

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

*Ministère de l'Enseignement Supérieur et de la Recherche Scientifique*

Université MOHAMED KHIDER BISKRA  
Faculté des sciences et des sciences d'ingénieur  
Département d'informatique

Mémoire en vue de l'obtention du diplôme de

MAGISTER EN INFORMATIQUE  
Option  
INTELLIGENCE ARTIFICIELLE ET IMAGES

*Thème*

Une approche connexionniste basée  
sur les cas sémantiques pour la  
traduction automatique du langage  
naturel  
«cas de la langue arabe à la langue française»

Présenté par  
REZEG Khaled

Jury

P<sup>r</sup> M. BOUFAIDA  
D<sup>r</sup> M.T. LASKRI  
D<sup>r</sup> N. DJEDDI  
D<sup>r</sup> A. HENNI

Président  
Rapporteur  
Examineur  
Examineur

Juin 2003

## REMERCIEMENTS

*Mes remerciements les plus vifs sont tout d'abord adressés à Monsieur Mohamed Tayeb LASKRI, Recteur, et Maître de conférence à l'université BADJI-Mokhtar (Annaba), pour m'avoir proposé et dirigé ce travail avec patience et disponibilité. Je tiens à lui exprimer ma profonde gratitude et mon profond respect.*

*Je tiens également à exprimer ma reconnaissance aux membre de jury :*

*Monsieur Mahmoud BOUFAIDA, Professeur à l'université de Constantine qui me fait l'honneur de présider mon jury.*

*Monsieur Noureddine DJEDDI, Maître de conférence à l'université de Biskra pour l'honneur qu'il me fait en acceptant d'examiner ce travail.*

*Monsieur Abderrazek HENNI, Directeur et Docteur d'Etat à l'Institut National d'Informatique I.N.I Oued Smar Alger pour l'honneur qu'il me fait en acceptant de faire partie du jury.*

*Je n'omettrai pas de remercier tout les enseignants de l'institut d'informatique de l'université de Biskra.*

*Je voudrais exprimer aussi ma gratitude envers tous ceux qui par leurs conseils ou leurs amitié, m'ont apporté leur soutien, en particulier : Med Faouzi ZERARKA, Toufik BENDAHMANE, Aissa BOUABDELLAH, ma tante REZEG Khadidja et d'autres.*

*Enfin, à ma famille pour son soutien indéfectible, ses encouragements, leur disponibilité et leur compréhension et sans lesquelles ce thèse n'aurait pas été possible.*

## DEDICACES

*Je dédie ce travail*

*A mes parents*

*A ma femme et mon fils Mohamed Slimane*

*A mon frère et sa famille*

*A mes sœurs Amel, Hassiba, Nadjoua, Soumia et sa famille*

*A mon oncle et sa grande famille*

*A mes tantes et leurs familles*

*A toute ma famille et ma belle famille*

*A tous mes amis*

## **RESUME :**

*La traduction automatique des textes d'origine humaine est une application fort complexe appelée à appréhender les univers textuels ouverts, sans aucune contrainte quant à leur nature ou à leur diversité. Pour résoudre cette problématique, plusieurs tentatives ont été entamées ayant à chaque fois pour objectif l'obtention d'une meilleure qualité de traduction. Mais devant les diverses ambiguïtés de la langue naturelle, ce problème de traduction est loin d'être simple à résoudre. En effet, en absence de la maîtrise du contexte la plupart des phrases sont ambiguës. C'est pour cette raison, que les recherches actuelles dans ce domaine se sont orientées d'abord vers la maîtrise du sens véhiculé par les phrases avant de procéder au processus même de la traduction. En effet, traduire sans comprendre mène directement vers un échec. Pour ce faire, et dans le but d'augmenter la qualité de traduction, nous proposons une approche neuronale pour la génération des différents cas sémantiques relatifs aux différents composants de la phrase pour cerner d'abord le sens et de générer ensuite la traduction dans la langue cible. Ce qui nous a permis d'obtenir des résultats satisfaisants, comparativement à des travaux similaires utilisant d'autres techniques. Effectivement, Les techniques connexionnistes sont robustes, elles ont la réputation d'offrir de bonnes capacités, elles sont parallèles et elles s'appuient sur les notions de calcul réparti et de raisonnement distribué.*

*Dans notre approche, nous utilisons un apprentissage supervisé par un réseau simplement récurrent (SRN) pour apprendre les techniques de traduction avec une performance similaire à celle des humains, ce qui constitue une première phase de notre système, celle de l'apprentissage. Une deuxième phase du système constitue la partie test et généralisation pour mesurer le degré de généralisation du réseau par rapport à ce qu'il a appris durant la phase d'apprentissage.*

*Le réseau de neurones utilisé, tient compte de la représentation approfondie en utilisant les cas sémantiques et la représentation surfacique en utilisant les éléments qui indiquent la forme de la phrase.*

## **MOT CLES :**

*langage naturel, traduction automatique, T.A.L.N., Apprentissage, généralisation, réseau de neurone, cas sémantique, dictionnaire, représentation approfondie, représentation surfacique.*

## **ABSTRACT :**

*The automatic translation of the human origin texts is a complex strong implementation called to apprehend the open universe textuels without any constraints towards nature or to their diversity. To resolve this problematic, several trials have been made every time with the goal of obtaining the best quality of translation , but in front of the different ambiguities of the natural language, this problem of translation is too far to be solved. In fact, in absence of the language mastery of context, most of phrases are ambiguous. To this reason, the current researches in the domain were oriented at first to the mastery of the meaning vehiculed by phrases before proceeding to the translation process itself. In fact, translating without understanding leads directly to failure. In order to elevate the translation quality, we suggest a neuronal approach for the generation of the different semantic cases related to the different parts of the phrase to wicle the sens first and to generate the translation in the aimed language later. It is what allowed us to obtain satisfying results in comparison to the same works using other technics. Effectively, the connexionnist technics have good capacities of training and generalization, hardiness, mainly tolerance breakdowns of it, and parallel treatment possibility.*

*In our approach we use a supervised training by a simple recurrent neurons network (SRN) to learn translation technics with a performance similar to the human. Our system contain two phases, the first one is the training phase, the second is the test part and generalization to measure the network generalization degree of the network according to what was learnt during the training phase.*

*The uses neurons network hold account the deepened representation using the semantic cases and the surface representation using elements wich indicate the shape of phrase.*

## **KEY WORDS**

*natural language, automatic translation, T.A.L.N. , training, generalization, Neural Network, semantic cases, dictionary, deepened representation, surface representation.*

## ملخص :

الترجمة الآلية للنصوص تمثل تطبيقا غاية في التعقيد ، تهدف لترجمة النصوص بكل الأشكال دون التقييد بطبيعتها وتنوعها. و من أجل الوصول لترجمة آلية ذات نوعية جيدة عدة أعمال جادة أنجزت ، لكن غموض و صعوبة اللغة جعل من حل هذا الإشكال بعيد المنال.

للتحكم بشكل جيد في عملية الترجمة الأوتوماتيكية باحثي هذا المجال ركزوا أعمالهم على فهم الجملة في لغتها الأصلية أولا ثم إيجاد ما يقابلها في اللغة المراد الترجمة إليها، لأنه بكل بساطة الترجمة بدون فهم تؤدي إلى فشل مؤكد.

من خلال عملنا هذا نهدف إلى إيجاد نظام ترجمة آلية ذات مستوى حسن بإستعمال نظام شبكات الخلايا العصبية الإصطناعية التي تمتاز بقدرة التعلم و التعميم ، و قدرة مواجهة الضلال و القدرة على المعالجة المتوازية، معتمدين على نظرية حالات المعنى ( نظرية فيلمور 1968 ) التي تسمح بتحديد معنى الجملة من خلال تحديد دور كل عنصر من عناصرها، ثم القيام بالترجمة في اللغة المقابلة مما سمح لنا الحصول على نتائج مقبولة مقارنة بتقنيات أخرى. و تم تجسيد نظامنا على أساس التمثيل العميق للجملة (représentation approfondie) مستعملين حالات المعنى و التمثيل الظاهري للجملة (représentation surfacique) الذي تحدده مجموعة من العناصر التي تبين شكل الجملة الخارجي.

## الكلمات المفتاحية :

اللغة الطبيعية ، الترجمة الآلية ، المعالجة الآلية للغة الطبيعية ، التعلم ، التعميم ، شبكة الخلايا العصبية الإصطناعية ، حالات المعنى ، منجد ، التمثيل العميق ، التمثيل الظاهري.

## Sommaire

Résumé	01
Introduction générale .....	01
<b>Chapitre I : Les réseaux de neurones</b>	
1. Introduction .....	04
2. Bref historique .....	04
3. Fondement biologique .....	05
4. Le neurone artificiel ( Formel) .....	06
5. Structure et définition d'un réseau de neurones .....	08
6. L'apprentissage .....	09
6.1-Règle d'apprentissage .....	09
6.2-L'algorithme d'apprentissage .....	10
6.3-Les types d'apprentissage .....	10
6.3.1-L'apprentissage supervisé .....	11
6.3.2-L'apprentissage non supervisé .....	11
6.3.3-L'apprentissage par renforcement .....	11
6.3.4-Exemple d'algorithme d'apprentissage .....	11
7. Généralisation .....	13
8. Les principaux modèles connexionnistes et les méthodes d'apprentissage .....	13
8.1-Les réseaux à couches .....	13
8.1.1-Les réseaux à deux couches .....	13
8.1.1.1-Le perceptron .....	13
8.1.2-Les réseaux multicouches .....	14
8.1.2.1 Le perceptron Multi-couches .....	14
8.2- Les réseaux récurrents .....	15
8.2.1-Les réseaux à compétition .....	15
8.2.2-Les réseaux dynamiques .....	15
8.3-Les réseaux récurrents à couches .....	15
8.3.1- Le réseau de Jordan .....	15
8.3.2- Le réseau d'Elman .....	16
8.3.2.1-Les Variantes du SRN .....	18
9. Mise en oeuvre des réseaux de neurones .....	18
10. Avantages et inconvénients des réseaux de neurones .....	19
10.1-Les avantages .....	19
10.2 Les inconvénients .....	19
11. Conclusion .....	20
<b>Chapitre II : Traitement automatique du langage naturel T.A.L.N.</b>	
1. Introduction .....	21
1.1- Communication Homme-Machine .....	21
1.2- Compréhension Automatique .....	21
1.3- Bref Historique du T.A.L.N. ....	21
2. Le traitement automatique du langage naturel .....	22
3. Les niveaux d'analyse .....	23
3.1-Niveau morphologique .....	23
3.2-Niveau lexical .....	24

3.3- Niveau syntaxique .....	24
3.3.1-Grammaires formelles .....	24
3.3.2-Grammaires transformationnelles .....	24
3.3.3- Réseaux de transitions (Augmented Transition Network A.T.N) .....	25
3.3.4 Analyse à base de connaissances .....	25
3.4-Niveau sémantique .....	25
3.4.1-La logiques des prédicats .....	25
3.4.2- Les réseaux sémantiques .....	26
3.4.3-"Scripts" et "Frames" .....	27
3.4.3.1-Les Scripts .....	27
3.4.3.2-Les Frames .....	28
3.4.4- Les Grammaires de cas .....	29
3.4.4.1-Théorie de Charle fillmore .....	30
3.4.4.2-La structure argumentale .....	30
3.4.4.3-Liste et définition des cas .....	30
3.4.5-Classification des noms de Chafe .....	32
3.4.5.1-Topologie des noms .....	32
3.4.6 Les méthodes des dépendances conceptuelles .....	33
3.5 Niveau pragmatique .....	34
4.Architecture hiérarchique .....	35
5.Le traitement connexionniste du langage naturel .....	36
6.Conclusion .....	37

### Chapitre III : Connexionisme et le traitement automatique du langage naturel C.N.L.P

1. Introduction .....	38
2. Représentation des connaissances .....	38
2.1- Caractéristiques d'une représentation .....	38
2.2- Les types de représentation .....	39
2.2.1-Représentation locale .....	39
2.2.2-Représentation distribuée .....	39
2.2.3-Représentation semi- distribuée .....	40
2.3-Les problèmes de la représentation .....	40
2.3.1-La productivité .....	40
2.3.2-La systémacité .....	40
2.3.3-La compositionnalité .....	41
2.3.3.1-Assemblage .....	41
2.3.3.2-La référence .....	42
3.L'importance du contexte dans le traitement du langage naturel .....	42
4.Les systèmes connexionnistes pour le traitement automatique du langage naturel .....	43
4.1-Traitement de surface .....	44
4.1.1-Système PARNISP .....	44
4.1.2-Réseau totalement récurrent de Kamimura .....	45
4.2-Désambiguïsation lexicale .....	46
4.2.1-Système de désambiguïsation lexicale de Benello .....	46
4.2.2-Modèle de la désambiguïsation lexicale en langue française du Michalak et Rolbert .....	47

4.3- Traitement de la structure syntaxique de la phrase .....	47
4.3.1-Modèle de Kawamoto .....	47
4.3.2-Réseau simplement d'Elman .....	48
4.3.3- Modèle de Jodouin .....	50
4.3.4-Système XERIC .....	50
4.3.5-Système hybride de Kwasny et Faisal .....	52
4.3.6- Travaux de Descroques .....	52
4.3.7-Réseau d'Archambault .....	53
4.4- Analyse sémantique .....	54
4.4.1-Réseau récurrent de Victorri, Raysz et Konfe .....	55
4.4.2-Réseau de Ide et Veronis .....	56
4.4.3-Modèle de Wermeter et Lehner .....	57
4.4.4-Système SAILI .....	58
4.4.5-Réseau NASK .....	58
4.4.6-Système de Koon HC .....	60
4.4.7-Système de Risto Miikkulainen .....	62
4.4.8-Système SPEC .....	63
5-Conclusion .....	65

#### Chapitre IV : La traduction automatique

1-Introduction .....	66
2-Bref Historique .....	66
3-La traduction .....	67
3.1- Traduction et compréhension .....	67
3.2. Traduction et Profondeur .....	68
3.3- Traduction, Ambiguïté et Surgénération .....	69
3.3.1-Ambiguïté .....	69
3.3.2-Types d'ambiguïtés .....	70
3.3.2.1-L'ambiguïté lexicale .....	70
3.3.2.2-L'ambiguïté de structure de surface .....	70
3.3.2.3-L'ambiguïté de structure profonde .....	70
3.3.2.4-L'ambiguïté sémantique .....	70
3.3.2.5-L'ambiguïté Intrinsèque .....	70
3.3.3- Surgénération .....	71
3.3.3.1- Surgénération et traduction .....	71
4-La traduction automatique .....	72
4.1-Méthode de traitement de la traduction automatique .....	72
4.1.1-La méthode directe .....	72
4.1.2-La méthode par transfert .....	72
4.1.3- La méthode par interlangue par langage pivot .....	72
4.2-Exemple de système de traduction automatique .....	72
5-Les logiciels de traduction .....	73
5.1 Exemple de logiciels de traduction .....	74
6-Conclusion .....	76

#### Chapitre V : Traducteur automatique du langage naturel par une     approche connexionniste T.A.L.N.A.C.

1-Introduction .....	77
2-Structures de la phrase .....	78

3-Description du système T.A.L.N.A.C .....	78
3.1- Représentation Interne de la phrase .....	78
3.2- Représentation surfacique de la phrase .....	80
3.3- Représentation des connaissances .....	81
4-Architecture du système T.A.L.N.A.C .....	82
4.1- Topologie du réseau .....	82
4.1.1- La couche d'entrée .....	83
4.1.2- La couche de sortie .....	85
4.2- Fonctionnement du réseau .....	86
4.2-1- Propagation .....	86
4.3- Sous système d'apprentissage .....	88
4.3-1- Le module de construction de corpus d'apprentissage .....	88
4.3.1.1- Les formules utilisées dans l'algorithme d'apprentissage ..	89
4.3.1.2- Les paramètres d'apprentissage .....	90
4.3.1.3- Le temps d'apprentissage .....	90
4.4- Sous système de test ou de généralisation .....	91
4.4.1- Module de construction de Frames et de pré traitement .....	91
4.4.2- Module de test et généralisation .....	92
4.4.3- Module d'interprétation .....	92
4.5- Le dictionnaire .....	92
4.6- Exemple explicatif .....	93
5-Résultats expérimentaux .....	95
5.1- Résultats des test de l'apprentissage .....	96
5.1.1- Interprétation des résultats des test de l'apprentissage .....	100
5.2- Résultats des test de la généralisation .....	100
5.2.1- Interprétation des résultats des test de la généralisation .....	100
6-Conclusion .....	101
Conclusion générale .....	102
Bibliographie .....	106
Glossaire .....	109

Annexe A : Description du système

Annexe B : Résultats des statistiques d'apprentissage

Annexe C : Résultats des statistiques de la généralisation