ce fait, un corpus de phrase parallele pour les paires de langues Anglais-Espagnole, Japonais-Coreen et Anglais-Japonais a été constituer. Ce corpus contient pour la paire Anglais-Japonais 480000 phrase et pour les paires de langues Anglais-Espagnole, Japonais-Coreen 160000 phrases dont les audio de ces phrases ont été générés pour chaque paire de langue avec API Google text-to-speech.

Ce travail a utilisé une approche mixte ou les trois niveaux d'une approche en cascade a été établie et relie par un transcodeur pour en faire une approche directe à trois composent. Pour le premier composent, est constitué d'un encodeur qui permet d'extraire dans les MFCC des données audios des paires de langue, qu'un réseau neuronal Transformeur est utilisée pour effectuer la reconnaissance de mot à (ASR). Le second composant reçoit à travers un transcodeur les états cachés du premier composant pour la traduction text to texte permettant de convertir les phrase d'une langue à l'autre à travers un autre réseau neuronal Transformeur. Le troisième composant reçoit également à travers un transcodeur les états cachés du deuxième composant, pour le text-to-speech permettant de convertir les textes dans une langue en parole de la même langue au travers aussi d'un réseau neuronale Transformeur puis un décodeur pour sortir les prononciations.

La perfection du système résultant a été évaluée en plusieurs étapes. Une première a été de récupérer les données des audio transcrites en texte et de les comparer à celui d'un autre modèle ASR ayant de bon résultat avec des réseaux neuraux récurrent (RNN) pour constater que les scores WER de cette nouvelle approche était meilleur. Une seconde a été de récupérer les résultats de la traduction des textes de la langue source en texte de la langue cible et de les comparer aussi au résultat d'une stratégie base sur des réseaux de neural récurent (RNN) du donné de bon résultat et ici aussi de constater les scores BLEU et METEOR de la nouvelle stratégie son meilleur, et ceci pour les langues syntaxiquement proches et syntaxiquement éloigne

Une troisième étape a été donc de comparer les traductions audio obtenues dans la langue cible avec ceux de quatre autres stratégies à savoir : une stratégie en cascade base sur les réseaux de neurone récurent (RNN), une autre stratégie en cascade (Transformer), un système de traduction de Google basée sur les RNN et les autres aussi de Google base sur les Transformer. Le constat est resté le même, car les scores BLEU de 44, METEOR de 59.3 et les scores BLEU de 41 et METEOR de 56.6, de la nouvelle stratégie son meilleur, respectivement pour les langues syntaxiquement proches et syntaxiquement éloigne.

Tab. 3.4 – Tableau récapitulatif de ces approches

Papier	description	approche	langue	corpus	prétraitement	Critères d'évaluations	performances
Hindi-English Systeme de traduction automatique Speech-to-Speech pour les expressions des voyageurs	Réalisation de traduction «speech to speech» du couple de langue Hindi-English pour permettre au touriste de communiquer avec les riverains	Cascade	English	Corpus de 306 phrase	Méthode statistique	MOS	3.21
Traduction directe speech-to-speech avec un modèle sequence-to-sequence	traduction «speech to speech» du couple de langue Espagnol-English	directe	Espagnol-English	Corpus de texte «conversationnel» et le corpus de Fisher de conversations téléphoniques	RNN -attention	Score BLEU	42.7
Traduction automatique speech-speech entre des langues non connues et non écrite	traduction «speech to speech» pour les paires de langues Francais-Anglais et Japonais-Anglais pour permettre à des touristes de communiquer avec les riverains	directe	Francais-Anglais et Japonais-Anglais	Corpus BTEC contenant 162318 phrases	RNN -attention	Score BLEU	25
Traduction directe Speech to Speech to Speech basé sur un réseau Transformeur avec un Transcodeur	traduction « speech to speech » pour les paires de langues syntaxiquement proche et de langue syntaxiquement eloigner pour permettre à des touristes de communiquer avec les riverains	directe	Anglais-Espagnole, Japonais-Coreen et Anglais-Japonais	Corpus BTEC contient pour la paire Anglais-Japonais 480000 phrase et pour les paires de langues Anglais-Espagnole, Japonais-Coreen 160000 phrases	TNN	Score BLEU	44

3.3 Étude comparative

3.3.1 Catégorisation des travaux

Les différents travaux que nous avons collectés se sont basée sur plusieurs techniques à travers différentes approche. Nous avons effectué une catégorisation selon les approches et selon les techniques.

Tab. 3.5 – Tableau de comparaison des langues

Techniques	cascade	directe
statistique	n1	...
VAE	...	n3
RNN	...	n2 (LSTM, Attention),n3 (Attention)
TNN	n4 (self-Attention)	...

De l'analyse de ce tableau, nous déjà constater que pour de meilleure performance, la majorité de ces travaux se sont basés sur les techniques de Deep Learning mais plus encore se sont basés ou inspiré des techniques de RNN, surtout pour les approches de traduction directe.

3.3.2 Forces et limites de cette technique RNN

De l'analyse de ces travaux, nous pouvons dire que le choix des techniques, base sur les réseaux d'attention des réseaux RNN, est dû au fait qu'elles permettent de palier au manque de ressource très prononcé dans le domaine pour certaines langues, qu'elles réduisent le temps de calcul et offre de meilleur résultat surtout pour les langues à faible ressource en permettant un meilleur établissement de relation de logique entre les mots des phrases, mais plus encore utilisé dans les approches de traduction directe pour éviter les cumule des erreurs au cours du processus. Mais par contre, elles demandent beaucoup de ressource de calcul. L'approche directe ne permet pas d'un autre coter de promouvoir l'intégration dans le numérique des langues à faibles ressources et rend très difficile l'établissement de relation de logique entre les mots des phrases. Cependant, la technologie des réseaux de neurone Transformeur utilisé dans les travaux n3, semble permettre de pouvoir outrepasser ces limite.

3.3.3 Étude comparative entre nos langues et celle étudiée dans les travaux

3.3.3.1 Tableau des langues étudiées dans les travaux

Des différents travaux que nous avons recensés, ceux ayant les meilleurs résultats ont effectué leur recherche sur plusieurs langues, à savoir : l'Hindi, l'anglais, l'espagnole, le français, le japonais, le Coréen. À partir des différentes paires de langue étudier dans les différents travaux, nous pouvons les classés en deux catégories, à savoir les langues syntaxiquement proches entre eux pour les langues telles que

l'anglais, l'espagnole, le français, ou syntaxiquement éloigné tel que comme l'ensemble des langues syntaxiquement proche par rapport à l'Hindi, le japonais et le Coréen.

Nous avons alors relevé les caractéristiques de ces langues dans le tableau suivant :

Tab. 3.6 – Tableau des caractéristiques des langues papiers analysé

Langue	Hindi	Anglai	Espagnol	Francais	Japonais	Coréen
Voyelles	13 [88]	5 [89]	6 [90]	6	5 [91]	21 [92]
Consonnes	36	21	25	20	19	19
-	-	-	-	-	-	
Type sylabique	CV, CVC, V, VC, VCC, CCV, CCVC, CVCC [60]	V VCC, CVCCCC, CCCVC, CV, CCV, CCCV, VCCC, CVC, CVCC, CCVC, CCVCC, CCVCCC, CCCVCC, CCCVCCC, VC, CVCCC, CCVCCCC [61]	CV, CVC, V CCV [62]	CV, CVC, V, VC, VCC, CCV, CCVC, CVCC, CCVCC, CCCVC, CCCVCC [60]	V, CV, vV, CvV, N, T, R [63]	V, CV, VC, CVC, CVCC [93]
Typologie syntaxique	SOV [94]	SVO [95]	SVO [95]	SVO [95]	SVO [95]	SOV [96]
Expression des Temps	Passée : था. Futur : गा , गी , गे [97]	Conjugaison totale avec le sujet	Conjugaison totale avec le sujet [98]	Conjugaison totale avec le sujet	Conjugaison non en fonction des sujets, mais en fonction de trois formes de verbes et de la forme d'expression[99]	Passée : Verb + 았어 ou 었어 Future : Verb + 르 ou 을거야 [100]
Caractères particuliers	Ri, ii, au, ah, aa, uu, am, cha, shha, gya, kha, chha, tha, pha, jha, dha, bha, ksh, nja, sha[101]	-	ch, ñ, ll,rr	-	-	ng, ch,kk, tt, pp, ss, jj, ya, oe, yeo, yo, yu, eu, ae, yae, ye, wa, wae, oe, wo, wi, ui[102]

3.3.3.2 Situation de nos langues par rapport aux langues des travaux

Une première comparaison des alphabets, de l'expression du temps et des caractères spéciaux de nos langues et ceux des travaux nous permet de les classées avec les langues syntaxiquement éloignées de ceux syntaxiquement proches. De plus, nos langues sont des langues à faible ressource. Comme le coréen et l'Hindi, pour lesquelles des résultats satisfaisant et même très satisfaisant ont été obtenus. Cependant, malgré le fait que le Fongbé et le Yoruba sont de type SVO comme l'Anglais, le français et l'Espagnol et le Bariba/Baatoonu de type SOV comme l'Hindi, le japonais et le coréen, il est à noter que différemment de toutes ces langues des travaux, nos langues d'ici sont tonales.

3.3.4 Approche pour la traduction automatique de nos langues

Au regard des différentes observations précédemment cité, tels que le caractère langues à faibles ressources de nos langues et du fait qu'elles sont syntaxiquement éloignées de la majorité ou mieux de la quasi-totalité des langues des travaux parcourues et des résultats obtenus des différentes approches de ces travaux, l'approche directe basée sur les réseaux d'attention des méthodes de Deep Learning plus précisément le réseau Transformeur, conviendrait plus pour la traduction automatique de nos langues. En effet, le réseau de neurone TNN, nécessite moins de puissance de calculs pour s'entraîner et s'adapte mieux au matériel d'apprentissage automatique moderne, accélérant l'obtention des résultats (plus efficace que le CNN et le RNN en cascade, pas de travaux de l'approche directe trouver).

Cependant, en considérant le besoin de ressource pour nos langues en vue de pouvoir permettre à nos sociétés de profiter des nouvelles opportunités des TIC, tel que la classification des documents, le résumé de document, le Part-Of-Speech Tagging qui permet de déterminer la classe morphosyntaxique de chaque mot à partir de connaissances lexicales et du contexte de son emploie, le Named-Entity Recognition (reconnaissance d'entités nommées en français) qui permet de reconnaître dans un texte un certain type de concepts catégorisables dans des classes telles que noms de personnes, noms d'organisations ou d'entreprises, noms de lieux, quantités, distances, valeurs, dates et en considérant ainsi les performances de l'approche en cascade basée sur le réseau transformeur accompagné de transcodeur, cette approche pourra donc nous aider à répondre à ces besoins au travers du Speech to text.

3.4 Discussion

3.4.1 Rappels

La traduction automatique a permis de nos jours de briser les barrières que représentaient les différences de langues dans les échanges, ceci dans bon nombre de pays en voie de développement. Cela a permis à ces pays une meilleure visibilité et une meilleure expansion socioculturelle et économique. Malheureusement, à notre connaissance, il n'y a actuellement aucun travail faisant la synthèse de toutes ces techniques et donnant des directives sur comment créer un système speech to speech pour nos langues locales africaines et spécifiquement béninoises qui sont donc restées en marge de ces prouesses tech-

nologiques. Alors, nous nous sommes fixés comme objectif d'effectuer L'état de l'art des méthodes de traduction automatique speech to speech en Deep Learning afin de répertorier les problèmes à leur éclosion pour les langues locales du Bénin, en Afrique en général et de proposer des directives futures pour une mise en œuvre plus efficace et effective.

À cet effet, nous avons utilisé une approche méthodologique basée sur la méthodologie PRISMA qui s'est déroulée en deux temps. Dans un premier temps, nous nous sommes basés sur les technologies de la bibliométrie qui nous ont permis de déterminer pour le domaine de la traduction automatique, les mots clés désignant les centre d'intérêt majeur, les technologies en d'actualité de façon générale et aussi spécifiquement pour les langues a faible ressource. Cette approche a été utilisée dans un autre domaine pour les travaux de Jose M. Alonso [64].

Dans un second temps, en nous basant sur les orientations obtenue grâce à l'analyse bibliométrique, nous avons procédé à une revue lecture systématique de l'état de l'art dans le domaine de la traduction automatique Speech To Speech. Cette approche a été utilisée dans un autre domaine pour les travaux de Mario A [65] et de Prachi Kadam [66]

3.4.2 Forces et limites de nos travaux

3.4.2.1 Forces (Les contributions)

Cette étude est une première dans son genre de revue de littérature des approches de Speech To Speech appliques a nos langues locales. Elle se décline comme suit :

- une revue des langues locales avec une étude comparative d'un point de vue des techniques NLP.
- un état de l'art faisant l'état de santé des recherches en Speech To Speech. Cela nous a permis de voir différente technique et approche de traduction automatique Speech To Speech. L'évaluation des force et faiblesse de ces techniques nous ont permis de voir que la technique dit Transformers qui offre plus d'avantage pour les deux, celle en cascade comme celle directe. Cela a été aussi démontré dans les travaux de Takatomo Kano [59].
- une revue de littératures conduisant à une sélection de 4 papiers de référence à travers les technologies de la bibliométrie comme cela a été le cas pour les travaux de Jose M. Alonso [64] et d'Hamid Darvish [67], et la méthodologie PRISMA, recommandant des approches référence et des méthodes de deep learning pour nos langues.

3.4.2.2 Difficulté et limites

Au cours de cette étude, nous avons rencontré certaine difficulté que nous pouvons décliner dans les points suivants :

- une difficulté liée a la revue de nos langues aux vues de l'absence de ressources suffisantes.
- une difficulté d'acquérir des outils adéquats pour faire une revue de littérature globale.
- une difficulté liée a la quasi-absence des travaux en traduction automatique et sur la majorité de langue en Afrique.

Aussi certaines limites ont été observées à savoir :

- que la recherche bibliométrique est uniquement lié à scopus qui également présente des défauts dans la structuration des publications,
- la revue concernant les langues peut être largement étendu en s'associant à des linguistes pour affiner l'étude.
- que la bibliométrie qui elle-même est souvent critiquée, car basée sur des approches de texte-mining qui présente aussi des défauts et même le fait que parfois, elle manque de profondeur, car c'est l'essentiellement l'abstract, le titre, les mots clés qui sont analysés alors que les résultats peuvent être entachés de beaucoup de problème.

Conclusion De ce chapitre, nous retenons que les méthodes Transformers sont plus indiquées pour la traduction automatique de nos langues locales au vu de leur caractéristique.

Conclusion et perspectives

L'objectif de ce mémoire de recherche était d'effectuer l'état de l'art des méthodes de traduction automatique speech to speech en apprentissage profond afin de répertorier les problèmes à leur éclosion pour les langues locales du Bénin, en Afrique en général et de proposer des directives futures pour une mise en œuvre plus efficace et effective.
Dans un premier temps, nous avons utilisé la bibliométrie pour faire une revue globale et avons eu une indication de l'importance des papiers les uns par rapport aux autres dans le domaine, nous avons déterminé aussi les communautés existantes, la collaboration entre auteurs et les principaux sujets au tours desquels les recherches sont effectuées.

Dans un deuxième temps, fort de cette bibliométrie, nous avions sélectionné un certain nombre de papiers que nous avions analysés (description, exigences, forces, faiblesse) et ensuite, fort de ces exigences, nous avions pu identifier les défis avec nos langues locales (les difficultés).
Au vu des caractères peu dote de nos langues, du fait qu'elles sont très éloignées des langues occidentales très dotées, nous avons proposée l'emploie des techniques neuronales Transformers pour la réalisation des systèmes de traduction automatique. Les systèmes de traduction automatique ont montré leurs utilités de par le monde et prouver les nombreux avantages quelle offre dans les pays d'enveloppés. Alors, la création de système de traduction automatique de nos langues locales se trouve être une nécessité pour notre développement et pour une meilleure intégration de nos langues dans le numérique.

Bibliographie

[1] https://www.ekino.com/articles/introduction-nlp-partie-i consulté le 04 Août 2021.

[2] Historique de NLP https://www.ekino.com/articles/introduction-to-nlp-part-i consulté le 04 Decembre 2021.

[3] LIGAN Charles *Quelle stratégie pour l'aménagement du statut des langues béninoises ?*,Université de Lokossa (Bénin)

[4] HOMBERT, J.M. G. PHILIPPSON *The linguistic importance of language isolates : the African case*,Dynamique du Langage (CNRS, Université de Lyon), (2009)

[5] Igue, A. M. *Grammaire Yorùbá de base abrégée*,Center for Advanced Studies of African Society (CASAS), monograph 238 (2009)

[6] John Oscar Raoul AOGA *CONCEPTION DU CORPUS DE PAROLE POUR LA SYNTHESE VOCALE TEXT-TO-SPEECH DU YORUBA ET APPLICATION DE LA METHODE DE SELECTION D'UNITES*, MEMOIRE DIPLOME D'ETUDE APPROFONDIE,2012 -2013

[7] J. Greenberg. *Languages of africa*,La Haye Mouton, page 117, (1996)

[8] C. Lefebvre and A.M. Brousseau *A grammar of fonge*,de gruyter mouton. page 608, (2001)

[9] Cossi B. GNANGUENON *Analyse syntaxique et sémantique de la langue "f□n" au Bénin en Afrique de l'Ouest, pour la création d'un dictionnaire bilingue en langues fon et français*,THÈSE DE DOCTORAT EN SCIENCES DU LANGAGE, (2014)

[10] Frejus Adissa Akintola Laleye *Contributions à l'étude et à la reconnaissance automatique de la parole en Fongbe*,THESE, Université du Littoral Côte d'Opale ; Université d'Abomey-Calavi (Bénin), (2016)

[11] A. B. AKOHA *Syntaxe et lexicologie du fon-gbe : Bénin*,Ed. L'harmattan, page 368, 2010.

[12] Jean-Marie HOMBERT *LES SYSTEMES TONALS DES U N W E S AFRICAINES : TYPOLOGIE ET DIACHRONIE*

[13] Sayane Gouroubéra *Une approche morphophonologique du système verbal du Baatonum*,Université d'Abomey-Calavi (Département des Sciences du Langage et de la Communication),(2005)

[14] Alan V. Oppenheim, Ronald W. Schafer *From Frequency to Quefrency : A History of the Cepstrum*, IEEE, 2004.

[15] Andros Tjandra , Sakriani Sakti , Satoshi Nakamura *SPEECH-TO-SPEECH TRANSLATION BETWEEN UNTRANSCRIBED UNKNOWN LANGUAGES*,Nara Institute of Science and Technology, Japan, RIKEN, Center for Advanced Intelligence Project AIP, Japan, (2019)

[16] Alexandre Bérard, Olivier Pietquin, Christophe Servan, and Laurent Besacier *Listen and translate : A proof of concept for end-to-end speech-to-text translation*,CoRR, vol.abs/1612.01744, (2016)

[17] Olivier Kraif *Corpus parallèles, corpus comparables : quels contrastes ?*, Informatique et langage [cs.CL]. Université de Poitiers, 2014.

[18] Véronis J. *From the Rosetta Stone to the information society : A survey of parallel text processing*, In Véronis, J. (Ed.), Parallel Text Processing, Dordrecht, Netherlands, Kluwer Academic Publishers, § 1, 24 p., (2000)

[19] Pierre Isabelle *La bi-textualité : vers une nouvelle génération d'aides à la traduction et la terminologie*, META, Outremont, PQ, XXXVII, 4, pp. 721-731, (1992)

[20] Olivier Kraif *Corpus parallèles, corpus comparables : quels contrastes ?*, Dossier en vue de l'Habilitation à diriger des recherches, (2015)

[21] Teubert, W. *Comparable or Parallel Corpora ?*,International Journal of Lexicography, 9 (3) : 238-264., (1996)

[22] Weaver, W *Translation. Machine translation of languages*, Computational linguistics, 16(2), 14–23,1955.

[23] Brown, P., Cocke, J., Pietra, S. D., Pietra, V. D., Jelinek, F., Mercer, R., Roossin, P *A statistical approach to language translation, In Proceedings of the 12th conference on Computational linguistics-Volume 1*, Association for Computational Linguistics,pp 71–76,1988. (, August). (). .

[24] Brown, P. F., Cocke, J., Della Pietra, S. A., Della Pietra, V. J., Jelinek, F., Lafferty, J. D., ⋯ Roossin, P. S. *A statistical approach to machine translation*, Computational linguistics, 16(2), 79–85,1990.

[25] Michael Collins, Philipp Koehn, Ivona Kucerova *ClauseRestructuringforStatisticalMachineTranslation*, Developed in collaboration with the Association for the Development of Education in Africa (ADEA)UNESCO Institute for Lifelong Learning,2010.

[26] Clément Réverdy *Modèles de Markov Cachés (HMM) pour de la reconnaissance de gestes humains*, Apprentissage [cs.LG]. 2014.

[27] Reseau de neurines https://www.universalis.fr/encyclopedie/reseaux-de-neurones-formels/ consulté le 04 Decembre 2021.

[28] Bengio, Y., Ducharme, R., Vincent, P. *Proceedings of NIPS*, (2001).

[29] Kannan A., Kurach K., Ravi S., Kaufmann T., Tomkins A., Miklos B., ... Ramavajjala,V. *Smart reply : Automated response suggestion for email*, In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 955-964), (2016).

[30] Caruana R. *Multitask Learning. Autonomous Agents and Multi-Agent Systems*, (1998).

[31] Collobert R., Weston J *A unified architecture for natural language processing,In Proceedings of the 25th International Conference on Machine Learning*,(pp. 160−167) (2008).

[32] Mikolov T., Corrado G., Chen K., Dean J. *Efficient Estimation of Word Representations in Vector Space*, Proceedings of the International Conference on Learning Representations, (2013)

[33] Anthony HARTLEY, Andrei POPESCU-BELIS *Chapitre 13, Chapitre 13 L□évaluation des systèmes de traduction automatique*,(2020)

[34] Snover S., Dorr B., Schwartz R., Micciulla M. et Makhoul J. *A study of translation edit rate with targeted human annotation*, Proceedings of Association for Machine Translation in the Americas, pages 223−231, (2006)

[35] Tillmann C., Vogel S., Ney H., Zubiaga A., Sawaf H. *Accelerated DP-based search for statistical translation*, In proceedings of Eurospeech, (1997)

[36] Matthew Snover, Dorr B., Schwartz R., Micciulla L., Makhoul J. *A study of translation edit rate with targeted human annotation*, In proceedings of AMTA, (2006)

[37] Matthew Snover, Nitin Madnani, Bonnie Dorr, and Richard Schwartz *Fluency, Adequacy, or HTER ? Exploring Different Human Judgments with a Tunable MT Metric*, In proceedings of the Fourth WMT at the 12th EACL, Athens, Greece, (2009)

[38] Papineni K., Roukos S., Ward T., Zhu W. *BLEU : a method for automatic evaluation of machine translation*, In proceedings of ACL, (2002)

[39] G. Doddington *Automatic evaluation of machine translation quality using n-gram co-occurrence statistics*, In proceedings of HLT/NAACL, (2002)

[40] Satanjeev Banerjee and Alon Lavie *METEOR : An automatic metric for MT evaluation with improved correlation with human judgement*, IIn proceedings ofWorkshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or summarization, ACL, (2005)

[41] De Bellis, N. *Bibliometrics and Citation Analysis : From the Science Citation Index to Cybermetrics.*,Scarecrow Press, Lanham (2009)

[42] Vargas-Quesada, B., Moya-Anegón, F. *Visualizing the Structure of Science.*,Springer, Heidelberg (2007).

[43] Cobo, M., López-Herrera, A., Herrera-Viedma, E., Herrera, F. *Science mapping software tools : review, analysis, and cooperative study among tools.*,J. Assoc. Inf. Sci. Tech. 62, 1382−1402 (2011)

[44] Van Eck, N., Waltman, L. *Software survey : vosviewer, a computer program for bibliometric mapping.* ,Scientometrics 84, 523−538 (2010)

[45] Salton, G., Bergmark, D. *A citation study of computer science literature.*,IEEE Trans. Prof. Commun. 22, 146−158 (1979)

[46] Wasserman, S., Faust, K. *Social Network Analysis : Methods And Applications (Structural Analysis in the Social Sciences).*, Cambridge University Press, Cambridge (1994)

[47] Serrano, E., Quirin, A., Botia, J., Cordón, O. *Debugging complex software systems by means of pathfinder networks.*, Inf. Sci. 180(5), 561−583 (2010)

[48] Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Muñoz-Fernández, F. *A new technique for building maps of large scientific domains based on the cocitation of classes and categories.*, Scientometrics 61(1), 129−145 (2004)

[49] Pancho, D., Alonso, J., Cordón, O., Quirin, A., Magdalena, L. *FINGRAMS : visual representations of fuzzy rule-based inference for expert analysis of comprehensibility.*, IEEE Trans. Fuzzy Syst. 21(6), 1133−1149 (2013)

[50] di Battista, G., Eades, P., Tamassia, R., Tollis, I. *Graph Drawing : Algorithms for the Visualization of Graphs.*, Prentice Hall, Upper Saddle River (1998)

[51] Kobourov, S.G. *Force-directed drawing algorithms. In : Tamassia, R. (ed.) Hand- book of Graph Drawing and Visualization.*, CRC Press, Boca Raton (2012)

[52] Hugh Waddington , Howard White , Birte Snilstveit , Jorge Garcia Hombrados , Martina Vojtkova , Philip Davies , Ami Bhavsar , John Eyers , Tracey Perez Koehlmoos , Mark Petticrew , Jeffrey C. Valentine, Peter Tugwell *How to do a good systematic review of effects in international development : a tool kit.*, Journal of Development Effectiveness, 4 :3, 359-387, (2012) ,

[53] KARITA S ;WANG X ;WATANABE S ;YOSHIMURA T ;ZHANG W ;CHEN N ;HAYASHI T ;HORI T ;INAGUMA H ;JIANG Z ;SOMEKI M ;SOPLIN NEY ;YAMAMOTO R *A Comparative Study on Transformer vs RNN in Speech Applications*, IEEE Automatic Speech Recognition and Understanding Workshop 2019

[54] DUONG L ;ANASTASOPOULOS A ;CHIANG D ;BIRD S ;COHN TL *An Unsupervised Probability Model for Speech-to-Translation Alignment of Low-Resource Languages*, Cornell University Sep 2016

[55] JIA Y ;JOHNSON M ;MACHEREY W ;WEISS RJ ;CAO Y ;CHIU CC ;ARI N ;LAURENZO S ;WU Y *Leveraging Weakly Supervised Data to Improve End-to-End Speech-to-Text Translation*, Cornell University Nov 2018

[56] Mrinalini K,Vijayalakshmi P *Hindi-English Speech-to-Speech Translation System/or Travel Expressions*,2015

[57] Ye Jia, Ron J. Weiss, Fadi Biadsy, Wolfgang Macherey, Melvin Johnson, Zhifeng Chen, Yonghui Wu *Direct speech-to-speech translation with a sequence-to-sequence model*

[58] Andros Tjandra, Sakriani Sakti, Satoshi Nakamura *SPEECH-TO-SPEECH TRANSLATION BETWEEN UNTRANSCRIBED UNKNOWN LANGUAGES*, 2019

[59] Takatomo Kano, Sakriani Sakti, Satoshi Nakamura *TRANSFORMER-BASED DIRECT SPEECH-TO-SPEECH TRANSLATION WITH TRANSCODER*, 2021 IEEE Spoken Language Technology Workshop (SLT),25 March 2021.

[60] Adèle J ATTEAU, Annie M ONTAUT *Phonologie du HINDI et de l'OURDOU*, 2017.

[61] Shahabullah, Rahman, Khan *Syllabification of English Words by Pashto Speakers*, 2020

[62] Claartje Levelt, Ruben Van de Vijver *Syllable types in cross-linguistic and developmental grammars*

[63] Hideo Torii *A Comparative Analysis of the English and the Japanese Syllables*, 2004.

[64] Jose M. Alonso ;Ciro Castiello ;Corrado Mencar ; *A Bibliometric Analysis of the Explainable Artificial Intelligence Research Field*, International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IPMU 2018 : Information Processing and Management of Uncertainty in Knowledge-Based Systems. Theory and Foundations pp 3−15, 18 May 2018

[65] Mario A. Rojas-Sánchez, Pedro R. Palos-Sánchez , José A. Folgado-Fernández *Systematic literature review and bibliometric analysis on virtual reality and education*, Education and Information Technologies ,27 juin 2022.

[66] Prachi Kadam, Nayana Petkar, Shraddha Phansalkar Dr. *A Systematic Literature Review With Bibliometric Meta-Analysis Of Deep Learning And 3D Reconstruction Methods In Image Based Food Volume Estimation Using Scopus, Web Of Science And IEEE Database*, Library Philosophy and Practice (e-journal) ,24 juin 2020.

[67] Hamid Darvish *Bibliometric Analysis using Bibliometrix an R Package*, Journal of Scientometric Research,2019,8,3,156-160.

,

[68] Langue selon Ferdinand de Saussure https://www.coursum3.org/itic/?wpdf_download_file=/home/ichigo1vs/www8/wp-content/uploads/cours/ITIC/Sciences%20du%20Langage%20parcours%20Sciences%20du%20langage/l1%20science%20du%20language/Sociolinguisitique/Linguistique_definitions.pdf consulté le 04 Decembre 2022.

[69] Famille nigero-congolaise https://www.axl.cefan.ulaval.ca/afrique/benin.htm consulté le 04 Septembre 2021.

[70] Famille nigero-congolaise https://www.axl.cefan.ulaval.ca/monde/famnigero-congolaise.htm consulté le 04 Septembre 2021.

[71] Expression des temps en Fongbe https://www.oocities.org/fon_is_fun/French/fr_grammar_And_pronounciation.htm, consulté le 04 Decembre 2020.

[72] Expression des temps en Yoruba http://learn101.org/fr/yoruba_verbes.php, consulter le 04 Décembre 2020.

[73] Notion de langue proche https://www.cairn.info/revue-ela-2004-4-page-393.html consulté le 04 Janvier 2021.

[74] Notion d'utterance https://www.engati.com/glossary/utterance, consulter le 04 Janvier 2021.

[75] Notion MFCC —Mel Frequency Cepstral Co-efficients https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd, consulter le 04 Janvier 2021.

[76] Echelle de Mel https://readinganswer.fr/quelles-sont-les-caracteristiques-de-laudio/, consulter le 04 Janvier 2021.

[77] Natural Language Processing (NLP) What it is and why it matters https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html consulté le 27 Aout 2021.

[78] Adit Deshpande, Deep Learning Research Review Week 3 : Natural Language Processing https://adeshpande3.github.io/adeshpande3.github.io/Deep-Learning-Research-Review-Week-3-Natural-Language-Processing consulté le 12 Octobre 2021.

[79] Historique de la traduction automatique https://towardsdatascience.com/machine-translation-a-short-overview-91343ff39c9f consulté le 04 Août 2021. ======

[80] Plateforme de traduction automatique https://www.science-et-vie.com/article-magazine/traduction-automatique-lia-est-en-train-de-faire-tomber-la-barriere-de-la-langue consulté le 04 Août 2020.

[81] Biggest Open Problems in Natural Language Processing https://medium.com/dair-ai/deep-learning-for-nlp-an-overview-of-recent-trends-d0d8f40a776d consulté le 04 Septembre 2021.

[82] Presentation de LSTM GRU https://penseeartificielle.fr/comprendre-lstm-gru-fonctionnement-schema/ consulté le 04 Août 2020.

[83] Presentation de l'attention https://inside-machinelearning.com/mecanisme-attention/ consulté le 04 Août 2021.

[84] Reseau de neurines https://ledatascientist.com/a-la-decouverte-du-transformer/ consulté le 04 janvier 2022.

[85] Definition de Mos https://www.twilio.com/docs/glossary/what-is-mean-opinion-score-mos, consulter le 04 Décembre 2020.

[86] Bibliometrix https://warin.ca/shiny/bibliometrix/ consulté le 10 janvier 2022.

[87] VOSviewer https://ocean.sagepub.com/research-tools-database/vosviewer consulté le 10 janvier 2022.

[88] Voyelle hindi https://hindi.swiftutors.com/hindi-basics.html consulté le 10 juillet 2021.

[89] Voyelle Française https://www.theschoolrun.com/what-are-vowels-and-consonants consulté le 10 juillet 2021.

[90] Voyelle espagnol https://www.mimicmethod.com/spanish-pronunciation-ultimate-guide/ consulté le 10 juillet 2021.

[91] Voyelle japonnais https://www.linguajunkie.com/japanese/japanese-consonants consulté le 10 juillet 2021.

[92] Voyelle korean https://linguistics.byu.edu/classes/Ling450ch/reports/Korean3.html consulté le 10 juillet 2021.

[93] Type sylabique du korean https://exploredprk.com/learn-korean/lesson-2-korean-syllables/ consulté le 10 juillet 2021.

[94] Typologie syntaxique de l'hindi https://www.hindipod101.com/blog/2020/08/07/hindi-word-order/ consulté le 10 juillet 2021.

[95] Typologie syntaxique du Français anglais https://www.uio.no/studier/emner/hf/ikos/EXFAC03-AAS/h06/larestoff/linguistics/Chapter6(H06).pdf, consulté le 04 Decembre 2021.

[96] Typologie syntaxique du korean https://www.koreanclass101.com/blog/2020/08/07/korean-word-order/ consulté le 10 juillet 2021.

[97] Expression du temps en hindi https://www.hindipod101.com/blog/2021/07/08/hindi-tenses/ consulté le 10 juillet 2021.

[98] Expression du temps en espagnol https://blog.lingoda.com/en/basic-verb-conjugation-in-spanish-for-beginners/ consulté le 10 juillet 2021.

[99] Expression du temps en japonais https://blog.lingodeer.com/japanese-verb-conjugation-guide/ consulté le 10 juillet 2021.

[100] Expression du temps en korean https://www.howtostudykorean.com/unit1/unit-1-lessons-1-8/unit-1-lesson-5/ consulté le 10 juillet 2021.

[101] Caracteristique de l'hindi https://hindi.swiftutors.com/hindi-basics.html consulté le 10 juillet 2021.

[102] Caracteristique du Korean https://thinkzone.wlonk.com/Language/Korean.htm consulté le 10 juillet 2021.

Table des matières

TABLE DES MATIÈRES