

**Pourquoi les *fonctions lexicales* seraient-elles acceptables
pour le Traitement Automatique des Langues ?**

Alhassan Ahmed Abdelbary Abdellatif (*)

Résumé

L'objectif de ce papier est de présenter, tout d'abord, dans ses grandes lignes, le cadre formel de notre projet doctoral qui a pour objet la modélisation formelle des *Constructions à Verbe Support* (désormais CVS) françaises-arabes, postulé dans le cadre théorique de la *Théorie Sens-Texte* (désormais TST) (Mel'čuk, 1997 ; Polguère, 1998). Nous souhaitons insister sur le fait que cet exposé n'est pas un survol impartial du domaine des fonctions lexicales (désormais FL), loin de là. Il s'agit clairement d'un point de vue personnel sur la question. De là, la suite de cet article se divise en deux grandes parties: premièrement, nous tenons de revenir un peu sur la notion de FL en la dégagant du formalisme auquel elle est associée. Puis, nous proposons des pistes précises pour l'utilisation des FL pour le *Traitement Automatique des Langues* (désormais TAL).

Mots-clés: Théorie Sens-Texte (TST) – Fonctions Lexicales (FL) – système lexical – modélisation formelle – base de données lexicales – réseau lexical.

1. Introduction

En français, on dit *faire une promenade*, en anglais *take a journey*, en espagnole *dar un paseo*, en allemand, *Eine Reise*

(*) Cet article est extrait de la thèse doctorale intitulée : “Une modélisation formelle des constructions à verbe support en français et en arabe : une étude sémantico-syntaxique”, sous la direction du M. Denis Maurel- Université de Tours – France & M. Mohamed El-Hannach – Université Sidi Mohamed Ben Abdallah –Fès / Maroc & M. Hazem Ali Kamal eddine– Université de Sohag & Mme Salwa Hussein– Université de Héliwan..

machen et en arabe *يَقُومُ بِرَحْلَةٍ*. Toutes ces constructions bizarres, qui sont en même temps connues depuis longtemps dans chaque langue, ont besoin d'une description formelle calculable, logique et exhaustive bien structurée. En effet, la linguistique ne découvre pas une solution satisfaisante et bienséante pour traiter ces constructions ci-mentionnées d'une façon cohérente et rationnelle. L'approche Sens-Texte, à son tour, propose un modèle fonctionnel de ce phénomène: *les fonctions lexicales*.

En fait, le concept de FL repose essentiellement sur l'hypothèse que les cas de cooccurrence lexicale restreinte du type de ces constructions sont généraux. En même temps, le caractère variable et imprévisible de la cooccurrence lexicale est avéré. Au moyen des FL, le linguiste modélise les choix lexicaux d'une façon systématique et rigoureuse fait spontanément par le locuteur. D'ailleurs, on ne peut pas assurer d'une façon stricte que l'utilisation des FL représente exactement ce qui se passe dans le cerveau des locuteurs, mais constatons que la systématisme et la cohérence de la description témoigne de la validité de cette modélisation.

Pour modéliser les liens qu'entretiennent les lexies de la langue, les FL utilisent une soixantaine de fonctions lexicales. On distingue, d'ailleurs, deux grands groupes de fonctions lexicales : les FL paradigmatisques vs les FL syntagmatiques. Les FL, définies dans le cadre de la TST, sont déjà connues comme un outil pour décrire les relations sémantico-syntaxiques, paradigmatisques et syntagmatiques, entre les lexies. D'ailleurs, la *Lexicologie Explicative et Combinatoire* [LEC] constitue la composante lexicale de la TST. Elle vise, d'une part, à mettre en évidence les rapports lexicaux entre les lexies, et à démontrer leur combinatoire lexicale et syntaxique, d'autre part.

Les FL forment donc un système formel structuré qui doit être appris comme tel. En outre, comme ce système est assez complexe pour être compris comme un bloc, il nous semble

tout d'abord important de le concevoir en profondeur d'un fragment isolé. Nous proposons donc des pistes pour introduire cette notion fondamentale d'une bonne façon : qui nous permette d'en saisir en profondeur la valeur, et l'usage des contextes variés.

2. Problématique de la recherche

La problématique de l'article est donc liée à l'acceptabilité des FL dans le domaine du TAL. Pour clarifier davantage cette problématique, l'article pourrait explorer des questions telles que :

- Qu'est-ce qu'une FL dans le contexte du TAL ? Il pourrait être nécessaire de définir ce concept pour les lecteurs qui ne sont pas familiers avec le domaine.
- Quels sont les avantages potentiels d'utiliser des FL dans le TAL ? Comment pourraient-elles améliorer la compréhension et le TAL ?
- Quels sont les défis et les limites associés à l'utilisation de FL dans le TAL ? Peuvent-elles être applicables à tous les types de langues et de contextes ?
- Quelles méthodologies ou approches sont utilisées pour évaluer l'efficacité des FL dans le TAL ? Quels sont les résultats de telles évaluations ?
- Comment les FL se comparent-elles à d'autres méthodes ou techniques du TAL ? Y a-t-il des scénarios spécifiques où elles se démarquent particulièrement ?

3. Objet de recherche et mise en contexte

Dans cet article, nous essaierons d'aborder la question des FL qui sont réservées à la description lexicographique des langues naturelles. Nous tenterons, d'ailleurs, de préciser quelles sont les FL qui seraient acceptables pour décrire formellement les CVS. Ces FL s'intéressent à repérer les différents liens paradigmatiques (les dérivations sémantiques),

et syntagmatiques (les relations base-collocatif) qui sont valides pour la modélisation formelle de ces constructions.

Cette modélisation pourrait aboutir à une appréhension des constructions pour être prêtes/préparées à s'intégrer dans un réseau lexical. Notre objet offre donc une courte présentation aux différentes FL selon les principes théoriques et descriptifs de la *Lexicologie Explicatif et Combinatoire* (LEC), puis à identifier les différentes étapes à suivre pour la rédaction des entrées lexicales dans le DiCo.

4. Les approches de Traitement Automatique des Langues

(TAL)

Le TAL est un domaine de recherche qui vise à développer des algorithmes et des systèmes capables d'analyser, de comprendre et de générer du langage naturel de manière automatique. Il existe plusieurs approches pour le TAL, chacune ayant ses propres avantages et limites. Il est important de noter que les approches de TAL sont souvent combinées pour répondre aux besoins spécifiques des applications. Par exemple, un système de traduction automatique peut utiliser des approches statistiques pour la traduction de base, mais incorporer des éléments de la sémantique et des connaissances expertes pour améliorer la qualité de la traduction. D'autre part, les approches qui traitent du TAL sont nombreuses comme suit : Manning & Schütze (1999), Jurafsky & Martin (2020), Charniak (1997), Goldberg (2017), Bird & Klein & Loper (2009) :

[1] Approches basées sur des règles : Ces approches sont basées sur la définition de règles grammaticales et linguistiques pour analyser et générer du langage naturel. Les systèmes basés sur des règles sont souvent très précis, mais ils sont également très coûteux en temps et en ressources, car ils nécessitent l'expertise de linguistes et de spécialistes de la langue pour créer les règles.

[2] Approches statistiques : Ces approches statistiques utilisent des techniques d'apprentissage automatique pour extraire des modèles à partir de grands ensembles de données linguistiques. Ces approches sont moins précises que les systèmes basés sur des règles, mais elles sont plus efficaces pour traiter de grands volumes de données et peuvent être utilisées pour des langues pour lesquelles il n'existe pas de ressources linguistiques suffisantes.

[3] Approches hybrides : Ces approches combinent des éléments des approches basées sur des règles et statistiques pour combiner la précision des systèmes basés sur des règles avec la capacité de traitement de grands volumes de données des approches statistiques.

[4] Approches basées sur l'apprentissage en profondeur : Ces approches utilisent des réseaux de neurones artificiels pour apprendre à comprendre et à générer du langage naturel de manière automatique. Ces approches sont très prometteuses pour le TAL car elles ont montré des résultats significatifs dans des domaines tels que la traduction automatique et la reconnaissance vocale.

[5] Approches basées sur la sémantique distributionnelle : Ces approches se concentrent sur la signification des mots et des phrases plutôt que sur leur forme grammaticale. Elles utilisent des modèles sémantiques pour représenter la signification des mots et des phrases, et peuvent être utilisées pour des tâches telles que l'analyse de sentiment, la réponse automatique aux questions, et la compréhension du langage naturel. Elles utilisent, d'ailleurs, des modèles statistiques pour capturer les relations sémantiques entre les mots dans un corpus de textes. Ils sont souvent utilisés pour des tâches de traitement de texte telles que la représentation vectorielle des mots, la mesure de similarité entre les mots et l'analyse des sentiments.

[6] Approches basées sur les connaissances : Ces approches utilisent des connaissances expertes ou des bases de connaissances pour comprendre et générer du langage naturel. Par exemple, les systèmes de question-réponse peuvent utiliser des bases de connaissances pour répondre à des questions spécifiques. Ces approches sont souvent très précises mais nécessitent des ressources importantes pour la création et la maintenance des bases de connaissances.

[7] Approches basées sur l'apprentissage en profondeur : Ces approches utilisent des réseaux de neurones artificiels pour faire apprendre à partir de données d'apprentissage et sont souvent utilisées pour des tâches de compréhension du langage naturel telles que la classification de texte, la reconnaissance d'entités nommées, et la traduction automatique. Ces approches ont révolutionné le domaine du TAL ces dernières années en permettant des améliorations significatives des performances des systèmes.

[8] Approches basées sur l'apprentissage par renforcement : Ces approches utilisent une méthode d'apprentissage par renforcement pour apprendre à un agent à interagir avec son environnement de manière à maximiser une récompense donnée. Ces approches sont souvent utilisées pour des tâches de dialogue, telles que la réponse automatique aux questions, et peuvent être combinées avec d'autres approches de TAL pour améliorer les performances du système.

En gros, les approches de TAL sont nombreuses et variées, et chaque approche a ses avantages et ses limites en fonction de la tâche spécifique à accomplir. Le choix de l'approche dépendra des besoins spécifiques de chaque application de TAL et des ressources disponibles. D'autre part, les avancées récentes dans les domaines de l'apprentissage en profondeur et de l'apprentissage par renforcement ont permis des améliorations significatives dans les performances des systèmes de TAL, ouvrant la voie à des applications encore plus avancées dans le domaine.

5. Les FL et le TAL

Constatons ici que le domaine de TAL a suscité depuis plusieurs années les intérêts des linguistes : Béchet & Bouillon (2003), Mellet, Fabre & Gleizes (2005), Habash & Rambow (2005), A. Belaïd & Y. Belaïd (2005), Mesfar (2006), Tarpin-Bernard (2007), Boitet & Blanchon (2007), Gross & Tasmowski (2009), Silberztein & Masson (2012), Jacquet (2014), Sagot (2014), Abdaoui & Diop (2015), Crabbé & Iordachioaia (2015), Jacquelard & Namer (2016), Langlais & Patry (2017), etc.

D'ailleurs, les FL ont une grande importance dans le TAL, qui continue de susciter l'intérêt des chercheurs et des professionnels dans le domaine, car elles permettent de mieux comprendre et d'analyser les textes, ce qui est essentiel pour de nombreuses perspectives pour des applications innovantes et utiles, telles que la traduction automatique, l'analyse de sentiments, la génération de textes, etc.

Cependant, l'utilisation des FL en TAL soulève également des défis et des limites, notamment en ce qui concerne leur identification et leur traitement automatique, ainsi que la difficulté de les modéliser de manière formelle.

Pour surmonter ces défis, de nombreuses recherches ont été menées pour développer des approches innovantes pour la reconnaissance et l'analyse des FL, telles que l'utilisation de l'apprentissage automatique, des réseaux de neurones, des modèles probabilistes. En outre, il est important de prendre en compte les défis éthiques liés au TAL, en particulier en ce qui concerne la protection de la vie privée et la lutte contre la discrimination et les préjugés. Les systèmes de TAL doivent être conçus de manière à garantir que les données personnelles des utilisateurs sont protégées et que les résultats obtenus sont impartiaux et justes.

Il convient également de noter que les FL ne sont pas universelles et peuvent varier selon les langues et les cultures. Ainsi, les typologies présentées pourraient être différentes pour

d'autres langues, ce qui souligne l'importance de mener des études linguistiques approfondies pour chaque langue en particulier.

En outre, les FL ne sont qu'un aspect du TAL, qui est un domaine multidisciplinaire qui implique également d'autres aspects tels que la morphologie, la syntaxe, la sémantique. L'analyse des FL peut également être utilisée pour étudier et comprendre les différences et les similitudes entre les langues, ainsi que les processus cognitifs impliqués dans la production et la compréhension des langues naturelles. Cette approche peut contribuer à une meilleure compréhension de la nature humaine et de notre capacité à communiquer et à partager des idées.

Il est important de souligner donc que le TAL est un domaine en constante évolution, et que de nouvelles approches et techniques pour l'analyse des FL et d'autres aspects linguistiques continueront d'être développées dans les années à venir, avec l'objectif de créer des systèmes de TAL encore plus performants et efficaces.

Pour évaluer leur exploitabilité dans le TAL, nous lançons dans cet article quatre critères des FL : sa lisibilité, son degré de formalisation, le degré de validité de son contenu informatif, et sa richesse en information. la distinction de ces critères nous permettrait d'évaluer l'exploitabilité des FL dans TAL. Dans ce qui suit, nous présentons brièvement ces quatre critères.

5.1. Lisibilité de fonctions lexicales

La question de la lisibilité des FL pour les systèmes de TAL est un sujet de débat et de recherche dans le domaine. Bien que les FL puissent apporter des informations importantes sur la structure syntaxique et sémantique des phrases, leur complexité et leur variabilité peuvent rendre difficile leur intégration dans les systèmes de TAL.

Certaines approches de TAL ont cherché à utiliser des ressources lexicales telles que WordNet et FrameNet pour améliorer la précision des tâches de traitement de texte, en

utilisant des méthodes telles que la désambiguïstation lexicale et la reconnaissance de la collocation. Cependant, ces approches nécessitent des ressources lexicales bien développées et des modèles de traitement de texte sophistiqués pour être efficaces.

D'autres approches ont cherché à utiliser des méthodes d'apprentissage automatique pour capturer les propriétés des FL et leur utilisation dans le langage naturel. En outre, certains chercheurs ont proposé des approches hybrides qui combinent les ressources lexicales et l'apprentissage automatique pour améliorer la performance des systèmes de TAL. Par exemple, une approche de traitement de texte appelée "bootstrapping" combine l'utilisation de règles lexicales avec l'apprentissage automatique pour améliorer la précision de l'analyseur syntaxique.

D'autre part, la lisibilité des FL pour les systèmes de TAL dépendra de la complexité et de la variabilité des FL elles-mêmes, ainsi que de la sophistication des modèles de traitement de texte utilisés pour les intégrer dans les systèmes de TAL. Les progrès dans le développement de ressources lexicales et de modèles de traitement de texte peuvent contribuer à améliorer la lisibilité des FL pour les systèmes de TAL.

Dans l'ensemble, bien que la lisibilité des FL pour les systèmes de TAL soit un défi, leur potentiel pour améliorer la précision et la qualité des analyses syntaxiques et sémantiques en fait une ressource précieuse pour le développement de systèmes de TAL plus avancés. Cependant, leur complexité et leur variabilité rendent difficile leur utilisation par les systèmes de TAL sans traitement préalable. Les ressources lexicales et les approches hybrides peuvent aider à surmonter ces défis de lisibilité, et la recherche continue dans ce domaine peut contribuer à améliorer l'utilisation des FL pour le TAL.

5.2. Son degré de formalisation

Constatons que les FL peuvent être enseignées, observées et même employées comme un outil conceptuel indépendamment de leur formalisme d'encodage traditionnel. Effectivement, les FL sont des outils essentiels pour l'analyse et la modélisation du langage naturel. Leur formalisation mathématique permet de préciser leur fonctionnement et d'élaborer des algorithmes pour leur utilisation dans des domaines variés, tels que la traduction automatique, la reconnaissance de la parole ou encore l'analyse sémantique.

Il est à noter que la TST, dont il a été question précédemment, utilise également les FL comme outil d'analyse et de description du langage. Les FL y sont considérées comme des unités sémantiques qui permettent de représenter les relations sémantiques entre les mots d'un texte, ainsi que les relations syntaxiques qui les unissent.

Dans ce cadre, la formalisation mathématique des FL permet de décrire de façon rigoureuse les relations sémantiques et syntaxiques qu'elles expriment, ainsi que leur utilisation dans la construction de phrases et de textes. La TST a ainsi permis de développer des outils informatiques pour l'analyse et la production de textes, en utilisant les FL comme brique de base de la représentation sémantique et syntaxique du langage.

Le degré de formalisation des FL varie en fonction des approches et des modèles utilisés. Certaines approches de TAL utilisent des listes de FL spécifiques, qui sont souvent définies de manière informelle ou semi-formelle en fonction des caractéristiques linguistiques des langues considérées. D'autres approches utilisent des modèles formels basés sur la théorie linguistique, qui peuvent offrir une description plus précise et systématique des fonctions lexicales. Ces modèles formels peuvent être basés sur des grammaires formelles, des ontologies sémantiques ou des réseaux de concepts.

En outre, la formalisation des FL peut également dépendre du niveau de granularité choisi pour la description linguistique.

Par exemple, certaines approches peuvent définir des FL à un niveau très détaillé, en les distinguant selon les traits morphologiques, syntaxiques et sémantiques pertinents pour la tâche de TAL considérée. D'autres approches peuvent choisir une granularité plus grossière, en définissant des catégories plus générales de fonctions lexicales qui regroupent plusieurs sous-types de fonctions.

En fin de compte, le degré de formalisation FL dépendra des objectifs de la tâche de TAL considérée et des ressources et modèles disponibles. Une formalisation plus précise peut offrir une meilleure performance pour certaines tâches, mais peut nécessiter des ressources linguistiques et des modèles plus sophistiqués.

En revanche, une formalisation moins précise peut être plus facile à mettre en place et à utiliser, mais peut entraîner une baisse de la qualité de la performance pour certaines tâches plus complexes. Dans tous les cas, le degré de formalisation des fonctions lexicales devra être évalué en fonction des besoins spécifiques de la tâche de TAL considérée et des ressources disponibles pour la mise en place du système.

Il convient de noter que la formalisation des FL peut être un sujet de débat entre les linguistes et les ingénieurs en TAL. Les linguistes peuvent avoir des attentes différentes en termes de description linguistique et de précision sémantique, tandis que les ingénieurs peuvent être plus préoccupés par les performances du système et la faisabilité de la mise en place de la technologie. C'est pourquoi il est important d'avoir une collaboration étroite entre les experts en linguistique et les ingénieurs en TAL pour développer des modèles de FL adaptés aux besoins spécifiques de la tâche de TAL considérée.

Pour conclure, la formalisation des FL est un enjeu important pour le TAL. Les FL jouent un rôle crucial dans l'analyse syntaxique et sémantique des phrases et sont utilisées dans de nombreuses applications de TAL telles que la traduction automatique, la reconnaissance de la parole, la

génération automatique de textes, etc. Leur formalisation peut varier en fonction de la tâche de TAL considérée et des ressources disponibles, et peut faire l'objet de débats entre les linguistes et les ingénieurs en TAL. Il est donc important d'avoir une collaboration étroite entre ces experts pour développer des modèles de FL adaptés aux besoins spécifiques de chaque tâche de TAL. Pour aller plus loin, on pourrait consulter les suivants: Mel'čuk (1997), Polguère (1998), Schenk (2006). Pustejovsky (1991). Biber & Conrad & Reppen (1998). Jurafsky & Martin (2019). Manning & Schütze (1999).

5.3. Degré de validité de son contenu informatique

Bien que le Modèle Sens-Texte (désormais MST) soit universel selon les outils descriptifs et les formalismes proposés qui pourraient s'appliquer à toutes les langues de la même façon, soulignons que les FL permettent de décrire la cooccurrence lexicale de n'importe quelle langue avec des moyens identiques, en favorisant ainsi la comparaison entre les langues. Voici quelques valeurs des FL: **Magn** (intensificateur) : **Magn** (passion) = ardente, brulante, et **Oper**₁ (un des verbes supports) pour plusieurs langues de structure bien différente : français, russe, chinois, arabe, allemand, etc.

Le degré de validité du contenu informatique lié aux FL dépend de la qualité des données et des modèles utilisés pour les représenter et les traiter. Les modèles de FL ont été largement utilisés en linguistique théorique et en TAL, et ils ont souvent été validés par des experts linguistes. Cependant, leur utilisation dans des applications pratiques nécessite une évaluation rigoureuse pour s'assurer de leur validité et de leur précision.

Les approches statistiques et l'utilisation de grands corpus de données peuvent aider à augmenter la validité des modèles de FL en permettant une analyse empirique de leur comportement dans des contextes réels. En outre, la comparaison avec d'autres modèles de TAL et l'évaluation de leur performance dans des

tâches spécifiques peuvent également aider à mesurer leur validité.

En outre, le degré de validité peut également dépendre du contexte d'utilisation et des objectifs spécifiques de l'application. Par exemple, un modèle de FL qui fonctionne bien pour l'analyse de textes littéraires peut ne pas être adapté pour l'analyse de textes techniques ou juridiques.

Le degré de validité du contenu informatique lié aux FL peut également dépendre de la méthode d'évaluation utilisée. Il existe plusieurs méthodes d'évaluation, telles que l'évaluation manuelle, l'évaluation semi-automatique et l'évaluation automatique, chacune ayant ses avantages et ses inconvénients.

Il est donc important de prendre en compte ces facteurs lors de l'évaluation du degré de validité du contenu informatique lié aux FL, afin de s'assurer qu'il est adapté aux besoins spécifiques de l'application et qu'il est capable de fournir des résultats précis et fiables.

L'évaluation manuelle peut être très précise, mais elle est également très coûteuse en termes de temps et de ressources humaines. L'évaluation semi-automatique peut être plus rapide et moins coûteuse, mais elle peut être moins précise que l'évaluation manuelle. L'évaluation automatique est la plus rapide et la moins coûteuse, mais elle peut être limitée par la qualité des données d'entraînement et la complexité des modèles.

Grosso modo, pour évaluer correctement le degré de validité du contenu informatique lié aux FL, il est important de sélectionner la méthode d'évaluation la plus appropriée en fonction des objectifs spécifiques de l'application et des ressources disponibles.

6. Liste des fonctions lexicales fréquemment utilisées pour la modélisation formelle des CVS dans un réseau lexical

En fait, en étudiant les *Constructions à Verbe Support*, il y a cinq points essentiels qui résument tout ce que nous avons

toujours voulu sur l'utilisation des lexies ou leur mise en discours (Mel'čuk, 2004, 203) :

- Un *verbe support* est sémantiquement vide ;
- Il n'existe que trois types de *verbes supports* possibles ;
- Il y a deux types majeurs de sens qui s'ajoutent fréquemment à des *verbes supports* : les sens phasiques et les sens causatifs ; ces sens forment avec les *verbes supports* des FL complexes ;
- Il existe d'autres sens (surtout, intensification) qui portent sur L mais s'expriment ensemble avec les *verbes supports* et forment avec eux des configurations de FL ;
- Il existe une famille de verbes sémantiquement pleins qui ont le même comportement syntaxique que les *verbes supports* : les verbes de réalisation.

Pour modéliser formellement ces *verbes supports*, selon les principes des FL, on se trouve devant trois fonctions principales, distinguées par la position syntaxique qu'occupe la base relativement au collocatif comme suit:

[1] **Oper_i** : verbe support prenant x comme premier complément et l'expression de l'actant i de x comme sujet grammatical :

Oper₁ (*suprématie*) = *détenir* [ART ~]

Oper₁ (*remarque*) = *faire* [ART ~]

Oper₁ (*méfait*) = *perpétrer* [ART ~]

Oper₂ (*danger*) = *courir* [ART ~]

Oper₂ (*applaudissements*) = *recueillir* [ART ~]

[2] **Func_i** : verbe support prenant x comme sujet grammatical et l'expression de l'actant i de x comme premier complément ; **Func₀** (c'est-à-dire, $i = 0$) signifie que le verbe support ne prend pas de complément exprimant un actant de x . Il s'agit donc d'un verbe support conversif de **Oper₁** :

Func₀ (*pluie*) = *tomber*

Func₀ (*tonnerre*) = *gronder*

Func₁ (*responsabilité*) = *incomber* [à NX]

Func₂ (*danger*) = *menacer* [NY]

Func₂ (*interdiction*) = *frapper* [NY]

[3] **Labor_{ij}** : verbe support prenant *x* comme second complément, l'expression de l'actant *i* de *x* comme sujet grammatical et l'expression de l'actant *j* de *x* comme premier complément :

Labor₁₂ (*estime*) = *avoir, tenir* [NY en ~]

Labor₁₂ (*silence*) = *passer* [NY sous ~]

7. **Apport de la modélisation formelle des CVS** :

En effet, l'apport de la modélisation formelle des CVS peut être examiné sous plusieurs angles comme suit :

1. **Représentation linguistique** : La modélisation des CVS permet de représenter de manière plus précise la structure syntaxique des CVS dans une langue. Cela contribue à une meilleure compréhension des relations grammaticales entre les éléments d'une phrase.

2. **Traitement automatique des Langues (TAL)** : En modélisant les CVS, on pourrait améliorer les performances des systèmes de TAL, tels que la traduction automatique, la recherche en information, et l'analyse de sentiment. Comprendre les CVS est crucial pour traiter de manière adéquate les expressions idiomatiques et les nuances de sens.

3. **Enseignement des langues** : La modélisation des CVS peut être utile dans l'enseignement des langues, en particulier pour les apprenants étrangers. Comprendre comment les verbes support fonctionnent dans différentes constructions peut aider à maîtriser la syntaxe d'une langue.

4. **Analyse sémantique** : La modélisation des CVS peut contribuer à une analyse sémantique plus fine en permettant de capturer les nuances de sens associées à des constructions spécifiques. Cela est particulièrement important pour les langues dans lesquelles les verbes support jouent un rôle prépondérant dans la création de significations complexes.

5. Développement de ressources linguistiques : La modélisation des CVS peut conduire au développement de ressources linguistiques spécifiques, telles que des lexiques spécialisés ou des bases de données annotées, qui peuvent être utilisées pour la recherche linguistique, l'ingénierie linguistique et d'autres applications liées au langage.

En gros, la modélisation des CVS offre une perspective utile pour comprendre et représenter les particularités syntaxiques et sémantiques associées aux verbes support dans une langue donnée, avec des applications potentielles dans divers domaines liés au langage.

8. Conclusion

Les approches de TAL ont connu des avancées significatives ces dernières années grâce aux progrès technologiques dans les domaines de l'apprentissage automatique et du traitement de grandes quantités de données. Les systèmes de TAL sont de plus en plus utilisés dans divers domaines tels que la traduction automatique, l'analyse de sentiments, la reconnaissance vocale et la réponse aux questions. Cependant, il reste encore des défis importants à relever pour améliorer davantage la précision et la fiabilité de ces systèmes.

Parmi les défis auxquels sont confrontées les approches de TAL, on peut citer la polyvalence et la complexité du langage naturel, la variabilité du langage en fonction de facteurs tels que le contexte et l'intonation, la difficulté de modéliser les connaissances implicites et les nuances de sens, et les limitations des ressources de données disponibles.

Pour surmonter ces défis, les chercheurs travaillent actuellement sur le développement de modèles de TAL plus sophistiqués qui peuvent mieux capturer la complexité et la nuance du langage naturel, ainsi que sur l'exploitation de sources de données plus vastes et plus diverses, telles que les données multimodales (texte, audio, vidéo) et les données provenant de sources sociales et en ligne. Dans l'ensemble, les

approches de TAL continuent de faire l'objet d'un développement et d'une recherche intensifs, et leur utilisation est susceptible de se généraliser encore davantage à l'avenir.

Les FL sont un domaine important de la linguistique qui joue un rôle crucial dans le TAL. Leur typologie et leur utilisation en TAL sont des sujets de recherche qui ont suscité beaucoup d'intérêt ces dernières années. Ces fonctions ouvrent de nombreuses perspectives pour des applications innovantes et utiles, ainsi que pour une meilleure compréhension de la nature humaine et de la communication.

Cependant, malgré les avancées récentes dans le domaine du TAL, l'analyse des FL demeure un domaine complexe et difficile à traiter pour les systèmes de TAL. En effet, la variabilité des FL en langues et les nombreuses ambiguïtés liées à leur utilisation rendent la tâche difficile pour les algorithmes de traitement automatique.

Par conséquent, de nombreux chercheurs continuent de travailler sur le développement de nouvelles techniques et de nouvelles approches pour améliorer la précision et l'efficacité de l'analyse des FL en langues. Parmi ces approches, on peut citer l'utilisation de méthodes d'apprentissage automatique, de réseaux de neurones et d'autres approches basées sur l'intelligence artificielle.

Malgré les défis et les limites actuels, l'analyse des FL est un domaine de recherche important qui offre des perspectives intéressantes pour améliorer les systèmes de TAL, ainsi que pour une meilleure compréhension de la structure et de la fonction des langues.

Enfin, il convient de noter que l'analyse des FL est un domaine en constante évolution, qui est alimenté par les avancées technologiques et les nouvelles perspectives théoriques en linguistique et en sciences cognitives. Il est donc probable que de nouvelles approches et de nouvelles techniques émergeront dans les années à venir, offrant ainsi de

nouvelles possibilités pour l'analyse et l'utilisation des FL dans le TAL.

9. **Bibliographie**

- [1] Abdaoui, A., & Diop, S. (2015). « Traitement automatique des langues pour la traduction arabe-français ». *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation*, 22.
- [2] Béchet, N., & Bouillon, P. (2003). *Traitement automatique du langage naturel, informatique et linguistique*. Paris: Hermès.
- [3] Belaïd, A., & Belaïd, Y. (2005). « Arabic language processing: State of the art ». In *International Conference on Arabic Language Resources and Tools*, 1 – 8.
- [4] Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge University Press.
- [5] Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.
- [6] Boitet, C., & Blanchon, H. (2007). « Étude des désambiguïisations morphosyntaxiques et lexicales en contexte pour la traduction automatique statistique ». *Traitement Automatique des Langues*, 48(2), 53 – 82.
- [7] Charniak, E. (1997). *Statistical language learning*. MIT press.
- [8] Crabbé, B., & Iordachioaia, G. (2015). « Le traitement automatique des langues pour le multilinguisme en ligne : enjeux, méthodes et applications ». *Revue française de linguistique appliquée*, XX(1), 59 – 80.
- [9] Goldberg, Y. (2017). « Neural network methods for natural language processing ». *Synthesis Lectures on Human Language Technologies*, 10(1), 1 – 309.
- [10] Gross, G., & Tasmowski, L. (2009). « Le traitement automatique des langues naturelles et les sciences cognitives ». *Langages*, (173), 111 – 122.
- [11] Habash, N., & Rambow, O. (2005). « Arabic tokenization, part-of-speech tagging and morphological

disambiguation in one fell swoop ». In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 573 – 580.

[12] Jacqueland, F., & Namer, F. (2016). « Approche sémantique et corpus du calcul de redondance lexicale : vers un système de résolution des anaphores nominales en français ». *Revue d'Intelligence Artificielle*, 30(2), 203 – 236.

[13] Jacquet, G. (2014). « Traitement automatique des langues naturelles et Big Data : quelles applications pour les sciences humaines et sociales ? » *Les Cahiers du Numérique*, 10(2), 33 – 49.

[14] Jurafsky, D., & Martin, J. H. (2019). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. United Kingdom: Pearson.

[15] Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing*. Pearson.

[16] Kordoni, V., & Rosner, M. (2011). *Lexical resources and natural language processing*. Londres: Springer.

[17] Langlais, P., & Patry, A. (2017). « Traitement automatique des langues naturelles et traduction automatique ». *Revue d'Intelligence Artificielle*, 31(4), 335 – 359.

[18] Manning, C- D. & Schütze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

[19] Mel'čuk, I. (1993, 1994, 1996, 1997, 2000). *Cours de Morphologie Générale (5 volumes)*. Montréal : Presses de l'Université de Montréal, CNRS éditions.

[20] Mel'čuk, I. (1997). « Lexical functions: A tool for the description of lexical relations in a lexicon ». *Acta Linguistica Hungarica*, 44(3 – 4), 249 – 276.

[21] Mel'čuk, I. (1997). *Vers une linguistique sens-texte : leçon inaugurale faite le vendredi 10 janvier 1997*. Paris : Collège de France.

[22] Mel'čuk, I. (2004). « Verbes supports sans peine ». *Lingvisticae Investigationes*, 27: 2, 203 – 217.

- [23] Mellet, S., Fabre, C., & Gleizes, M.-P. (2005). « Traitement automatique des langues naturelles : vers une nouvelle génération de systèmes de dialogue homme-machine ». *Revue d'Intelligence Artificielle*, 19(4-5), 485 – 506.
- [24] Mesfar, M. A. (2006). « Le traitement automatique des langues dans l'enseignement et la recherche en traduction : entre la spécialisation et l'interdisciplinarité ». *Traductologie*, 11, 55 – 78.
- [25] Polguère, A. (1998). « Constructions à verbe support et fonctions lexicales ». *Langue française*, (120), 18 – 32.
- [26] Polguère, A. (1998). « Les fonctions lexicales en TAL ». In J. François (Ed.), *Traitement Automatique des Langues Naturelles : TALN 98*, 421 – 430. Université d'Avignon et des Pays de Vaucluse.
- [27] Polguère, A. (1998). La Théorie Sens-Texte. *Dialangue*, vol. 8 – 9, 9 – 30.
- [28] Pustejovsky, J. (1991). « The syntax of event structure ». *Cognition*, 41(1 – 3), 47 – 81.
- [29] Sagot, B. (2014). « Le traitement automatique des langues naturelles : quelques enjeux et défis de la reconnaissance automatique de la parole ». *Informatique des Organisations et Systèmes d'Information et de Décision*, 5(4), 37 – 65.
- [30] Schenk, N. (2006). De l'importance de la formalisation des lexies pour la fouille de textes. *Document Numérique*, 9(2), 83 – 98.
- [31] Silberztein, M., & Masson, C. (2012). « Une ressource lexicale pour la gestion des anaphores et la rédaction de textes en français ». *Traitement Automatique des Langues*, 53(2), 119 – 151.
- [32] Tarpin-Bernard, F. (2007). « Le traitement automatique des langues naturelles au service de la traduction assistée par ordinateur ». *Traduction, Terminologie, Rédaction*, 20(1), 79 – 94.