



SIMULATION PROBABILISTE D'ACQUISITION DES LANGUES SECONDES (ALS)

Thèse présentée comme exigence partielle
du doctorat en informatique cognitive

Par Badis Merdaoui

Février 2018

<http://r-libre.teluq.ca/1389>

TELUQ

SIMULATION PROBABILISTE D'ACQUISITION DES LANGUES SECONDES
(ALS)

THÈSE
PRÉSENTÉE
COMME EXIGENCE PARTIELLE
DU DOCTORAT EN INFORMATIQUE COGNITIVE - DIC

PAR
BADIS MERDAOUI

FÉVRIER 2018

*À mes parents, mes grands-parents, mes beaux-parents, mes frères et sœurs, ma femme
et mes enfants.
À toute ma grande famille.*

REMERCIEMENTS

En arrivant à la fin de ce travail, je voudrais, du fond du cœur, témoigner mes sentiments de gratitude à l'égard de toutes les personnes qui ont aidé, de près ou de loin, à la réalisation de cette thèse.

Mes remerciements vont surtout à mes directeurs de thèse, Richard Hotte (Directeur du Centre de recherche en informatique cognitive et environnements fonctionnels et professeur titulaire en technologie de l'information dans le Département Science et Technologie de la TELUQ) et Daniel Lemire (Professeur titulaire en informatique dans le Département Science et Technologie de la TELUQ), qui m'ont encadré et soutenu pendant toutes ces années, de la germination de mes idées jusqu'à la réalisation et la rédaction.

J'adresse mes remerciements aux professeurs qui vont participer à l'évaluation de ce travail.

À ma famille et mes amis, merci de m'avoir soutenu durant cette thèse et d'avoir été si présents.

Au nom d'Allah, le Tout Miséricordieux, le Très Miséricordieux.

Lis, avec le nom de ton Seigneur qui a créé. (1)

Qui a créé l'homme d'Alak. (d'embryon) (2)

Lis, et ton Seigneur est infiniment Noble. (3)

Saint Coran: 96:1,2 et 3

Le savoir parfait appartient à Allâh, et notre dernière invocation est qu'Allâh, Seigneur des Mondes soit Loué, et que prière et salut soient sur notre Prophète, ainsi que sur sa famille, ses compagnons et ses frères jusqu'au jour de la résurrection.

TABLE DES MATIERES

LISTE DES FIGURES.....	X
LISTE DES TABLEAUX.....	XI
RESUME	13
CHAPITRE I	
DÉFIS D'ACQUISITION DES LANGUES SECONDES.....	15
1.1 Problématique.....	15
1.2 Acquisition des verbes d'une LS.....	17
1.3 Acquisition des structures actanciennes d'une LS.....	22
1.4 Acquisition des sens des mots d'une LS.....	24
1.5 Acquisition des rôles sémantiques des actants d'une LS.....	27
1.7 Questions de recherche et portée de l'étude.....	31
CHAPITRE II	
UN MODELE PROBABILISTE D'ACQUISITION DES LANGUES SECONDES	34
2.1 Introduction.....	34
2.2 Constructions probabilistes et ALS.....	38
2.2.1 Représentation sémantique et LS.....	39
2.2.2 Paires scène/énoncé et intrants du modèle.....	41
2.2.3 Catégorisation des cadres.....	44
2.3 Calcul de similarité et élaboration de constructions.....	46
2.4 Comportement du modèle dans le cas d'ALS.....	54
2.5 Conclusion.....	56
CHAPITRE III	
SIMULATIONS PROBABILISTES D'ACQUISITION DES STRUCTURES ACTANCIENNES DE VERBES DE LS.....	59
3.1 Introduction.....	59
3.2 Théories d'acquisition des structures actanciennes d'une LS.....	63
3.3 Modèles computationnels d'acquisition des structures actanciennes d'une LS..	66

3.4	Adaptation du modèle bayésien à l'acquisition des structures actanciennes de LS	71
3.5	Extraction de données.....	74
3.5.1	Extraction de données pour les intrants en anglais.....	74
3.5.1.1	Extraction des intrants en anglais.....	76
3.5.1.2	Outil automatique pour l'extraction des primitives sémantiques en anglais.....	77
3.5.2	Extraction de données pour les intrants en français.....	78
3.5.2.1	Extraction des intrants en français.....	79
3.5.2.2	Outil automatique pour l'extraction des primitives sémantiques en français.....	80
3.6	Simulations probabilistes.....	81
3.6.1	Associations des éléments syntaxiques et sémantiques dans les structures actanciennes de la LS.....	90
3.6.2	Simulation de la production dans le cas de la LS.....	94
3.6.3	Surgénéralisation des règles dans le cas de l'ALS.....	98
3.6.4	Généralisations appropriées.....	100
3.7	Discussion des résultats.....	101
3.8	Conclusion.....	104
CHAPITRE IV		
ACQUISITION DES ROLES SEMANTIQUES D'UNE LS.....		107
4.1	Introduction.....	107
4.2	Théories d'acquisition des rôles sémantiques.....	110
4.3	Modèles computationnels et rôles sémantiques multilingues.....	113
4.4	Adaptation du modèle bayésien à l'acquisition des rôles sémantiques de LS.....	116
4.5	Extraction de données.....	120
4.5.1	Extraction de données pour les intrants en anglais.....	120
4.5.2	Extraction de données pour les intrants en français.....	121
4.6	Simulations probabilistes.....	121
4.7	Discussion des résultats.....	129
4.8	Conclusion.....	130
CHAPITRE V		

PRÉFÉRENCES DE SÉLECTION DES VERBES ET PROPRIÉTÉS SÉMANTIQUES IMPOSÉES PAR CES DERNIERS DANS L'ALS	132
5.1 Introduction.....	132
5.2 Théories et approches d'acquisition des préférences de sélection des verbes.	135
5.3 Modèles computationnels d'acquisition des préférences de sélection des verbes d'une LS.....	137
5.4 Adaptation du modèle bayésien à l'acquisition des préférences de sélection des verbes de LS	139
5.5 Extraction de données.....	142
5.5.1 Extraction de données pour les intrants en anglais.....	142
5.5.2 Extraction de données pour les intrants en français.....	142
5.6 Simulations probabilistes d'acquisition des préférences de sélection d'une LS	142
5.7 Discussion des résultats	152
5.8 Conclusion.....	154
CHAPITRE VI	
CONTRIBUTIONS ET PERSPECTIVES FUTURES.....	156
6.1 Contributions	156
6.1.1 Modèle computationnel	158
6.1.2 Processus d'ALS.....	159
6.1.3 Dimensions cognitives et informatiques.....	161
6.2 Intrants du modèle et évolution des constructions.....	163
6.2.1 Corpus linguistiques et intrants du modèle.....	163
6.2.2 Évolution des constructions.....	165
6.2.3 Système conceptuel mature et ensemble de constructions de la L1	166
6.2.4 Généralisation et surgénéralisation.....	167
6.2.5 Profils sémantiques des actants et des verbes.....	168
6.3 Futures perspectives.....	170
6.3.1 Acquisition de la morphologie grammaticale d'une LS	170
6.3.2 Le rôle de la motivation dans l'ALS.....	171
6.3.3 Influence des verbes très fréquents dans l'ALS	173
6.3.4 Désambiguïsation de sens dans le cas d'ALS.....	174
BIBLIOGRAPHIE	176

APPENDICE A	
CORPUS ANGLAIS.....	193
APPENDICE B	
CORPUS FRANCAIS	223

LISTE DES FIGURES

Figure	Page
Figure 2.1: Modèle computationnel d'ALS.	47
Figure 3.1: Simulation d'ALS chez les jeunes enfants bilingues.....	83
Figure 3.2: Acquisition du français LS dans un âge tardif et avec un faible quotient d'intrants.	84
Figure 3.3: Acquisition l'anglais LS dans un âge tardif et avec un faible quotient d'intrants.	84
Figure 3.4: Acquisition du français LS avec différents quotients LS/L1.....	85
Figure 3.5: Acquisition de l'anglais LS avec différents quotients LS/L1.....	86
Figure 3.6: Acquisition du français LS aux différentes étapes d'apprentissage.	87
Figure 3.7: Acquisition de l'anglais LS aux différentes étapes d'apprentissage.	88
Figure 3.8: Arg Max de la meilleure construction pour des différentes paires scène/énoncé des verbes de « mouvement provoqué ».....	93
Figure 3.9: Prédiction du modèle syntaxique pour le verbe 'Descendre'.	95
Figure 3.10: Prédiction des modèles syntaxiques L1 et LS.	98
Figure 4.1: Acquisition des rôles sémantiques des actants du verbe 'admire'.....	123
Figure 4.2: Acquisition des rôles sémantiques des actants du verbe 'frighten'.	123
Figure 4.3: Acquisition des rôles sémantiques des actants du verbe 'miss'.	124
Figure 5.1: Prédiction des profils sémantiques pour 182 intrants de l'anglais LS à des étapes différentes d'apprentissage.	152

LISTE DES TABLEAUX

Tableau	Page
Tableau 2.1: Structure extraite à partir de la paire scène/énoncé.....	43
Tableau 3.1: Les différentes catégories des verbes.	60
Tableau 3.2: Des paires scène/énoncé des verbes de « mouvement provoqué ».	93
Tableau 5.1: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'prendre' et 'donner'.	146
Tableau 5.2: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'manger' et 'piloter'.	147
Tableau 5.3: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'achever' et 'demander'.	148
Tableau 5.4: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'entendre' et 'jouer'.	149

LISTE DES ABRÉVIATIONS, DES SIGLES ET DES ACRONYMES

ALS	Acquisition de Langues Secondes
CL	Compétence Linguistique
DDC	Degré De Certitude
L1	Langues Maternelle
LS	Langue Seconde
SL1	Ensembles de Constructions de la Langues Maternelle
SLS	Ensembles de Constructions de la Langue Seconde
SVO	Sujet-Verbe-Objet

RESUME

Cette thèse s'intéresse aux défis de simulation des mécanismes cognitifs d'acquisition des langues secondes (ALS) par des modèles computationnels. Dans notre société mondialisée, l'ALS est incontournable. Chez un individu l'importance des langues secondes (LS) est proportionnelle à son usage dans sa vie personnelle et professionnelle et dans le développement des sociétés dans lesquelles il vit. Un exemple tangible est l'intégration des nouveaux immigrants dans leurs sociétés d'accueil et milieux professionnels. De même, des milliers de personnes, dans le monde, passent des tests tels que SL/EFL, TOEFL et TOEIC pour leur admission dans des universités, des écoles supérieures ou, encore, dans des entreprises internationales.

L'offre de formation à l'apprentissage des langues est un marché fort varié et lucratif. On retrouve partout des écoles de langues et les approches pédagogiques pour faciliter les apprentissages vont des plus traditionnelles (Berliz, YMCA, UQAM) aux plus innovantes (Gymglish, Rosetta Stone, Assimil, Babel) utilisant soit la présence, la ligne ou, encore, les deux combinées comme modes de formation. Mais ces formations sont-elles vraiment efficaces ? Les systèmes qui les supportent permettent-ils de maximiser l'apprentissage des langues en assurant un apprentissage rapide, et, surtout, durable dans le temps ?

L'analyse, la compréhension et la modélisation des processus cognitifs d'ALS restent un problème ouvert, et de ce fait, la simulation des différents mécanismes de cette acquisition par des modèles computationnels nous semble un défi à relever.

En vue d'optimiser ce type d'apprentissage, nous avons conçu un modèle computationnel qui s'inspire des processus cognitifs humains.

Plusieurs travaux (Fazly, Alishahi et Stevenson, 2010; Erk, 2007; Andrews et Vigliocco, 2010) tentent de cerner et d'analyser l'apprentissage des langues en construisant des modèles qui capturent les structures complexes servant de base à l'organisation des langues comme les sens des mots simples et les structures actanciennes. Toutefois, très peu de modèles ont été consacrés à la modélisation des LS en vue d'optimiser leur apprentissage (Rappoport et Sheinman, 2005 ; Matusevych et al., 2016).

D'une part, les modèles probabilistes et Bayésiens sont adaptés à la modélisation de la langue naturelle et de son apprentissage parce que la nature incertaine de la langue naturelle est fondée sur des informations inexacts, imprécises, incomplètes et non mesurables (Weber et Jouffe, 2003; Draper 2004). D'une autre part, notre modèle est un modèle probabiliste qui simulera les processus cognitifs humains pris en compte dans l'ALS; que sont, entre autres l'acquisition des sens des mots, l'acquisition des structures actanciennes des verbes, l'acquisition des rôles sémantiques des actants et l'acquisition des préférences de sélection des verbes, en s'appuyant seulement sur l'activité langagière (*Usage-Based Theory*) (Tomasello, 2003).

Le modèle pourra être utilisé comme outil de support à l'apprentissage pour l'acquisition de l'anglais comme LS chez des apprenants dont la langue maternelle (L1) est le français et vice-versa.

Mots-clés : Acquisition des langues secondes, sens des mots, structures actanciennes des verbes, modèles probabilistes, modèles bayésiens.

Keywords: Acquisition of second languages, meaning of words, Verb argument structure, Probabilistic models, Bayesian networks.

CHAPITRE I

DÉFIS D'ACQUISITION DES LANGUES SECONDES

1.1 Problématique

L'informatique cognitive est un domaine de recherche fondamentale et appliquée visant à la fois la conception de technologies cognitives de nature informatique et la conception de technologies informatiques au service des sciences cognitives. Les questionnements qui préoccupent les chercheurs d'aujourd'hui en cette dernière relativement à la problématique de l'apprentissage des langues sont les trois suivants (Tomasello, 2003; Randall, 2007; Alishahi et Stevenson, 2010; Rappoport et Sheinman, 2005 ; Matusévych et al., 2014) :

- la compréhension et l'analyse des processus cognitifs pris en compte dans l'apprentissage de la langue naturelle;
- le développement des modèles computationnels qui s'appuient sur ces processus pour simuler l'apprentissage de la langue naturelle par une machine;
- l'élaboration d'outils efficaces d'aide à l'apprentissage selon les théories cognitives de l'apprentissage de la langue naturelle.

Tels sont les problèmes ouverts dont les chercheurs en informatique cognitive s'inspirent pour formuler leurs questions et concevoir leurs projets de recherche.

Du côté cognitif, l'ALS se distingue de l'acquisition de la L1 par certains facteurs dont, entre autres, la motivation, l'âge, la durée, l'intensité, le système conceptuel, l'identification avec la langue, la concentration, l'autonomie langagière, la L1, le contexte socioculturel, etc. (Randall, 2007 ; Matusévych et al., 2016)

La prise en considération de tous ces facteurs et variantes rend complexe la modélisation informatique des processus d'acquisition de la LS car elle nécessite une étude et une analyse précise.

De ce fait, ce projet de recherche trouve sa motivation principale dans l'importance des LS et dans les besoins des chercheurs en sciences cognitives et en informatique à vouloir analyser, comprendre et simuler les processus cognitifs d'ALS (Randall, 2007).

Les processus cognitifs d'acquisition, de mémorisation et de production des langues naturelles sont au coeur de nombreux projets de recherches universitaires. Pouvoir intégrer ces processus dans la machine en élaborant des modèles computationnels capables d'apprendre, de produire et de générer des structures complexes de la langue naturelle est un vrai défi à relever (Fazly, Alishahi et Stevenson, 2010; Erk, 2007).

En outre, élaborer des systèmes efficaces d'ALS et des outils d'aide à l'apprentissage qui s'appuient sur des théories cognitives solides pour des apprenants de LS est un domaine peu développé par rapport à celui de l'apprentissage de la L1 (Rappoport et Sheinman, 2005; Matusévych et al., 2016).

Plusieurs modèles simulent l'apprentissage du L1 dans différents aspects tels que : l'acquisition du lexique bilingue et des catégories lexicales (Li et Farkas, 2002), l'acquisition des structures actanciennes (Gildea, 2002; Alishahi et Stevenson, 2008), l'acquisition des sens des mots (Griffiths, 2007; Andrews et Vigliocco, 2010; Yu, Ballard, et Aslin, 2005), l'acquisition des préférences de sélection et des rôles sémantiques des actants (Séaghdha, 2010; Erk, 2007; Fazly, Alishahi et Stevenson, 2010). Cependant, peu de chercheurs se préoccupent de la modélisation de l'ALS en étudiant les différents aspects qui distinguent cette dernière (Rappoport et Sheinman, 2005 ; Matusévych et al., 2013).

Motivés par l'importance des LS et de leur acquisition, par le rôle qui peut être joué par des modèles informatiques fondés sur des théories cognitives dans l'avancement des

systèmes d'apprentissage des LS, et, finalement, par les besoins à vouloir analyser, comprendre et simuler les processus cognitifs sollicités dans l'apprentissage des LS, nous avons élaboré un modèle cognitif computationnel pour l'ALS. Nous avons simulé l'acquisition des sens des mots et des structures actanciennes des verbes avec un modèle probabiliste. Ce dernier a rendu les processus cognitifs tels que l'apprentissage des rôles sémantiques des actants et les structures actanciennes des verbes plus explicites.

1.2 Acquisition des verbes d'une LS

Un des défis des apprenants d'une LS est de produire une phrase grammaticalement et sémantiquement correcte. Cela peut être influencé d'une façon positive ou négative par la L1. Dans le cas du «transfert positif» (Benson, 2002; Khaled et Hossein, 2013) la L1 aide à la maîtrise du LS. Dans certains cas, l'ALS est très rapide à ses débuts vu que l'apprenant connaît déjà des congénères (vrais amis) et des modèles syntaxiques qui sont partagés entre les deux langues. À titre d'exemple, lorsqu'il s'agit de la même règle du pluriel dans les deux langues L1 et LS. A l'encontre, l'application des modèles syntaxiques inappropriés de la L1 sur la LS, ou l'utilisation des faux amis peuvent provoquer des erreurs de surgénéralisation. De même, lorsque le verbe ou l'une de ses propriétés sémantiques est utilisé dans un contexte inadéquat, l'erreur de surgénéralisation se produit (Benson, 2002).

Même s'il s'agit de deux langues qui ont la même typologie syntaxique telles que l'anglais et le français, certains nombres de différences peuvent provoquer des interférences dans la production de la LS. Dans cette thèse nous avons étudié l'interaction entre l'anglais et le français, où chaque langue peut être utilisée comme LS ou L1 dans différentes expérimentations. Bien que ces deux langues appartiennent à la même typologie syntaxique qui est Sujet-Verbe-Objet (SVO) et possèdent la même structure profonde (structure-D), nous avons plusieurs cas où le déplacement du verbe reste différent entre ces deux langues (Benson, 2002; Pollock, 1989; Chomsky, 1968; Chomsky, 1991) :

- en français, on place, contrairement à l'anglais, les verbes conjugués au présent, avant les adverbes et les négations, comme dans (Pollock, 1989) :
 - o Adam voit souvent Ève / Adam ne voit pas Ève,
 - o Adam often sees Eve / Adam does not see Eve,
- les quantificateurs flottants (Tous/Chacun), en français, à l'opposé de l'anglais, sont placés après le verbe conjugué au présent (Ayoun, 2005) :
 - o Les élèves aiment tous Ève / Les élèves reçoivent chacun un prix,
 - o The students all love Eve / The students each receive a prize,
- En français, la quantification à distance peut se placer, facultativement, avant ou après le participe passé, tandis qu'en anglais elle est placée constamment après ce dernier (Dekydtpotter, L. Rex et Rachel, 2000) :
 - o Adam a fait beaucoup de devoirs / Adam a beaucoup fait de devoirs,
 - o Adam has done a lot of homework,
- En français, l'adverbe de lieu 'y' est indispensable et doit se placer avant le verbe, alors qu'en anglais l'adverbe 'There' est optionnel et se place après le verbe (Chomsky, 1991) :
 - o J'y vais / Voulez-vous y aller ? / Ils y étaient,
 - o I'm going (there) / Do you want to go (there)? / They were there.

Indépendamment des différences associées au déplacement du verbe, les apprenants du LS doivent tenir compte de certaines différences syntaxiques telle que la position de l'adjectif dans la phrase. En anglais, les adjectifs précèdent les noms. En français, les adjectifs peuvent être placés avant ou après le nom, en fonction de leur type et leur signification. Ils suivent généralement le nom qu'ils modifient lorsqu'ils le classent dans une certaine catégorie (adjectifs de la forme, la couleur, la personnalité et de l'humeur, ...). En outre, les participes présents et participes passés utilisés comme adjectifs sont toujours placés après le nom :

- une voiture rouge / une petite maison / une table carrée / une histoire intéressante / un débat passionné ,
- a red car / a little house / a square table / an interesting story / a lively debate.

À l'opposé, certains adjectifs qui décrivent la taille, la beauté morale ou physique, l'âge ou l'ordre, sont placés, comme en anglais, avant le nom :

- une jolie fille / un jeune homme / une nouvelle maison,
- a pretty girl / a young man / a new house.

D'autres adjectifs changent de sens selon qu'ils précèdent ou suivent le nom qu'ils modifient (Pollock, 1989; Chomsky, 1991; Ayoun, 2005). Cependant, nous n'avons pas considéré ce dernier cas dans nos expérimentations. L'objectif principal est d'élucider l'adaptabilité des apprenants d'une LS dans le cas de différence entre les typologies syntaxiques, en s'appuyant, seulement, sur les intrants, et sans instruction explicite. Nous avons adopté, dans nos expérimentations, des intrants bruyants et incomplets, pour se rapprocher de la réalité des apprenants.

Les apprenants d'une LS acquièrent les règles qui régissent l'utilisation des formes linguistiques, mais aussi les exceptions et les dissemblances avec leur L1. Les mécanismes avec lesquels les apprenants réussissent à découvrir et utiliser les modèles généraux, et distinguer les exceptions font l'objet de nombreux débats et études (Matusévych et al., 2015).

L'apprentissage des verbes d'une LS est très important parce qu'il constitue la partie du discours principale de chaque langue. En outre, l'acquisition des typologies, des modèles syntaxiques et des exceptions constitue un champ d'expérimentation très favorable dans le domaine d'ALS. Pour que la phrase soit grammaticalement correcte, le verbe doit être cohérent syntaxiquement et sémantiquement avec ses arguments. L'application des structures actanciennes (ordre, types, rôles et propriétés sémantiques des arguments) de la L1 sur une LS sans prendre en considération les différences, conduit à des surgénéralisations qui se corrigent en recevant plus d'intrants de la LS. Par exemple, la phrase « I'm going » est une phrase complète en anglais, tandis qu'en français, la phrase « Je vais » a besoin d'un autre argument pour qu'elle soit complète « Je_{<Agent>} vais à l'institut_{<Destination>} ».

La considération des spécificités d'une LS par rapport à la L1 dans le cas des structures actanciennes est indispensable. Plus spécifiquement, l'admission de l'ordre des arguments

selon leurs fonctions grammaticales (Adverbe, Négation, Quantificateur flottant, Adjectif, Quantification à distance, Adverbe de lieu ...) ou leurs rôles sémantiques (Agent, Objet, Destination, Bénéficiaire, Source, But, Temps, ...).

Améliorer l'acquisition des verbes de la LS en s'appuyant seulement sur l'utilisation de la langue elle-même « *Usage-Based* » a déjà prouvé son efficacité comme modèle computationnel d'apprentissage (Tomasello, 2003; Behrens, 2009; Eskildsen, 2008; Matusévych et al., 2016)

L'efficacité de ce modèle est la même que celle du mécanisme d'apprentissage utilisé pour l'acquisition de la langue maternelle. Elle est fondée surtout sur le concept scène/énoncé où l'apprenant voit une scène et entend un énoncé qui déclenche son mécanisme de 'compréhension' et par la suite celui de 'production'. Ce dernier passe par une courbe d'apprentissage 'U' (Saville-Troike, 2006; Carlucci, Case, 2013) qui commence avec une étape de performance élevée, une étape de surgénéralisation des règles (les compétences descendent vers une position inférieure sur l'axe des ordonnées), puis une troisième étape où la compétence monte encore une fois à une position plus élevée, grâce à trois facteurs :

- fréquence d'utilisation,
- nombre d'intrants,
- concept de la "preuve positive", où l'apprenant commence à recevoir plus d'entrées qui corrigent sa surgénéralisation.

La fréquence d'utilisation et le nombre d'intrants permettent aux apprenants de développer leur LS en utilisant l'interlangue (Corder, 1975 ; Richards, 1972 ; Selinker, Swain, et Dumas, 1975) ou la langue intermédiaire qui est, en linguistique, une phase de transition dans l'apprentissage ou la production d'une langue cible. Celle-ci peut se produire dans le cas de l'ALS.

Le rôle des processus d'ALS est celui de l'amélioration progressive et le raffinement de l'interlangue. L'objectif d'un apprenant est de minimiser l'écart entre les formes de l'interlangue et celles de la LS. Les apprenants adultes d'une LS ont déjà un système de conceptualisation mature et sont capables d'établir un raisonnement symbolique

explicite (Rappoport et Sheinman, 2005). En outre, le temps disponible pour l'acquisition de la L1 est largement supérieur à celui de l'ALS.

Les modèles élaborés dans le cadre de l'apprentissage de la langue naturelle sont catégorisés selon les éléments langagiers à apprendre. À partir de ces éléments, nous avons des modèles d'apprentissage des structures actanciennes des verbes (Matusevych et al., 2016, 2013; Gildea, 2002; Alishahi et Stevenson, 2008), des sens des mots simples (Griffiths, 2007; Fleischman et Roy, 2005), des liens lexicaux bilingues (Li et Farkas, 2002) et des phrases en LS (Rappoport et Sheinman, 2005).

Deux théories principales se concurrencent sur la nature de l'apprentissage de la langue naturelle, y compris les verbes de la LS. Les théories nativistes de l'acquisition de la langue affirment l'existence d'un système inné, chez l'être humain, responsable de l'apprentissage de la langue naturelle. Ce dernier est composé de symboles, de règles et de relations qui sont responsables des processus d'apprentissage de la langue naturelle et d'acquisition des LS. Ce système est fondé sur la théorie de la Grammaire Universelle et considère la langue comme un nombre infini de phrases générées par un ensemble limité de règles et de symboles. Pour cette théorie, les principes généraux et les éléments communs de la langue sont, antérieurement, encodés dans le cerveau de l'être humain sous forme d'une Grammaire Universelle. En conséquence, la relation entre les structures syntaxiques et la sémantique des différents composants de la langue est assurée par ce dispositif inné et les apprenants n'ont pas besoin de les apprendre (Chomsky 1965, 1968; Grimshaw, 1990).

Pour leur part, les théories opposantes, appelées théories de construction affirment qu'on apprend et acquiert à force de l'activité langagière elle-même « Usage-Based ». Un exemple concret est l'apprentissage des structures actanciennes des verbes. Cet apprentissage commence par l'acquisition des sens des verbes (mots simples) et ensuite l'apprentissage ou la construction des structures syntaxiques grâce aux processus de catégorisation et de généralisation appliqués sur des exemples observés (Fillmore, Kay et O'Connor, 1988; Goldberg, 1995).

La théorie dite « Usage-Based » qui est fondée sur l'activité langagière et l'acquisition des connaissances linguistiques à force de la pratique de la langue a gagné beaucoup d'intérêt, principalement, dans la modélisation de l'apprentissage de la langue naturelle. Les approches guidées par l'usage, ont été expérimenté dans les contextes de L1 (Perek, 2015 ; Alishahi et Stevenson, 2010) et LS (Matusevych et al., 2016), plus spécifiquement dans le cas d'acquisition des verbes et de leurs structures actanciennes.

Ces études et expériences ont montré que des surgénéralisations produites par des apprenants à partir des intrants auxquels ils ont été exposés, comme l'utilisation d'un verbe intransitif dans une structure transitive. Ces surgénéralisations ont été corrigées en recevant plus d'intrants. En outre, les modèles computationnels, utilisés, parallèlement, dans ces expérimentations ont montré exactement le même comportement que celui des apprenants en suivant la même courbe d'apprentissage en U. De ce fait, le degré d'ancrage d'une structure linguistique est lié à la fréquence d'utilisation, le niveau d'accessibilité, l'aisance de récupération, et la prééminence cognitive globale (Blumenthal-Dramé, 2012).

1.3 Acquisition des structures actanciennes d'une LS

L'apprentissage des structures actanciennes des verbes est un problème non trivial qui a été modélisé au cours de plusieurs travaux. En effet, dans un premier temps, on retrouve un modèle computationnel pour l'acquisition des structures actanciennes chez les enfants (Alishahi et Stevenson, 2008). Ce modèle simule l'acquisition des structures actanciennes chez les enfants en déterminant les mécanismes appropriés d'apprentissage qui peuvent saisir des règles générales à partir d'exemples et d'usages spécifiques.

Dans un deuxième temps, on retrouve un modèle statistique pour l'apprentissage non supervisé des structures actanciennes des verbes (Pardo et Nunes, 2005) qui est un modèle statistique génératif pour induire des structures actanciennes des verbes, automatiquement, pour un ensemble représentatif des verbes.

Dans un troisième temps, nous retrouvons un modèle qui établit un lien entre l'ordre des mots dans une phrase et leurs sens (Chang, 2004). Enfin, dans un dernier temps, on retrouve un autre modèle d'acquisition de la syntaxe dans le cadre de la construction de la grammaire avec une validation linguistique en anglais et en japonais (Dominey et Inui, 2004). Cependant, ce modèle ne peut pas acquérir des connaissances liées aux structures actanciennes et les généraliser sur d'autres verbes.

En linguistique, une structure actancielle est une phrase qui a une relation syntaxique avec le verbe dans une clause (Geeraerts et Cuyckens, 2007). Tant en anglais qu'en français, les deux actants syntaxiques les plus importants dans une structure actancielle sont le sujet et l'objet direct. Ils sont appelés actants de base. Le sujet est indispensable dans toutes les clauses bien formées tandis que l'objet direct est utilisé dans le cas des verbes transitifs. Les verbes intransitifs n'acceptent pas d'autres actants que le sujet. À l'opposé, les verbes transitifs acceptent un actant d'objet en option. D'autres verbes acceptent un troisième actant qui est l'objet indirect (Geeraerts et Cuyckens, 2007). Le nombre d'actants qu'un verbe peut recevoir ou doit recevoir pour être saturé est appelé valence. Un verbe est saturé s'il produit une phrase syntaxiquement correcte (Perek, 2015).

Certains verbes ont la valence 1, autrement dit, ils n'ont besoin que d'un sujet. Les verbes avec la valence 2 ont besoin d'un sujet et d'un objet direct. En outre, les verbes avec la valence 3 ont besoin d'un sujet, d'un objet direct et d'un objet indirect.

Indépendamment des actants syntaxiques, les verbes ont, aussi, des actants sémantiques. Ces derniers peuvent être inférés de la signification ou du sens de la proposition. En outre, trois actants sémantiques principaux existent en anglais et en français. Ces actants sont l'agent, le thème et le but ou le bénéficiaire. D'autres actants secondaires peuvent être ajoutés, tels que : la destination, la source, la relance, l'état, l'expérimentateur, l'instrument, et l'agent secondaire (Jackendoff, 1990).

Ces actants sémantiques indiquent les rôles sémantiques ou les relations entre les différents acteurs qui participent à un événement donné. La nature de ces rôles sémantiques et la manière avec laquelle ils sont appris restent des questions ouvertes, notamment en raison de l'ambiguïté qui peut exister entre ces derniers et leurs formes syntaxiques.

Certaines théories affirment que ces rôles sont établis à partir d'un ensemble prédéfini de symboles, de règles et de relations sémantiques acquis d'une façon innée et permettant aux apprenants d'établir des relations entre les rôles sémantiques et les structures syntaxiques (Chomsky, 1981; Pinker, 1989).

D'autres théories, fondées sur *Usage-Based*, énoncent que les apprenants acquièrent ces relations par une utilisation soutenue de la langue (Fisher, 1996; Shayan et Gershkoff, 07).

1.4 Acquisition des sens des mots d'une LS

La notion de la 'représentation' est au centre de la problématique cognitive. C'est grâce aux représentations qu'un agent cognitif est en mesure d'utiliser des règles computationnelles qui lui permettent d'appréhender le monde extérieur (Bernard, 2005). Cette représentation montre le lien principal entre cet agent cognitif – l'être humain – et son environnement. En effet, cette dernière permet à un agent cognitif d'établir un lien entre un référent (objet externe de son environnement), un signifiant (signe ou symbole de cet objet) et un signifié (concept ou représentation mentale). En outre, il peut catégoriser, classifier ou généraliser une nouvelle connaissance (Meunier, 1989 ; Meunier, 1999). Avec le concept scène/énoncé, un enfant peut établir une représentation interne ou avoir une compréhension d'un sens spécifique d'un mot simple ou d'une scène complète.

L'apprentissage des sens des mots simples a été simulé par plusieurs modèles computationnels, tels que : Modèle d'espace vectoriel structuré pour la représentation des sens des mots dans un contexte (Erk et Pado, 2008); modèles vectoriels de la

composition sémantique (Mitchell et Lapata, 2008); déduction d'un modèle probabiliste de la mémoire sémantique à partir des règles d'association des mots (Andrews et Vinson, 2008); étude computationnelle des techniques d'apprentissage des sens des mots (Siskind, 1996); contexte intentionnel dans l'apprentissage de la langue (Fleischman et Roy, 2005); le rôle de l'intention incorporée dans l'acquisition lexicale (Yu, Ballard, et Aslin, 2005).

Ces modèles ont été fondés sur différentes théories d'apprentissage des sens des mots, comme, d'une part, celle de l'observation intersituationnelle (*Cross-Situational Learning*) (Pinker, 1989).

Cette théorie affirme que les sens des mots peuvent être acquis à partir des observations des situations différentes. Ces observations permettent de détecter le point commun (sens) entre toutes les situations (scènes) où le mot a été utilisé. À titre d'exemple, si un apprenant entend «Il y a le jouet», «Tiens le jouet», « Tournes le jouet », « Je veux le jouet», « Lance le jouet », « Le jouet d'Adam », «Où est votre jouet? », Etc., le mot 'jouet' vient à exister comme un énoncé potentiel pour une utilisation future lorsque l'apprenant doit indiquer certaine classe d'objets comme une sous-fonction d'un énoncé (Tomasello, 2003; Behrens, 2009). Dans ce cas, l'objectif de l'apprenant est la compréhension de l'énoncé au lieu de la compréhension d'un seul mot en particulier. Mais cela n'empêche pas la compréhension du mot simple dans son contexte général.

Dans le cas d'acquisition de la L1, l'unité linguistique de base avec laquelle les enfants commencent n'est pas le mot, mais l'énoncé (Tomasello, 2003). Un énoncé exprime un motif de communication (impératif, déclaratif, informatif et autres), par le biais d'une forme d'expression émotionnelle (visage, voix, etc.). Par un énoncé «Bravo! » qui vient avec un applaudissement, un enfant comprend tout un acte communicatif, même si la forme est tout simplement un seul mot exprimé avec une certaine expression émotionnelle. Dans cette situation, ce que l'apprenant tente de faire le plus urgent est de comprendre l'intention « globale » de la communication, derrière l'énoncé; en même temps, il essaye de déterminer les fonctions communicatives particulières qui forment l'énoncé. Cela est dû au fait qu'on ne peut pas déterminer une sous-fonction sans la

perception de la fonction globale. Cela donne une importance considérable au concept scène/énoncé dans l'apprentissage des verbes d'une LS.

Toutefois, la théorie de l'observation intersituationnelle n'est pas valide dans tous les cas, car pour un événement donné, et selon la qualité de l'intrant (incomplet, bruyant, ...) l'apprenant peut hésiter entre deux mots qui décrivent la même situation (Gleitman, 1990).

D'autre part, une deuxième théorie de la structure conceptuelle (Jackendoff, 1990) énonce que la construction d'une structure conceptuelle à partir d'une nouvelle scène ou un nouvel événement est la manière avec laquelle un agent cognitif élabore des représentations internes de son environnement, et par conséquent comprend le sens global de la scène ou le sens spécifique du mot simple. De plus, un apprenant de la langue établit un autre lien entre une phrase (description linguistique) qui décrit l'événement et sa structure conceptuelle (représentation). Toutefois, ce dernier peut ne pas détecter correctement le vrai sens de cette description linguistique. Ceci est peut-être dû aux différents aspects de la scène qui peuvent créer de l'incertitude concernant le sens de l'énoncé, ou tout simplement de l'énoncé lui-même, qui peut-être incomplet ou bruyant. Ce problème, appelé l'indétermination référentielle, est évoqué par plusieurs chercheurs tels que (Quine, 1960) et (Gleitman, 1990).

Ceci est un phénomène lié à l'apprentissage de la langue naturelle où l'apprenant (observateur) n'arrive pas à sélectionner le bon ou le vrai sens d'un énoncé lié à un événement donné (environnement). Devant une scène de chasse, par exemple, un apprenant peut avoir plusieurs structures conceptuelles en lien avec cette dernière. En outre, lorsqu'il entend un énoncé qui décrit cette scène, il n'est pas évident pour lui de déterminer quelle est la représentation conceptuelle la plus adéquate à cet énoncé, autrement dit le sens de ce dernier.

Parmi les modèles computationnels qui ont été élaborés pour étudier ou simuler le rôle de l'indétermination référentielle dans l'apprentissage des sens des mots, on retrouve celui de l'apprentissage des sens des mots par des agents incorporés (Wellens, 2008), du contexte intentionnel dans l'apprentissage de la langue (Fleischman et Roy, 2005), et,

finalement, de l'inférence d'un modèle probabiliste de la mémoire sémantique à partir des normes d'association des mots (Andrews et Vinson, 2008). Ce dernier est un modèle probabiliste qui peut établir une représentation intuitive des connaissances sémantiques acquises à partir des données liées aux associations des mots et de leurs caractéristiques. À chaque stimulus, le modèle retourne une liste de mots qui lui correspond.

1.5 Acquisition des rôles sémantiques des actants d'une LS

Les rôles sémantiques tels que l'agent, le thème et le bénéficiaire sont des éléments principaux dans l'apprentissage des langues et plus spécifiquement l'ALS. Ces rôles indiquent les relations entre les divers acteurs qui jouent des rôles différents dans une scène ou une situation donnée. D'autres actants secondaires peuvent être ajoutés, tels que la destination, la source, le but, la relance, l'état, l'expérimentateur, l'instrument, et l'agent secondaire (Jackendoff, 1990).

Les rôles sémantiques peuvent mieux expliquer le comportement lexical des verbes par rapport à la sous-catégorisation syntaxique de ces derniers (intransitifs, transitifs directs, transitifs indirects, ...). Dans son modèle d'attribution de rôles sémantiques dans une analyse syntaxique (Peter, 1997) a confirmé que la fonction principale de l'analyse syntaxique est l'attribution des rôles sémantiques aux actants du verbe. D'autres modèles d'analyse syntaxique, comme la théorie de la valence et la théorie de l'analyse en prédicat ont montré le lien important entre les fonctions syntaxiques (sujet, objet, complément, ...) et les rôles sémantiques (agent, patient, but, ...), ce qui crée une adjonction syntaxico-sémantique (Radimsky, 2012). Que ce soit en anglais ou en français, il y a un ordre normatif selon lequel les rôles sémantiques doivent être affectés. Dans la phrase suivante, qui est à la voix active, "Adam a donné une fleur à Ève.", les rôles sémantiques des actants doivent suivre l'ordre <Agent> <Patient> <Bénéficiaire>, tandis qu'à la voix passive l'ordre devient <Bénéficiaire> <Patient> <Agent> (en ajoutant certaines prépositions "a", "de"). Cependant, dans le cas des fonctions syntaxiques, le <Sujet> garde sa première place dans les deux situations.

Certains modèles computationnels (Perek, 2015; Alishahi et Stevenson, 2008) ont été adaptés pour qu'ils s'adressent directement à l'apprentissage et l'utilisation des rôles sémantiques au lieu de représenter ces derniers par des étiquettes explicites en supposant que les apprenants établissent déjà le lien entre les structures actanciennes et les rôles sémantiques. Ces modèles associent chaque actant à un rôle sémantique. Chaque rôle est une distribution de probabilité sur un ensemble de propriétés sémantiques que chaque actant peut prendre. De même, chaque rôle peut être appris à partir des exemples ou des cas de son utilisation.

Ces modèles ont été adaptés en supposant que l'apprenant peut déduire certaines propriétés sémantiques à partir des actants. De même, ils supposent qu'il est capable d'établir des associations probabilistes entre les propriétés sémantiques et les positions syntaxiques des actants d'un côté, et les primitives sémantiques du prédicat d'un autre côté. Ces associations sont généralisées dans les constructions pour former des notions plus abstraites des rôles sémantiques. Ces notions abstraites dépendent des positions des actants et des primitives sémantiques du prédicat.

Les propriétés des entrées sont très similaires à celles utilisées pour simuler l'acquisition des constructions en utilisant des paires scène/énoncé. La seule différence est de remplacer les rôles et les catégories sémantiques (ontologie hiérarchique) dans le cadre d'un prédicat par les propriétés sémantiques de chaque actant. Par exemple, dans la phrase 'Adam made cake', l'expression du prédicat ne contient pas des informations sur les rôles et les catégories sémantiques, mais des propriétés sémantiques des actants, tirées de *WordNet*¹ (George, 1995), comme :

{ man, male, person, individual, ... } pour Adam et { baked goods, food, solid, substance, ... } pour cake.

Cette nouvelle présentation permettra à un modèle probabiliste d'apprentissage, de déduire les rôles sémantiques des actants au lieu de supposer qu'ils ont déjà encodés

¹ <http://wordnet.princeton.edu/>

quelque part, ce qui est plus proche de cas des apprenants. Ce modèle probabiliste produit une expression (un cadre) de structure actancielle pour chaque entrée ou paire scène/énoncé. Chaque cadre contient des informations concernant le mot principal, la valence, les primitives sémantiques, le modèle syntaxique et les propriétés sémantiques de chaque actant.

Par ailleurs, dans deux langues différentes tels que le français et l'anglais, deux mots ont les mêmes propriétés sémantiques s'ils ont le même référent, comme dans le cas de "chat" et "cat". Toutefois, plusieurs confusions sont possibles, plus spécifiquement dans le cas des verbes. Pour un verbe, le référent est plutôt une 'action' ou un 'événement' qu'un objet concret. Durant une scène de combat, il peut y avoir une confusion de la part d'un agent cognitif pour le verbe 'défendre'. L'apprenant peut affecter les propriétés sémantiques du verbe 'attaquer' au verbe 'défendre' à cause de cette confusion de scène. D'autre part, des faux amis peuvent produire le même genre d'ambiguïté. Un apprenant peut affecter les propriétés sémantiques du verbe 'to assist' < aider, activer, manœuvrer > au verbe 'assister à' dont les propriétés sémantiques < participer, voir, visiter >.

1.6 Acquisition des préférences de sélection des verbes d'une LS

Les verbes ont des préférences concernant les propriétés sémantiques de leurs actants. Par exemple, le verbe 'eat' a des préférences pour les actants avec des propriétés telles que comestibles et solides, alors que 'drink' a des préférences pour les actants avec les propriétés 'liquide' et 'potable', tandis que l'inverse de ces préférences (propriété 'potable' pour 'eat' et propriétés 'comestible' et 'solide' pour 'drink') n'est pas accepté.

Certains chercheurs (Fodor et Katz, 1964) considèrent les préférences de sélection comme des restrictions sur l'application des prédicats sur les actants. Ces restrictions sont représentées par des fonctions booléennes avec des propriétés sémantiques, tels que 'human' ou 'higher animal' pour les sujets d'un verbe par exemple. D'autres estiment (Resnik, 1993) que la connaissance des mots ou des concepts est représentée par une hiérarchie sémantique de classes, et les processus statistiques sont utilisés pour

apprendre les préférences de sélection à partir des exemples. Par exemple, les préférences de 'eat' ne sont pas représentées comme des propriétés d'une classe, mais plutôt comme une distribution de probabilité sur ces propriétés.

L'étude de (Gleitman et Gillette, 1995) a démontré que les apprenants adultes peuvent identifier à quatre-vingts pour cent un verbe en s'appuyant sur sa structure actancielle (actants et leurs positions), toutefois, l'absence des informations sur les positions ou sur les actants n'aide pas les apprenants à identifier le verbe. De même, cette étude affirme que les préférences fournissent des informations importantes sur les sens des verbes inconnus.

Cela montre que les propriétés sémantiques des actants d'un verbe ou ces préférences de sélection sont aussi importantes dans l'apprentissage des sens des verbes que les associations sémantiques entre ce dernier et l'ensemble de ses actants ou de ses propriétés syntaxiques. En outre, les préférences de sélection jouent un rôle très important dans de nombreux aspects du traitement de la langue. En fait, elles influencent sur l'apprentissage des structures actanciennes (actants et positions) d'un verbe et peuvent enlever leur ambiguïté. En conséquence, il est très important de clarifier le processus, par lequel l'apprentissage des préférences de sélection est effectué.

Des modèles computationnels (Perek, 2015; Alishahi et Stevenson, 2008) ont été conçus pour pouvoir apprendre des structures actanciennes et des préférences de sélection des verbes dans le contexte de L1. Ils extraient des structures et les rassemblent dans des constructions selon leurs similarités. Pour chaque verbe, ces modèles associent une position d'un actant avec une distribution de probabilité sur l'ensemble des propriétés du profil sémantique du verbe.

Les entrées de ces modèles sont générées automatiquement. Un cadre est créé pour chaque verbe. Ce dernier contient le verbe, sa valence, ses propriétés sémantiques et le modèle syntaxique de la phrase où il se trouve. La structure extraite est stockée dans l'entrée lexicale du mot principal (verbe ou préposition), l'objectif est de tenir compte

des propriétés sémantiques des actants d'un verbe dans une phrase donnée. De même, grouper les expressions des verbes utilisés dans des constructions, puisque les constructions les plus incarnées devraient avoir une plus grande influence sur le profil et l'apprentissage du verbe.

Notre modèle probabiliste a montré que dans un contexte de LS, l'apprentissage des préférences de sélection suit une courbe U en passant par une étape de surgénéralisation des règles. Cette surgénéralisation est causée par l'utilisation des préférences de sélection des faux amis anglais-français. Prenons l'exemple du verbe 'advance' qui est utilisé dans un contexte de progression, et dont les préférences de sélection 'progressif' et 'développable' et qui peut être utilisé à la place du verbe 'avancer' dont les préférences de sélection 'destination' et 'direction'.

1.7 Questions de recherche et portée de l'étude

Tant du point de vue des sciences cognitives que de l'informatique, la conception d'un modèle d'apprentissage des LS se doit de tenir compte d'un certain nombre de questions dont les principales sont les suivantes :

- Quels sont les facteurs les plus importants à prendre en considération dans la modélisation de l'ALS ?
- Les mécanismes d'apprentissage du L1 peuvent-ils être pris en compte dans l'ALS (Clark, 2009; Matuskevych et al., 2016) ?
- Comment un système de conceptualisation mature (chez les apprenants adultes et les enfants moins jeunes) peut influencer, positivement ou négativement l'ALS ?
- Quels sont les processus cognitifs qui relient les sens des verbes à des formes syntaxiques valides ?
- Comment les théories cognitives d'apprentissage de la langue naturelle et d'ALS peuvent améliorer voire rendre plus efficaces les systèmes d'apprentissage et les outils d'aide à l'apprentissage orientés vers des apprenants d'une LS ?

Toutefois, la question principale de notre projet de recherche est comment déterminer les éléments clés d'un modèle computationnel d'ALS, capable de rendre plus explicites certains processus cognitifs sur lesquels se fonde cette acquisition?

Les processus cognitifs sont les mécanismes mentaux qui permettent à un agent cognitif d'interagir avec son environnement. Plusieurs mécanismes peuvent représenter ces processus dont la perception, l'attention, la mémoire, l'apprentissage de la langue, l'oubli, l'émotion, etc. (Gazzaniga, 1995). Les processus cognitifs abordés dans ce projet sont : la compréhension et la production de la langue, l'apprentissage des sens des mots, des structures actanciennes, des rôles sémantiques et des préférences de sélection et, finalement, la surgénéralisation et l'autocorrection par preuve positive.

La réponse à cette question permettra, éventuellement, de trouver des solutions à plusieurs problèmes liés à l'apprentissage des langues et au traitement automatique de la langue naturelle.

Parmi ces problèmes, notons, entre autres, l'élaboration des systèmes d'apprentissage et d'aide à l'apprentissage des langues, la traduction automatique, la recherche d'informations, la performance des systèmes de questions/réponses, la génération de textes, etc.

Nous allons étudier l'influence de chaque facteur d'apprentissage sur les autres facteurs (par exemple, l'influence de l'acquisition des sens des mots sur l'acquisition des structures actanciennes) et nous allons prendre en considération l'influence du système conceptuel mature chez les apprenants adultes. Ce dernier contient les représentations sémantiques de nature symbolique qui peuvent être liées à la perception, au raisonnement, et au langage (Lakoff, 1994).

En tenant compte du système conceptuel, nous allons élaborer un modèle probabiliste d'ALS qui est capable de simuler l'ALS en s'appuyant seulement sur l'utilisation de la langue elle-même (Tomasello, 2003; Behrens, 2009; Eskildsen, 2008). En fait, la théorie

« *Usage-Based* » est au cœur de la question suivante : L'apprentissage et l'acquisition de la langue naturelle sont-ils de nature symbolique ou connexionniste?

Notre modèle computationnel sera fondé sur plusieurs théories et aspects cognitifs, dont la théorie « *Usage-Based* », le système conceptuel, la compréhension et la production de la langue et les processus cognitifs d'apprentissage.

CHAPITRE II

UN MODELE PROBABILISTE D'ACQUISITION DES LANGUES SECONDES

2.1 Introduction

Établir un modèle cognitif computationnel d'ALS, fondé sur les théories de construction telle que la théorie « Usage-Based », nous permet de simuler les processus cognitifs humains d'ALS. Le modèle sera capable d'appliquer des mécanismes de catégorisation et de généralisation sur des intrants observés et extraits à partir des corpus linguistiques pour prédire des modèles syntaxiques, des structures actanciennes ou des propriétés sémantiques du verbe. Cela nous donnerait également le moyen de comprendre comment un apprenant peut généraliser des associations entre la syntaxe et le sens (sémantique) du verbe et de ces actants en se basant sur l'unité fondamentale du langage qui est « la construction » (Fillmore, 1988 ; Goldberg, 1995). Ces constructions sont générées par le modèle/apprenant à partir d'un nombre limité d'intrants.

Selon Goldberg (1995) « **C** est une **construction** si et seulement si **C** constitue un couple forme-sens $\langle F_i, S_i \rangle$ tel qu'il existe un aspect de F_i ou de S_i qui ne soit pas être strictement prédictible à partir des composants de **C** ou d'autres constructions préalablement établies »

Notre modèle computationnel place l'échelle d'observation au niveau des interactions entre les connaissances générales et spécifiques au cours des différentes étapes d'acquisition pour mesurer le progrès de l'apprenant. Des interactions entre le modèle syntaxique et les propriétés sémantiques du verbe, dans un contexte d'utilisation de la

langue et d'acquisition des structures actanciennes, constituent le milieu d'évolution, de la compréhension et la production de la langue.

Selon les théories de construction, le développement des compétences linguistiques liées aux structures actanciennes ne peut pas s'appuyer seulement sur les connaissances spécifiques au niveau des verbes individuels, mais surtout sur les connaissances générales associées à la construction. Par exemple, dans une construction ditransitive telle que "Il lui passe le ballon", avec un sens de « transfert provoqué de possession » et un modèle syntaxique avec deux compléments d'objet, le sens de la structure actancielle et son modèle syntaxique limitent le choix des mots qui peuvent être utilisés comme actants et, de même, limitent la valence du verbe. En outre, ces deux facteurs déterminent le sens de la construction en fournissant un contenu sémantique qui est fusionné avec le sens le plus spécifique du verbe et augmentent les capacités de catégorisation (Perek, 2015).

Comment un modèle syntaxique est associé à la sémantique (sens) d'une structure actancielle ? Les théories de construction montrent qu'il existe un lien étroit entre le sens de la construction à qui appartient la structure actancielle et l'utilisation de cette construction, en particulier les verbes qui la forment. Notre modèle, et en observant l'usage des verbes qui constituent une construction, fournit de l'information sur le sens général de cette dernière, en simulant l'acquisition d'une LS par un apprenant. Cependant, nous allons montrer que ce lien étroit entre le sens et la syntaxe dépend directement de la fréquence d'utilisation du verbe au niveau de la construction, et qu'il y a des verbes qui peuvent dominer l'interprétation du sens grâce à leur fréquence d'utilisation, tels que : "aller", "donner" et "mettre".

Plusieurs études ont analysé l'acquisition des constructions dans le cas d'une LS, plus spécifiquement les constructions de « mouvement provoqué » (Hickmann et Henriette, 2010 ; Ellis et Cadierno, 2009). Ces constructions de « mouvement provoqué » et d'autres propriétés liées aux verbes posent un défi aux apprenants de l'anglais LS dont la

L1 est le français et vice versa. Même si le français et l'anglais ont la même typologie syntaxique qui est Sujet-Verbe-Objet (SVO) et possèdent la même structure profonde (structure-D). Selon d'autres classifications, les deux langues appartiennent à deux familles différentes (Talmy, 1985, 2000). L'anglais qui est une langue germanique à cadrage satellitaire se diffère du français qui est une langue romane à cadrage verbal, plus spécifiquement dans l'élaboration des constructions de « mouvement provoqué ». En anglais, les verbes expriment souvent le mode de déplacement, tandis que la direction est distinguée par un affixe : “She run into”, “He walk across”, “He go out”, “She fall down”. Alors qu'en français, la majorité des verbes de mouvement révèlent la direction, cependant que la manière de déplacement est signalée par un complément tel qu'un gérondif ou une construction adverbiale, ou elle peut ne pas être énoncée : “Il sort en courant”, “Elle entre”, “Elle traverse à la nage”, “Il monte à pied” (Henriette, 2010 ; Benson, 2002; Pollock, 1989; Chomsky, 1968, 1991). Certaines études (Henriette, Hickmann, et Annie-Claude, 2008) ont analysé le lien entre ces dissimilarités typologiques et l'ALS. Ils ont pris comme sujets d'expérience des apprenants anglophones du français LS. Les résultats ont montré que, durant les premières étapes d'apprentissages, les apprenants de LS organisent l'information en fonction de leur L1. Ainsi, ils expriment, comme dans leur L1, le mode de déplacement et la cause dans le verbe, en ajoutant la direction dans l'affixe : “Il sort à l'extérieur”. Durant les étapes avancées d'apprentissage, les apprenants expriment la manière dans un gérondif : “Il sort en courant”. Nous allons montrer dans les chapitres suivants que celui-ci est un comportement présumé par notre modèle computationnel. Durant ses différentes étapes d'apprentissage, le modèle va évoluer en augmentant la densité sémantique, et va se corriger en produisant des phrases comme « He run in » au lieu de « He enters by running », juste en recevant plus d'intrants.

Notre modèle applique le principe de construction de (Goldberg, 1995), et qui a été adopté par plusieurs autres chercheurs (Perek, 2015 ; Alishahi et Stevenson, 2010 ; Matusевич et al., 2016). De même que (Alishahi et Stevenson, 2010) et (Matusевич et al., 2016) notre modèle infère des associations probabilistes entre le modèle syntaxique

et les propriétés sémantiques d'un verbe et celles de ces actants. Les constructions sont considérées comme des structures qui regroupent des verbes qui ont une similarité élevée des propriétés syntaxiques et sémantiques. Cette similarité est calculée à partir d'un modèle bayésien en utilisant les valeurs de ces propriétés. En outre, le modèle fournit et utilise des prédictions probabilistes pour calculer les valeurs possibles des propriétés syntaxiques et sémantiques dans les différentes étapes d'apprentissage. Cette caractéristique, qui le rend capable de simuler l'apprentissage de LS selon le nombre d'intrants offert dans chaque étape d'apprentissage et fournit de l'information probabiliste sur l'effet de l'usage de la langue sur son apprentissage.

Durant l'apprentissage du LS, une interférence entre les connaissances spécifiques associées au verbe et les connaissances générales associées à la construction permettra à l'apprenant de généraliser ces connaissances sur des nouveaux verbes (Ellis, 2013). Cela donne la possibilité de produire des nouvelles phrases qui n'ont pas été vues dans les intrants. Dans ce cas, l'apprentissage est considéré comme un processus de prédiction dans lequel les propriétés ou les composants manquants dans des tâches de production ou de compréhension de la langue sont fixés aux valeurs les plus probables et selon les caractéristiques disponibles. Ainsi, la production de phrases est vue comme une prédiction de sa structure syntaxique la plus probable, étant donné les caractéristiques sémantiques d'une représentation, d'une observation ou d'une scène. À l'inverse, la compréhension est vue comme la prédiction des rôles et des propriétés sémantiques des composants d'un énoncé donné. Cette prédiction est de nature incertaine et elle est fondée sur des informations inexacts et imprécises. Pour cela, cette approche probabiliste bayésienne s'impose en permettant de conceptualiser un modèle d'acquisition de langues secondes en modélisant les différentes tâches d'apprentissage comme des processus de prédiction.

Nous allons présenter dans ce chapitre les différents composants de notre modèle bayésien. Nous montrons dans la section 2.2 une représentation des constructions probabiliste dans le cas de LS. Cette représentation est une adaptation de la théorie de

construction de (Goldberg, 1995) et du modèle d'apprentissage de L1 de (Alishahi et Stevenson, 2008), à l'apprentissage de LS. Nous allons montrer la notion de la représentation sémantique dans le cas de LS. En outre, nous expliquons les intrants de notre modèle et les paires scène/énoncé que nous avons élaboré à partir des corpus linguistiques. De même, la catégorisation des structures. Dans la section 2.3 nous décrivons le calcul de similarité et l'élaboration de constructions. La section 2.4 est consacrée au comportement du modèle dans le cas d'ALS. Finalement, nous allons conclure dans la section 2.5.

Le modèle sera adapté en fonction des éléments d'apprentissage liés aux actants du verbe. Nous allons nous concentrer sur l'acquisition des structures actancielles du verbe dans le cas de LS dans le chapitre 3. Nous montrons également l'acquisition des rôles sémantiques des verbes de LS dans le chapitre 4. Le chapitre 5 sera consacré à l'acquisition des préférences de sélection des verbes de LS. Dans chaque chapitre et pour tout élément d'apprentissage, nous analysons les résultats des simulations en comparant le procédé de notre modèle avec celui des apprenants de LS dans différents expérimentations et contextes.

2.2 Constructions probabilistes et ALS

Selon la sémiotique ou l'étude des signes et de leur signification, c'est grâce à la 'représentation' qu'un agent cognitif est en mesure d'utiliser des règles computationnelles et est capable d'établir un lien entre un référent, un signifiant et un signifié. Cet agent cognitif pourra, par la suite, catégoriser, classifier ou généraliser des nouvelles connaissances (Bernard, 2005; Meunier, 1989 ; Meunier, 1999). D'un autre côté, Meunier (1989) confirme qu' « Avant d'être une forme sémiotique -une langue-, le langage est défini comme un "comportement", une "faculté ", un "agir" qui permet à l'humain soit d'"exprimer sa pensée", de "représenter le monde", soit de "communiquer avec les autres". »

De ce fait, le concept scène/énoncé est très important dans la modélisation d'ALS. Ce dernier exprime cette "représentation du monde" qui permettra à l'apprenant de

comprendre son environnement et saisir les sens des mots et des phrases. Selon le concept scène/énoncé, l'apprenant voit une scène et entend un énoncé qui déclenche ses mécanismes de 'compréhension' ou de 'production'. De même qu'aux modèles de (Perek, 2015 ; Alishahi et Stevenson, 2010 ; Matusevych et al., 2016) nous avons représenté les intrants de notre modèle comme des paires scène/énoncé. Dans ce cas, l'énoncé représente la description linguistique entendue par l'apprenant, tandis que la scène est la représentation interne d'un événement ou d'une observation qui s'est passée, parallèlement, devant de ce dernier. Pour la circonstance, les apprenants d'une LS (enfants moins-jeunes et adultes) sont supposés avoir un système de conceptualisation mature qui les permet d'identifier le sens de la phrase ou de l'étiquette linguistique qui décrit la scène, de même que les sens des mots simples. Ce système de conceptualisation est le facteur qui constitue la différence principale entre les jeunes enfants qui apprennent leur L1 et les apprenants d'une langue seconde (Lakoff, 1994; Rappoport et Sheinman, 2005).

À partir de ces paires scène/énoncé, le modèle construit des structures ou des cadres qui le permettent de simuler la "représentation interne" chez un agent cognitif. Ensuite, il les catégorise selon leur degré de similarité dans des constructions. Ces constructions de nature dynamique, plus spécifiquement durant les phases d'apprentissage, expriment les différents niveaux d'apprentissage chez un apprenant conformément au nombre d'intrants reçu par le modèle.

Dans les sous-sections suivantes, nous allons illustrer le lien qui existe entre la représentation sémantique et la LS, l'élaboration des intrants du modèle à partir des paires scène/énoncé et finalement la catégorisation des cadres.

2.2.1 Représentation sémantique et LS

Comme nous l'avons déjà mentionné, les apprenants d'une LS (enfants moins-jeunes et adultes) possèdent déjà un système de conceptualisation mature qui leur permet d'identifier le sens des phrases liées à une scène. Ce qui constitue la différence

fondamentale entre les jeunes enfants qui apprennent une L1 et les apprenants d'une LS. Ce système permet d'établir une représentation sémantique d'un mot simple, d'une phrase ou d'un texte, consiste à en déterminer le sens (Lakoff, 1994). Toutefois, pour la circonstance d'ALS, un mot (nom ou verbe) entendu par un apprenant, peut être lié à un mot de la L1 qui est par son tour associée à une représentation sémantique/conceptuelle (Michael, 2005). Cette équivalence translinguistique peut être en faveur ou en défaveur de l'ASL. Dans le cas des langues ayant des cultures différentes, cela peut générer plusieurs confusions, particulièrement pour des mots ou des expressions polysémiques, qui ont des sens multiples ou des significations différentes (Kerrie et John, 2008). Cette polysémie affecte l'ALS lorsque les apprenants interprètent les sens des expressions linguistiques de la LS en activant les concepts linguistiques de la L1. Par exemple, le mot peinture peut exprimer l'activité de la peinture, le matériau utilisé dans cette activité, le résultat de cette dernière, son type, une œuvre d'art, etc. (Bernard et Catherine, 1996). Dans cette circonstance, le choix du concept linguistique adéquat, dépendamment du contexte, pose un défi pour un apprenant de LS. Différents modèles ont décrit la représentation sémantique des sens des mots et des expressions linguistiques : Modèles vectoriels de la composition sémantique (Mitchell et Lapata, 2008); influence de la L1 sur LS dans l'interprétation du sens (Kerrie et John, 2008); étude computationnelle des techniques d'apprentissage des sens des mots (Siskind, 1996); contexte intentionnel dans l'apprentissage de la langue (Fleischman et Roy, 2005) ; le rôle de l'intention incorporée dans l'acquisition lexicale (Yu, Ballard, et Aslin, 2005); l'observation intersituationnelle et l'apprentissage du vocabulaire du LS chez des apprenants adultes (Ludington, 2013).

Au nombre de théories d'apprentissage des sens des mots, nous avons celle de l'observation intersituationnelle (*Cross-Situational Observation*) (Pinker, 1989). Cette dernière confirme qu'une représentation sémantique peut être élaborée à partir des observations des situations différentes. Ces observations permettent de détecter le sens commun entre toutes les situations. Dans le cas d'ALS (Ludington, 2013) a montré l'effet positif de la variété des contextes d'apprentissage (intersituationnelle) dans

l'acquisition de LS, plus spécifiquement celle des sens des mots. En comparant le degré d'apprentissage entre deux groupes d'apprenants de LS, ceux qui ont appris le vocabulaire dans différents contextes (différents textes et histoires) ont eu une meilleure performance par rapport à ceux qui l'ont appris dans un contexte unique.

Cependant, les apprenants de LS ne réussissent pas toujours à établir des associations entre une étiquette linguistique et une représentation interne à cause du décalage ou de l'asynchronisation entre la scène et l'énoncé. De même, l'absence de l'action attachée au référent peut causer un problème, par exemple, la prononciation du verbe « ouvrir » là où l'action d'ouverture est absente ne peut pas expliquer la scène. Dans certains autres cas, enlever l'ambiguïté et l'indétermination référentielle lorsqu'il s'agit de plusieurs structures conceptuelles qui peuvent être associées à une observation donnée, est indispensable. Cette indétermination référentielle peut empêcher l'apprenant d'associer correctement un référent à une expression linguistique ou découvrir l'objet qui doit être associé à un concept. Nous allons montrer comment notre modèle passera par différentes étapes d'apprentissage avant qu'il soit capable d'éliminer certaines indéterminations référentielles, en supposant qu'il possède déjà une représentation sémantique adéquate pour les mots simples.

2.2.2 Paires scène/énoncé et intrants du modèle

Pour simuler l'apprentissage des apprenants d'une LS et leurs représentations sémantiques, et par conséquent pouvoir expliquer les mécanismes d'ALS, nous allons adopter le concept scène/énoncé dans la conception des intrants de notre modèle computationnel. Ce concept a été utilisé par plusieurs autres modèles d'apprentissage (Perek, 2015 ; Alishahi et Stevenson, 2010) et LS (Matushevych et al., 2016). Nous allons montrer le passage de la paire scène/énoncé à une structure qui représentera un intrant du modèle. L'énoncé est représenté par la phrase sous sa forme texte ou audio, tandis que la scène est représentée par un prédicat, des arguments et des propriétés. Le prédicat exprime le type d'événement ou l'état décrit par l'énoncé. Les participants à cet événement ou état sont exprimés par les arguments du prédicat. Les prédicats

représentent souvent des verbes, et dans certains cas, des adjectifs ou des prépositions. La phrase “Adam boit de l’eau” peut-être représentée par la paire :

$$\left\{ \begin{array}{l} \text{Scène : BOIRE}_{[\text{acte, absorber}]}(\text{ADAM}_{\langle \text{agent} \rangle \langle \text{animé} \rangle}, \text{EAU}_{\langle \text{thème} \rangle \langle \text{potable} \rangle}) \\ \text{Énoncé : Adam boit de l’eau.} \end{array} \right.$$

Et d’une façon générale, un énoncé d’un prédicat avec deux arguments, peut-être représenté par la paire:

$$\left\{ \begin{array}{l} \text{Scène : PREDICAT}_{[\text{Primitives_sémantiques}]}(\text{ARG1}_{\langle \text{Rôle_sémantique} \rangle \langle \text{Traits_sémantiques} \rangle}, \\ \text{ARG2}_{\langle \text{Rôle_sémantique} \rangle \langle \text{Traits_sémantiques} \rangle}) \\ \text{Énoncé : ARG1 prédicat ARG2.} \end{array} \right.$$

Dans nos futures présentations des structures, nous allons utiliser le verbe à l’infinitif. La morphologie verbale dans le cas d’ALS sera traitée dans d’autres perspectives de cette thèse.

Pour transférer la paire scène/énoncé à une structure/cadre (intrans) nous allons adopter une définition de schéma XML. Cette définition procure, à proportion, un exemple de fichier des entrées au format XML. Les balises de la structure conceptuelle du document XML sont représentées dans le tableau 2.1, qui montre les composants élémentaires et composés (ensemble d’éléments, comme le cas des Rôles_sémantiques).

Élément structurel
Nom du prédicat : Verbe ou préposition.
Fréquence du prédicat dans le corpus linguistique
Structures : Ensemble de structures
Structure : cadre avec plusieurs paramètres
Structure ID : Identificateur de la structure dans le cadre d’un prédicat.
Représentation sémantique

Fréquence Représentation sémantique
Primitives sémantiques du verbe
Nombre d'arguments
Rôle sémantique d'un actant
Trait sémantique d'un actant
Primitives sémantiques d'un verbe ou d'un actant
Modèle syntaxique

Tableau 2.1: Structure extraite à partir de la paire scène/énoncé.

Les intrants du modèle sont des structures ou des cadres qui ont été extraites à partir des paires scène/énoncé et qui sont exprimées comme des documents XML.

Cette structure représente des éléments syntaxiques et sémantiques qui peuvent former une entrée du lexique mental et permettent de trouver ou sélectionner le bon mot dans un contexte de production ou le classifier et l'indexer dans un contexte de compréhension. L'entrée du lexique mental permet, en outre, de créer une représentation interne en associant une forme phonologique ou morphologique aux éléments syntaxiques et sémantiques. Une telle représentation interne participe à la construction du lexique de l'apprenant et plus spécifiquement son interlangue qui peut se produire dans le cas de l'ALS (Corder, 1975 ; Richards, 1972 ; Selinker, Swain, et Dumas, 1975 ; Adone, 2003; Lumsden, 1999; Lefebvre, 2006).

Conformément aux éléments d'apprentissage liés aux actants du verbe, nous allons adapter ces structures en conséquence de la modélisation d'acquisition des structures actanciennes du verbe, l'acquisition des rôles sémantiques des verbes et finalement l'acquisition des préférences de sélection des verbes de LS.

Selon le principe de catégorisation cognitive (Rosch, 1973; Lakoff, 1987) ces structures peuvent être regroupées sous des constructions, dépendamment de leurs similitudes. Le modèle calcule la similarité d'une nouvelle structure avec les anciennes constructions et la catégorise avec la plus similaire. Si la similarité avec toutes les constructions ne dépasse pas le seuil prédéfini, une nouvelle construction sera créée, spécifiquement, pour ce cadre. Dans le calcul de similarité, nous prenons en considération toutes les propriétés sémantiques et syntaxiques du cadre d'un côté et de l'ensemble de cadres qui forment la construction d'un autre côté.

2.2.3 Catégorisation des cadres

Les constructions sont des unités symboliques de la langue qui relient les propriétés morphologiques, syntaxiques et lexicales avec des fonctions sémantiques particulières (Lakoff, 1987; Langacker, 2000; Goldberg, 1995; Tomasello, 2003; Ellis et Cadierno, 2009). Plusieurs aspects communs entre l'acquisition de la L1 et de la LS peuvent être repérés et compris à partir de l'analyse des corpus linguistiques, de l'aspect cognitif de l'acquisition des constructions et des principes de catégorisation. Par conséquent, les approches basées sur l'utilisation sont de plus en plus influentes dans la recherche d'ALS (Ellis et Cadierno 2009, Ellis, 2013). Toutefois, l'ALS a un autre aspect de complexité par rapport à la L1. Reconstruire une langue est plus complexe que son induction initiale, car, au cours du développement, les constructions de la LS sont en concurrence directe avec celles de la L1 des apprenants, et celles-ci peuvent représenter de manière différente la même scène, ce qui favorise l'influence translinguistique (Ellis et Cadierno 2009).

Apprendre une LS implique d'apprendre à interpréter le monde comme les natives de cette langue. Autrement dit, apprendre des façons alternatives de penser pour parler ou à repenser avec la LS pour parler. Les théories de transfert telles que l'hypothèse de l'analyse contrastive (Selinker, Swain, et Dumas, 1975 ; Adone, 2003; Lumsden, 1999; Lefebvre, 2006) estiment que l'ALS est peut être plus facile lorsque les constructions de la LS sont semblables à celles de la L1. Dans ce cas, les constructions de la L1 peuvent

servir de base pour les constructions de la LS. Dans l'autre cas, l'acquisition de la LS peut être freinée par le modèle de la L1 (Ellis et Cadierno 2009). Que ce soit avec des locuteurs natifs de la L1 ou des apprenants de la LS, plusieurs études ont été menées sur l'influence de l'effet d'amorçage structurel sur l'acquisition de différents aspects de la langue telles que les structures actanciennes, les constructions datives, les préférences de sélection des verbes, les phrases interrogatives, etc. (Anita, 2014). Un effet d'amorçage se produit lorsqu'un apprenant utilise une amorce déjà observée pour produire un objet similaire. Lorsque ce dernier réutilise une structure syntaxique, rapidement et sans erreur, c'est le cas d'un amorçage syntaxique ou structurel positif. Pour ce qui est de l'ALS, certaines études (McDonough, 2006; McDonough et Trofimovich, 2013) ont montré que cet amorçage peut être positif ou négatif, dépendamment du niveau d'apprentissage et de la fréquence des structures.

La catégorisation des structures de LS est peut-être influencée positivement ou négativement par les facteurs précédemment mentionnés, tels que l'amorçage structurel, l'influence translinguistique et la fréquence d'utilisation. Cette catégorisation consiste, comme le cas de la L1, à regrouper les structures dans des constructions selon un calcul de similarité des propriétés syntaxiques et sémantiques.

Les éléments considérés dans le calcul de similarité sont ceux qui ont été mentionnés dans le tableau 2.1, tels que les primitives sémantiques du verbe, le nombre d'arguments, les rôles sémantiques des actants, les traits sémantiques, les primitives sémantiques d'un verbe ou d'un actant et le modèle syntaxique. Toutefois, les verbes les plus utilisés et qui ont une fréquence importante dans les corpus linguistiques tels que "Faire", "Aller" et "Dire" peuvent influencer ce calcul de similarité en imposant des valeurs spécifiques pour certaines constructions. À titre d'exemple, nous pouvons voir un modèle syntaxique transitif qui domine certaines constructions suite à la fréquence des prédicats regroupés dans ces dernières. Dans le cas d'ALS, cette similarité qui est calculée à partir d'un modèle bayésien en utilisant les valeurs des propriétés syntaxiques et sémantiques devra prendre en considération quelques aspects généraux (L1) et spécifiques (LS) :

- les verbes qui peuvent avoir des valences différentes (possibilité d'être transitif ou intransitif, selon le contexte et le sens), comme le cas des verbes 'Monter', 'Parler'. Etc.
- l'amorçage structurel négatif; à titre d'exemple l'utilisation de l'auxiliaire avoir au lieu de l'auxiliaire être à cause de l'homophonie de la forme de l'auxiliaire avoir (ai) et celle de l'auxiliaire être (es) (Q : tu es allé? ; R : Oui, j'ai allé)
- le placement des verbes conjugués au présent (Pollock, 1989) :
 - o Adam voit souvent Ève / Adam often sees Eve,
- les quantificateurs flottants (Tous/Chacun) (Ayoun, 2005) :
 - o Les élèves aiment tous Ève / The students all love Eve
- Le placement de la quantification à distance (Dekydspotter, L. Rex et Rachel, 2000) :
 - o Adam a fait beaucoup de devoirs / Adam has done a lot of homework,
- le cas de l'adverbe de lieu 'y'/'There' (Chomsky, 1991) :
 - o J'y vais / I'm going (there),
- l'élaboration des constructions de « mouvement provoqué ».

2.3 Calcul de similarité et élaboration de constructions

À partir des paires scène/énoncé, notre modèle construit des structures qui lui permettent de simuler la "représentation interne" chez un agent cognitif. Ensuite, il les catégorise selon leur degré de similarité dans des constructions. Ces constructions expriment les différents niveaux d'apprentissage chez un apprenant conformément au nombre d'intrants reçus par le modèle.

Notre modèle applique le principe de construction de Goldberg (1995) et infère des associations probabilistes entre le modèle syntaxique et les propriétés sémantiques d'un verbe et de ces actants. Les constructions regroupent des cadres qui ont une similarité élevée de leurs propriétés syntaxiques et sémantiques. Cette similarité est calculée à partir d'un modèle bayésien en utilisant les valeurs de ces propriétés. Les cadres peuvent être regroupés sous des constructions selon leurs similarités. Le modèle calcule la

similarité d'un nouveau cadre par rapport aux constructions existantes; ensuite il le catégorise avec la construction la plus similaire. Si la similarité avec toutes les constructions ne dépasse pas le seuil prédéfini, une nouvelle construction sera créée pour ce cadre. Dans le calcul de similarité, nous prenons en considération toutes les propriétés sémantiques et syntaxiques du cadre d'un côté et de l'ensemble des cadres qui forment la construction d'un autre côté. De même, le modèle calcule et par la suite utilise des prédictions probabilistes des valeurs possibles des variables syntaxiques et sémantiques, dans les différents niveaux d'apprentissage. Cette caractéristique, elle lui permet de simuler l'apprentissage de la LS selon le nombre d'intrants offert dans chaque étape d'apprentissage, ce qui permet d'analyser l'information probabiliste sur l'effet de l'usage de la langue sur son apprentissage.

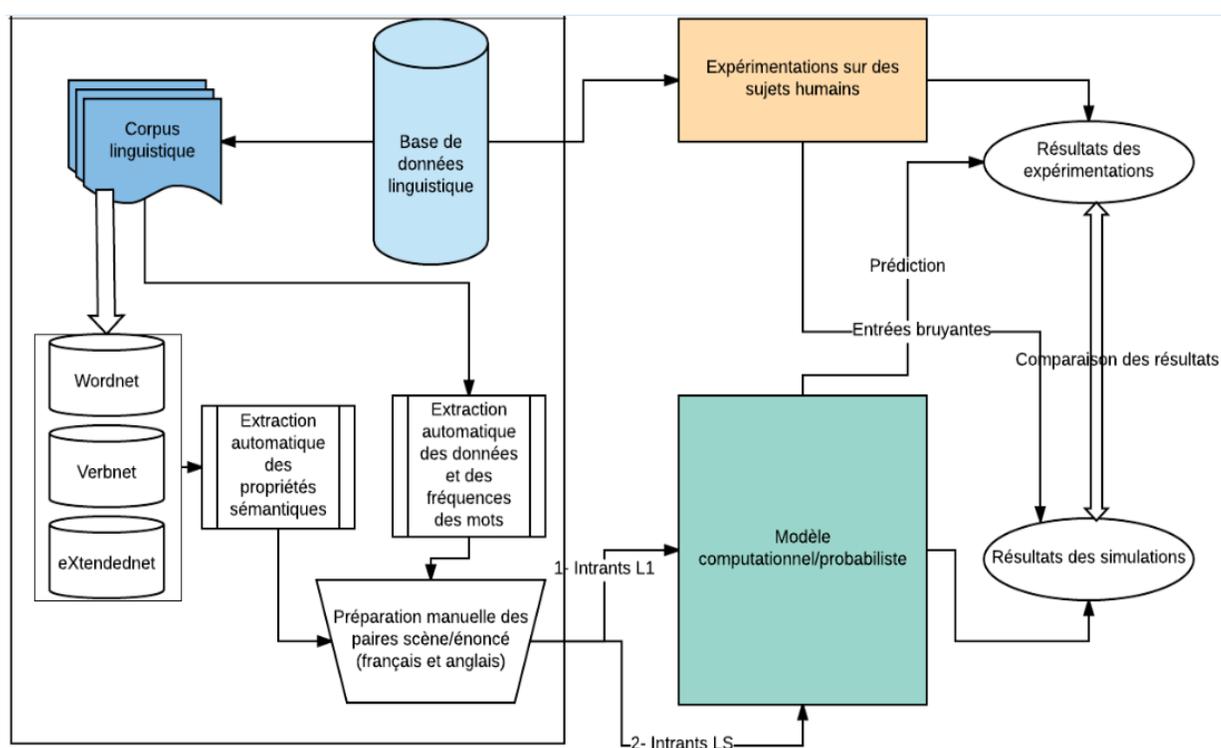


Figure 2.1: Modèle computationnel d'ALS.

Nous adoptons une approche bayésienne non supervisée extraite d'un modèle de catégorisation humaine proposé par Anderson (1991) et qui a été adoptée par d'autres

modèles de L1 (Perek, 2015 ; Alishahi et Stevenson, 2010) et LS (Matusevych et al., 2016). Ce modèle regroupe les cadres en constructions. Les intrants du modèle ne sont pas étiquetés et par conséquent les constructions ne sont pas prédéfinies, mais plutôt déterminées selon la similarité des cadres observés.

Nous appliquons la formule de Bayes et l'argument du maximum (en parcourant toutes les constructions s), noté argmax , pour trouver la construction la plus similaire.

Dans le cas d'ALS, deux ensembles de constructions vont être parcourus par le modèle, SL1 et SLS. SL1 est plus stable durant les simulations tandis que le SLS est dynamique et s'incrémente avec le temps. Dans notre modèle, le SLS est l'équivalent d'un système de conceptualisation. À une étape avancée d'apprentissage, le modèle s'oriente automatiquement vers l'ensemble SLS.

L' argmax est le point ou l'ensemble des points en lesquels une expression atteint sa valeur maximale. Le modèle utilise un algorithme de regroupement incrémental qui trouve la meilleure construction pour un nouveau cadre. Sinon, une nouvelle construction sera créée pour ce dernier :

$$\text{ConstructionPlusSimilaire}(S) = \underset{s}{\text{argmax}} (\log(P(S|C))) \quad (2.1)$$

Où s varie sur les indices de toutes les constructions, y compris $\mathbf{0}$ pour représenter une nouvelle construction. D'un autre côté, et tant que nous adoptons un modèle computationnel, nous allons utiliser la log-probabilité. L'utilisation de la log-probabilité indique la représentation des probabilités dans un espace logarithmique, au lieu d'une intervalle standard $[0,1]$. Cela donne des avantages pratiques côté précision et stabilité numérique, en raison de la façon dont les ordinateurs traitent les nombres réels. En outre, ces derniers peuvent effectuer l'addition plus efficacement que la multiplication, ce qui augmente la performance du modèle.

Le modèle calcule la similarité d'un nouveau cadre **C** par rapport à une construction **S**. En utilisant la règle de Bayes, et en éliminant la probabilité a priori du cadre $P(C)$ qui est constante pour toutes les constructions, on aura :

$$P(S|C) = P(S)P(C|S)/P(C) \propto P(S)P(C|S) \quad (2.2)$$

La probabilité a priori $P(S)$ indique le degré d'ancrage de la construction **S**. Ce degré d'ancrage rend la construction plus favorable pour un nouveau cadre. En outre, l'usage de la langue permet de modifier la construction en ajoutant de nouveaux modèles (Exemple : ajouter un cadre qui contient une forme transitive d'un verbe dans une construction qui avait seulement la forme intransitive) ou en corrigeant des modèles existants. Pour représenter cet ancrage, nous calculons la probabilité a priori d'une construction selon la fréquence des éléments syntaxiques et sémantiques de ces cadres dans le nouveau cadre observé. Ainsi, la probabilité a priori de chaque construction varie au cours de l'apprentissage, en proportion des nouvelles entrées.

La probabilité a priori de la construction s , pour $s > 0$, est calculée selon la formule d'Anderson (1991) utilisée dans son modèle de catégorisation humaine :

$$P(S) = c \cdot n_s / (c \cdot n + (1 - c)) \quad (2.3)$$

Où, n est le nombre total des cadres observés et n_s est le nombre de cadres qui forment la construction **S**. c représente la probabilité de couplage qui est une probabilité constante. Elle détermine le degré auquel les éléments préfèrent être regroupés. Son utilisation permet d'éviter la forte tendance d'un cadre à préférer ou à se préserver d'une construction existante. Lorsque deux objets proviennent de la même catégorie, cette probabilité ne dépend pas du nombre d'objets observés jusqu'à présent. Cette probabilité de couplage est très adéquate dans notre cas, là où le modèle va élaborer et utiliser deux ensembles de constructions (SL1 et SLS). Naturellement le modèle va progresser vers l'utilisation de chaque ensemble dans son contexte convenable. Pour cela, le SLS sera favorisé dans le cas d'ALS. Par conséquent, nous allons adopter la probabilité de

couplage $c = 0.5$ lorsque $\mathbf{s} \in \mathbf{SLS}$. En outre, nous utilisons la probabilité de couplage indiquée par Anderson (1991) $c = 0.3$ lorsque $\mathbf{s} \in \mathbf{SL1}$ pour orienter le modèle vers l'élaboration des constructions de la LS comme le cas d'un apprenant.

La probabilité d'une nouvelle construction ($s = 0$) :

$$P(0) = (1 - c)/(c \cdot n + (1 - c)) \quad (2.4)$$

Ainsi, la probabilité a priori pour une construction est proportionnelle à la fréquence de ses cadres tandis que la probabilité a priori d'une nouvelle construction est inversement proportionnelle au nombre total des cadres observées.

Ainsi, la probabilité conditionnelle d'un cadre C est la somme (en parcourant toutes les valeurs possibles de \mathbf{i}) des logarithmes des probabilités conditionnelles de ces propriétés :

$$P(C|S) = \sum_i \text{Log}(P_i(v|S)) \quad (2.5)$$

Tel que v est la valeur de la propriété \mathbf{i} du cadre C et $P_i(v|S)$ est la probabilité que la propriété \mathbf{i} ait la valeur v dans la construction S . Cette probabilité est estimée en utilisant la formule du maximum de vraisemblance lissée. La probabilité de chaque propriété est la suivante :

$$P_i(v|s) = c(t_v^s(\mathbf{i}) + \alpha)/(n_s + \alpha \cdot d_i + (1 - c)) \quad (2.6)$$

Tel que n_s est le nombre de cadres de la construction \mathbf{s} et $t_v^s(\mathbf{i})$ est le nombre de cadres de \mathbf{s} qui ont la valeur v pour la propriété \mathbf{i} . Le paramètre de lissage d_i représente le nombre de valeurs que la propriété \mathbf{i} peut avoir. Le paramètre α est une constante ajustée selon les expérimentations. Nous utilisons le principe de lissage additif (Manning, Raghavan et Schütze, 2008; Jurafsky et Martin, 2008), qui est une technique utilisée pour lisser les données catégorielles en ajoutant une valeur constante à chaque occurrence d'événement. Par l'ajout de la constante α aucun événement n'aura une

probabilité nulle. En outre, nous adoptons la même probabilité de couplage utilisée dans l'équation (2.3).

Selon (2.6) et si la propriété i n'existe pas dans la construction s , cette dernière aura une valeur minimum, car $t_v^s(i)$ devient 0. Le modèle crée une nouvelle construction si la propriété i est absente dans toutes les constructions existantes.

Pour une nouvelle construction on aura :

$$P_i(v|S) = c \cdot \alpha / (\alpha \cdot d_i + (1 - c)) \quad (2.7)$$

Prenons l'exemple de la section (2.2.2). La phrase "Adam boit de l'eau" qui est représentée par la paire :

$$\left\{ \begin{array}{l} \text{Scène : BOIRE}_{[\text{acte, absorber}]}(\text{ADAM}_{\langle \text{agent} \rangle \langle \text{animé} \rangle}, \text{EAU}_{\langle \text{thème} \rangle \langle \text{potable} \rangle}) \\ \text{Énoncé : Adam boit de l'eau.} \end{array} \right.$$

Dont la structure générale est:

$$\left\{ \begin{array}{l} \text{Scène} \quad : \quad \text{PREDICAT}_{[\text{Primitives_sémantiques}]}(\text{ARG1}_{\langle \text{Rôle_sémantique} \rangle \langle \text{Traits_sémantiques} \rangle} \\ \quad \quad \quad \text{, ARG2}_{\langle \text{Rôle_sémantique} \rangle \langle \text{Traits_sémantiques} \rangle}) \\ \text{Énoncé} : \text{ARG1 prédicat ARG2.} \end{array} \right.$$

Dans des situations d'apprentissage, y compris la compréhension et la production, le modèle calcule les valeurs des propriétés syntaxiques et sémantiques manquantes dans les cadres, en modifiant les constructions existantes ou en formant de nouvelles constructions. Dans cet exemple, nous considérons la liste des primitives sémantiques suivante : Primitives_sémantiques = {"cause", "act", "move", "possess", "changestate", "become", "perceive", "contact", "manner", "playfully", "consume", "rest", "fall1", "descend", "create", "appreciate", "cogitate", "desire", "demand", "learn", "tell", "affirm", "look", "observe", "attend", "beware", "attempt", "absorb", "produce", "appear"}, dans ce cas :

La propriété i : Primitives_sémantiques.

$t_v^s(i)$: Le nombre de cadres de la construction s qui ont la propriété ‘Primitives_sémantiques’ est égale a ‘absorb’ ou ‘act’.

$d_i = 30$.

α : cette valeur peut être ajustée selon les expérimentations en prenant une valeur très petite selon le nombre d_i de chaque propriété. Par exemple, si le scénario de l’expérimentation computationnelle utilise le nombre de la propriété ‘Primitives_sémantiques’ (30), le nombre de modèles syntaxiques disponibles (20) et le nombre de verbes extraits du corpus anglais (24) et selon la formule

$$\alpha = \prod_i 1/d_i \quad (2.8)$$

α aura la valeur $1/(30 \times 20 \times 24)$ ce qui donne $\alpha \approx 10^{-4}$

Pour la circonstance, soit dans une situation de compréhension ou de production, si la valeur de la ‘Primitives_sémantiques’ est inconnue partiellement ou totalement, le modèle aura besoin de la trouver en calculant la probabilité de chaque valeur possible pour trouver celle qui est la plus adéquate :

$\left\{ \begin{array}{l} \text{Scène : BOIRE}_{[X,Y]}(\text{ADAM}_{\langle \text{agent} \rangle \langle \text{animé} \rangle}, \text{EAU}_{\langle \text{thème} \rangle \langle \text{potable} \rangle}) \\ \text{Énoncé : Adam boit de l'eau.} \end{array} \right.$

$$\text{ValeurPlusAppropriée}(C) = \underset{i}{\operatorname{argmax}} \left(\sum_s \log((t_v^s(i) + \alpha)/(n_c + \alpha d_i)) P(S|C) \right) \quad (2.9)$$

La valeur la plus appropriée de la propriété manquante i est l’argument du maximum (en parcourant toutes les valeurs possibles de i) des sommes (en parcourant toutes les constructions existantes) des logarithmes des produits de la probabilité conditionnelle du cadre C sachant une construction s et la probabilité conditionnelle de cette construction sachant le cadre C . Dans notre cas, les constructions qui ont des cadres qui contiennent

des verbes ou des actants dont les primitives sémantiques ont les valeurs [acte, absorber] auront des probabilités plus élevés, ce qui exprime l'enracinement de l'information chez l'apprenant à force d'utilisation des verbes qui appartiennent à la même catégorie sémantique et syntaxique. Le modèle passe par une courbe d'apprentissage 'U' (Saville-Troike, 2006; Carlucci, Case, 2013) qui commence avec une étape de performance élevée, une étape de surgénéralisation, puis une dernière étape de performance élevée. Trois facteurs principaux déterminent ces étapes de compétence linguistique: la fréquence d'utilisation, le nombre d'intrants, et le concept de la "preuve positive". Pour évaluer cette compétence linguistique (**CL**) d'un apprenant de LS nous testons le modèle en réalisant et observant des simulations sur une tâche de production/compréhension où l'une des valeurs des propriétés sémantiques ou syntaxiques d'un cadre (intrants de test) est supprimée. Par la suite, nous comparons la valeur calculée par le modèle avec la valeur originale. La précision de calcul de la valeur manquante pour la propriété i (**PCi**), est estimée en fonction de la concordance entre la valeur originale et la probabilité/valeur calculée par le modèle. Dans notre exemple, **PCi** où $i =$ 'Primitives_sémantiques' et m est le nombre d'éléments de cet ensemble [X, Y] dans le cadre C, qui est 2 :

$$PCi = \sum_m \text{ValeurPlusAppropriée}(C) / m \quad (2.10)$$

Pour les propriétés élémentaires comme le modèle syntaxique ou le prédicat lui-même le **PCi** est tout simplement :

$$PCi = \text{ValeurPlusAppropriée}(C) \quad (2.11)$$

Dans chaque simulation, la compétence linguistique **CL** est calculée selon les propriétés linguistiques utilisées dans cette dernière, telles que Prédicat, Rôle_Sémantique, Trait_Sémantique, Primitives_sémantiques, ou Modèle_Syntaxique :

$$CL = \sum_n PCi / n \quad (2.12)$$

où n est le nombre de propriétés linguistiques évaluées dans une simulation donnée.

2.4 Comportement du modèle dans le cas d'ALS

Comme nous avons déjà mentionné dans la section 2.1, la LS se distingue de la L1 par l'influence de cette dernière dans l'ALS, positivement ou négativement. Pour illustrer cet effet, nous allons suivre les étapes suivantes :

- Préparer deux lexiques (ensembles d'entrées) pour notre modèle. Le premier est en anglais et le deuxième est en français. Nous allons donner plus de détails sur ces deux lexiques dans le chapitre suivant,
- passer le premier (anglais ou français, dépendamment de la simulation) lexique à notre modèle en lui laissant passer par les différentes étapes d'apprentissage pour avoir un degré de compétence élevé et équivalent à celui de L1,
- passer le deuxième lexique pour simuler l'ALS sachant que les constructions acquises par notre modèle sont explicites et peuvent être examinées dans chaque étape d'apprentissage,
- expérimenter le modèle avec des ensembles d'entrées de taille différentes pour simuler les trois étapes d'apprentissage selon la courbe 'U' et examiner et analyser le comportement du modèle dans chaque étape,
- suivre le comportement du modèle qui va lui permettre de diminuer la "zone grise" entre les deux ensembles de construction de LS et L1 en réalisant des expérimentations avec des éléments tels que :
 - la comparaison du déplacement du verbe entre ces deux langues (Benson, 2002; Pollock, 1989; Chomsky, 1968; Chomsky, 1991). Par exemple, en français, on place, contrairement à l'anglais, les verbes conjugués au présent, avant les adverbes et les négations, comme dans (Pollock, 1989) :
 - Adam voit souvent Ève / Adam ne voit pas Ève,
 - Adam often sees Eve / Adam does not see Eve,
 - les quantificateurs flottants (Tous/Chacun), en français, à l'opposé de l'anglais, sont placés après le verbe conjugué au présent (Ayoun, 2005) :

- Les élèves aiment tous Ève / Les élèves reçoivent chacun un prix,
- The students all love Eve / The students each receive a prize,
- en français, la quantification à distance peut se placer, facultativement, avant ou après le participe passé, tandis qu'en anglais elle est placée constamment après ce dernier (Dekydspotter, L. Rex et Rachel, 2000) :
 - Adam a fait beaucoup de devoirs / Adam a beaucoup fait de devoirs,
 - Adam has done a lot of homework,
- en français, l'adverbe de lieu 'y' est indispensable et doit se placer avant le verbe, alors qu'en anglais l'adverbe 'There' est optionnel et se place après le verbe (Chomsky, 1991) :
 - J'y vais / Voulez-vous y aller ? / Ils y étaient,
 - I'm going (there) / Do you want to go (there)? / They were there.
- la position de l'adjectif dans la phrase. En anglais, les adjectifs précèdent les noms. En français, les adjectifs peuvent être placés avant ou après le nom, en fonction de leur type et leur signification,
- l'élaboration des constructions de « mouvement provoqué ». En anglais, les verbes expriment souvent le mode de déplacement, tandis que la direction est distinguée par un affixe : "She run into". Alors qu'en français, la majorité des verbes de mouvement révèlent la direction, cependant que la manière de déplacement est signalée par un complément tel qu'un gérondif ou une construction adverbiale, ou elle peut ne pas être énoncée : "Il sort en courant,
- le comportement attendu par le modèle est de diminuer la "zone grise" qui existe entre les deux ensembles de construction en arrivant à une étape d'apprentissage plus avancée et en recevant plus d'intrants. Dans le cas de « mouvement provoqué », il doit arriver à produire "Il sort" ou "Il sort en courant" au lieu de "Il sort à l'extérieur". Le modèle comme le cas de l'apprenant, arrive à une étape où il va constater qu'il n'est pas nécessaire d'ajouter la direction dans l'affixe. Il trouvera que le verbe de mouvement de la LS 'Sortir' révèle déjà la direction contrairement

au verbe ‘Go’ de la L1, et élaborera, probablement, une nouvelle construction pour ce cas,

- Le nombre de constructions de L1 parcourues par le modèle sera diminué dans chaque étape d’apprentissage. Par exemple, dans les équations (2.1) ou (2.9), et dans une étape avancée d’apprentissage, les constructions SLS seront utilisées fréquemment pour catégoriser les nouveaux cadres, contrairement aux premières étapes d’apprentissage où les constructions SL1 vont attirer plus de cadres,
- les paramètres des équations mentionnés dans la section précédente seront ajustés dans les chapitres suivants selon les simulations de chaque élément d’apprentissage.

2.5 Conclusion

Nous avons présenté la méthodologie adoptée pour simuler l’ALS avec un modèle bayésien non supervisé qui utilise le principe de catégorisation humaine proposé par Anderson (1991) et qui a été utilisé dans d’autres études (Perek, 2015 ; Alishahi et Stevenson, 2010) et LS (Matusевич et al., 2016). Notre modèle catégorise les cadres de LS en constructions. Les intrants du modèle ne sont pas prédéfinis, mais classés selon la similarité des cadres observés.

Les simulations probabilistes seront utilisées pour montrer l’effet de l’usage la langue elle-même dans l’acquisition des verbes de LS. En outre, pour confirmer le passage de cette acquisition par une courbe d’apprentissage ‘U’ (Saville-Troike, 2006; Carlucci, Case, 2013) qui commence avec une étape de performance élevée, la où l’apprenant produit/répète des phrases correctes avec le peu d’entrées qu’il a reçues. Par la suite, une étape de surgénéralisation, où l’interlangue (Corder, 1975 ; Richards, 1972 ; Selinker, Swain, et Dumas, 1975) est dans sa phase préliminaire. Dans cette phase, la zone entre les constructions de la L1 et la LS n’est pas encore bien déterminée. Une grande zone grise entre les deux ensembles de constructions. Cette zone sera diminuée avec le temps en recevant plus d’intrants de la LS. La zone grise existe est exprimée par un chevauchement des deux ensemble de constructions. Ce chevauchement est illustré par

l'utilisation de certains éléments des constructions de la L1 dans l'élaboration des constructions de la LS ou dans la production ou la compréhension de cette dernière. Un exemple plus tangible et fréquent est l'utilisation d'un modèle syntaxique de la L1 pour l'élaboration d'une construction de la LS, ou la production d'une phrase de la LS selon le principe du Genre (masculin/féminin) de la L1. Ultérieurement, l'apprenant arrive à une étape finale de la performance élevée, grâce aux nombres d'intrants, la fréquence d'utilisation et le principe de la "preuve positive". Dans cette étape la zone grise est diminuée à un point qui montre le degré de la compétence linguistique (CL) atteint par l'apprenant.

Selon son âge, l'apprenant distingue consciemment ou inconsciemment les constructions de la L1 de celle de la LS. Un apprenant adulte ou un enfant de plus de 4 ans sait qu'il s'agit d'une ASL malgré les chevauchements qui peuvent exister entre la L1 et la LS. Un enfant de moins de 4 ans, élabore les deux ensembles de constructions parallèlement ce qui nous amène beaucoup plus vers un enfant parfaitement bilingue ou un enfant avec deux L1 au lieu de parler d'une L1 et d'une LS.

Dans nos simulations probabilistes comme dans l'ALS chez un apprenant, avec chaque entrée ou utilisation d'un nouveau verbe, l'apprenant mémorise les caractéristiques sémantiques et syntaxiques de ce verbe sous forme de constructions qui est en fait une catégorisation des cadres en fonction de leurs similarités. Cette mémorisation des connaissances linguistiques chez les adultes est sous forme de constructions de différents niveaux de complexité et d'abstraction. Les constructions peuvent avoir des éléments concrets et particuliers comme des mots et des idiomes, de même, des classes d'objets plus abstraites ou des constructions mixtes de parties concrètes et abstraites de la langue. Ces constructions peuvent être simultanément représentées et stockées sous des formes différentes, et à différents niveaux d'abstraction. Par exemple, élément concret : tables ou élément abstrait ou règle : [Nom] + (morphème + s) = pluriel (Ellis et Cadierno, 2009). Dans la catégorisation des cadres, le modèle calcule des probabilités qui permettent d'obtenir les valeurs les plus appropriées aux propriétés sémantiques et

syntaxiques manquantes dans le cas de la production, la compréhension ou l'élaboration de nouvelles constructions.

Dans les trois chapitres suivants nous allons montrer et expliquer l'application de notre modèle à l'acquisition des structures actanciennes des verbes dans le cas de LS, l'acquisition des rôles sémantiques des verbes de LS et l'acquisition des préférences de sélection des verbes de LS. Dans chaque chapitre, nous allons présenter les intrants du modèle, ainsi que la structure du cadre utilisé pour modéliser la tâche d'apprentissage. En outre, chaque chapitre présente des expériences computationnelles qui démontrent la correspondance entre le comportement du modèle et celui des apprenants de LS dans l'apprentissage des régularités linguistiques, le passage par les trois étapes de la courbe d'apprentissage 'U', et dans l'utilisation de ces connaissances dans des différentes tâches linguistiques.

CHAPITRE III

SIMULATIONS PROBABILISTES D'ACQUISITION DES STRUCTURES ACTANCIELLES DE VERBES DE LS

3.1 Introduction

Les structures actanciennes sont au centre de toute théorie de la grammaire. En raison de la relation étroite entre leurs caractéristiques sémantiques et syntaxiques, les structures actanciennes des verbes d'une LS sont considérées comme une approche théorique très innovante pour les études intéressées par l'évaluation des connaissances des apprenants de LS. L'analyse des structures actanciennes des verbes d'une LS surpasse l'analyse de la position de l'actant pour déterminer la fonction, le contenu et le sens (Souza, 2011). La compréhension des structures actanciennes permet d'avoir certains nombres de caractères des événements et des états interprétés par une expression linguistique (Jackendoff, 2002). Par ailleurs, le verbe est l'élément lexical ou le noyau qui supporte la phrase et autour duquel s'articulent les actants. Ainsi, l'analyse des structures actanciennes d'une LS est une analyse de la façon dont les verbes se comporteront dans cette langue lorsqu'ils expriment des événements et des états par la structuration des propositions. Cette expression des événements et des états est considérée comme la transition ou le lien entre les représentations mentales des concepts et les structures morphosyntaxiques. Elle représente l'interface entre la sémantique et la syntaxe. Par le fait, en analysant les structures actanciennes, nous devons distinguer les caractéristiques sémantiques et syntaxiques des actants et des phrases. Pour la circonstance d'une phrase et selon la situation de l'"être" qui est éventuellement en action ou en repos, nous pouvons distinguer les cas suivants :

Il fait l'action	Il reçoit l'action	Il exerce l'action sur lui même	Repos
Verbe actif	Verbe passif	Verbe pronominal réfléchi	Verbe neutre

Tableau 3.1: Les différentes catégories des verbes.

Dépendamment de chaque cas, le verbe peut être ‘transitif’, ‘intransitif’ ou ‘ditransitif’, ce qui détermine le nombre d’actants que ce dernier va accepter (Ladevie-Roche, 1853).

Harley (2006) a utilisé une approche qui permet d’expliquer les théories des structures actanciennes. Il met en évidence les notions de la représentation interne, la signification et le concept des mots simples et des phrases. Il distingue deux catégories de concepts, ceux qui peuvent exister dans un énoncé indépendamment des autres concepts, et les autres qui exigent une relation avec d’autres concepts. Lorsqu’un énoncé comme “Pingouin” ou “Ceci est un pingouin” se produit, le concept ou la représentation interne de ce dernier peut exister indépendamment. Cependant, un énoncé à l’instar de “lancer” qui représente une action, doit avoir un aspect relationnel ou participatif. Il doit avoir un nombre d’arguments (actants) qui portent certaines caractéristiques syntaxiques et sémantiques et qui peuvent répondre à « qui lance quoi ? ». Dans le modèle syntaxique « X lance Y », X doit avoir des caractéristiques sémantiques qui permettent à cet actant de réaliser l’action de “lancer”. Entre les deux concepts suivants :

- Concept1 : Volitionnel, sensible, animé, indépendant.
- Concept2 : Affecté, indépendant, comestible.

Le concept1 est mieux approprié pour l’actant X puisqu’il répond aux caractéristiques sémantiques de “lancer” qui sont : jeter, ficher, envoyer, émettre, flanquer, lâcher, etc.

De ce fait, un concept peut ainsi être décrit comme une représentation sémantique qui peut conduire à l’interprétation des relations, des états et des événements. La sélection des structures actanciennes d’un verbe est déterminée par ses traits sémantiques. Les théories de la sémantique lexicale prétendent que les verbes spécifient les rôles sémantiques des actants possibles. Ces rôles sémantiques ou rôles thématiques

(Jackendoff, 1972) tels que «agent», «patient» ou «thème», «expérience», «instrument» et «bénéficiaire» montrent les propriétés sémantiques indispensables dans la sélection des structures actanciennes. Ces arguments sémantiques doivent correspondre aux fonctions syntaxiques. Dans l'exemple suivant qui montre la polysémie de la fonction de sujet :

{ Le joueur lance la balle.
 { La balle a été lancée.
 { La machine lance la balle.

Dans les trois phrases, les sujets ont des sens différents, le joueur accomplit intentionnellement une action, la balle est le sujet de l'exécution de l'action et la machine est un élément qui aide à réaliser l'action. Relativement aux rôles sémantiques, il s'agit particulièrement d'« agent », de « patient » et d'« instrument ». Ainsi, un rôle sémantique peut occuper différentes fonctions syntaxiques.

En ce qui concerne structures actanciennes, nous avons deux théories principales qui sont les théories projectionnistes et constructivistes (Levin et Rappaport-Hovav, 2005). Dans les approches projectionnistes (Levin, 2004), l'acquisition des structures actanciennes est estimée comme un encodage distinctif des caractéristiques sémantiques des verbes. Ces approches, privilégiant le lexique, ont alors accordé beaucoup d'importance au statut des verbes comme responsables des structures syntaxiques de leurs phrases. En conséquence, les verbes qui ont des structures actanciennes similaires partagent soit la même représentation sémantique, ou la même structure d'événement. Ainsi, pour la structure d'événement [[X ACT] CAUSE [BECOME [Y <STATE>]]], pour les verbes qui expriment un état ou un changement d'état comme 'warm' ou 'dry'. La structure d'événements de ces verbes est une dimension sémantique perceptible à la syntaxe. Autrement dit, cette dimension sémantique détermine les composants syntaxiques X et Y qui participeront à la phrase. Par ailleurs, les autres dimensions sémantiques d'une catégorie de verbes peuvent constituer des représentations sémantiques qui discernent chaque verbe sous la même catégorie.

D'autre part, dans le cas des approches constructives (Goldberg, 2006), le lexique et la syntaxe sont indissociables. Si un modèle ou des éléments syntaxiques appartiennent à une construction, cette dernière doit forcément avoir des propriétés sémantiques qui correspondent le mieux à ces composants. Les structures actanciennes sont considérées comme des constructions, en conséquence, une interaction entre le modèle et les composants syntaxiques d'un côté et les caractéristiques sémantiques d'un autre côté est indispensable pour l'élaboration de la structure actancielle. Cela confirme que l'élaboration de cette dernière ne peut pas se limiter seulement au sens du verbe. Pour un apprenant d'anglais LS, en entendant, pour la première fois, le verbe "devour" dans la phrase « They devour spaghetti with chopsticks » il pourra facilement déduire le sens du verbe "devour" à partir de la construction (caractéristiques sémantiques et syntaxiques) de la phrase. L'apprenant peut, naturellement, déduire les traits en commun entre le nouveau verbe "devour" et le verbe "eat" qu'il connaît déjà et qui existe dans une construction préalablement établie (Goldberg, 2006). En effet, la principale différence entre l'approche projectionniste et l'approche constructive tient au fait que des arguments ou des rôles sémantiques qui peuvent être liés au verbe "devour" tels que "agent", "patient" et "manner" ne sont pas préprogrammés dans une entrée lexicale, mais plutôt ils peuvent être déduits du contexte général et du sens de la construction. La construction, son sens et ces caractéristiques sont tous établis à partir des intrants de la langue elle-même et qui permettent par la suite la généralisation de cette information pour trouver des composants manquants ou créer des nouvelles instances ou des nouvelles phrases. Les expérimentations sur des données d'acquisition de L1 et LS restent un domaine d'étude important avec l'objectif de révéler les mécanismes d'acquisition de la langue grâce à des modèles computationnels.

Dans la section suivante, nous allons montrer les théories des structures actanciennes dans le cas d'ALS. Dans la section 3.3 nous présentons quelques modèles computationnels existants et qui adoptent les différentes théories liées aux structures actanciennes. Nous discutons l'utilisation de notre modèle bayésien proposé dans le chapitre II dans l'acquisition des structures actanciennes dans la section 3.4. Nous présentons les corpus,

les outils et les méthodes d'extraction de données dans la section 3.5. La section 3.6 est une illustration des expérimentations fondées sur des intrants préalablement préparés. Dans la section 3.7 nous allons discuter les résultats et les différentes étapes d'acquisition des structures actanciennes. Finalement, nous discuterons les différents points révélés par l'application de notre modèle computationnel sur l'acquisition des structures actanciennes de la LS dans la section 3.8.

3.2 Théories d'acquisition des structures actanciennes d'une LS

La différence entre l'AL1 et l'ALS a fait émerger des grands débats théoriques dont la résolution est actuellement partielle. Les défis d'ALS sont considérés plus complexes que ceux d'AL1. Certaines études ont illustré les raisons pour lesquelles les enfants développent leur L1 avec un grand succès (Werker et Tees, 1984). Dans le cas d'AL1, les enfants développent en parallèle leur système de conceptualisation avec leur L1. Toutefois, le fait que les apprenants d'une LS connaissent déjà une L1 et possèdent un système de conceptualisation mature qui a été conçu au travers la L1, cela peut influencer positivement ou négativement leur LS et plus spécifiquement l'acquisition des structures actanciennes de cette dernière (Ellis, 2006b). Dans cette optique (Ellis, 2014) a montré à travers la théorie "Usage-based" que les apprenants de la langue seconde peuvent accueillir leur LS selon les principes d'analyse universelle. Cette théorie montre la façon avec laquelle les apprenants de la LS traitent les intrants LS. Elle illustre la nature des connexions forme-sens et les cartographies des structures syntaxiques établies par les apprenants. Les principes d'analyse universelle permettent aux apprenants de LS d'analyser et de transformer l'information offerte par les intrants LS en représentation interne sous forme d'éléments syntaxiques et sémantiques. En outre, ces principes aident les apprenants à généraliser les composants syntaxiques et sémantiques extraits à partir des intrants pour produire de nouvelles phrases (Benati, 2013). D'autre part, les théories de compétition (MacWhinney, 1987, 2015) mentionnent que les apprenants adoptent leurs stratégies d'AL1 pour l'ALS. Cependant, aucune de

ces théories ne comble tout le domaine des questions liées à l'ALS. De même, la plupart de ces théories n'ont pas été appuyées par des modèles computationnels qui peuvent confirmer leurs hypothèses. Selon les théories de compétition, les multiples façons d'accord des actants sont constituées en associations et pas indépendamment du contexte et de chaque actant (Kail, 1991). D'autres théories (Ellis, 2006a, 2013) ont proposé un modèle associatif-cognitif dont le principe est que l'ALS est régi par les mêmes processus cognitifs d'apprentissage associatif qui soutiennent les autres types du savoir humain. L'apprentissage associatif est fondé sur l'association de deux stimulus qui se produisent simultanément. Un stimulus peut être inconditionnel, neutre ou conditionnel. Un stimulus inconditionnel provoque une réponse inconditionnelle (réponse par réflexe et qui ne nécessite aucun apprentissage) tandis qu'un stimulus neutre ne provoque aucune réponse ou ne déclenche pas la réponse prévue. En associant un stimulus conditionnel à un stimulus neutre (conditionnement), ce dernier devient un stimulus conditionnel et provoque une réponse conditionnelle (Pavlov, 1927). La théorie (Ellis 2006a, 2013) confirme que l'ALS est fondé sur deux composants fondamentaux qui sont la grammaire de construction (Goldberg, 1995) et l'apprentissage à partir des intrants/exemples. L'acquisition de constructions permet d'associer la forme à la fonction linguistique et de créer une représentation interne de cette liaison. Le modèle associatif-cognitif (Ellis 2006a) a montré qu'une fois la L1 est enracinée l'ALS devient limitée, cependant, cette limitation peut être réduite par le renforcement de l'interlangue par plus d'exemples de LS. Dans un autre ordre d'idées, les théories nativistes de la L1 et la LS croient que l'être humain est équipé de certaines connaissances innées de la langue et que le nombre d'intrants et la fréquence avec laquelle l'apprenant les reçoit ne jouent pas un rôle primordial dans l'acquisition de la L1 et celle de l'ALS (Chomsky, 1965, 1981, Pinker, 1989). Chomsky (1965), affirme que l'être humain possède déjà des structures abstraites de la langue dans sa capacité de langage inné et, par conséquent, ne les apprend pas au travers des intrants. De même, il confirme que ces structures abstraites de la langue et certains éléments communs entre les langues sont codés chez l'être humain sous la forme d'une grammaire universelle (GU). Les théories nativistes

mentionnent que le nombre d'intrants reçu par l'apprenant est tellement pauvre qui ne lui permet jamais de construire des principes de la langue ou des structures syntaxiques abstraites. Par ailleurs Pinker (1984), affirme que dans le cas d'acquisition des structures actanciennes, les rôles sémantiques des verbes sont des propriétés dans des représentations sémantiques des sens des verbes. La position de ces propriétés dans la représentation sémantique change le sens du verbe, et la structure abstraite de cette dernière est auparavant codée dans le cerveau humain. Ceci implique que les règles innées permettent à l'apprenant d'établir des associations correctes entre les éléments syntaxiques et sémantiques d'une structure actancielle à partir du sens de verbe. De ce fait, les théories nativistes affirment que l'acquisition d'une langue est une adaptation des règles innées et des structures abstraites aux intrants reçus par l'apprenant. Cependant, les théories fondées sur le principe de l'usage de la langue « usage-based », et en se basant sur des études expérimentales et des modèles computationnels, confirment que le nombre et la fréquence des intrants et d'exemples d'une langue donnée, reçus par un apprenant, jouent un rôle principal dans l'acquisition de cette dernière (Akhtar, 1999; Tomasello, 2000). L'évolution de l'apprenant avec le temps et à travers l'observation de nouveaux exemples confirme que les structures actanciennes sont apprises grâce à la fréquence et au nombre d'intrants. La capacité de l'apprenant de pouvoir généraliser de nouvelles constructions avec le temps montre que cette évolution n'est pas un simple ajustement des intrants par rapport aux règles innées, mais toute une élaboration de nouvelles structures abstraites et de nouveaux modèles. Les théories « usage-based » affirment que ce ne sont pas les règles linguistiques et les structures abstraites de la langue qui sont innées et codées dans le cerveau, mais plutôt la capacité de création des règles, modèles et structures à partir d'exemples (Ellis 2013; MacWhinney, 1987). D'autres théories affirment que l'ALS passe par une courbe d'apprentissage 'U' (Saville-Troike, 2006; Carlucci et Case, 2013) qui commence avec une étape de performance élevée, par la suite une étape de surgénéralisation des règles (les compétences descendent vers une position inférieure sur l'axe des coordonnées), et finalement une troisième étape où la compétence monte encore une fois à une position plus élevée,

grâce à la fréquence d'utilisation, le nombre d'intrants et le principe de la "preuve positive", où l'apprenant commence à recevoir plus d'intrants qui corrigent ses surgénéralisations (MacWhinney, 2004, Bowerman, 1988). Selon Long (1996), il existe deux types d'intrants dans l'ALS, la preuve positive et la preuve négative. La preuve positive est tous les intrants que les apprenants reçoivent dans leur environnement y compris les conversations, les énoncés, l'école (Chaudron, 1988), le média, etc. En revanche, les preuves négatives orientent les apprenants en montrant ce qui n'est pas correct dans la langue (Long, 1996, White, 1990). Par ailleurs, certaines théories (Tomasello, 2003) indiquent que l'acquisition des structures actanciennes a un noyau central qui est le verbe, autour duquel se forment les différentes propriétés qui composent les constructions. Toutefois, certaines autres théories (Fillmore et al., 1988, Langacker, 2000) fondées sur le principe de la grammaire de construction confirment que l'acquisition des structures actanciennes est le résultat de l'élaboration des associations entre la forme et le sens en composant des constructions abstraites qui associent une forme syntaxique au sens, indépendamment des verbes. Ces théories affirment que l'acquisition de ces constructions (associations) est grâce aux processus de catégorisation et de généralisation adoptés par les apprenants (Goldberg, 1995, Bencini et Goldberg 2000). Cette capacité de catégorisation et de généralisation chez les apprenants, et qui permet la compréhension des nouveaux énoncés ou la production de nouvelles phrases, s'améliore avec le temps par l'acquisition de nouveaux exemples. Les théories fondées sur l'usage en général, et celles fondées sur les constructions des structures actanciennes en particulier, ont récemment reçu beaucoup d'attention dans la communauté d'ALS. Dans la section suivante, nous allons présenter quelques modèles computationnels fondés sur les différentes théories d'ALS.

3.3 Modèles computationnels d'acquisition des structures actanciennes d'une LS

L'apprentissage des structures actanciennes des verbes est un défi pour les apprenants de LS. Cet apprentissage a été représenté par plusieurs modèles computationnels.

Cependant, très peu de modèles computationnels ont simulé l'ALS et plus spécifiquement l'acquisition des structures actanciennes d'une LS. En effet, dans un premier temps, Rappoport et Sheinman (2005) ont élaboré un modèle computationnel qui construit des structures actanciennes et génère des nouvelles phrases, pour une langue seconde, à partir d'un ensemble limité d'exemples et selon le principe de catégorisation de concepts. Ce modèle n'utilise aucune règle explicite et il est capable d'acquérir des nouvelles connaissances liées aux structures actanciennes et les généraliser en produisant des nouvelles phrases. Il simule, d'une façon computationnelle, les processus d'apprentissage. En outre, il intègre un système conceptuel dans un algorithme d'apprentissage. Le modèle était conçu pour des élèves japonais d'anglais LS et il est utilisé comme outil d'aide à l'apprentissage pour ces derniers. Le modèle est évalué en assistant ces élèves dans leur apprentissage d'anglais LS. Les chercheurs évoquent la notion d'interlangue ou de la langue intermédiaire qui est, en linguistique, une phase de transition dans l'apprentissage ou la production d'une langue cible. Celle-ci peut se produire dans le cas de l'ALS. Le rôle des processus d'ALS est celui de l'amélioration progressive et le raffinement de l'interlangue en minimisant l'écart entre les formes de l'interlangue et celles de la LS. Le modèle a considéré principalement que les apprenants adultes d'une langue seconde ont déjà un système de conceptualisation mature et qu'ils sont capables d'établir un raisonnement symbolique explicite. En outre, ce dernier estime que le temps disponible pour l'AL1 est largement supérieur à celui d'ALS. La théorie linguistique sur laquelle ce travail est fondé est celle de la grammaire de construction (Goldberg, 1995), dans laquelle, une construction C, constituée d'un couple forme-sens $\langle F_i, S_i \rangle$ où F_i et S_i ne sont pas forcément prédictible à partir des composants de C ou d'autres constructions précédemment réalisées. Ce travail est proche de celui de (Tomasello, 2003) qui présente une théorie de la grammaire de construction pour la langue maternelle selon laquelle les enfants apprennent les constructions comme des «îles» qui sont progressivement généralisées et fusionnées. Le but de ce travail était d'établir un modèle computationnel d'ALS et qui couvre certains aspects liés à ce phénomène cognitif, tels que l'apprentissage à partir d'un ensemble très limité

d'exemples, l'influence d'un système conceptuel mature, la simulation grâce aux algorithmes d'apprentissages et, enfin, l'utilisation de ce modèle comme un outil d'aide à l'apprentissage pour les élèves d'anglais LS. Les principaux composants de ce modèle computationnel sont le modèle de l'apprenant, le modèle de connaissances linguistiques et les algorithmes d'apprentissage. Le modèle de l'apprenant prend en considération l'existence d'un système conceptuel mature, il n'accepte que les concepts qui existent avant le début de la phase d'apprentissage. En conséquence, l'algorithme d'apprentissage n'accepte pas les éléments liés à la langue seconde, tels que le genre et le nombre des noms. À l'opposé, les élèves peuvent établir un raisonnement sur les relations entre les concepts. Toutefois, le modèle conceptuel est statique durant les phases d'apprentissage et n'évolue pas en parallèle, autrement dit, pas d'apprentissage dans ce contexte. En outre, les chercheurs supposent que l'apprenant possède une mémoire parfaite et peut invoquer les algorithmes d'apprentissage sans erreur. Le modèle de connaissances linguistiques est fondé sur la grammaire de construction, il utilise un ensemble de constructions et de relations. La grammaire de construction est adoptée pour représenter l'interlangue d'une LS. Elle permet de représenter des connaissances comme les formes syntaxiques des mots concrets, les phrases et les constructions. La capacité générative de la langue permet de remplacer les actants par des constructions. De ce fait, le modèle peut remplacer (the child goes to school) par <x> <v> <z> ou <X> <v> <Z> où les caractères minuscules représentent des simples mots et les majuscules représentent des constructions. En outre, les algorithmes d'apprentissage attribuent un degré de certitude (DDC) à chaque construction où cette dernière peut être un actant d'une autre construction. Le principe de ces algorithmes est d'identifier la similarité entre deux ou plusieurs exemples pour pouvoir généraliser des nouvelles connaissances liées aux actants et aux constructions sur d'autres verbes en utilisant la programmation logique inductive. Le modèle final est un système d'apprentissage par induction fondé sur la catégorisation structurée qui est une force motrice dans la perception et les processus cognitifs (Medin et Ross, 2005). Toutefois, le choix d'un modèle conceptuel et des connaissances L1 statiques, l'absence de propriétés

sémantiques dans les intrants de la LS et la représentation de la L1 seulement avec des mots et des relations ne permettent pas au modèle computationnel d'avoir une cohésion entre les données L1 et LS. Ce manque de cohésion crée un écart avec le monde réel et rend difficile la mise en place de métriques qui permettent de suivre l'évolution de l'apprenant telles que le nombre d'intrants LS par rapport au L1, la fréquence des verbes plus utilisés dans la LS, ou l'âge de début d'apprentissage.

Dans un deuxième temps, nous avons le modèle computationnel d'acquisition des structures actanciennes d'une LS (Matusевич et al., 2016). Ce modèle simule l'acquisition des structures actanciennes chez des apprenants allemands qui apprennent l'anglais LS. Ce dernier adopte le modèle (Alishahi et Stevenson, 2008) qui détermine les mécanismes appropriés d'apprentissage qui peuvent saisir des règles générales à partir des intrants, des exemples et d'usages spécifiques. Le modèle analyse l'influence de certains paramètres tels que l'âge de début d'ALS, le nombre d'intrants et la distribution de fréquence (présence des verbes très fréquents comme "aller" et "mettre" dans le corpus) dans l'ALS. Ce dernier, a montré que l'augmentation du nombre d'intrants LS par rapport au L1 favorise le développement de la LS. De même, il a montré qu'une distribution équilibrée de la fréquence des verbes avantage l'ALS, toutefois, les simulations n'ont pas montré un effet positif similaire pour l'âge de début d'ALS. Cependant, certains éléments peuvent être perfectionnés au niveau du modèle (Matusевич et al., 2016), pour avoir des résultats plus concluants :

- Le modèle (Matusевич et al., 2016) adapte celui de (Alishahi et Stevenson, 2008) qui a été orienté principalement vers la simulation d'acquisition de la L1 sans prendre en considération l'aspect d'ALS. Dans le cas où le modèle élabore et utilise deux ensembles de constructions (SL1 et SLS) la probabilité de couplage, pour ce dernier, est la même pour SL1 et SLS ce qui ne favorise pas la catégorisation d'un nouveau cadre de la LS dans SLS,
- le modèle (Matusевич et al., 2016) a utilisé un petit ensemble de 6 verbes allemand et 6 verbes anglais. La taille limitée de l'ensemble de verbes rend

difficile de conclure avec une grande certitude l'influence de certaines variables telles que la fréquence des verbes et des actants ou le nombre d'intrants dans l'ALS,

- le modèle a utilisé les corpus SALSA (Burchardt et al., 2006) et PropBank qui contiennent les sous-corpus TIGER (Brants et al., 2004) à partir duquel les cadres en allemand ont été extraits et Treebank (Palmer, Gildea, & Kingsbury, 2005) pour les cadres en anglais. Comme les auteurs mentionnent, les deux corpus ne contiennent pas le type de langage que les enfants ou les apprenants de LS reçoivent, toutefois, les corpus ont été adoptés seulement parce qu'ils contiennent les annotations nécessaires. TIGER, par exemple, comprend 40.000 phrases de journaux annotés syntaxiquement. Le type de langage qui est différent de celui des apprenants adultes et enfants moins jeunes de LS, n'avantage pas la simulation d'ALS qui est fondée plus spécifiquement sur des intrants bruyants ce qui permet la comparaison du modèle avec un réel apprentissage.

Dans un troisième temps, nous retrouvons certaines études qui ont établi un lien entre l'acquisition des structures actanciennes d'une LS et l'interlangue telles que (Souza, 2011; Seibert, 2012; Bruhn et Montrul, 1996; Cuervo, 2007). Souza (2011) s'est intéressé au développement des représentations des structures actanciennes d'une LS de l'apprenant en constituant des liens entre les formes syntaxiques et les propriétés sémantiques. Les résultats ont montré que les dissemblances entre le portugais brésilien et l'anglais au niveau des alternances de transitivité des verbes de mode de mouvement affectent l'acquisition des structures actanciennes de l'anglais LS, car elles ne le sont pas représentés dans leurs structures/constructions interlangue. En outre, les résultats ont indiqué que les alternances de mouvement provoqué ont été acquises par les apprenants aux étapes ultérieures du développement de leur interlangue. Cependant, l'étude elle s'est fondée sur l'analyse des résultats des expérimentations sur deux groupes de sujets (apprenants brésiliens qui apprennent l'anglais LS et apprenants anglais qui apprennent le portugais LS) sans adopter un modèle computationnel qui peut valider ces résultats. Enfin, dans un dernier temps, on retrouve beaucoup de modèles qui ont étudié

l'acquisition des structures actanciennes de la L1. Le modèle (Alishahi et Stevenson, 2008) simule l'acquisition des structures actanciennes chez les enfants en déterminant les mécanismes appropriés d'apprentissage qui peuvent saisir des règles générales à partir des exemples et des usages spécifiques. Celui de (Pardo et Nunes, 2005) est un modèle statistique génératif pour l'apprentissage non-supervisé des structures. Le modèle (Chang, 2004) établit un lien entre l'ordre des mots dans une phrase et leurs sens. Tandis que celui de (Dominey et Inui, 2004) s'est intéressé à l'acquisition de la syntaxe dans le cadre de la grammaire de construction avec une validation linguistique en anglais et en japonais, toutefois, il ne peut pas acquérir des connaissances liées aux structures actanciennes et les généraliser sur d'autres verbes.

Ces différents modèles et d'autres confirment que le rapport de modélisation computationnelle de l'acquisition des structures actanciennes de la LS est insuffisant par rapport au celui de la L1. La modélisation de l'ALS permettra de comprendre les processus cognitifs élémentaires d'ALS impliqués dans cette dernière.

3.4 Adaptation du modèle bayésien à l'acquisition des structures actanciennes de LS

Dans cette étude, nous utilisons une version adaptée du modèle bayésien présenté dans le chapitre précédent. Ce modèle élabore des constructions abstraites à partir des intrants qui ont été construits depuis des paires scène/énoncé. Ces constructions sont utilisées ultérieurement dans des situations de production ou de compréhension de la langue en créant des associations probabilistes entre les composants syntaxiques et sémantiques pour calculer la probabilité de l'élément syntaxique ou sémantique le plus adéquat. Le modèle est fondé sur le principe de la grammaire de construction (Goldberg, 1995) et l'apprentissage à partir des intrants/exemples. Le modèle recevra des intrants qui peuvent être en L1 ou en LS et développe et distingue les processus de traitement qui peuvent être en commun ou spécifiques à chaque langue. Ceci est en ligne avec le modèle de compétition unifié, qui affirme que la L1 et LS partagent certains processus cognitifs, tandis que d'autres différences entre les niveaux de compétence L1 et LS

peuvent être liées à d'autres facteurs tels que l'enracinement de la L1 (MacWhinney, 2015). Le modèle de compétition est une théorie d'acquisition de la langue et du traitement des phrases qui affirme que la langue est apprise à travers la compétition entre les processus cognitifs en présence d'un environnement linguistique riche d'exemples. (MacWhinney, 2015). Dans le cas d'acquisition des structures actanciennes d'une LS, nous partons de l'hypothèse que l'apprenant de la LS possède un système de conceptualisation mature qui lui permet de comprendre le sens des mots simples, le contexte général des événements et des énoncés, les rôles joués par les arguments, dans un événement ou une scène, tels que agent, thème et destinataire. En outre, comme nous l'avons déjà mentionné dans le chapitre précédent, chaque verbe qui est représenté par un prédicat est complété par un ensemble de primitives sémantiques tels que cause, mouvement, etc.

Les prédicats représentent des verbes, des adjectifs ou des prépositions. Le modèle représente chaque intrant par un cadre qui est, en fait, une collection de caractéristiques que l'apprenant peut induire de l'énoncé ou de l'événement observé. Ces caractéristiques comprennent le prédicat, ses propriétés sémantiques, sa valence, les arguments et leurs propriétés sémantiques, et le modèle syntaxique. Dans le cas de l'ALS, certaines fonctionnalités sémantiques partagent des primitives ou des valeurs entre la L1 et la LS, tandis que d'autres fonctionnalités lexicales ou syntaxiques prennent des valeurs spécifiques soit à la L1 ou à la LS.

La phrase "Adam boit de l'eau" peut-être représentée par la paire :

$$\left\{ \begin{array}{l} \text{Scène : BOIRE}_{[\text{acte, absorber}]}(\text{ADAM}_{\langle \text{agent} \rangle \langle \text{animé} \rangle}, \text{EAU}_{\langle \text{thème} \rangle \langle \text{potable} \rangle}) \\ \text{Énoncé : Adam boit de l'eau.} \end{array} \right.$$

Généralement, un énoncé d'un prédicat avec deux arguments, peut-être représenté par la paire:

$$\left\{ \begin{array}{l} \text{Scène : PREDICAT}_{[\text{Primitives_sémantiques}]}(\text{ARG1}_{\langle \text{Rôle_sémantique} \rangle \langle \text{Traits_sémantiques} \rangle} \\ \text{, ARG2}_{\langle \text{Rôle_sémantique} \rangle \langle \text{Traits_sémantiques} \rangle}) \\ \text{Énoncé : ARG1 prédicat ARG2.} \end{array} \right.$$

Nous allons utiliser le verbe à l'infinitif pendant l'élaboration des intrants du modèle. Nous intéresserons à la morphologie verbale dans le cas d'ALS dans d'autres perspectives de cette étude.

Nous avons, auparavant, présenté les composants élémentaires d'un cadre (intrans) tels que la structure, la fréquence et les primitives sémantiques dans le tableau 2.1 (section 2.2.2 du chapitre II).

En ce qui concerne les rôles sémantiques, ils décrivent le sens d'un actant par rapport à une action ou un événement exprimé par le verbe au sein d'une phrase. Dans notre modèle nous adoptons les rôles sémantiques définis par (Thomas, 2007) :

- Agent : entité accomplissant volontairement une action,
- Thème : Subit l'action, mais ne change pas son état,
- Patient : siège de l'accomplissement d'une action,
- Stimulus : entité où se manifeste un état physique ou mental, Expérient: entité affectée par l'accomplissement d'une action ou la manifestation d'un état physique ou mental, Instrument: entité servant involontairement à l'accomplissement d'une action, Bénéficiaire : entité recevant un profit ou un dommage du fait d'un procès,
- Destinataire: entité pour laquelle on change la propriété d'un objet, Cause : entité déclenchant l'accomplissement d'une action, Résultat : état final à l'issue d'un procès, Source : état initial avant l'accomplissement d'une action ou d'un mouvement, Destination : état final après l'accomplissement d'un mouvement,
- But : état vers lequel est censée progresser une action ou la raison pour laquelle une action est effectuée,
- Lieu : circonstances spatiales d'une action,
- Temps : circonstances temporelles d'une action,
- Mode : la manière dont une action est menée.

Cette liste des rôles sémantiques n'est pas complète, et plusieurs chercheurs proposent d'autres rôles ou négligent certains. Il n'est pas évident d'établir une liste universelle de rôles sémantiques, car leur définition dépend en effet de la subtilité de l'analyse linguistique. Pour certains chercheurs, les rôles d'agent et d'instrument peuvent être indiqués par un seul rôle. Par exemple, dans "l'aiguille a éclaté le ballon", certains linguistes traitent l'aiguille comme agent et d'autres comme un instrument.

En ce qui concerne les primitives sémantiques du verbe, nous adoptons l'ensemble suivant : « Agir, déplacer, consommer, percevoir, apprécier, jouir, réfléchir, demander, désirer, apprendre, posséder, changer, causer, dire, affirmer, regarder, observer, assister, se méfier, tenter, tomber, absorber, produire, apparaître, assister, devenir, créer, comprendre, discerner, étiqueter, nom, appeler, demander, être, rester, examiner, explorer, jouer ». Certaines primitives sémantiques ont été extraites de (Levin et Rappaport, 2005), tandis que d'autres primitives qui peuvent indiquer les significations plus spécifiques des verbes telles que absorber, jouir et tenter sont extraites de la base de données lexicale *WordNet*. Les ensembles d'intrants d'anglais et du français partagent plusieurs primitives sémantiques, alors que des primitives telles qu'étiqueter, nom, appeler et demander existent seulement dans l'ensemble des intrants du français. Finalement, pour chaque actant, une liste adéquate de propriétés sémantiques est extraite à partir de Wordnet et *Wolf*² (Sagot et Fišer, 2008).

3.5 Extraction de données

3.5.1 Extraction de données pour les intrants en anglais

Pour les intrants en anglais, nous avons choisi deux corpus d'ALS à partir de la base de données *TalkBank*³. Ces corpus de données sont *PAROLE*⁴ (Hilton, 2008) et *BELC*⁵

² <https://gforge.inria.fr/frs/download.php/file/33496/wolf-1.0b4.xml.bz2>

³ <http://talkbank.org/data/local.html>

⁴ <http://talkbank.org/browser/index.php?url=SLABank/PAROLE/>

⁵ <http://talkbank.org/browser/index.php?url=SLABank/BELC/>

(Muñoz, 2006). Le corpus PAROLE (*Parallel Oral Foreign Language*) a été conçu et compilé pour étudier les caractéristiques des différents niveaux de compétence de la LS. PAROLE se compose de productions orales de 68 jeunes apprenants adultes (19-23 ans) de trois langues étrangères qui sont l'anglais, le français et l'italien, ainsi qu'un corpus de productions de 27 locuteurs natifs qui est utilisé comme référence.

Les transcriptions et les enregistrements audio sont les résultats des tâches réalisées par les sujets apprenants et natifs. Ces tâches sont des conversations, des tests, des questionnaires et des résumés vidéo visualisés par les sujets.

Les participants au corpus LS (54 femmes, 14 mâles) sont 33 apprenants d'anglais (24 français-L1, 9 allemand-L1; âge moyen 21); 12 apprenants du français (5 espagnol-L1, 3 chinois-L1, 2 suédois-L1, 1 polonais-L1, 1 Anglais-L1; Âge moyen 23) et 23 apprenants d'italien (tous français-L1; âge moyen 19).

Les participants au corpus L1 (20 femmes et 7 mâles) sont 9 anglais-L1 (âge moyen 21); 8 français-L1 (âge moyen 22) et 10 Italien-L1 (âge moyen 23). Dans cette étude nous nous concentrons seulement sur le corpus anglais de PAROLE.

La deuxième partie de nos intrants en anglais a été extraite à partir du corpus BELC (Barcelona English Language Corpus). BELC est un projet qui examine les effets de l'âge sur l'acquisition de l'anglais LS. Des sujets dont l'âge varie entre 8 et 39 ans et dont la L1 est soit le français ou l'espagnol ont été inclus dans la conception du corpus. Les sujets ont été repartis entre un groupe central de 11 ans, un groupe de 8 ans, un groupe de 14 ans et un dernier groupe de plus de 18 ans. Le nombre approximatif d'heures d'apprentissage d'anglais LS était entre 416 et 826 heures. Des tests et des tâches ont été réalisés à différents intervalles de temps (Temps1 = 200 heures, Temps2 = 416 heures, Temps3 = 726 heures et Temps4 = 826 heures). Les tâches accomplies dans chaque étape sont la composition écrite, le récit oral, des conversations et des jeux de rôle. Cette étude a pris en considération les paramètres suivants :

- AD = âge de début d'apprentissage,
- AT = âge au test,
- N = nombre de sujets,
- OES = exposition scolaire.

3.5.1.1 Extraction des intrants en anglais

À partir de ces deux corpus, nous avons suivi les étapes suivantes pour créer les intrants anglais de notre modèle:

- Nous avons extrait depuis les corpus PAROLE et BELC tous les temps des 24 verbes les plus fréquents.
- Nous avons utilisé la commande *FREQ*⁶ de l'application *CLAN*⁷. pour calculer les fréquences des verbes.
- Nous avons ajouté ces verbes au lexique d'entrée.
- Nous avons associé à chaque verbe sa fréquence totale et un symbole unique pour représenter son prédicat sémantique,
- Nous avons assigné à chaque verbe un ensemble de structures d'arguments possibles et des fréquences associées. Ces informations sont compilées manuellement par l'examen de toutes les utilisations du verbe dans toutes les conversations du corpus,
- Nous avons utilisé la commande *KWAL*⁸ de l'application *CLAN* pour extraire toutes les occurrences du verbe et compiler manuellement ses cadres,
- Nous avons utilisé des outils que nous avons créés précédemment pour extraire automatiquement les primitives sémantiques des verbes et des actants, selon le contexte,

⁶ <http://aphasia.talkbank.org/transcribe/clanglossary.txt>

⁷ <http://talkbank.org/clang/>

⁸ <http://aphasia.talkbank.org/transcribe/clanglossary.txt>

- Pour chaque entrée lexicale, nous avons défini :
 - le nombre de cadres (FIDX),
 - les primitives sémantiques du verbe (FTRS),
 - la fréquence (FREQ),
 - le modèle syntaxique (PATT),
 - les rôles sémantiques de chaque actant (ARGS),
 - les catégories sémantiques de chaque actant (COND)
 - les primitives sémantiques de chaque actant (PROL),
- Nous avons ajouté des prépositions utilisées dans les corpus des conversations au lexique,
- Nous avons sélectionné un nombre de termes nominaux pour chaque catégorie sémantique à partir des mots utilisés dans le corpus,
- Comme nous l'avons précédemment mentionné dans la section 2.2.2 du chapitre II, une transformation automatique des entrées lexicales au schéma XML s'effectuera par le modèle pour une normalisation de la structuration des données et l'élaboration des structures par le modèle.

3.5.1.2 Outil automatique pour l'extraction des primitives sémantiques en anglais

Nous avons créé un outil automatique (C#, .NET) pour l'extraction des primitives sémantiques en anglais depuis *Wordnet*, *VerbNet*⁹ et *WordNet eXtended*¹⁰. Nous avons téléchargé une base de données (SQL serveur) contenant des données (*Wordnet 3.0*, *VerbNet 2.3*, *XWordnet 2.0-1.1*)¹¹ qui sera utilisée par cet outil. Nous avons construit une liste de primitives sémantiques pour chaque mot à partir des Synonymes/Hyponymies de WordNet selon son contexte dans le texte (corpus). Nous avons filtré manuellement les

⁹<http://verbs.colorado.edu/verb-index/>

¹⁰ <http://www.hlt.utdallas.edu/~xwn/about.html>

¹¹ <http://wordnetsqlserver.googlecode.com/svn/trunk/Database/>

Synonymes/Hyponymies (sens 1, sens 2, ...) qui correspondent au contexte du verbe dans le corpus.

La liste suivante montre les 24 verbes les plus fréquents extraits à partir des corpus anglais de PAROLE et BELC.

Verbe	Fréquence	Verbe	Fréquence	Verbe	Fréquence
Go	405	Study	81	Drink	52
Eat	267	Get	79	Prepare	48
See	197	Say	75	Arrive	44
Like	179	Come	70	Listen	38
Play	113	Watch	65	Make	36
Think	105	Try	54	Put	31
Want	97	Give	53	Take	28
Look	93	Fall	53		

3.5.2 Extraction de données pour les intrants en français

Le corpus français de TalkBank est très limité et contient très peu de dialogues et de phrases. C'est pourquoi nous avons choisi le corpus du français parlé de *Kate Beeching*¹² (Beeching, 2013). Celui-ci contient les transcriptions de 95 interviews de différentes longueurs, enregistrées à Paris et en Bretagne avec des participants dont la langue maternelle est le français. Les sujets de discussion comprennent une gamme de fonctions linguistiques différentes telles que le transfert d'informations sur une région, des instructions, des récits, des arguments sur les relations familiales, le racisme, la politique ou la société de l'informatique. Les sujets de conversation ont suscité l'intérêt des locuteurs. Les conférenciers, dont 45 hommes et 50 femmes, étaient âgés de 7 à 88 ans et avec différents niveaux de scolarité. Les détails des entretiens donnent les caractéristiques démographiques, sociologiques et contextuelles des sujets telles que l'identité, la durée en minutes, le sexe, l'âge et le niveau d'éducation pour chaque

¹² <https://www.llas.ac.uk/resources/mb/80>

interview et chaque sujet. Nous avons utilisé ce corpus pour créer manuellement les entrées lexicales en français.

3.5.2.1 Extraction des intrants en français

Nous avons suivi les étapes ci-dessous pour créer nos entrées lexicales en français :

- Nous avons extrait du corpus français (Kate, 2013) tous les temps des 32 verbes les plus fréquents,
- Nous avons développé un outil (C#, .NET) qui calcule automatiquement la fréquence des mots dans le corpus français,
- Nous avons ajouté ces verbes au lexique d'entrée,
- Nous avons associé à chaque verbe sa fréquence totale et un symbole unique pour représenter son prédicat sémantique (par exemple le symbole du prédicat ALLER pour le verbe aller),
- Nous avons assigné à chaque verbe un ensemble de structures d'arguments possibles et des fréquences associées. Ces informations sont compilées manuellement en parcourant toutes les apparitions du verbe dans toutes les conversations du corpus,
- Nous avons extrait, manuellement, toutes les occurrences du verbe et compilé manuellement ses cadres,
- Nous avons utilisé des outils que nous avons créé avec (C#, .NET) pour extraire automatiquement les primitives sémantiques des verbes et des actants, selon le contexte,
- Pour chaque entrée lexicale, nous avons défini :
 - le nombre de cadres (FIDX),
 - les primitives sémantiques du verbe (FTRS),
 - la fréquence (FREQ),
 - le modèle syntaxique (PATT),
 - les rôles sémantiques de chaque actant (ARGS),

- les catégories sémantiques de chaque actant (COND)
- les primitives sémantiques de chaque actant (PROL),
- Nous avons ajouté des prépositions utilisées dans ces conversations au lexique,
- Nous avons sélectionné à partir du corpus un nombre de termes nominaux pour chaque catégorie sémantique.
- Une transformation automatique des entrées lexicales au schéma XML s'effectuera par le modèle pour une normalisation de la structuration des données et l'élaboration des structures par le modèle.

3.5.2.2 Outil automatique pour l'extraction des primitives sémantiques en français

L'outil d'extraction des traits sémantiques des mots en français est fondé sur Wolf. En outre, nous utilisons CRISCO (Gaëlle, 2011) pour extraire plus d'hyperonymes/synonymes et compléter ceux extraits avec Wolf. WordNet est considéré comme plus complet que sa version française (Wolf), qui est encore en construction et qui n'est pas encore terminé. Nous avons téléchargé le fichier Wolf xml et nous l'avons converti en base de données SQL Server pour construire notre outil pour extraire des fonctionnalités sémantiques françaises (nous sélectionnons SF manuellement correspondant au contexte du mot).

La liste suivante montre les 32 verbes les plus fréquents extraits à partir du corpus "français parlé" de (Kate, 2013).

Verbe	Fréquence	Verbe	Fréquence	Verbe	Fréquence
Faire	1294	Partir	202	Manger	74
Aller	693	Aimer	181	Essayer	70
Dire	560	Arriver	183	Revenir	69
Voir	417	Passer	178	Regarder	51
Vouloir	376	Croire	165	Chercher	44
Savoir	335	Donner	159	Visiter	40
Travailler	244	Mettre	154	Boire	35
Venir	237	Parler	152	Jouer	31

Penser	226	Appeler	143	Tomber	24
Trouver	219	Rester	134	Écouter	23
Prendre	203	Apprendre	87		

3.6 Simulations probabilistes

Nous présentons des simulations probabilistes qui montrent la capacité de notre modèle à catégoriser les constructions de la L1 et la LS, à apprendre les structures actanciennes d'une LS et à généraliser des règles ou des modèles. Ces capacités vont permettre au modèle de générer de nouvelles structures. Ces structures sont créées à partir du calcul de la probabilité de la propriété sémantique ou syntaxique la plus appropriée, autrement dit, le calcul de l'argument du maximum des valeurs possibles pour une propriété tel que montré dans la formule 2.9 du chapitre II.

Pour chaque expérimentation computationnelle et chaque scénario d'ALS, le modèle construit et traite graduellement une séquence de cadres. Des paires scène/énoncé sont aléatoirement générées à partir des entrées lexicales montrées dans la section précédente. L'appendice A.2 montre un exemple des paires scène/énoncé générées automatiquement par l'outil de génération d'intrants. Cet outil génère ces paires selon les fréquences des verbes dans le corpus pour déterminer les probabilités de sélection de chaque verbe et propriété. L'outil génère les deux cinquièmes des paires créées pour la LS avec un élément syntaxique ou sémantique manquant pour permettre au modèle de calculer la compétence linguistique LS (CL) selon l'équation 2.12 du chapitre précédent. Les intrants de l'anglais et le français échangent le rôle de L1 et LS conformément aux simulations. Cet outil peut générer des intrants bruyants (manque de primitives sémantiques ou d'éléments syntaxiques) ou non bruyants. De même, nous pouvons le configurer pour qu'il accepte ou élimine les redondances.

Les premières simulations montrent le rôle d'ALS chez les jeunes enfants (moins de 5 ans) qui sont considérés comme des enfants bilingues. Dans ce cas le modèle reçoit simultanément les intrants bruyants du français et d'anglais et construit en parallèle les constructions de chaque langue. Une fois que les constructions sont élaborées par le

modèle, nous testons ce dernier avec cinq cadres dont des éléments manquants qui sont : le modèle syntaxique, les primitives sémantiques du prédicat, le rôle sémantique de l'un des actants, les primitives sémantiques de l'un des actants et les traits sémantiques de l'un des actants. Dans une deuxième étape, le modèle calcule la CL (2.12) selon ces cinq éléments. Lors des autres simulations, on commence par les intrants de la L1, ensuite, et suivant le type d'expérimentation, nous passons au modèle les intrants de la LS, dans soit un apprentissage anticipé ou tardif pour analyser l'influence de l'âge de début d'ALS (AD). Dans certaines autres simulations, le modèle reçoit au début les intrants L1, par la suite il prend en parallèle les intrants L1 et LS pour mesurer le facteur du quotient LS/L1 (Q) dans l'ALS. Nous passons un nombre de cadres différents à chaque étape d'apprentissage (100,200,300,400 et 500) pour calculer la CL et tester les autres paramètres tels que l'âge de début d'apprentissage, le quotient LS/L1 (Q) et l'acquisition des modèles syntaxiques. Les données que nous collectons sont calculées à partir de l'observation du développement des constructions dans dix simulations différentes. Autrement dit, pour chaque expérimentation nous calculons la moyenne de dix simulations différentes avec des ensembles d'intrants différents. Chaque expérimentation permet d'évaluer une étape et un facteur spécifique d'apprentissage.

La figure 3.1 montre les courbes d'apprentissage L1 et LS pour une simulation d'apprentissage pour des enfants bilingues (moins de 5 ans) qui reçoivent le même nombre d'intrants pour la L1 et la LS dès le début, et dans différentes étapes d'apprentissage ($Q_{LS/L1} = 1$; $AD = 0$).

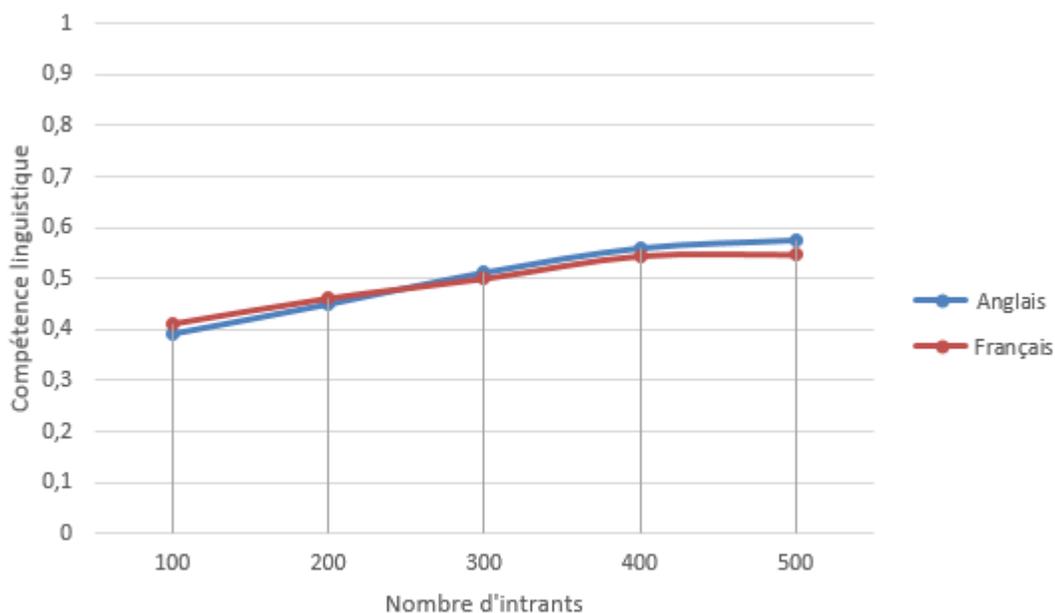


Figure 3.1: Simulation d'ALS chez les jeunes enfants bilingues.

Tandis que les figures 3.2 et 3.3 illustrent les courbes d'apprentissage pour la L1 et la LS pour des apprenants qui commencent l'ALS à un âge relativement tardif et avec un quotient inférieur à celui de la L1 ($Q_{LS/L1} = 1/5$; $AD = 300$ cadres), en échangeant les rôles de la L1 et la LS entre l'anglais et le français. Dans ce cas, au moment où nous présentons le nombre de (300, 400 et 500) intrants pour la L1 nous présentons (60, 80 et 100) respectivement, pour la LS.

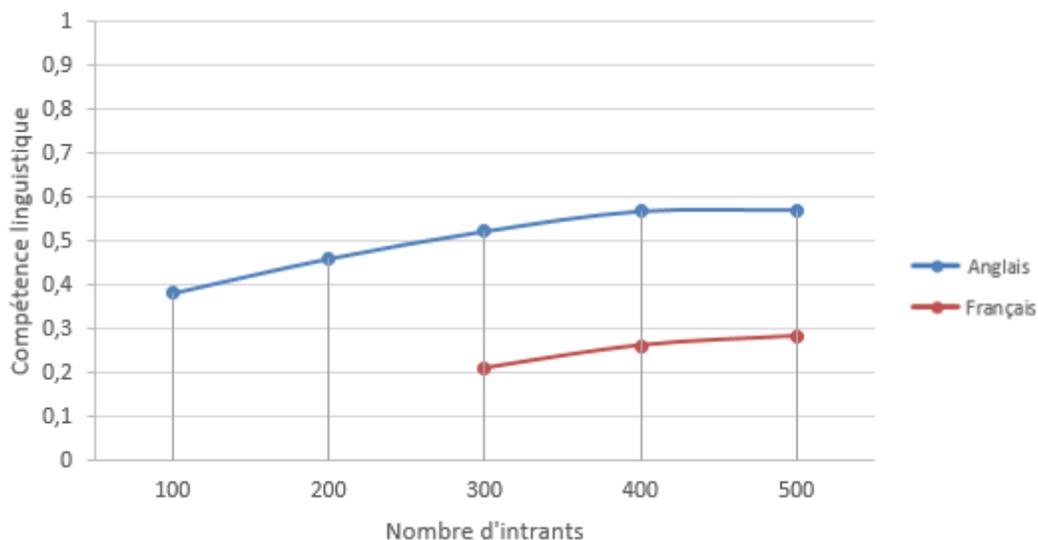


Figure 3.2: Acquisition du français LS dans un âge tardif et avec un faible quotient d'intrants.

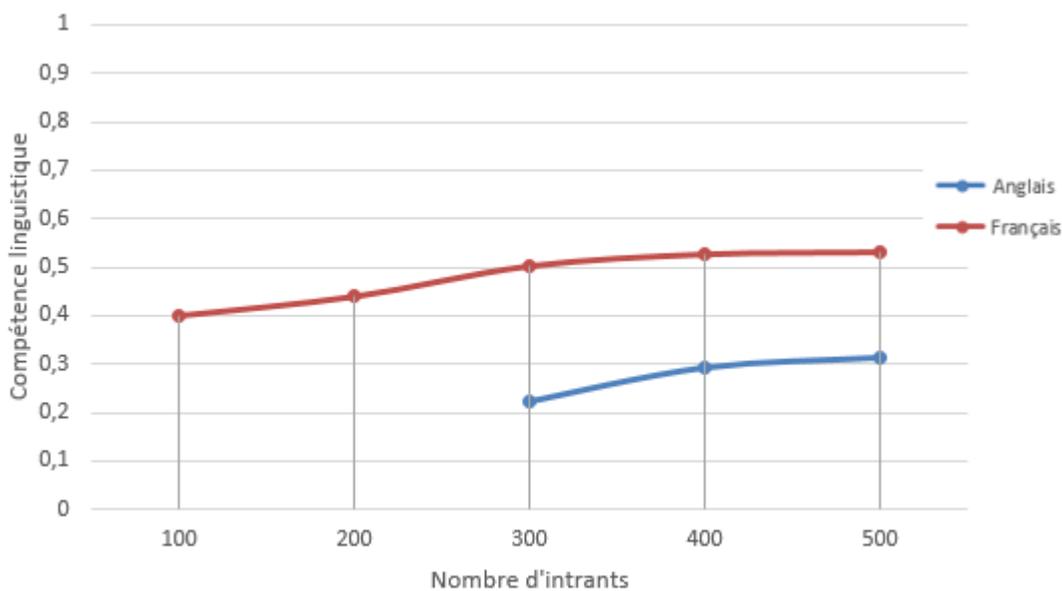


Figure 3.3: Acquisition l'anglais LS dans un âge tardif et avec un faible quotient d'intrants.

La figure 3.1 indique que les apprenants exposés à des intrants anglais et français avec la même quantité et dès le début révèlent un progrès presque équivalent pour les deux

langues, et cela malgré les différences entre l'anglais (langue germanique) et le français (langue romane), précédemment mentionnées dans le premier chapitre.

Cependant, les apprenants qui ont été exposés tardivement aux intrants de la LS (figures 3.2 et 3.3) et avec un faible quotient (1/5) ont montré un progrès inférieur de la LS par rapport à la L1 ($\approx 1/2$).

En considérant seulement le quotient LS/L1, nous avons mené certaines expérimentations ($Q \text{ LS/L1} = 1/10, 1/5, 1/2, 3/4$ et 1) pour étudier l'impact de ce facteur sur la CL. Dans toutes les simulations, nous avons exposé les intrants L1 et LS dès le début et en parallèle. Le nombre de cadres pour la L1 était 500 dans toutes les simulations. L'anglais et le français ont joué les rôles de L1 et LS respectivement.

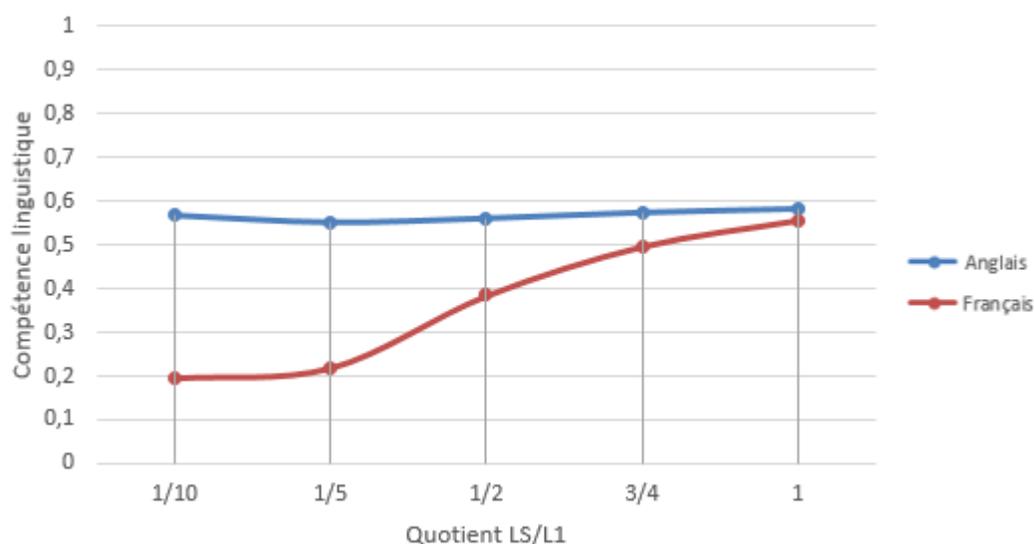


Figure 3.4: Acquisition du français LS avec différents quotients LS/L1.

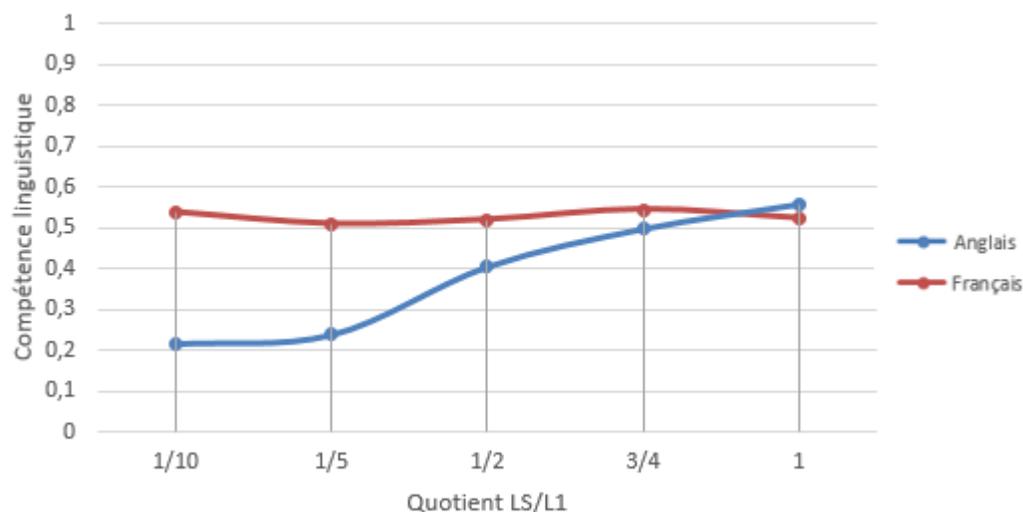


Figure 3.5: Acquisition de l'anglais LS avec différents quotients LS/L1.

Les figures 3.4 et 3.5 montrent que la CL et le Q LS/L1 sont directement proportionnels, où la valeur de la CL de la LS s'accroît de manière significative lorsque Q LS/L1 augmente. Ces résultats confirment ceux du modèle de (Matusévych et al., 2014) et les expérimentations de (Munoz, 2011). (Matusévych et al., 2014) a montré le rôle de volume d'intrants dans des simulations d'ALS pour l'anglais et l'allemand, tandis que (Munoz, 2011) a expérimenté les facteurs du quotient LS/L1 et le contact avec des locuteurs natifs dans l'ALS. Par conséquent, notre modèle confirme le rôle important du volume d'intrants LS dans l'ALS.

L'impact de l'âge de début d'ALS sur la compétence LS a été simulé par l'exposition des apprenants aux données LS aux différentes étapes d'apprentissage (100, 200, 300, 400 et 500 cadres). Les expérimentations ont été effectuées avec un nombre constant de cadres, et un quotient invariable aussi ($Q \text{ LS/L1} = 1$). À chaque point d'apprentissage de la L1, nous fournissons au modèle un nombre d'intrants LS égale aux nombres d'intrants L1 qui a été déjà traité par le modèle jusqu'à cette étape d'apprentissage. À titre d'exemple, nous alimentons le modèle au point d'apprentissage 200 avec 100 cadres d'anglais et 200 cadres du français pour le cas d'acquisition du français LS à l'étape 200, car ce dernier a déjà traité 100 cadres d'anglais L1 à une étape précédente. Cela

permet de séparer le facteur de nombre d'intrants traités par le modèle et se concentrer seulement sur le facteur de l'étape d'apprentissage, car au point 200, le modèle traite exactement le même nombre d'intrants pour l'anglais et le français, mais à deux étapes différentes.

Dans les figures 3.6 et 3.7 la courbe de la L1 est à titre indicatif, car chaque étape d'apprentissage possède sa propre courbe L1 qui est, en fait, proche des autres courbes L1 des différents points d'apprentissage. En effet, les étapes d'apprentissage concernent beaucoup plus la LS, tandis que la L1 commence toujours de la première quantité d'intrants qui est 100 cadres.

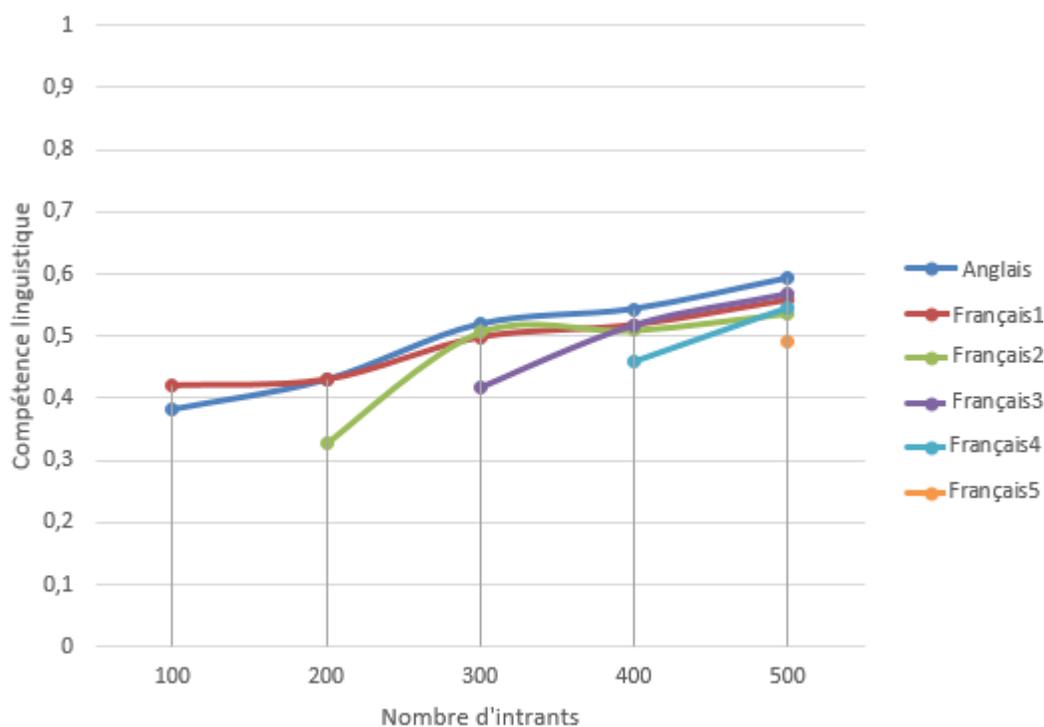


Figure 3.6: Acquisition du français LS aux différentes étapes d'apprentissage.

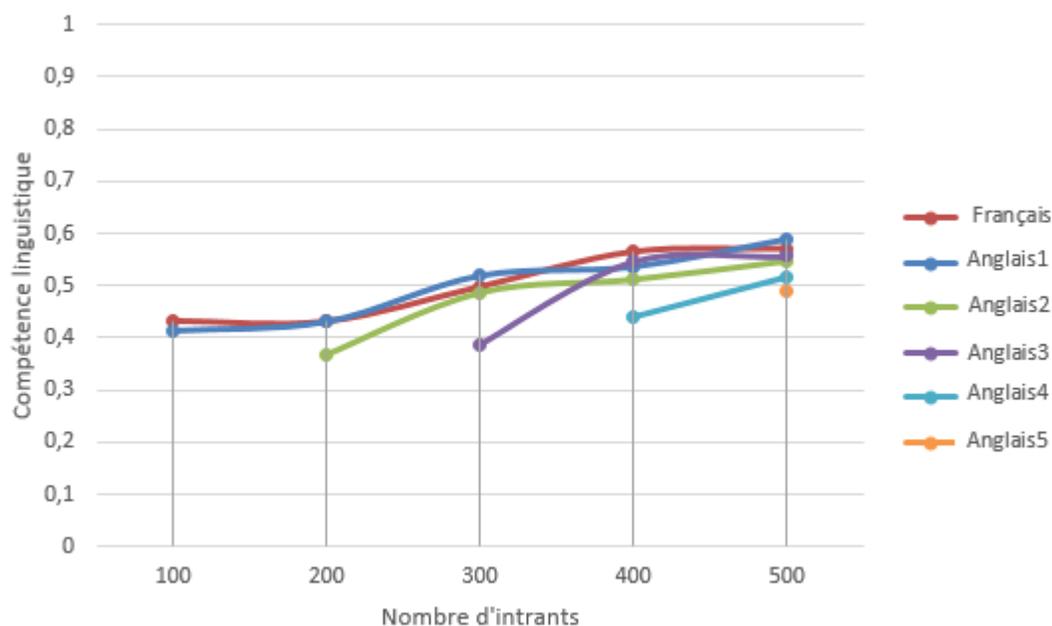


Figure 3.7: Acquisition de l'anglais LS aux différentes étapes d'apprentissage.

Pareillement pour le français LS (figure 3.6) ou l'anglais LS (figure 3.7), les résultats montrent qu'à chaque étape d'apprentissage, l'ALS commence avec une CL proportionnellement faible, par la suite et dans l'étape suivante elle atteint un niveau proche du cas de l'ALS qui commence en parallèle avec la L1 et dès le début. Cela peut s'expliquer par le fait que l'acquisition de la L1 qui commence à une étape précédente élabore un ensemble de constructions qui peut attirer certains cadres de la LS au début de l'étape d'apprentissage, ce qui minimise le nombre de constructions de la LS. Dans un point postérieur d'apprentissage, les constructions LS commencent à capter plus de cadres de leur catégorie ce qui illustre le progrès de la LS dans cette étape. À titre d'exemple, la courbe verte montre que le niveau de la CL a commencé avec un faible taux à l'étape 200, cependant il s'est amélioré dans l'étape suivante (300 cadres) de 0,386 à 0,547 dans le cas de l'anglais LS et de 0,418 à 0,517 dans le cas du français LS.

Ces résultats confirment partiellement ceux de (Monner et al., 2013) où les chercheurs ont utilisé un modèle du réseau de neurones pour montrer l'impact de l'enracinement de la L1 et le développement de la mémoire selon les processus cognitifs impliqués dans

cet ancrage, dans l'acquisition du genre grammatical d'une LS. Cette étude montre qu'un âge de début d'ALS tardif rend difficile l'ALS. Les auteurs ont attribué ces difficultés à l'enracinement de certaines règles et processus de la L1.

Un autre modèle du réseau de neurones (Zhao & Li, 2010) avait des résultats qui confirment ceux de notre modèle. (Zhao & Li, 2010) ont étudié les interactions lexicales bilingues dans le cas d'un apprentissage non supervisé. Ils confirment que des représentations de deux lexiques différents peuvent se développer lorsque les deux langues sont apprises simultanément. Cependant, la structure de la représentation dépend fortement du début d'ALS. Cela est dû au fait que les représentations LS sont perturbées avec une ALS tardive. En conséquence, ils affirment qu'une acquisition de deux langues implique une forte concurrence dans un contexte interactif ce qui influence la flexibilité d'ALS selon l'étape d'apprentissage. Cependant, nous croyons et selon les résultats de notre modèle que ces obstacles d'ALS peuvent être éliminés, dans une étape d'apprentissage ultérieure, avec un volume important d'exemples de LS. Dans une première étape d'apprentissage, l'ALS ne sera pas favorisée à cause de deux facteurs principaux, qui sont les constructions de la L1 qui attirent certains premiers cadres de la LS et les dissemblances entre la L1 et L2 telles que le déplacement du verbe dans les deux langues ou la position de l'adjectif dans la phrase. À un stade ultérieur d'apprentissage, les constructions déjà établies de la LS, le nombre de nouveaux intrants et les similarités entre la L1 et la LS telles que les structures grammaticales communes ou les verbes qui possèdent les mêmes primitives sémantiques vont permettre aux apprenants d'atteindre un niveau supérieur de la LS. Ces facteurs jouent un rôle majeur dans la correction de l'apprentissage de la LS et l'obtention d'un niveau de CL de la LS plus élevé.

Notre étude explore l'influence de quelques facteurs tels que l'âge de début d'ALS, le volume d'intrants ou le quotient LS/L1 dans l'ALS. Les résultats des expérimentations concordent, dans certains aspects, avec les résultats des recherches récentes. Ces résultats ont permis de confirmer que l'âge de début d'ALS peut jouer un rôle important

dans cette acquisition. Cependant, le nombre d'intrants et le quotient LS/L1 restent les éléments les plus décisifs dans cette opération d'apprentissage. D'autres études (Place & Hoff, 2011; Marchman et al., 2010) confirment que les apprenants de la LS peuvent être bilingues s'ils reçoivent au minimum entre dix pour cent et vingt-cinq pour cent comme quotient LS/L1. Cependant, ce niveau d'exposition ne garantit pas un bilinguisme fonctionnel (De Houwer, 2007). En ce qui concerne l'âge de début d'ALS, certaines études (Gardner et al., 1959) minimisent le rôle de ce facteur en le comparant avec la motivation de l'apprenant et en affirmant qu'il n'est jamais trop tard pour apprendre une nouvelle langue. Ce facteur de motivation sera parmi les perspectives de cette thèse.

Nous allons voir dans les simulations suivantes des expérimentations et analyses plus détaillées qui traitent les structures actanciennes de la LS et leur rôle dans le développement de l'interlangue.

3.6.1 Associations des éléments syntaxiques et sémantiques dans les structures actanciennes de la LS

Nous avons réalisé d'autres simulations sur l'acquisition des constructions dans le cas d'une LS, plus spécifiquement les constructions de « mouvement provoqué » (Hickmann et Henriette, 2010 ; Ellis et Cadierno, 2009). Bien que le français et l'anglais ont la même typologie syntaxique qui est Sujet-Verbe-Objet (SVO) et possèdent la même structure profonde (structure-D), les deux langues appartiennent à deux familles différentes (Talmy, 1985, 2000). L'anglais à cadrage satellitaire diffère du français à cadrage verbal, plus spécifiquement dans l'élaboration des constructions de « mouvement provoqué ». En anglais, les verbes expriment souvent le mode de déplacement, tandis que la direction est distinguée par un affixe. Cependant qu'en français, les verbes de mouvement révèlent la direction dans la plupart des cas. La manière de déplacement est signalée par un complément tel qu'un gérondif ou une construction adverbiale, ou elle peut ne pas être énoncée (Henriette, 2010 ; Benson, 2002; Pollock, 1989; Chomsky, 1968, 1991).

Nos simulations ont été menées en nous référant aux expérimentations de (Henriette, et al., 2008) qui ont analysé le lien entre ces dissimilitudes typologiques et l'ALS. Ils ont pris comme sujets d'expérience des apprenants anglophones du français LS. Les apprenants ont été divisés en trois groupes : niveau débutant, niveau intermédiaire et niveau avancé. De notre côté, nous avons considéré les énoncés qui contiennent les primitives sémantiques [MODE] et [DIRECTION] pour les verbes de « mouvement provoqué » ou leurs actants. Ces énoncés ont été produits par des apprenants de différents niveaux du français LS :

- EM : Elle marche [MODE] dans la maison.
- ED : Elle entre [DIRECTION] dans la maison en courant [MODE]
- ED : Elle traverse [DIRECTION] la route à pied [MODE].

Par la suite, nous avons créé des représentations des scènes qui correspondent à ces énoncés et dans lesquelles le mode de déplacement est exprimé par le verbe (EM) dans le premier énoncé tandis que dans le deuxième et le troisième le verbe exprime la direction (ED) et le mode est révélé par les actants.

- **S1** : MARCHE_[acte,déplacer](ELLE_{<agent><animé>},DANS_[](MAISON_{<Destination>})
<Destination>)
- **S2** : ENTRE_[acte,déplacer](ELLE_{<agent><animé>}, DANS_[](MAISON_{<Destination>})
EN_[](COURANT_{<Mode>})<Mode><Destination>)
- **S3** : TRAVERSE_[acte,déplacer](ELLE_{<agent><animé>}, (ROUTE_{<Destination>})
A_[](PIED_{<Mode>})<Mode><Destination>)

Ces énoncés et représentations de scène constituent la base des intrants utilisés pour l'acquisition des structures actanciennes des verbes de « mouvement provoqué » d'une LS par les apprenants de cette dernière. Dans des simulations similaires aux apprenants anglophones du français LS qui entendent des verbes qui expriment le mode ou la direction, chacune des scènes 2 et 3 est combinée avec des énoncés 'mode' et 'direction' pour former quatre paires scène/énoncé, S2-ED, S2-EM, S3-ED et S3-EM. Les paires

d'entrées S2-ED et S3-ED correspondent à la sélection appropriée du verbe qui exprime la direction selon la LS tandis que les paires scène/énoncé S2-EM et S3-EM correspondent à la sélection inappropriée du verbe qui exprime le mode selon la L1. Dans une simulation de référence, nous utilisons une paire avec un verbe qui révèle le mode de déplacement dans la L1 et la LS. Les cinq paires scène/énoncé résultantes sont présentées dans le tableau 3.2. Nous présentons séparément ces intrants à notre modèle. Ce dernier extrait un cadre C de la paire scène/énoncé présentée et calcul la probabilité de la meilleure construction S de l'ensemble SLS (Constructions de la LS) pour ce cadre P(S|C). Si le cadre est compatible avec une construction existante, cela indique que la combinaison scène/énoncé correspond à une association entre la syntaxe et le sens dans le modèle, sera catégorisée dans cette construction. Si la paire scène/énoncé a une probabilité faible, le modèle crée une nouvelle construction pour cette dernière ce qui confirme que cette nouvelle association est étrange au modèle. Dans les résultats des expérimentations et selon les paires scène/énoncé appropriées ou inappropriées, les probabilités élevées correspondent aux apprenants du niveau avancé ou intermédiaire tant que les valeurs faibles correspondent aux débutants. La figure 3.8 montre les valeurs des probabilités P(S|C) pour des volumes d'intrants différents (10, 50, 100, 200, 300, 400 et 500) avec une moyenne de 10 simulations. Les verbes de « mouvement provoqué » existent dans l'ensemble d'intrants du français tels que revenir, aller, tomber, arriver, passer, partir et venir, et qui expriment particulièrement le déplacement, constituent le noyau autour duquel se composent les constructions des nouveaux verbes et qui n'ont pas été vu auparavant dans l'ensemble d'intrants. Ces verbes peuvent avoir des structures actanciennes, des primitives sémantiques et des modèles syntaxiques différents.

S1-EM	EM : Elle marche [MODE] dans la maison.	S1 :MARCHE _[acte,déplacer] (ELLE <agent><animé>,DANS _[] (MAISON <Destination>))
S2-ED	ED : Elle entre [DIRECTION] dans la maison en courant [MODE]	S2 : ENTRE _[acte,déplacer] (ELLE<agent>< animé>, DANS _[] (MAISON<Destination> EN _[] (COURANT<Mode><Mode>)<Destination>))

S3-ED	ED : Elle traverse [DIRECTION] la route à pied [MODE].	S3 : TRAVERSE _[acte,déplacer] (ELLE _{<agent>< animé>} , (ROUTE _{<Destination>} A _[] (PIED _{<Mode>})<Mode>)<Destination>)
S2-EM	EM : Elle court [MODE] vers [DIRECTION] la maison.	S2 : COURT _[acte,déplacer] (ELLE _{<agent>< animé>} , VERS _[] (MAISON _{<Destination>})<Destination>)
S3-EM	EM : Elle marche [MODE] à travers [DIRECTION] la route.	S3 : MARCHÉ _[acte,déplacer] (ELLE _{<agent>< animé>} , (À TRAVERS _[] (ROUTE _{<Destination>})<Destination>)

Tableau 3.2: Des paires scène/énoncé des verbes de « mouvement provoqué ».

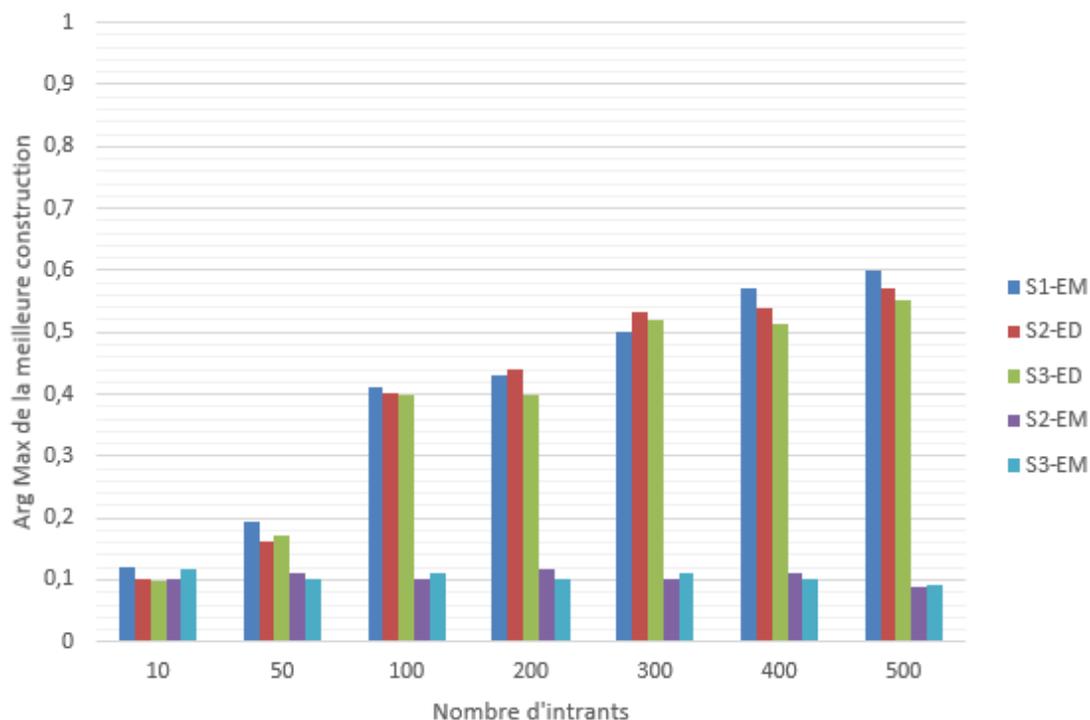


Figure 3.8: Arg Max de la meilleure construction pour des différentes paires scène/énoncé des verbes de « mouvement provoqué ».

La différence de l'arg max entre les deux types de paires scène/énoncé (appropriées et inappropriées) simule la capacité de l'apprenant à choisir la scène adéquate pour un énoncé selon les règles des structures actancielles de la LS déjà apprises. Autrement dit, cela montre la capacité des apprenants LS d'établir des nouvelles associations entre la syntaxe et la sémantique de la LS selon les intrants de cette dernière et en ignorant les associations de la L1. La figure 3.8 montre que les paires scène/énoncé appropriées à la LS ont eu un score proportionnel avec le nombre d'intrants. Dans ce cas, une construction LS dont les caractéristiques sémantiques et syntaxiques sont plus proches des caractéristiques du cadre va attirer ce dernier. Pour les paires scène/énoncé appropriées à la L1 et inappropriées à la LS, les simulations montrent que leurs probabilités sont restées faibles et stables même avec l'augmentation du nombre d'intrants de la LS. Cela veut dire qu'aucune construction LS aura le seuil qui lui permettra d'attirer le cadre d'une paire inappropriée à la LS.

En outre, ces résultats confirment, d'un autre côté, ceux de (Henriette, et al., 2008) qui affirment que les constructions inappropriées vont disparaître dans un stade avancé d'apprentissage.

Dans les prochaines simulations, nous montrons de plus la capacité des apprenants à établir des associations entre les éléments syntaxiques et sémantiques des structures actancielles de la LS, avec des généralisations productives.

3.6.2 Simulation de la production dans le cas de la LS

Avec ces expérimentations nous analysons la capacité du modèle de produire une phrase LS avec un nouveau verbe. Les apprenants LS peuvent utiliser un nouveau verbe de mouvement provoqué dans une construction qui n'a pas été observée pour ce dernier auparavant. Pour simuler cette capacité d'apprentissage, nous avons utilisé un exemple de (Henriette et al., 2008) : - Popi descend[Direction] la colline.

Après l'élaboration des constructions du français LS par notre modèle nous avons lui exposé la paire scène/énoncé suivante :

- S : DESCEND_[acte,déplacer](POPI_{<agent>< animé>}, COLLINE_{<Destination>})
- E :

Le verbe ‘Descendre’ n'a pas été précédemment vu par le modèle dans un énoncé, car il n'existe pas dans la liste initiale des verbes qui forment l'ensemble SLS. Nous savons que le principe de ‘Descendre’ est sa signification sont similaires à ceux que les apprenants du français LS (Henriette et al., 2008) ont déjà vue dans certaines constructions avec des verbes de « mouvement provoqué » tels que revenir, aller, arriver et venir. À cause de l'absence de l'énoncé dans la paire précédente, le modèle extrait un cadre sans la propriété qui représente le modèle syntaxique. Ensuite, ce dernier calculera cette propriété à partir des similarités avec les constructions LS déjà établies. Puisque la représentation de la scène possède la signification d'une utilisation transitive de type SVO, nous prévoyons que notre modèle prédit le modèle syntaxique transitif, "arg1 verbe arg2", avec les même rôles et primitives sémantiques des arguments, et cela même si le verbe ‘Descendre’ n'a jamais été vu auparavant dans l'ensemble des intrants et des constructions préalablement établies.

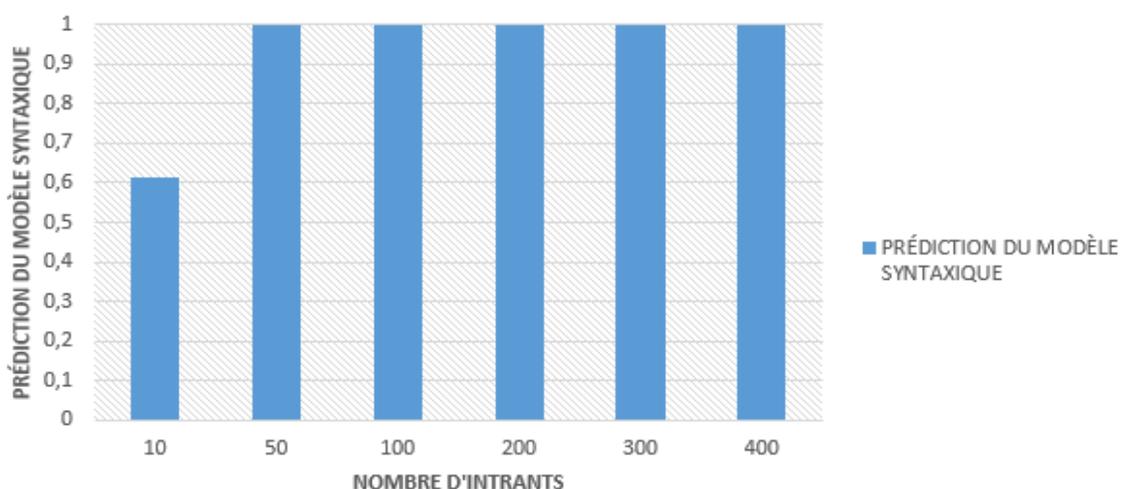


Figure 3.9: Prédiction du modèle syntaxique pour le verbe ‘Descendre’.

Le modèle a calculé les probabilités du modèle syntaxique pour le verbe ‘Descendre’ avec différentes quantités d’intrants. Les résultats sont la moyenne de 10 simulations différentes avec des ensembles d’intrants générées aléatoirement. Ces résultats montrent que le modèle syntaxique transitif a été calculé avec une probabilité égale à 1 à partir d’un nombre d’intrants de 50. Cela peut être expliqué par la fréquence de certains verbes, qui possèdent les mêmes caractéristiques, du « mouvement provoqué » tels que ‘aller’ dans le corpus initial et dans les paires d’intrants. De même, ces résultats montrent que notre modèle peut généraliser certaines connaissances syntaxiques et sémantiques, préalablement présentées au niveau des constructions pour calculer la valeur de certains composants linguistiques. Cette capacité est similaire à celle des apprenants LS, qui peuvent généraliser leurs règles sur des nouveaux verbes de la LS.

- Pour affirmer cette caractéristique du modèle, nous avons réalisé une expérimentation similaire à l’une de Ayoun (2005) avec des exemples extraits de Pollock (1989). Dans cette expérimentation, des apprenants francophones de l’anglais LS ont été mis dans des situations de production de phrases avec des déplacements de verbes différents de leur L1. À titre d’exemple, en français, on place, contrairement à l’anglais, les verbes conjugués au présent, avant les adverbes et les négations, comme dans (Pollock, 1989) :
- Adam voit souvent Ève / Adam ne voit pas Ève,
- Adam often sees Eve / Adam does not see Eve,

Après l’élaboration des constructions LS par notre modèle et avec des nombres d’intrants différents nous avons, ensuite, lui fourni la paire d’entrée suivante :

- **S:** SEE_[Percevoir,atteindre](ADAM_{<agent >>animé>},OFTEN_[](EVE_{< expérimentateur >> animé>})< expérimentateur >> animé>)
- **E:** Adam often sees Eve

La représentation de la scène S ressemble à des scènes observées avec des énoncés transitifs (SV[ADVERBE]O) dans la L1, mais l’énoncé E de la LS à un ordre de mots

différent (S[ADVERBE]VO). Après l'élaboration des constructions, nous avons présenté cette paire d'entrée au modèle, mais sans l'énoncé. En second lieu, le modèle calcule les probabilités pour avoir un arg max du modèle syntaxique le plus approprié à la scène S. Nous avons effectué 10 simulations pour chaque nombre de paires d'entrées de la L1, puis de la LS (dans ce cas l'anglais LS) et avons calculé en moyenne la probabilité des modèles SV[ADVERBE]O (L1) et S[ADVERBE]VO (LS). Comme on peut le voir dans la figure 3.10, à partir de 200 intrants (LS) le modèle prédit avec une certitude égale à 1 le modèle syntaxique S[ADVERBE]VO au lieu de SV[ADVERBE]O. Nous remarquons que ce transfert graduel du choix du modèle syntaxique L1 au modèle syntaxique LS est équivalent aux différents niveaux d'apprenants (débutant, intermédiaire et avancé) dans Ayoun (2005). Lorsque le nombre de paires d'entrées est faible, le modèle est plus susceptible d'imiter les apprenants débutants de la LS qui produisent un modèle syntaxique équivalent à celui de la L1. Lorsque le nombre de paires d'entrées augmente jusqu'à 200 ou plus, le modèle calcule celui de la LS avec un score de cent pour cent, ce qui est similaire à ce que font les apprenants du niveau avancé. Ces résultats montrent que le modèle montre un avancement avec plus d'exposition aux paires d'entrées, comme le cas des apprenants de la LS.

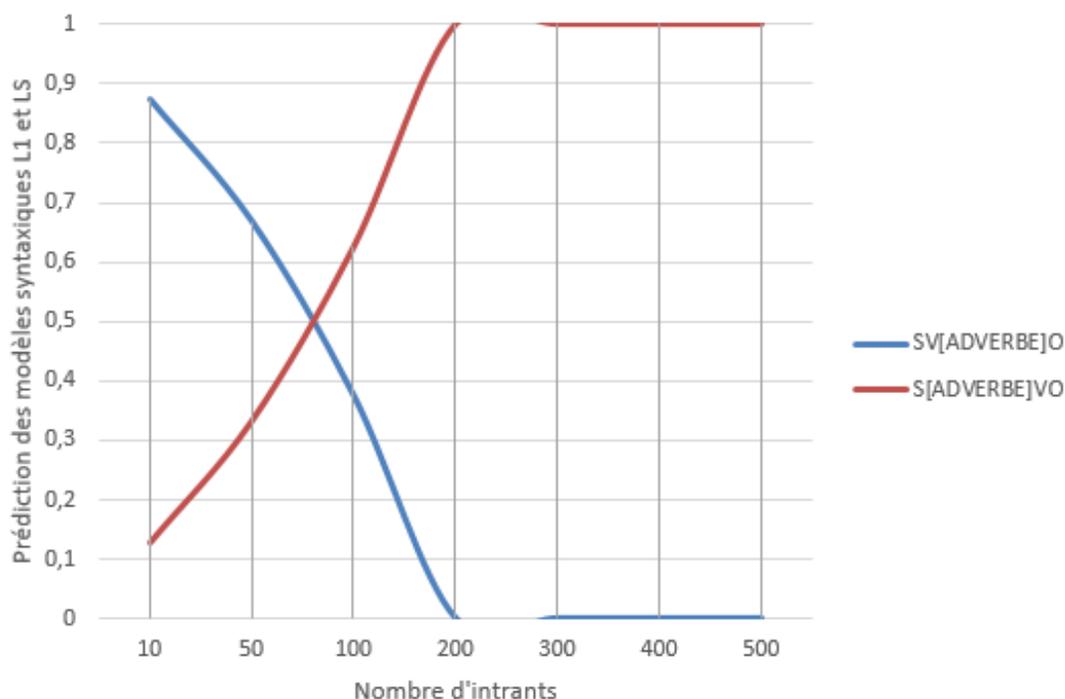


Figure 3.10: Prédiction des modèles syntaxiques L1 et LS.

3.6.3 Surgénéralisation des règles dans le cas de l'ALS

En linguistique, la surgénéralisation est l'application d'une règle grammaticale dans des situations non appropriées. Plusieurs études affirment que les apprenants débutants et intermédiaires de la LS sont fortement influencés par leur L1 (Juffs 2000, White 2003). Cependant, ces derniers finissent par adopter les règles de la LS, avec quelques intrants (preuves positives) supplémentaires. À titre d'exemple, les apprenants anglophones (débutants) du français LS supposent que les constructions à double objet sont autorisées en français, pour exprimer une relation de transfert de possession (White 1987, 1991), tandis que les apprenants francophones (débutant) de l'anglais LS n'utilisent pas ces constructions en anglais (Mazurkewich, 1984).

En anglais, il existe souvent deux formes que les verbes ditransitifs de transfert de possession peuvent prendre:

- John gave/sent a book to Mary.

- John gave/sent Mary a book.

Ces exemples ont été utilisés dans certaines expérimentations de (Cabrera et Zubizarreta, 2005). Bien que (Cabrera et Zubizarreta, 2005) ont étudié le phénomène de la surgénéralisation des règles dans le cas des apprenants anglophones qui apprennent l'espagnole LS et vice versa, certains exemples peuvent être appliqués dans le cas du français et de l'anglais LS. Nous avons imité l'acquisition du français LS par des apprenants anglophones. Notre modèle montre qu'au début d'apprentissage, ces derniers font une surgénéralisation avec des verbes tels que 'donner' et 'envoyer' en les utilisant dans des constructions à double objet comme le cas de leur L1. Dans les premières étapes d'ALS, des phrases similaires à "Adam a donné Ève un livre." ont été générées avec une probabilité proche de 0.5. Par la suite, et avec plus d'intrants de la LS, cette probabilité a été diminuée à 0. Dans nos expérimentations nous avons testé le comportement du modèle dans la production d'un modèle syntaxique pour un verbe ditransitif de la LS tel que 'donner'. Une fois que le modèle a été exposé aux paires d'intrants, nous exposons par la suite une paire de test dans laquelle l'énoncé était laissé vide.

- **S:** DONNER_[causer, posséder](ADAM_{<agent ><animé>}, (LIVRE_{<thème>} A_[](EVE_{<bénéficiaire><animé>})_{<bénéficiaire >}))
- **E:**

Le modèle calcule la probabilité du modèle syntaxique après l'élaboration des constructions de la LS. Le calcul de cette probabilité est effectué avec une moyenne de 10 simulations. Dans ces 10 simulations, le modèle a montré un comportement de surgénéralisation et, par la suite, de récupération à partir des nouveaux intrants (preuve positive) en excluant les constructions à double objet dans le cas du français LS. Au début, notre modèle prédit des modèles syntaxiques comme "arg1 verbe arg3 arg2", ensuite (à partir de 150 intrants) ces derniers ont laissé la place aux modèles syntaxiques plus appropriés à la LS "arg1 verbe arg2 à arg3".

3.6.4 Généralisations appropriées

Généralement, les apprenants de la LS généralisent des règles de la L1 ou la LS sur des nouveaux verbes de cette dernière en élaborant des nouvelles constructions et en utilisant des propriétés sémantiques particulières. Pour illustrer ce point, nous testons notre modèle avec le verbe de « mouvement provoqué » ‘push’ qui n’a pas été vu précédemment par ce dernier. Le modèle calculera les propriétés sémantiques appropriées de ce verbe à partir des autres constructions de « mouvement provoqué » élaborées auparavant. Nous exposons à notre modèle 500 paires d'intrants de la LS pour lui emmener à une étape avancée d'apprentissage, ensuite nous présentons la paire suivante, dans laquelle il manque les primitives sémantiques du prédicat et les rôles sémantiques des arguments :

- **S** : PUSH_[?](HE_{<?><?>}, ACROSS_[?](STREET_{<?><?>})_{<?>})
- **E** : He pushes the ball across the street.

La représentation partielle de la scène correspond à un apprenant de la LS qui entend pour la première fois le verbe ‘Push’, et qui a besoin de déterminer les associations entre les arguments syntaxiques et sémantiques. Compte tenu de cette entrée partielle, le modèle doit calculer les primitives sémantiques manquantes, en fonction de l'énoncé. Sur une moyenne de 10 simulations, le modèle prédit les primitives sémantiques pour le verbe ‘Push’ avec une probabilité de 0.61 pour la primitive ‘causer’, 0.55 pour la primitive déplacer, 0.18 pour la primitive ‘changer’, et des valeurs infimes pour les autres primitives. Les rôles prédits pour les trois arguments, avec des probabilités associées, sont agents (1), thème (1) et destination (1).

Ainsi, le modèle attribue des propriétés sémantiques qui captent le mouvement causal destiné au nouveau verbe ‘Push’ avec un score de cent pour cent pour les rôles sémantiques et des scores proportionnellement élevés pour les primitives sémantiques qui représente les verbes de « mouvement provoqué ». Dans ce cas, nous pouvons confirmer que les primitives sémantiques prédites pour une utilisation inhabituelle sont

aussi hautes que celles d'une utilisation typique. Ceci montre que le modèle peut effectuer des généralisations appropriées pour des nouveaux verbes.

3.7 Discussion des résultats

L'acquisition des structures actanciennes de la LS a été étudiée dans le but de déterminer les facteurs principaux qui influencent l'ALS et de cerner les processus cognitifs impliqués dans cette dernière.

Dans les différentes étapes d'apprentissage, les apprenants de la LS améliorent leurs capacités de généralisation des règles et modèles, de compréhension et de production en recevant de nouveaux intrants de la LS avec un quotient adéquat avec la L1. Les difficultés qui peuvent impacter cette acquisition telle que la surgénéralisation des règles de la L1 ou la LS sont particulièrement intéressantes dans le cas d'acquisition des structures actanciennes. Cette acquisition peut être un élément très complexe de la langue pour les apprenants de la LS à cause de ce genre de surgénéralisation. En outre, pour maîtriser les structures actanciennes d'une LS, les apprenants doivent détecter les points communs et les dissemblances avec leur L1 pour que cette acquisition devienne plus efficace et pour éviter d'appliquer des modèles de la L1 sur la LS.

Les simulations probabilistes ont montré que l'acquisition des structures actanciennes de la LS, comme le cas de la L1, peut être représentée comme une association probabiliste entre la syntaxe (la forme) et la signification (le sens). Cette association reflète des généralisations de règles et des modèles sur l'utilisation des verbes et des actants. Dans le cas du modèle probabiliste, ces associations sont représentées par des constructions élaborées au fil du temps, et qui peuvent refléter les étapes qui correspondent bien à celles des apprenants de la LS, telles que la compréhension des énoncés et des scènes observées, la généralisation des règles sur de nouvelles situations, les erreurs liées à la surgénéralisation et la correction à partir des preuves positives. Ces expérimentations ont permis d'illustrer certains processus sous-jacents d'acquisition des structures actanciennes de la LS et leur utilisation dans l'ALS.

En recevant des paires scène/énoncé de la LS, les apprenants développent des représentations de constructions qui peuvent être utilisées dans les tâches de la compréhension et de la production. L'utilisation de ces constructions dans ces différentes tâches peut être observée dans le comportement de notre modèle en considérant des propriétés comme les composants sémantiques et syntaxiques et des paramètres comme le nombre d'intrants. Ces connaissances spécifiques sous forme de propriétés et ces paramètres permettent au modèle d'imiter l'ALS et de généraliser des règles au cours de l'apprentissage. Cette simulation peut expliquer les effets observés et illustrer certains processus d'apprentissage. Généralement, notre modèle illustre l'acquisition des structures actanciennes par les apprenants d'une LS en élaborant des constructions qui imitent des représentations internes des associations entre les composants sémantiques et syntaxiques de la langue.

L'utilisation appropriée des connaissances établies sous forme de constructions dans la production est constituée dans le modèle à partir de 50 paires d'intrants, comme le montrent nos expérimentations qui imitent celles de (Henriette et al., 2008). Des propriétés linguistiques comme un modèle syntaxique simple tel que "arg1 verbe arg2" peut être prédites avec une certitude égale à 1 à partir de 50 intrants, tandis que d'autres éléments plus compliqués tels que la position de l'adverbe SV[ADVERBE]O, peuvent être prédit à partir de 200 intrants comme le montre nos expérimentations qui simulent celles de (Ayoun, 2005).

D'un autre côté, la généralisation à partir des constructions sur une nouvelle entrée observée peut conduire à une généralisation erronée des modèles précédemment élaborés, autrement dit une surgénéralisation des règles.

Au niveau de l'ALS, plusieurs études ont été réalisées pour illustrer et expliquer ce phénomène de surgénéralisation (Juffs 2000, White 2003). Ces études ont montré que les apprenants de la LS peuvent corriger ce genre d'erreurs seulement en recevant des preuves positives (plus d'intrants de la LS), car la correction avec des preuves négatives

n'est pas toujours accessible. Cette correction est effectuée grâce à la construction de nouvelles fonctionnalités sémantiques et syntaxiques pour créer des exceptions analogiques qui permettent d'encadrer les généralisations précédemment établies (MacWhinney, 2004). Dans le cas de notre modèle, une tâche de production est présente comme une prédiction du meilleur modèle syntaxique parmi un ensemble de modèles possibles. Le degré d'enracinement des constructions est proportionnel avec la fréquence des verbes et des actants dans les paires d'intrants. Cet enracinement représente un taux de probabilité supérieure qui permettra aux apprenants de la LS d'activer les processus de correction des erreurs de la surgénéralisation des règles grâce aux nouveaux exemples traités. Theakston (2004) affirme ce principe en montrant que les apprenants adultes et enfants commettent des erreurs de surgénéralisation plus avec les verbes moins fréquents. D'autant plus que le fait d'avoir une fréquence basse pour certains verbes signifie à la fois une probabilité plus faible pour leur utilisation ainsi que des probabilités plus élevées pour leurs concurrents.

L'enracinement des constructions en utilisant des cadres avec une fréquence considérable, joue un rôle principal dans la généralisation des règles et modèles (Goldberg, 2006). Si une forme générale correspond mieux aux propriétés sémantiques et syntaxiques d'une paire d'entrée, elle sera utilisée même en présence d'une connaissance différente liée à cette dernière, autrement dit, la probabilité de chaque utilisation dépend directement de la fréquence des structures actanciennes dans les paires d'intrants. Par conséquent, plus l'utilisation est enracinée (plus fréquente), plus qu'elle soit sélectionnée dans les comparaisons probabilistes. De ce fait, lorsqu'un cadre spécifique à un verbe assez enraciné est traité, cela entraîne une diminution des erreurs de surgénéralisation pour ce dernier. Cependant, si aucun cadre n'est suffisamment enraciné, le modèle utilise une construction générale compatible avec la scène ou élabore une nouvelle construction, dépendamment des probabilités calculées.

D'un autre côté, le principe de la « preuve positive » permet aux apprenants de corriger les erreurs de surgénéralisation, toutefois, il préserve la capacité de ces derniers à

générer des composants syntaxiques et sémantiques dans des situations de production et de compréhension pour des nouveaux verbes et modèles. Cependant, bien que la majorité des utilisations des nouveaux verbes et modèles soient acceptables, beaucoup d'autres ne le sont pas. En effet, les propriétés sémantiques et syntaxiques fondamentales de la structure actancielle et sa compatibilité avec la scène jouent un rôle principal dans le résultat de cette tâche linguistique. À titre d'exemple, dans le cas des faux amis, un apprenant francophone de l'anglais LS peut entendre l'énoncé "I demanded an explanation." avec une scène qui semble appropriée, peut le confondre avec le verbe 'demander' de sa L1. Cette confusion conduit bien aisément à une catégorisation du cadre du verbe 'to demand' dans des constructions avec des caractéristiques syntaxiques et sémantiques similaires à 'demander'. L'élimination de ce genre de généralisation nécessite des représentations sémantiques qui peuvent distinguer les composants les plus élémentaires de la scène et qui peuvent faire la différence dans ce genre de situation. Parmi les perspectives de cette thèse, l'élaboration de ce type de représentations sémantiques pour permettre à notre modèle d'atteindre un niveau de distinction sans avoir besoin de « preuves négatives ».

3.8 Conclusion

Nous avons utilisé un modèle computationnel probabiliste pour simuler l'acquisition des structures actanciennes chez les apprenants de la LS. Nous avons adopté comme champ d'expérimentation le français et l'anglais où chaque langue peut être utilisée comme LS ou L1. Notre modèle est fondé sur la notion de construction, où des associations entre les composants sémantiques et syntaxiques des structures sont établies pour réaliser des tâches linguistiques appropriées (Goldberg, 1995). L'objectif était d'illustrer les processus d'apprentissage impliqués dans l'ALS et de montrer le rôle du volume d'intrants et le ratio LS/L1 dans cette acquisition, soit pour les enfants bilingues ou les apprenants adultes de la LS.

Nous avons isolé des facteurs tels que l'âge de début, le nombre d'intrants ou le quotient LS/L1 pour montrer l'influence de chaque facteur séparément. Les résultats des

expérimentations ont permis de confirmer le rôle de l'âge de début d'ALS. Cependant, le nombre d'intrants et le quotient LS/L1 sont les éléments les plus importants dans cette acquisition. Plusieurs autres recherches (Place & Hoff, 2011; Marchman et al., 2010) confirment que les apprenants de la LS peuvent être bilingues s'ils reçoivent au minimum entre dix pour cent et vingt-cinq pour cent comme quotient LS/L1. Tandis que d'autres études ont démontré que le facteur de motivation peut être l'élément le plus décisif, comparé à celui de l'âge de début d'ALS (Gardner & Lambert, 1959).

D'un autre côté, les résultats des simulations ont démontré l'émergence des constructions abstraites des structures actanciennes à partir de l'utilisation des verbes (intrants du modèle). Ces constructions ont été utilisées pour montrer la similarité entre les modèles et les règles établies par notre modèle et ceux qui ont été élaborés par les apprenants de la LS. En calculant les composants manquants dans des entrées de tests, tels que le modèle syntaxique et les primitives sémantiques, le modèle imite des tâches de production ou de compréhension qui peuvent être réalisées par les apprenants de la LS. Cette simulation est fondée sur l'interaction entre les composants sémantiques et syntaxiques des constructions précédemment établies d'un côté, et ceux de l'intrant de test d'un autre côté.

Les calculs bayésiens du modèle permettent de tenir compte de deux types de généralisation des règles et modèles qui sont la généralisation appropriée et la surgénéralisation dans le cas d'ALS. De nombreux principes ont été suggérés dans la littérature psycholinguistique pour expliquer ces types de généralisation, tels que les notions de la preuve positive et la preuve négative, les principes d'analyse universelle, l'enracinement des cadres, la concurrence entre constructions, la correction à partir des exemples et l'apprentissage à base d'utilisation. Notre modèle prend en considération tous ces facteurs à l'exception du facteur de la « preuve négative » qui peut être étudié dans le cadre d'un apprentissage formel et non dans celui de l'acquisition de la langue à base d'utilisation « usage based ».

Le modèle a démontré que l'intégration de ces notions dans les processus de généralisation et de correction des surgénéralisations peut expliquer le comportement d'apprentissage des apprenants de la LS.

En ce qui concerne les limitations du modèle, certains aspects de l'ALS peuvent être explorés dans des futurs travaux de cette thèse. À titre d'exemple, nous pouvons élaborer un modèle de simulation d'acquisition de mots simples comme dans le cas de (Erk et Pado, 2008) et l'intégrer dans notre modèle ce qui nous permet d'approcher la simulation réelle des apprenants de LS, au lieu de considérer qu'ils connaissent déjà le sens des mots simples. De même, l'étude de la morphologie du verbe peut être un aspect intéressant à explorer par l'extension de notre modèle à la considération de cet élément dans l'ALS. Dans le prochain chapitre, nous effectuons quelques simulations d'apprentissage des rôles sémantiques dans le cas de l'ALS.

CHAPITRE IV

ACQUISITION DES ROLES SEMANTIQUES D'UNE LS

4.1 Introduction

La langue naturelle est caractérisée par des correspondances systématiques et complexes entre la sémantique des verbes et les caractéristiques syntaxiques de leurs actants et structures (Thothathiri et Snedeker, 2011; Levin et Rappaport Hovav, 2005). Les rôles sémantiques servent à décrire ces correspondances. De même, ils illustrent les relations entre le verbe et les actants qui participent à l'événement. Les processus cognitifs adoptés par les apprenants d'une LS pour acquérir les rôles sémantiques de cette dernière restent encore à élucider. En effet, la façon avec laquelle les associations sont établies entre les rôles sémantiques et les positions grammaticales des actants portant sur ces derniers est un domaine à défricher, plus spécifiquement dans le cas de l'ALS. Dans ce chapitre nous effectuons quelques simulations probabilistes avec notre modèle computationnel pour illustrer l'acquisition des rôles sémantiques par les apprenants de la LS. Pour cela, nous avons adapté notre modèle pour pouvoir simuler cette acquisition en remplaçant les rôles sémantiques explicites tels que «agent», «patient» et «thème», par des primitives sémantiques élémentaires, ce qui nous permettra d'associer chaque actant à une distribution de probabilité des propriétés sémantiques que ce dernier peut prendre.

Nous allons montrer l'influence des différences principales et secondaires entre la L1 et la LS dans l'acquisition de ces rôles. En outre, nous allons illustrer l'importance des profils sémantiques des positions des actants et qui peuvent donner de l'information sur les caractéristiques spécifiques des verbes. Comme nous l'avons vu dans le chapitre précédent, nous allons confirmer le rôle du volume d'intrant dans la transformation

graduelle des propriétés des rôles sémantiques en représentations internes abstraites. Ces représentations expriment les propriétés générales des actants et des verbes, constituent l'outil des apprenants pour enraciner l'information liée à ces derniers et corrigent les erreurs de surgénéralisation à partir de la preuve positive (Jackendoff, 1990).

D'un autre côté, les rôles sémantiques peuvent mieux expliquer le comportement lexical des verbes par rapport à la sous-catégorisation syntaxique de ces derniers (intransitifs, transitifs directs, transitifs indirects, etc.) (Peter, 1997). En outre, le lien entre les fonctions syntaxiques (sujet, objet, complément, etc.) et les rôles sémantiques (agent, patient, but, ...) a été abordé dans plusieurs études pour montrer le rôle de cette corrélation dans la création de l'adjonction syntaxico-sémantique (Radimsky, 2012).

Que ce soit en anglais ou en français, les rôles sémantiques suivent un ordre prescriptif selon lequel ils doivent être affectés. Dans la phrase suivante "Adam a donné une fleur à Ève.", les rôles sémantiques des actants suivent l'ordre <Agent> <Patient> <Bénéficiaire>, tandis qu'à la voix passive ils peuvent suivre l'ordre <Bénéficiaire> <Patient> <Agent>. Néanmoins, dans le cas des fonctions syntaxiques, le <Sujet> garde sa première place dans les deux situations. Ces rôles sémantiques doivent correspondre aux fonctions syntaxiques. Dans l'exemple suivant qui montre la polysémie de la fonction de sujet :

{ Le joueur lance la balle.
 { La balle a été lancée.
 { La machine lance la balle.

Dans les trois phrases, les sujets ont des sens différents, en premier lieu c'est le joueur qui accomplit intentionnellement l'action, ensuite c'est la balle qui est le sujet de l'exécution de l'action et finalement c'est la machine qui aide à réaliser l'action. En ce qui concerne les rôles sémantiques, il s'agit notamment d'« agent », de « patient » et d'« instrument ». De ce fait, un rôle sémantique peut occuper différentes fonctions syntaxiques. En ce qui concerne les cas grammaticaux, l'anglais et le français partagent

certains éléments communs, toutefois, d'autres points de dissemblance existent entre les deux langues. Un cas grammatical est une caractéristique associée au nom, au pronom, à l'adjectif et au déterminant, et qui exprime la fonction syntaxique ou le rôle sémantique en relation avec le verbe (Maling, 1990). Dans les exemples suivants extraits de (White et al., 1999) les deux premiers exemples montrent la similarité entre le français et l'anglais dans le cas grammatical de certains actants, tandis que dans le troisième exemple, nous constatons qu'en anglais nous ne pouvons pas représenter de la même façon qu'en français les actants avec des expériences datifs.

- Susanne admire Jean / Susan admires John, (1)
- Le chien effraie Jean / The dog frightens John, (2)
- Marie manque a l'homme / Mary misses DAT the man / The man misses Mary. (3)

Avec ce type de différences entre le français et l'anglais dans les cas grammaticaux, la tâche de l'apprenant francophone de l'anglais LS consiste à comprendre la signification des caractéristiques lexicales des verbes et des actants de la LS. Une fois que cela est fait, le choix de la structure actancielle avec les rôles sémantiques appropriés se comporte de manière similaire à la L1. Cependant, ce choix s'effectue selon les caractéristiques lexicales et syntaxiques de la LS en utilisant les intrants de cette dernière.

Dans la section 4.2 nous allons illustrer quelques théories linguistiques de l'acquisition des rôles sémantiques et leur impact dans l'ALS. Les modèles computationnels existants liés aux rôles sémantiques seront examinés dans la section 4.3. La section 4.4 montre une adaptation de notre modèle computationnel pour l'orienter à l'acquisition des rôles sémantiques dans le cadre d'une LS. Nous allons montrer l'extraction de données dans la section 4.5, tandis que la section 4.6 sera consacrée à la présentation des résultats des simulations probabilistes. Dans la section 4.7 nous allons discuter les résultats de nos expérimentations en faisant le lien entre les processus d'acquisition des rôles

sémantiques. Une conclusion et un résumé des contributions seront présentés dans la section 4.8.

4.2 Théories d'acquisition des rôles sémantiques

Le rôle sémantique est un concept linguistique apparu à partir des années 1960 avec les travaux de (Gruber, 1965) sur les relations lexicales, de (Fillmore, 1968) sur la théorie de la grammaire des cas, et par la suite avec ceux de (Jackendoff, 1972) sur la théorie standard étendue. Ce concept a été utilisé pour expliquer la relation entre un prédicat et ses actants. En outre, ce dernier a été adopté pour expliquer l'indexation des actants sur le verbe (Chomsky, 1981; Pinker, 1984). La classification des rôles sémantiques varie selon les théories existantes. Ces rôles sont discrets selon certaines théories telles que la théorie précoce des structures actanciennes "Government and binding" (Chomsky, 1981; Williams, 1981). Un rôle sémantique discret n'a aucune relation hiérarchique avec les rôles des autres actants. Selon ce principe, chaque prédicat a une liste d'actants qui font partie d'une entrée lexicale d'une catégorie spécifique. Ces actants sont étiquetés avec des rôles sémantiques tels que 'agent', 'thème', 'but' et 'source' en fonction des hypothèses de Gruber (1965). Les actants selon (Williams, 1981) sont classés comme des actants internes et externes. À titre d'exemple dans "Agent see Theme" l'agent est considéré comme argument externe tandis que le thème est un argument interne. Il est important de noter que dans cette théorie, les rôles sémantiques n'ont pas de contenu ou de structure propre. Ils ne servent qu'à distinguer les arguments d'un prédicat.

Selon d'autres théories, les rôles sémantiques sont de nature hiérarchique (Jackendoff, 1972). Jackendoff (1972) et à partir de la liste des rôles thématiques de Gruber (1965) a donné l'hiérarchie appelée "Condition hiérarchique thématique" dont le principe est "La phrase passive doit être plus élevée dans la hiérarchie thématique que celle du sujet dérivé.":

agent > emplacement, source, but, > thème.

Cette théorie affirme que le rôle sémantique le mieux classé occupe la position du sujet. D'un autre côté, Goldberg (1995) affirme que la forme passive s'applique seulement aux verbes qui sont associés à deux rôles ou plus et dont l'un est plus élevé que les autres, en considérant la hiérarchie suivante :

agent, cause> destinataire, expérient> instrument> patient, thème> emplacement,
source, but

De même, il confirme que la construction passive nuit au rôle de participant le mieux placé associé au verbe.

Certaines autres théories (Dowty, 1989; Kibrik 1997) mentionnent le rôle sémantique généralisé. Selon ce principe, des rôles sémantiques généralisés ont été développés pour résoudre des problèmes liés aux concepts de rôles sémantiques discrets et hiérarchiques. Plusieurs études ont proposé des listes de rôles sémantiques différents en taille et en propriétés sémantiques. Cependant, il y a très peu de consensus sur l'ensemble de ces rôles. Une façon de résoudre ce problème est la notion de «prototypes sémantiques». Le principe de «prototypes sémantiques» est que les rôles sémantiques partagent des « ressemblances familiales » qui pourraient être généralisées pour formuler des rôles sémantiques prototypiques (Kailuweit, 2003). Kibrik (1997) a utilisé le concept de «prototypes sémantiques» pour proposer une typologie de la distribution d'arguments. Certains rôles sont soumis à une nouvelle généralisation dont le résultat est un "hyper rôle". Dans ce contexte l'"hyper rôle" sert à généraliser la catégorisation des actants dans une langue donnée. À titre d'exemple, l'utilisation des rôles tel que l'agent' et le 'patient' comme "hyper rôle" est une généralisation de ces catégories de codage à d'autres classes non prototypiques. Dowty (1989) confirme que les rôles sémantiques pour l'agent dérivent d'un ensemble de propriétés sémantiques élémentaires telles que l'affectivité, la volonté, la perception, le pouvoir de causer un changement d'état, l'existence indépendante et le mouvement. En outre, ceux du patient découlent d'autres propriétés telles que subit un changement d'état, thème supplémentaire, causalement

affecté par un autre, stationnaire et existence dépendante. Un rôle sémantique ne possède pas nécessairement toutes ces propriétés, et certains rôles ont plus de propriétés élémentaires que d'autres. Les propriétés proposées sont créées à partir de l'analyse des données linguistiques.

D'un autre côté, des expérimentations de (Kako, 2006) ont montré que les apprenants d'une langue attribuent des propriétés sémantiques élémentaires aux rôles grammaticaux tels que le sujet et l'objet, même pour des nouveaux verbes qui appartiennent à des nouvelles constructions. Ces expérimentations sont proches de celles de l'ALS en mettant les apprenants dans des situations de production avec des nouveaux verbes et des nouvelles constructions tels que : "The rom mecked the zarg". D'autres expérimentations (Tomasello, 2000; Lieven et al., 1997; McRae et al., 1997) ont confirmé que les rôles sémantiques pourraient être considérés comme des concepts spécifiques au verbe et par conséquent les apprenants de la langue acquièrent ces rôles progressivement avec des processus de catégorisation et de généralisation grâce à l'ensemble d'intrants qu'ils reçoivent. Cela peut confirmer que chez un apprenant (agent cognitif) la représentation interne d'un rôle sémantique n'est pas sous forme d'étiquette générale telle que 'agent' et 'thème', mais plutôt sous forme de propriétés élémentaires telles que l'affectivité et la volonté et sous forme de concepts spécifiques au verbe tels «transfert de possession», «cause» et « chose en cours » (Tomasello, 2000). Ces expérimentations ont montré que le fait de fournir des caractéristiques sémantiques d'un nouveau verbe rend plus facile, pour les apprenants, l'acquisition des similarités relationnelles entre les rôles sémantiques.

Plusieurs études ont proposé des listes différentes de rôles sémantiques. Cependant, il y a très peu d'accord sur l'ensemble de ces rôles. Cela s'explique par l'absence d'une définition cohérente des règles de liaison universelles (Kasper, 2008) qui peuvent attribuer chaque actant à un seul rôle sémantique selon les caractéristiques du verbe. La théorie du « linking » affirme qu'une très forte régularité existe dans la concordance entre les propriétés sémantiques et syntaxiques. Pour beaucoup de verbes, l'agent est le

sujet tandis que le thème est l'objet. Cependant, beaucoup d'autres théories (Kasper, 2008) ont montré qu'une simple liste de rôles sémantiques ne suffit pas pour expliquer cette concordance. La réalisation syntaxique peut être expliquée par une représentation détaillée de la sémantique lexicale du verbe. Ces théories affirment que les rôles sémantiques doivent respecter certaines caractéristiques telles que la définition opératoire, l'unicité et la distinction. Par conséquent, chaque actant d'un prédicat reçoit exactement un rôle et un rôle ne peut pas être attribué qu'à un seul actant.

Malgré les études précédemment mentionnées, beaucoup de particularités liées à l'acquisition des rôles sémantiques sont encore à explorer. Des simulations avec des modèles computationnels peuvent clarifier, considérablement, les processus cognitifs et les mécanismes d'apprentissage impliqués dans cette acquisition.

4.3 Modèles computationnels et rôles sémantiques multilingues

L'analyse, l'utilisation et l'acquisition des rôles sémantiques sont des questionnements profonds pour l'ALS, le traitement automatique de la langue naturelle (TALN) et le web sémantique. Plusieurs modèles ont été créés pour simuler l'acquisition des rôles sémantiques (Alishahi et al., 2008), l'analyse sémantique (Zouaq et al., 2008), l'étiquetage multilingue des rôles sémantiques (Anders et al., 2009) et l'annotation sémantique de textes multilingues (Bae et al., 2008). Parmi les rares modèles computationnels qui ont simulé l'acquisition des rôles sémantiques par des apprenants de la langue nous avons celui de (Alishahi, 2008). Ce dernier a été conçu pour acquérir des associations entre les caractéristiques sémantiques et les positions syntaxiques des actants d'un côté et les primitives sémantiques des verbes d'un autre côté. Ces associations probabilistes établies seulement sur les intrants du modèle permettent à ce dernier de généraliser des connaissances et des caractéristiques liées aux rôles sémantiques sans étiquetage supervisé. Cependant, ce modèle ne permet pas de représenter les caractéristiques d'un événement, tel que « causalement affectées », et qui peut être liée à un actant donné. Cette limitation n'a pas empêché le modèle d'apprendre des profils primitifs pour des rôles sémantiques à partir des intrants qu'il reçoit. Cette

caractéristique de généralisation des rôles sémantiques à partir de certains profils primitifs a été démontrée par d'autres recherches (Nation et al., 2003) qui s'intéressent à l'apprentissage des rôles sémantiques chez les jeunes enfants. Une autre limitation du modèle, est que la prédiction des rôles sémantiques utilise seulement les connaissances liées aux constructions, mais pas celles liées aux propriétés spécifiques des verbes tels que les préférences de sélection et les primitives sémantiques liées aux actants.

Un autre modèle Chang (2004) a été élaboré pour apprendre les associations entre les modèles syntaxiques et les rôles sémantiques à partir des exemples qui permettent la compréhension de la langue. Toutefois, ce dernier ne génère pas de nouvelles conceptions pour les rôles sémantiques ce qui ne permet pas l'apprentissage de ces rôles. Le modèle de (McClelland et Kawamoto, 1986) a simulé l'apprentissage des rôles sémantiques en représentant les propriétés sémantiques des actants par des vecteurs de primitives des noms et des verbes. Par la suite, le modèle associe des rôles sémantiques aux actants en utilisant ces vecteurs. D'autres modèles (Gildea et Jurafsky, 2002) ont été fondés sur l'étiquetage explicite d'un nombre limité de rôles sémantiques ce qui nuit à la simulation d'un vrai apprentissage de la L1 ou la L2.

En ce qui concerne l'ALS, (Dittmar, 1990) a étudié les aspects sémantiques de l'apprentissage des adultes en montrant le lien entre la sémantique développementale, la pragmatique (Dittmar, 1984) et la syntaxe. Une sémantique combinatoire, fondée sur la notion de l'interlangue, a été adoptée pour illustrer la "construction du sens" chez les apprenants de la LS en simulant un apprentissage réel. L'étude confirme que dans les premiers stades d'apprentissage les rôles sémantiques ne se laissent pas positionner dans une sémantique limitée ou dans une approche conceptuelle ce qui génère un statut d'ambiguïté. Ce dernier peut être résolu grâce au passage du potentiel et niveau de compréhension de l'implicite à l'explicite et de la corrélation du contexte situationnel à l'indépendance de ce dernier. L'acquisition des rôles sémantiques de la LS peut être caractérisée par le développement du sens au niveaux de l'interlangue grâce à la compétition des candidats d'une fonction communicative et par la négociation de sens

entre la L1 et la LS, ce qui distingue l'apprentissage de la langue entre natif et apprenant.

Dans le domaine du TALN certains modèles ont été conçus pour l'étiquetage des rôles sémantiques multilingues. Le modèle de (Anders et al., 2009) utilise des classificateurs indépendants qui identifient le sens du prédicat, les actants et les étiquettes de ces derniers. Ces classificateurs donnent des candidats adéquats parmi un groupe selon des critères de fonctionnalité ce qui a permis au modèle d'obtenir le deuxième meilleur score global d'étiquetage des rôles sémantiques, le premier score sur les données chinoises et allemandes et le deuxième score sur les données anglaises.

Le modèle de (Zouaq et al., 2008) a été fondé sur une architecture modulaire pour effectuer une analyse sémantique en utilisant les techniques du TALN. Cette analyse permettra d'attribuer un sens aux phrases et aux textes par l'annotation des phrases avec des rôles sémantiques grâce à des algorithmes de désambiguïsation. Pour effectuer cette annotation, une typologie des rôles sémantiques doit être choisie (Marquez et al., 2008). Par la suite, les rôles sémantiques seront définis dans des ontologies pour faciliter un raisonnement logique dans une étape ultérieure. En conséquence, cette analyse sémantique utilise une annotation des rôles, une grammaire de dépendances, une ontologie pour la définition des rôles sémantiques et un lexique sémantique pour l'annotation des rôles. Cette architecture modulaire a permis d'identifier les prédicats et les actants dans un texte donné et par la suite l'application d'un raisonnement logique sur les structures actanciennes, ce qui facilite l'attribution du sens dans une prochaine étape.

Pour simuler avec fidélité et précision l'acquisition des rôles sémantique pour une LS, nous allons remplacer l'étiquetage explicite des rôles sémantiques par des primitives sémantiques qui permettront à notre modèle de généraliser des profils sémantiques sur les actants et d'apprendre les associations entre ces derniers et les positions grammaticales.

4.4 Adaptation du modèle bayésien à l'acquisition des rôles sémantiques de LS

Nous adaptons notre modèle computationnel pour créer des constructions à partir des paires scène/énoncé et les utiliser dans des situations de production et de compréhension pour modéliser l'acquisition des rôles sémantiques chez les apprenants de la LS. Les constructions généralisent des liaisons probabilistes à partir des cadres qui correspondent aux représentations internes abstraites des différents aspects de la langue incluant les rôles sémantiques. Pour ces rôles, le modèle attribue un profil sémantique représenté par un vecteur de propriétés au lieu d'un étiquetage explicite des rôles sémantique. Le profil sémantique est calculé selon des associations probabilistes entre les propriétés sémantiques des actants, les primitives sémantiques du prédicat et les positions syntaxiques des éléments de la structure actancielle. En outre, les intrants du modèle sont des paires scène/énoncé où les rôles sémantiques des actants d'un cadre sont remplacés par les propriétés sémantiques de chaque actant. À titre d'exemple, dans la phrase 'Adam eat bread, l'expression du prédicat ne contient pas des informations sur les rôles et les catégories sémantiques, mais des propriétés sémantiques des actants, tirées de WordNet, comme :

{man, male, individual, person, organism, living thing, object, entity, causal agent} pour Adam et {baked goods, food, solid, substance, entity} pour bread.

Grâce à cette nouvelle présentation, notre modèle computationnel pourra calculer les probabilités qui peuvent aider à déduire les rôles sémantiques des actants au lieu de les représenter d'une manière explicite, ceci permet de simuler un apprentissage réel des rôles sémantiques par les apprenants de la LS. Le modèle génère des cadres à partir des paires scène/énoncé où chaque cadre contient des informations concernant le prédicat, la valence, les primitives sémantiques, le modèle syntaxique et les propriétés sémantiques de chaque actant. Le cadre extrait est stocké dans l'entrée lexicale du prédicat et initialisé avec une fréquence égale à 1.

D'un autre côté, dans deux langues différentes telles que le français et l'anglais, deux mots ont les mêmes propriétés sémantiques s'ils ont le même référent, comme dans le cas de "chat" et "cat". Cependant, les apprenants peuvent avoir différents types d'ambiguïté, particulièrement avec les verbes. Pour un verbe, le référent est plutôt une 'action' ou un 'événement' qu'un objet concret. Durant une scène de combat, l'apprenant peut affecter les propriétés sémantiques du verbe 'attaquer' au verbe 'défendre' à cause d'une confusion de scène. En outre, des faux amis peuvent produire le même genre d'ambiguïté. Un apprenant peut affecter les propriétés sémantiques du verbe 'to assist' < aider, activer, manœuvrer > au verbe 'assister à' dont les propriétés sémantiques < participer, voir, visiter >.

Pour calculer la probabilité de chaque propriété du cadre, nous allons utiliser l'équation 2.6 précédemment décrite dans le chapitre II.

Nous allons adopter la 'similarité cosinus' proposée et approuvée par (Salton, 1983) pour calculer la similarité d'un profil sémantique (représenté par le vecteur v) d'un cadre C et un cadre d'une construction S pour un actant donné. Pour cela, nous calculons l'arg max des similarités entre le vecteur qui représente le profil sémantique de cet actant dans le cadre C et chaque vecteur qui représente le même actant dans un cadre d'une construction S , et cela, pour chaque construction.

$$\text{ProfilSemantiquePlusSimilaire}(S) = \underset{s}{\text{argmax}} (\text{similarité}(v, S)) \quad (4.1)$$

Le score de similarité entre le profil sémantique d'un actant du cadre C et une construction S est :

$$\text{similarité}(v, S) = \sum_n^i \text{similarité}_i(v, S_{vi}) / n \quad (4.2)$$

Où S_{vi} est le profil sémantique de l'actant du cadre c_i (représenté par le vecteur v_i) de la construction S et n est le nombre de profils sémantiques dans la construction S :

Le score de similarité entre le profil du cadre C (représenté par le vecteur v) et un profil (représenté par le vecteur Sv) d'un cadre de la construction S est :

$$\text{similarité}(v, Sv) = \frac{\sum_i^n v_i s v_i}{\sqrt{\sum_i^n v_i^2} \sqrt{\sum_i^n s v_i^2}} \quad (4.3)$$

Le produit scalaire entre les deux vecteurs donne un score entre 0 (deux vecteurs indépendants; profils pas du tout reliés) et 1 (deux vecteurs similaires; profils semblables). Le cas des vecteurs opposés (score = -1) est écarté, car il n'est pas applicable dans ce contexte de similarité des profils sémantiques.

Pour comparer un cadre C qui a été extrait à partir d'une phrase comme "Adam eat bread" avec une construction qui contient des cadres fondés sur des phrases telles que :

- Mommy make sandwich.
- You play airplane.
- Dave see Mary.
- You want chocolate.

Nous allons élaborer une matrice des vecteurs qui correspondent aux profils sémantiques de la construction S, avec une dernière colonne qui correspond au profil de l'actant du cadre C.

	sandwich	airplane	Mary	chocolate	bread
snack food	1				
dish	1				
nutriment	1				
baked					1
goods					1
food	1			1	1
substance	1			1	1
aircraft		1			
craft		1			
vehicle		1			

conveyance		1			
instrumentality		1			
wife			1		
woman			1		
female			1		
person			1		
organism			1		
causal agent			1		
spouse			1		
domestic partner			1		
artifact					
brown				1	
cocoa				1	
solid				1	1

En utilisant l'équivalent de cette matrice et les équations 4.2 et 4.3, le modèle calculera le score de similarité entre le cadre C et la construction S. Par la suite, il utilise l'équation 4.1 pour comparer le score avec ceux des autres constructions, en sélectionnant le arg max.

Nous allons adapter l'équation 2.6 pour calculer la probabilité conditionnelle d'un profil sémantique représenté par un vecteur de propriétés v :

$$P_i(v|S) = c(\text{similarité}(v, S) + \alpha \cdot |v|) / (\sum_l^n \text{similarité}(v_l, S) + \alpha \cdot d_u + (1 - c))$$

(4.4)

Tel que similarité(v,S) est le score de similarité entre le profil sémantique représenté par le vecteur v et une construction S et similarité(v_u,S) est le score de similarité entre le

profil sémantique (vu) qui représente l'unité lexicale u , et une construction S . Le paramètre de lissage du représente le nombre total de valeurs que l'unité lexicale u peut avoir. En outre, pour aligner l'intégralité des distributions de probabilités à une proportion commune, nous ajustons chaque probabilité sur l'ensemble des unités lexicales. Le paramètre α est une constante ajustée selon les expérimentations. Nous utilisons le principe de lissage additif (Manning et al., 2008; Jurafsky et Martin, 2008), en ajoutant une valeur constante à chaque occurrence d'événement pour éviter d'avoir un événement avec une probabilité nulle. Puisque le modèle va progresser vers l'utilisation des cadres dans leur contexte convenable, les cadres de la LS seront favorisés dans le cas d'ALS. Par conséquent, nous allons adopter la probabilité de couplage $c = 0.5$ pour ces derniers. De même, nous utilisons la probabilité de couplage indiquée par Anderson (1991) $c = 0.3$ dans le cas des cadres de la L1 pour orienter le modèle vers l'élaboration des constructions de la LS comme dans le cas d'un apprentissage réel de la LS.

Pour calculer la probabilité conditionnelle d'un vecteur v nous allons utiliser la formule suivante :

$$Pv(va|s) = c(t_{va}^s(i) + \alpha) / (n_s + \alpha \cdot d_v + (1 - c)) \quad (4.5)$$

Où $t_{va}^s(i)$ est le nombre de cadres de la construction S qui ont un actant avec un profil sémantique i qui a la valeur va . En outre, n_s représente le nombre de cadres de la construction S .

4.5 Extraction de données

4.5.1 Extraction de données pour les intrants en anglais

Comme nous l'avons indiqué dans le chapitre précédent, nous avons extrait les intrants en anglais à partir de la base de données *TalkBank*, plus spécifiquement les corpus de données sont *PAROLE* (Hilton, 2008) et *BELC* (Muñoz, 2006).

À partir de ces deux corpus, nous avons suivi certaines étapes pour créer les intrants anglais de notre modèle. Une description détaillée de ces corpus et des étapes d'extraction de données a été présentée dans la section 3.5.1.

Pour l'extraction des primitives sémantiques en anglais, nous avons créé un outil automatique qui utilise des bases de données SQL serveur de *Wordnet*, *VerbNet* et *WordNet eXtended*. Nous avons construit une liste de primitives sémantiques pour chaque mot à partir des Synonymes/Hyponymies de WordNet selon son contexte dans le texte.

4.5.2 Extraction de données pour les intrants en français

Pour les intrants en français nous avons utilisé le corpus du français parlé de *Kate Beeching* (Beeching, 2013). Nous avons suivi certaines étapes pour créer les intrants qui permettront des simulations probabilistes d'acquisition des rôles sémantiques. Une description détaillée de ce corpus et des étapes d'extraction de données a été présentée dans la section 3.5.2. L'outil d'extraction des traits sémantiques des mots en français est fondé sur Wolf. En outre, nous avons utilisé CRISCO (Gaëlle, 2011) pour extraire plus d'hyperonymes/synonymes et compléter ceux qui ont été extraits avec Wolf.

4.6 Simulations probabilistes

Nous allons présenter des simulations probabilistes d'acquisition des profils sémantiques pour des positions d'actants dans le cas d'ALS. Ces profils sémantiques représentent des propriétés des rôles sémantiques conventionnels. Cependant, dépendamment des intrants de la L1 et la LS, et des primitives sémantiques du verbe, une position d'un actant peut avoir plus qu'un profil. D'un autre côté, ces simulations démontrent les processus cognitifs de généralisation des profils sémantiques chez les apprenants de la LS. De même, elles démontrent l'utilisation de ces profils dans des situations de production, de compréhension et de résolution d'ambiguïté entre les profils L1 et LS.

La formation des profils sémantiques pour les rôles chez les apprenants de la LS est liée à deux facteurs principaux qui sont les profils sémantiques précédemment établis pour la

L1 et la qualité et le nombre des intrants de la LS (Dittmar, 1990). Les apprenants de la LS établissent des nouvelles associations entre les positions des actants et les profils sémantiques de la LS, enracent les associations LS compatibles et homogène avec la L1 et reforment les associations LS qui ne correspondent pas à ceux de la L1.

Notre modèle simule l'acquisition des rôles sémantiques de la LS par la généralisation des profils sémantiques à partir des intrants et en calculant les probabilités des propriétés sémantiques appropriées à une position d'un actant donné.

Après une première étape d'apprentissage du français L1, nous avons exposé, graduellement, à notre modèle 500 paires d'intrants anglais LS, ensuite nous avons présenté les paires suivantes, extraites de (White et al., 1999), dans trois différentes expérimentations. Nous avons supprimé les rôles sémantiques des actants dans chaque expérimentation.

- **S:** ADMIRE_[adore, appreciate](ADAM_{<?>}, EVE_{<?>})
- **E:** Adam admires Eve
- **S:** FRIGHTEN_[affright,daunt](DOG_{<?>}, EVE_{<?>})
- **E:** The dog frightens Eve
- **S:** MISS_[want,yearn](ADAM_{<?>}, EVE_{<?>})
- **E:** Adam misses Eve

Le modèle calculera le profil (et par conséquent le rôle) sémantique pour chacun des actants, pour montrer l'apprentissage des propriétés sémantiques par le modèle pour une position d'actant à partir des constructions de la L1 et de la LS. Nous avons effectué des simulations pour des volumes d'intrants différents (10, 50, 100, 200, 300,400 et 500) et nous avons calculé la moyenne des résultats de 10 simulations différentes avec des intrants générés aléatoirement pour chaque simulation.

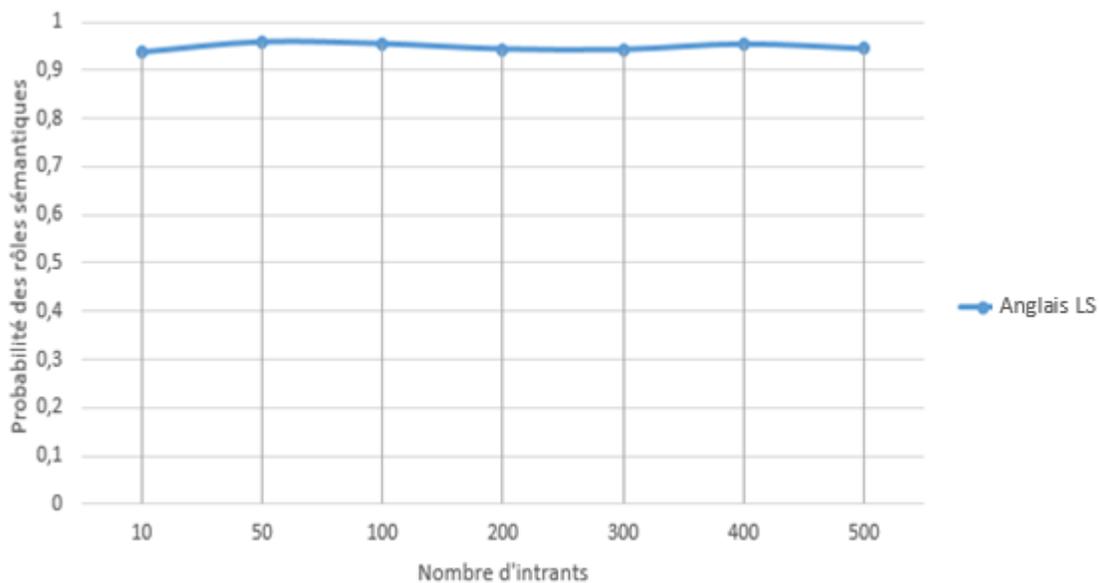


Figure 4.1: Acquisition des rôles sémantiques des actants du verbe 'admire'.

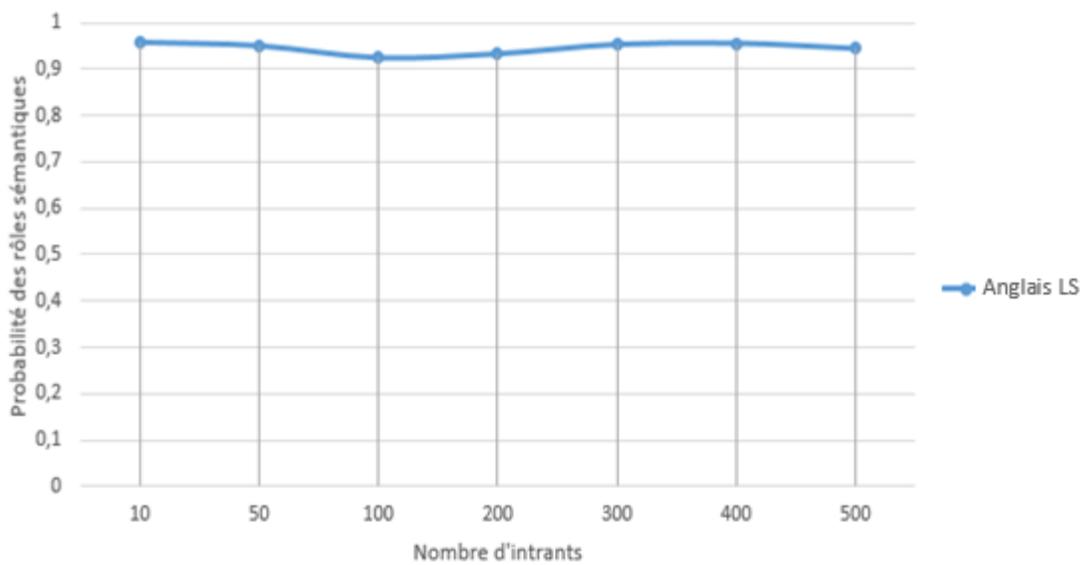


Figure 4.2: Acquisition des rôles sémantiques des actants du verbe 'frighten'.

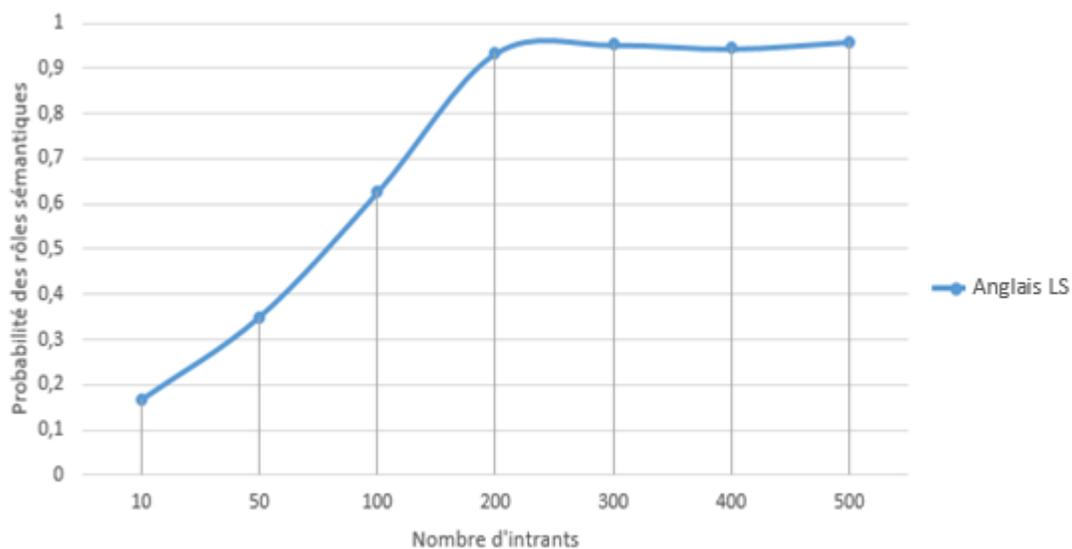


Figure 4.3: Acquisition des rôles sémantiques des actants du verbe ‘miss’.

Les figures 4.1 et 4.2 illustrent les courbes d’apprentissage des profils/rôles sémantiques des actants des verbes transitifs pour l’anglais LS. Ces verbes ont des caractéristiques sémantiques et syntaxiques similaires à ceux de la L1, ce qui explique le score élevé de la probabilité du profil sémantique indépendamment du nombre d’intrants. Les propriétés sémantiques calculées (dans une première étape d’apprentissage) dans cette expérimentation pour le premier actant du verbe ‘admire’ sont {male, person, organism, living thing, object, entity, causal agent}, tandis que les propriétés du deuxième actant sont {wife, woman, female, person, organism, living thing, object, entity, causal agent, spouse, relative, domestic partner}. En utilisant ces propriétés sémantiques et les primitives sémantiques du verbe, le modèle a pu calculer les rôles sémantiques du premier et du deuxième actant, respectivement (Expérient, Stimulus) avec une certitude de plus de 0.9 pour chaque simulation et avec des nombres d’intrants différents de la LS.

Pour le verbe ‘frighten’, les propriétés sémantiques calculées (dans la première étape d’apprentissage) pour les deux actants étaient {canine, carnivore, placental, mammal, vertebrate, chordate, animal, organism, living thing, object, entity} et { wife, woman, female, person, organism, living thing, object, entity, causal agent, spouse, relative,

domestic partner}. Les probabilités calculées ont donné les rôles sémantiques (Stimulus, Expérient), respectivement pour le premier et le deuxième actant. Ces résultats montrent que les rôles sémantiques des verbes transitifs de la LS qui ont des primitives sémantiques et propriétés communes avec la L1, peuvent être appris rapidement par les apprenants de la LS dès les premiers stades d'apprentissage. Dans ce cas, les intrants de la LS aident seulement à enraciner les propriétés de la L1, précédemment établies, pour les actants de la LS.

La figure 4.3 montre la courbe d'apprentissage des profils/rôles sémantiques des actants dans le cas d'un cas grammatical différent de la L1. À titre d'exemple, pour des apprenants dont la L1 est le français, l'expérient de la phrase « Ève manque à Adam » est en mode datif, autrement dit, il est l'objet indirect (préposition 'à') et il occupe la position du deuxième actant du verbe 'manquer'. Apprendre les rôles sémantiques du verbe 'miss' dans un contexte d'anglais LS constitue un défi pour ces apprenants, car l'expérient dans ce cas est en mode nominatif et il occupe la première position. Par conséquent, un apprenant francophone qui entend l'énoncé 'Adam misses Eve' et en considérant les primitives sémantiques du verbe 'miss', prédit les rôles sémantiques des actants avec une certitude très faible (0,16) au premier stade d'apprentissage. Par la suite, la probabilité des rôles sémantiques adéquats augmente graduellement et proportionnellement avec le nombre d'intrants jusqu'à une certitude supérieure à 0.9. Cette certitude montre une correction grâce à la preuve positive d'une surgénéralisation des propriétés de la L1 sur la LS. La correction de la surgénéralisation est enracinée par le nombre d'intrants de la LS. Cette expérimentation montre que l'apprentissage est effectué selon une courbe 'U', ou l'apprenant commence avec un score élevé qui reflète son succès de calcul des profils sémantiques de la LS pour des modèles identiques à ceux de sa L1. Cependant, ce score devient très faible en rencontrant des modèles différents de sa L1, tels que l'expérient qui peut être datif dans la L1 et nominatif dans la LS. Les résultats de l'expérimentation ont montré que ce score monte encore une fois avec plus d'exemples de la LS, et l'apprenant arrive à un niveau de maîtrise de la LS pareil à celui de la L1 dans une étape avancée d'apprentissage. En outre, cette

expérimentation confirme que d'autres facteurs indépendamment de celui des profils sémantiques jouent un rôle principal dans la prédiction des rôles sémantiques appropriés. Les primitives sémantiques du verbe, la position des actants dans la L1 et la LS, le cas grammatical de chaque actant et les prépositions peuvent tous influencer cet apprentissage. À titre d'exemple, les actants du verbe 'miss' peuvent avoir les mêmes propriétés sémantiques dans le cas de la L1 et la LS, cependant, ils n'ont pas le même cas grammatical ni les mêmes positions dans les deux langues.

Dans une deuxième simulation, nous avons calculé les profils sémantiques du premier et du deuxième actant du verbe 'like' avec différents actants de l'ensemble d'intrants. Après une première étape d'apprentissage du français L1, nous avons exposé, graduellement, à notre modèle 500 paires d'intrants anglais LS, ensuite nous avons présenté des paires scène/énoncé de la LS. Les rôles sémantiques des actants ont été supprimés dans toutes les expérimentations.

Avec les mots suivants (extraits de notre fichier mots-utilisés), qui jouent respectivement les rôles du premier et du deuxième actant, et selon leurs fréquences dans le corpus et l'ensemble d'intrants, les résultats des profils sémantiques passent du spécifique au plus général proportionnellement aux nombres d'intrants.

```
FIDX: 1
$1
:doggie:1
:you:17
:mommy:4
:we:1
:she:4
:he:3
:eve:5
:mary:1
:i:11
:they:1
$2
:music:4
```

```

:it:7
:you:7
:movie:5
:god:5
:soda:2
:elephant:3
:study:2
ENDF

```

Dans cette simulation, le profil sémantique du premier actant à passé de {person, organism, living thing, object, entity, causal agent} avec un nombre d'intrants de 10 à {entity, object, being, organism, living thing, animate thing, quality, artifact, whole thing, unit, animal, animate being, ...} avec un nombre d'intrant de 500, en passant par des étapes intermédiaires selon le nombre d'intrants du modèle.

De même, le profil sémantique du deuxième actant est passé de {artifact, object, entity, whole} avec un nombre d'intrants de 10 à {entity, object, physical object, artifact, whole, whole thing, unit, instrumentality, quality, communication, social relation, relation, abstraction, feature, belief, content, cognition, psychological feature, compound, substance, animal, organism, living thing, discipline, field, subject, subject area, subject field, ability, power, ...} avec un nombre d'intrants de 500.

Dans une troisième série d'expérimentations, et pour isoler le rôle des primitives sémantiques des verbes et celui des positions des actants, dans le cas de la LS, nous avons exposé à notre modèle les paires fondées sur des verbes psychologiques similaires à ceux de la L1 :

- **S:** FRIGHTEN_[affright,daunt](DOG_{<?>}, EVE_{<?>})
- **E:** The dog frightens Eve
- **S:** FEAR_[care,concern](DOG_{<?>}, EVE_{<?>})
- **E:** The dog fears Eve
- **S:** SCARE_[affright,daunt](EVE_{<?>}, DOG_{<?>})
- **E:** Eve scares the dog
- **S:** LIKE_[appreciate](EVE_{<?>}, ADAM_{<?>})

- **E**: Eve likes Adam
- **S**: PLEASE_[satisfy,charm](EVE_{<?>}, ADAM_{<?>})
- **E**: Eve pleases Adam

Nous avons obtenu les résultats suivants en calculant les probabilités des rôles sémantiques :

	10 intrants				500 intrants			
	Actant1		Actant2		Actant1		Actant2	
	Expérient	Stimulus	Expérient	Stimulus	Expérient	Stimulus	Expérient	Stimulus
The dog frightens Eve		> .9	> .9			> .9	> .9	
The dog fears Eve	> .9			> .9	> .9			> .9
Eve scares the dog		> .9	> .9			> .9	> .9	
Eve likes Adam	. > .9			> .9	> .9			. > .9
Eve pleases Adam		> .9	> .9			> .9	> .9	

Ces résultats montrent que le modèle prédit facilement les rôles sémantiques appropriés pour les verbes psychologiques similaires à ceux de la L1 avec une certitude entre 0.9 et 1 dès le premier stade d'apprentissage (10 intrants). En effet, cela montre l'aisance avec laquelle acquièrent les apprenants de la LS les rôles sémantiques pour les verbes qui ont les mêmes caractéristiques sémantiques et syntaxiques que ceux de leur L1. Le modèle est donc capable d'utiliser les associations apprises entre les propriétés sémantiques et les positions des actants de la L1 pour guider de manière appropriée l'acquisition de la LS.

4.7 Discussion des résultats

Les simulations présentées dans la section précédente ont démontré la capacité du modèle et par conséquent celle de l'apprenant de la LS à prédire des rôles sémantiques et à généraliser des profils sémantiques des actants en passant d'un ensemble de propriétés spécifiques à un ensemble plus général dans un stade avancé d'apprentissage. Les constructions de la LS qui ont des propriétés en commun avec celles de la L1 jouent un rôle important dans l'apprentissage de différents types de verbes de la LS tels que les verbes transitifs et psychologiques. À titre d'exemple, le profil sémantique d'un actant qui occupe la première position dans une construction transitive d'un verbe psychologique a les propriétés d'une entité animée et le modèle prédit son rôle sémantique (Expérient ou Stimulus) dès la première étape d'apprentissage de la LS en se fondant sur les primitives sémantiques du verbe qui sont les mêmes dans la L1.

Pour les verbes qui ont des propriétés différentes de la L1, tel qu'un actant qui peut être datif dans la L1 et nominatif dans la LS, le modèle commence avec un score très faible dans les premières étapes d'apprentissage et qui s'améliore avec le nombre d'intrants de la LS. Cette faible probabilité de prédiction des rôles sémantiques appropriés est le résultat de la surgénéralisation des caractéristiques syntaxiques non adéquates à la LS. Cette étape de surgénéralisation est suivie d'une étape de correction ce qui correspond à une courbe d'apprentissage 'U'.

En ce qui concerne les verbes psychologiques, les résultats sont compatibles avec ceux de (White et al., 1999), qui montrent l'aisance d'apprentissage des propriétés des verbes psychologiques dont les caractéristiques sont similaires à celles de la L1 pour des apprenants japonais et malgaches d'anglais LS. En testant la compétence linguistique de ces apprenants dans différentes étapes d'apprentissage, les résultats ont montré une grande différence dans l'acquisition des propriétés et des actants des verbes psychologiques selon la ressemblance de ces derniers avec la L1. Ces résultats sont, de même, compatibles avec ceux de (Dittmar, 1990) qui a étudié les aspects sémantiques de l'apprentissage des adultes en montrant les associations établies entre les propriétés

sémantiques syntaxiques. L'étude a montré que dans les premières étapes d'apprentissage les rôles sémantiques ne se laissent pas positionner dans une sémantique limitée ou dans une approche conceptuelle ce qui génère un statut d'ambiguïté, spécifiquement pour les verbes qui ont des caractéristiques syntaxiques différentes de la L1.

D'un autre côté, les résultats confirment que d'autres facteurs sont pris en considération dans l'acquisition des rôles sémantiques avec les profils sémantiques des actants. Les primitives sémantiques des verbes et les positions de ces derniers jouent un rôle primordial dans la prédiction des rôles sémantiques adéquats par les apprenants de la LS. Par conséquent, le modèle prédit aussi les rôles sémantiques des nouveaux actants en calculant leurs propriétés à partir des éléments précédemment mentionnés.

Les différentes étapes d'apprentissage pour le modèle (10, 50, 100, 200, 300, 400 et 500) montrent que le facteur de la compétence linguistique des rôles sémantiques dépend principalement des similitudes et des différences entre la L1 et la LS. Lorsque les verbes de la L1 et la LS possèdent les mêmes caractéristiques syntaxiques et sémantiques, le niveau de compétence linguistique représenté par la certitude de prédiction des rôles sémantiques est très élevé dès la première étape d'apprentissage (10 intrants). Contrairement à cela, dans le cas des verbes dont les différences syntaxiques et sémantiques sont visibles, une amélioration du niveau de compétence linguistique est montrée dans les premières étapes d'apprentissage et elle se stabilise à partir d'un nombre d'intrants égal à 200.

4.8 Conclusion

Nous avons montré par des expérimentations réalisées sur l'acquisition des rôles sémantiques de la LS le rôle des profils sémantiques, des positions des actants, le cas grammatical de chaque actant et les primitives sémantiques des verbes dans cette acquisition. En outre, nous avons illustré le rôle de la généralisation des profils sémantiques et la façon avec laquelle émergent les propriétés sémantiques vers des

propriétés plus générales proportionnellement avec le nombre d'intrants du modèle. En effet, les associations probabilistes entre ces propriétés et les positions syntaxiques des actants permettent au modèle d'apprendre des conceptions générales des rôles sémantiques sans avoir besoin d'un étiquetage explicite de ces derniers. La prédiction des rôles sémantiques appropriés dépend aussi des constructions de la L1 et de la similitude entre les propriétés des verbes de la L1 et de la L2. Dans le cas des cadres similaires, l'acquisition de ces rôles est plus rapide et le degré de certitude est élevé. Les propriétés des profils sémantiques ont été extraites à partir de Wordnet, VerbNet, WordNet eXtended, Wolf et CRISCO pour éviter un usage explicite des rôles sémantiques par le modèle. La généralisation des profils sémantiques permet d'apprendre les rôles sémantiques des verbes qui n'ont pas d'équivalents exacts dans la L1, tels que les verbes psychologiques dont les actants possèdent des cas grammaticaux différents dans la L1 et la LS. Cependant, cette généralisation doit être accompagnée d'autres éléments tels que les primitives sémantiques du verbe. Cependant, la prédiction des rôles sémantiques par le modèle est fondée sur les constructions établies préalablement et elle n'utilise pas les connaissances spécifiques au verbe telles que les préférences de sélection. Pour améliorer la capacité de notre modèle à utiliser les préférences de sélection d'un verbe et les propriétés sémantiques que ce dernier peut donner à ces actants, nous allons modifier le modèle et montrer des résultats des simulations dans le chapitre suivant.

En outre, comme perspective de cette thèse, nous pouvons étudier l'influence des verbes très fréquents dans l'acquisition des rôles sémantiques des autres verbes d'une LS. De même, l'influence des propriétés spécifiques à un événement telles que le changement d'état, le thème supplémentaire, l'actant causalement affecté par l'événement ou relatif au mouvement, peuvent être étudiés pour élucider de plus les processus cognitifs impliqués dans cette acquisition.

CHAPITRE V

PRÉFÉRENCES DE SÉLECTION DES VERBES ET PROPRIÉTÉS SÉMANTIQUES
IMPOSÉES PAR CES DERNIERS DANS L'ALS

5.1 Introduction

Les prédicats, et plus spécifiquement les verbes imposent des restrictions et montrent des préférences de sélection pour leurs actants (Resnik 1996; Chomsky 1968). À titre d'exemple, le verbe 'eat' a des préférences pour les actants qui ont des propriétés telles que 'comestible' et 'solide', alors que 'drink' a des préférences pour les actants avec les propriétés 'liquide' et 'potable'. L'inverse de ces préférences en donnant la propriété 'potable' pour 'eat' ou les propriétés 'comestible' et 'solide' pour 'drink' n'est pas accepté. En comparant les exemples suivants, un apprenant anglophone du français LS ne trouvera pas de difficulté pour sélectionner le 'Patient' (actant) le plus adéquat pour le verbe 'griller' (mêmes primitives sémantiques que 'grill'). Ce dernier sélectionnera avec un degré d'acceptation très élevé l'actant 'côtelettes', avec un degré d'acceptation moins élevé l'actant 'feu rouge' et avec un degré d'acceptation égal à 0 l'actant 'montagne'.

1. Elle grille des côtelettes.
2. Elle grille un feu rouge.
3. Elle grille une montagne.

La préférence de sélection de l'actant 'côtelettes' est déjà élevée à cause du système de conceptualisation de l'apprenant de la L1 qui contient auparavant des référents et des représentations internes qui correspondent aux concepts de la LS (Lakoff, 1994). Cependant, la préférence de sélection pour l'actant 'feu rouge' s'augmentera

progressivement avec le nombre d'intrants. Dans ce cas, l'apprenant découvrira avec le temps que le verbe 'grille' a les mêmes préférences de sélection que le verbe 'run' dans sa langue native, dans un contexte tel que 'She ran a red light'. En outre, la probabilité que cet apprenant reçoive un intrant qui appuiera le degré d'acceptation de l'actant 'montagne' pour le verbe 'grille' est très minime.

Selon (Fodor et Katz, 1964) les préférences de sélection sont enracinés avec des fonctions booléennes en utilisant des propriétés sémantiques telles que 'human' ou 'higher animal'. Ces préférences de sélection sont considérées comme des restrictions sur l'application des prédicats sur les actants. D'un autre côté, (Resnik, 1993) considère que l'apprentissage des mots et des concepts est représenté par une hiérarchie sémantique de classes, et les liaisons statistiques sont adoptées pour apprendre les préférences de sélection à partir des intrants. À titre d'exemple, les préférences de 'eat' ne sont pas représentées comme des propriétés d'une classe, mais plutôt comme une distribution de probabilité sur ces propriétés. Par conséquent, les modèles computationnels probabilistes sont les plus appropriés pour simuler l'acquisition des préférences de sélection en calculant le degré d'acceptation d'un actant par rapport à un prédicat.

L'étude de (Gleitman et Gillette, 1995) a démontré que les apprenants adultes peuvent prédire à quatre-vingts pour cent un verbe en s'appuyant sur sa structure actancielle et que les préférences de sélection apportent des informations importantes sur les sens des verbes inconnus. En effet, les propriétés sémantiques des actants d'un verbe et ces préférences de sélection permettent aux apprenants d'une LS d'acquérir les sens des verbes et les associations sémantiques entre ces verbes, leurs actants et les positions de ces derniers.

Le modèle (Resnik, 1996) simule l'acquisition des préférences de sélection par des apprenants en prenant en considération les intrants du modèle et le système conceptuel des apprenants. Cependant, ce modèle ne montre pas l'influence d'augmentation de

nombre d'intrants dans cette acquisition. De même, il suppose que cette dernière est réalisée grâce aux représentations hiérarchiques des propriétés sémantiques ce qui est un point de différence entre les théories existantes.

Dans ce chapitre, nous allons montrer des expérimentations probabilistes qui simulent l'acquisition des préférences de sélection dans le cas de l'ALS. Ces expérimentations sont une extension de celles réalisées dans le chapitre précédent pour simuler l'apprentissage progressif des profils et rôles sémantiques des actants selon le nombre d'intrants. Ces profils sont des propriétés spécifiques au verbe. Ils se transforment en représentations abstraites qui expriment les propriétés générales des actants. Nous avons adapté notre modèle pour qu'il prenne en considération les connaissances spécifiques aux verbes dans l'ALS, plus spécifiquement dans l'acquisition des rôles sémantiques et des préférences de sélection de verbes pour chaque actant. Le modèle sera capable de prédire et d'utiliser les préférences de sélection et de généraliser les propriétés observées sur des nouveaux actants. Les propriétés sémantiques sont extraites de l'ontologie lexicale WordNet. Pour les apprenants de la LS, l'apprentissage des préférences de sélection suit une courbe U en passant par une étape de surgénéralisation des règles. Cette surgénéralisation est causée par l'utilisation des préférences de sélection des faux amis anglais-français. Prenons l'exemple du verbe 'advance' qui est utilisé dans un contexte de progression, et dont les préférences de sélection 'progressif' et 'développable' et qui est peut-être utilisé à la place du verbe 'avancer' dont les préférences de sélection 'destination' et 'direction'. La section 5.2 présente certaines théories computationnelles liées aux préférences de sélection et leur lien avec l'ALS. Les modèles computationnels de préférences de sélection sont examinés dans la section 5.3. Dans la section 5.4, nous allons présenter une adaptation de notre modèle computationnel à la représentation, l'acquisition et l'utilisation des préférences de sélection des verbes. Nous allons montrer l'extraction de données dans la section 5.5 et nous allons présenter les résultats de nos expérimentations dans la section 5.6. Dans la section 5.7 nous allons discuter les résultats et les différentes étapes d'acquisition des

préférences de sélection. Nous allons présenter les résultats de nos contributions dans la section 5.8.

5.2 Théories et approches d'acquisition des préférences de sélection des verbes

Plusieurs théories ont tenté d'expliquer l'acquisition des préférences de sélection qui sont considérées comme des contraintes sur l'applicabilité des prédicats aux actants. Les théories sémantiques de (Fodor et Katz, 1964) ont considéré les préférences de sélection comme des restrictions booléennes pour limiter le nombre d'actants qui peuvent être choisis pour un prédicat. À titre d'exemple, bien que le mot 'Java' puisse être interprété comme le nom d'une île ou un langage de programmation, seul le dernier sens sera considéré dans la phrase 'Adam programme avec Java'. Dans ce contexte, le verbe 'programmer' exige que son actant ait la propriété 'langage de programmation'. Cependant, cette théorie reste limitée lorsque le contexte est très général et le choix d'actants est riche, pour identifier les caractéristiques sémantiques et les conditions pour lesquelles un actant est acceptable. Les théories lexicales de (Jackendoff, 1983; Pustejovsky, 1995) considèrent les préférences de sélection comme des propriétés intégrées dans la grammaire et liées au contexte du prédicat, ce qui rend la sélection de ces propriétés une tâche considérable. Les théories de (Resnik 1993; 1996) combinent les méthodes statistiques et à base de connaissance. C'est une approche probabiliste qui représente les préférences de sélection comme une hiérarchie de classes sémantiques prédéfinies et calcule leurs probabilités à partir des intrants d'un corpus. Elle prédit le comportement du prédicat et calcule la probabilité qu'un actant se produise comme argument d'un prédicat indépendamment de l'identité de ce dernier. À titre d'exemple, l'approche confirme que par défaut le sujet d'un verbe aura une probabilité plus élevée pour les propriétés { force, organization, social group, group, male, person, organism, living thing, object, entity, causal agent} que pour les propriétés {equine, odd-toed ungulate, ungulate, placental, mammal, vertebrate, chordate, animal, organism, living thing, object, entity}. Cependant, lorsque nous considérons l'identité du prédicat, le deuxième ensemble des propriétés aura une probabilité élevée si le verbe est 'hennir',

tandis que le premier ensemble aura une faible probabilité dans ce cas. Cette probabilité qui est sous forme d'entropie relative (Kullback et Leibler, 1951) est la différence entre une distribution conditionnelle/postérieure et une distribution préalable/antérieure. Elle est calculée pour mesurer la différence entre les distributions antérieures et postérieures. Par conséquent, les préférences de sélection ne sont pas booléennes, mais plutôt des distributions de probabilités sur différentes classes sémantiques. À titre d'exemple, les préférences de sélection d'un prédicat pour un actant donné ne sont pas représentées par un seul mot (vrais ou faux), mais par un vecteur de propriété (une distribution de probabilités). Certaines autres théories (Gleitman et Gillette; 1995) montrent que dans le cas des apprenants adultes, les préférences de sélection aident à prédire le sens et les primitives sémantiques des nouveaux verbes. Cette théorie est applicable dans le cas d'ALS où un apprenant adulte acquiert des nouveaux verbes qui ont des sens inconnus et des actants avec des profils sémantiques similaires à ceux de la L1. (Gleitman et Gillette, 1995) confirment que les propriétés syntaxiques et sémantiques sont indispensables dans cette acquisition. Par conséquent, l'absence de l'une de ces caractéristiques nuit à l'apprentissage des sens des verbes. En outre, ils affirment que les préférences de sélection des verbes sont très importantes dans l'acquisition des sens des nouveaux verbes en comparant avec les propriétés syntaxiques d'un prédicat ou les associations sémantiques entre ce dernier et ces actants. Ces préférences jouent un rôle important pas seulement dans l'acquisition des sens des verbes, mais aussi dans d'autres éléments du traitement de la langue tels que la clarification de l'ambiguïté potentielle qui résulte des faux amis entre la L1 et la LS et des autres ressemblances et différences entre ces dernières. Cette clarification est fondée sur la prédiction des actants les plus appropriés selon les préférences de sélection.

Pour rendre explicites les processus cognitifs d'acquisition de ces préférences dans le cas de la LS nous allons adapter notre modèle et réaliser quelques simulations probabilistes que nous allons présenter dans les sections suivantes.

5.3 Modèles computationnels d'acquisition des préférences de sélection des verbes d'une LS

Plusieurs modèles computationnels ont été fondés sur le principe de préférence de sélection pour simuler l'acquisition de ces dernières dans un contexte de la LS, la désambiguïsation des sens des verbes, l'acquisition de connaissances lexicales, l'acquisition automatique de sens, la classification des rôles sémantiques, la construction des ressources lexicales dans le cadre du TALN et la classification automatique des verbes.

(Xiaohua et al., 2010) ont proposé un modèle qui utilise un système d'étiquetage des rôles sémantiques pour modéliser l'utilisation des verbes dans un contexte spécifique. Le contexte d'utilisation du verbe rend l'acquisition des préférences de sélection une tâche non triviale. Le modèle a été développé pour approcher la question de l'acquisition des préférences de sélection des verbes pour les apprenants de l'anglais LS. Les chercheurs mentionnent que les verbes en anglais expriment des actions, des états d'être, des sentiments et des circonstances. À titre d'exemple, en comparant les verbes 'gagner' et 'obtenir', le premier peut indiquer un effort qui a été fourni par l'agent, contrairement au deuxième. En outre, pour la phrase "J'ai _ le médecin", plusieurs verbes tels que 'appeler', 'contacter', 'visiter' peuvent être utilisés dans ce contexte. D'autres verbes tels que 'manger' et 'boire' sont exclus. En ajoutant une phrase prépositionnelle telle que "par téléphone", le verbe 'visiter' sera exclu à son tour. D'un autre côté, les chercheurs mentionnent que d'autres études (Gui et Yang, 2002; Yi et al., 2008) ont révélé l'importance du contexte d'utilisation du verbe en indiquant que plus de trente pour cent des erreurs dans l'anglais LS par des apprenants chinois sont des erreurs liées aux préférences de sélection des verbes. Le modèle détecte et corrige automatiquement ce genre d'erreurs de sélection des verbes. Il utilise les caractéristiques locales telles que les associations n-grammes avec les fonctionnalités sémantiques qui modélisent explicitement le contexte d'utilisation du verbe. L'approche combine ces deux éléments dans un classifieur linéaire généralisé et les chercheurs confirment que cette dernière a atteint des performances très élevées.

Le modèle de (Patrick et al, 2006) capte les modèles des structures actanciennes en utilisant les préférences de sélection dont le but est de désambiguïser les sens des verbes. Les préférences de sélection permettent au prédicat de favoriser un petit ensemble d'actants pour chaque position de la structure actancielle. La désambiguïisation des sens des verbes est une tâche de classification qui précise le sens approprié de chaque verbe dans un contexte donné. Ce modèle utilise les adjonctions de verbes et les préférences de sélection fondées sur les fonctions multi-sémantiques pour enrichir les rôles sémantiques et désambiguïser les verbes. En conséquence, cette étude a démontré une amélioration considérable des résultats au niveau de désambiguïisation des sens des verbes.

(Messiant et al., 2010) ont conçu un modèle d'acquisition de lexique syntaxique et de catégorisation syntaxico-sémantique des verbes pour la langue française. Le modèle repère les données pour enrichir les lexiques existants, permet d'acquérir des données étiquetées et fournit des schémas de sous-catégorisation équilibrés. Ces schémas de sous-catégorisation sont plus spécifiques aux classes sémantiques des verbes français. Les chercheurs confirment que l'adoption des préférences de sélection a permis d'obtenir des classes différentes et plus précises. Dans le même contexte (Zapirain et al., 2013) ont élaboré un modèle pour la classification des rôles sémantiques. L'objectif était de diminuer l'influence de la limite des caractéristiques lexicales dans cette classification en adoptant les préférences de sélection apprises automatiquement. Les chercheurs confirment que l'utilisation des préférences de sélection permet au modèle de réduire le taux d'erreur de dix-sept pour cent avec une amélioration de quatre pour cent au niveau de l'étiquetage sémantique. En outre, cette utilisation a permis d'améliorer la performance du système dans des situations où les informations syntaxiques sont inadéquates ou incomplètes pour désambiguïser le rôle sémantique et le classer correctement.

Certains autres modèles (Li et Abe, 1998; Clark et Weir, 2002; Ciaramita et Johnson, 2000) représentent les préférences de sélection d'un verbe comme un ensemble de nœuds de WordNet avec une distribution de probabilités. Ils utilisent la distribution de probabilités pour repérer le meilleur ensemble pour chaque actant et verbe.

(Erk, 2007) propose un modèle qui considère les préférences de sélection d'une position d'un actant comme une somme pondérée des similarités entre cet actant et les actants observés pour cette position dans un corpus. Cependant, le rôle de nombre d'intrants dans le développement des représentations des préférences de sélection n'était pas démontré. En outre, la confirmation de la possession des apprenants de la LS d'une représentation hiérarchique complète des classes sémantiques peut exiger plus d'expérimentations computationnelles. De même, certaines autres recherches (Light et Greiff, 2002) confirment que les apprenants ne peuvent pas évaluer eux-mêmes les préférences de sélection acquises sans les preuves positives et négatives.

Les modèles différents des préférences de sélection des verbes ont été évalués souvent selon la désambiguïsation des sens des mots et plus spécifiquement ceux des verbes, l'étiquetage des rôles sémantique et l'acquisition des L1 et LS. En ce qui concerne l'ALS, des tâches telles que la simulation d'acquisition des préférences de sélection et l'analyse de processus impliqués dans la validation syntaxique des actants pour une classe de verbes, restent à éclaircir. Nous allons montrer quelques expérimentations liées à ces deux tâches dans les sections suivantes.

5.4 Adaptation du modèle bayésien à l'acquisition des préférences de sélection des verbes de LS

Nous allons utiliser notre modèle bayésien dans l'acquisition préférences de sélection des verbes de LS. Pour cela, nous allons adapter ce dernier pour qu'il puisse acquérir ces profils sémantiques. Le modèle apprend les structures actanciennes à partir des cadres des verbes, et en groupant ces derniers dans des constructions. Les actants de chaque verbe possèdent des profils sémantiques sous forme d'ensembles de propriétés sémantiques. Les propriétés sémantiques sont extraites de Wordnet 3.0, VerbNet 2.3 et XWordnet 2.0-1.1. Les cadres ont été extraits à partir des paires scène/énoncé qui ont été élaborées à partir des corpus PAROLE (Hilton, 2008), BELC (Muñoz, 2006) et corpus du français parlé de Kate Beeching (Beeching, 2013).

Un cadre d'un verbe contient le prédicat de ce dernier, le modèle syntaxique, les primitives sémantiques et les profils sémantiques des actants. Les cadres extraits pour un prédicat sont regroupés dans l'entrée lexicale de ce dernier. Les préférences de sélection d'un verbe pour une position d'un actant sont représentées par un profil sémantique qui est à son tour une distribution de probabilité sur les propriétés sémantiques appropriées. Deux éléments principaux sont utilisés dans le calcul des probabilités des préférences de sélection. Ces deux éléments sont l'enracinement des constructions d'un côté, et le profil sémantique du verbe d'un autre côté. Dans ce cas le vecteur des propriétés sémantiques (profil) n'est pas une généralisation des propriétés sémantiques des actants, comme c'était le cas dans la prédiction des rôles sémantiques de ces derniers, mais plutôt il est lié au verbe et il contient seulement les propriétés sémantiques des actants utilisés par ce dernier. D'un autre côté, les constructions les plus enracinées influencent directement le calcul des probabilités de ces profils sémantiques. Le profil d'un verbe v de la LS pour un actant **act** est calculé selon chaque propriété sémantique **ps** :

$$\text{Pact}(\text{ps}|v) = \sum_s P(s,v) \text{Pact}(\text{ps} |s,v) \quad (5.1)$$

Où, ps représente toutes les propriétés sémantiques possibles qu'un actant peut avoir, et s couvre toutes les constructions. La probabilité a priori $P(s,v)$ exprime la possibilité que la construction s contient le verbe v .

La probabilité $P(s,v)$, pour $s > 0$, est calculée selon la fréquence du verbe v dans la construction s , l'enracinement de la construction s par rapport au reste des constructions et la formule d'Anderson (1991) utilisée dans son modèle de catégorisation humaine :

$$P(s,v) = c \cdot n_s \cdot f_s^v / ((c \cdot n + (1 - c)) \cdot \sum_{s'} c \cdot n_{s'} / c \cdot n + (1 - c) \cdot f_s^v) \quad (5.2)$$

Pour aligner l'intégralité des distributions de probabilités a priori à une proportion commune, nous ajustons chaque probabilité sur l'ensemble de constructions. De même, une construction très enracinée joue un rôle principal dans le calcul de la probabilité a priori. L'enracinement de la construction s est calculé selon la formule n_s/n .

Tel que n est le nombre total des cadres observés et n_s est le nombre de cadres qui forment la construction S , tandis que f_s^v est la fréquence du verbe v dans la construction s . c représente la probabilité de couplage qui est une probabilité constante. Elle détermine le degré auquel les éléments préfèrent être regroupés. Son utilisation permet d'éviter la forte tendance d'un cadre à préférer ou à se préserver d'une construction existante. Lorsque deux objets proviennent de la même catégorie, cette probabilité ne dépend pas du nombre d'objets observés jusqu'à présent. Cette probabilité de couplage est très adéquate dans notre cas, là où le modèle va élaborer et utiliser deux ensembles de constructions (SL1 et SLS). Naturellement le modèle va progresser vers l'utilisation de chaque ensemble dans son contexte convenable. Pour cela, le SLS sera favorisé dans le cas d'ALS. Par conséquent, nous allons adopter la probabilité de couplage $c = 0.5$ lorsque $s \in \mathbf{SLS}$. En outre, nous utilisons la probabilité de couplage indiquée par Anderson (1991) $c = 0.3$ lorsque $s \in \mathbf{SL1}$ pour orienter le modèle vers l'élaboration des constructions de la LS comme le cas d'un apprenant.

La probabilité postérieure est calculée selon la formule suivante :

$$\text{Pact}(\mathbf{pr}|s, v) = c(t_{\mathbf{act}}^s(\mathbf{pr}, v) + \alpha) / (n_{sv} + \alpha \cdot d_{\mathbf{act-pr}} + (1 - c)) \quad (5.3)$$

Où $t_{\mathbf{act}}^s(\mathbf{pr}, v)$ est le nombre de cadres de la construction S qui contiennent le verbe v et qui ont un actant \mathbf{act} avec la propriété \mathbf{pr} . En outre, n_{sv} est le nombre de cadres qui font partie de la construction S et qui ont le verbe v . α , $d_{\mathbf{act-pr}}$ et $d_{\mathbf{act}}$ sont les facteurs de lissage. Où $d_{\mathbf{act-pr}}$ est le nombre de tous les actants possibles avec la propriété \mathbf{pr} .

La compatibilité entre un verbe v et un mot m pour que ce dernier puisse jouer le rôle de l'actant \mathbf{act} , est calculée avec la formule :

$$\text{COMPact}(v|m) = \log(\sum_s P(s, v) \text{Pact}(Sm|s, v)) \quad (5.4)$$

Où Sm est le vecteur des propriétés sémantiques pour le mot m .

5.5 Extraction de données

5.5.1 Extraction de données pour les intrants en anglais

De la même manière que pour l'acquisition des structures actancielles et les rôles sémantiques, nous avons extrait les intrants en anglais à partir des corpus de données *PAROLE* (Hilton, 2008) et *BELC* (Muñoz, 2006).

À partir de ces deux corpus, nous avons suivi certaines étapes pour créer les intrants anglais de notre modèle. Une description détaillée de ces corpus et des étapes d'extraction de données a été présentée dans la section 3.5.1.

Pour l'extraction des primitives sémantiques en anglais, nous avons créé un outil automatique qui utilise des bases de données SQL serveur de *Wordnet*, *VerbNet* et *WordNet eXtended*. Nous avons construit une liste de primitives sémantiques pour chaque mot à partir des Synonymes/Hyponymies de WordNet selon son contexte dans le texte.

5.5.2 Extraction de données pour les intrants en français

Pour les intrants en français nous avons adopté le corpus du français parlé de *Kate Beeching* (Beeching, 2013). Nous avons suivi les étapes mentionnées dans la section 3.5.2. L'outil d'extraction des traits sémantiques des mots en français est fondé sur Wolf. En outre, nous avons utilisé CRISCO (Gaëlle, 2011) pour extraire plus d'hyperonymes/synonymes et compléter ceux qui ont été extraits avec Wolf.

5.6 Simulations probabilistes d'acquisition des préférences de sélection d'une LS

Nous allons présenter des simulations probabilistes d'acquisition des préférences de sélection dans le cas d'ALS. Dans une première expérimentation, nous avons exposé à notre modèle 500 paires d'intrants de la L1 pour l'amener à une étape avancée d'apprentissage, ensuite nous lui avons présenté huit verbes transitifs du français LS, sans primitives sémantiques pour prédire le profil sémantique de l'objet direct (premier actant). Nous avons considéré le français comme LS et nous avons choisi deux verbes

qui ont un large intervalle de préférences de sélection (prendre, donner), deux verbes avec une ensemble très limité de préférences de sélection (manger, piloter), deux verbes qui sont considérés comme des faux amis entre le français et l'anglais (achever, demander) et deux verbes qui peuvent être utilisés dans des contextes différents dans la LS (entendre, jouer). Les profils sémantiques des actants de ces verbes sont affichés respectivement dans les tableaux 5.1, 5.2, 5.3 et 5.4. Dans ces tableaux, nous allons montrer les résultats après un entraînement avec 500 intrants de la LS et qui sont plus significatifs par rapport aux étapes précédentes d'apprentissage. Pour chaque expérimentation, nous calculons la moyenne de dix simulations différentes avec des ensembles d'intrants différents. En outre, pour chaque profil du premier actant, nous incluons les vingt-cinq premières propriétés sémantiques et qui possèdent généralement des probabilités représentatives. Ces tableaux montrent un classement des probabilités des propriétés sémantiques pour l'actant en première position de chaque verbe. Comme nous l'avons déjà vu dans le chapitre précédent, les propriétés sémantiques extraites à partir de WordNet peuvent être très générales telles que { entity, abstraction, object } ou très spécifiques telles que { baked, potable, comestible, game, digit, deliveryman }. Pour la majorité des verbes utilisés dans ces expérimentations, les propriétés générales sont associées à des probabilités plus élevées dans les profils sémantiques des actants qui représentent l'objet direct. Certains verbes tels que 'donner', 'demander' et 'prendre' ont des actants avec des profils sémantiques dont les propriétés générales représentent plus de cinquante pour cent. Dans le cas des verbes qui exigent des actants plus spécifiques tels que 'manger' et 'piloter' (tableau 5.2), les profils sémantiques contiennent plus de propriétés spécifiques telles que { food, nutriment, living thing, foodstuff, plant, repast, meal } ou { aircraft, airplane, aeroplane, team, battery }. Ces profils sémantiques ont été calculés selon les fréquences des verbes et des mots utilisés dans le corpus. À titre d'exemple, nous remarquons que la probabilité des actants liés à la propriété sémantique 'artifact' sont plus fréquents que ceux liés à la propriété 'team'. Le modèle calcule la probabilité de chaque propriété en prenant en considération l'usage des verbes ce qui reflète les situations d'apprentissages réelles des apprenants de la LS. Pour les

expérimentations avec des verbes qui représentent des faux amis entre le français et l'anglais tels que 'achever' et 'demander' les profils sémantiques calculés montrent que le modèle a pris en considération l'usage des verbes et des mots et non la similitude phonologique ou morphologique entre les verbes. Un exemple des profils sémantiques calculés respectivement pour 'achever' et 'demander' (tableau 5.3) est {artifact, artefact, communication, social relation, human action, human activity, relation, indication, activity, cause, state, instrumentality, instrumentation, event, being, living thing }, {job, business, production, product, task, assay, case study, book, paper, report, prepare, brochure, text, textbook, reference, mission, military mission, operation, support }. Ces profils montrent que les sens des verbes 'achever' et 'demander' sont ceux de 'to complete' et 'to ask' et pas les sens de 'to achieve' et 'to demand'. Dans des perspectives futures de cette thèse, nous allons adapter notre modèle pour qu'il puisse considérer les aspects phonologiques et morphologiques des verbes, ce qui nous permettra de simuler l'influence des faux amis dans le cas de calcul des profils sémantiques et des préférences de sélection des verbes en absence des primitives sémantiques de ces derniers. Dans le cas des verbes qui peuvent avoir des sens différents, selon le contexte d'utilisation tels que 'entendre' (Piron, 2004) et 'jouer', et après plusieurs stages d'apprentissages, le modèle calcule un profil sémantique représentatif des différents sens avec des probabilités qui montrent la fréquence des phrases qui représentent chaque contexte. À titre d'exemple, le verbe 'entendre' peut référer au sens relevant du domaine de l'audition ou de celui de la compréhension. Les préférences de sélection de ce dernier peuvent guider le choix du sens du verbe dans un contexte donné. (Piron, 2004) confirme que des modèles syntaxiques et des auxiliaires modaux spécifiques aident à spécifier le sens approprié du verbe selon le contexte d'utilisation. De même, pour le verbe 'jouer', que ce soit pour un contexte de jeu, de compétition, de théâtre ou autre, les préférences de sélection du verbe 'jouer' seront calculées selon son contexte d'utilisation en prenant en considération la fréquence mots utilisés avec ce dernier dans le corpus. Pour le verbe 'entendre', nous remarquons que la majorité des propriétés du profil sémantique sont des propriétés très spécifiques. En

outre, les propriétés qui ont un lien avec un contexte de 'l'audition' telles que { communication, auditory communication, speech, conversation, discussion, saying } ont eu des probabilités plus élevées que celles liées à un contexte de 'compréhension' telles que { issue, question, head, difficulty } ce qui reflète la fréquence des phrases liées à chaque contexte dans le corpus. En ce qui concerne le verbe 'jouer', des propriétés telles que { game, diversion, sport, athletic game, court game, board game, word game } ont eu des probabilités plus élevées que les propriétés { portrayal, personation, enactment, characterization } ce qui montre la fréquence du contexte de 'se distraire, s'amuser' ou 's'adonner à un jeu' par rapport à celui de 'imiter par jeu une personnalité ou une fonction'.

Prendre		Donner	
propriété	probabilité	propriété	probabilité
entity	0.021	entity	0.020
object	0.019	abstraction	0.020
whole	0.019	whole	0.019
whole thing	0.019	object	0.019
abstraction	0.018	thing	0.018
physical object	0.018	substance	0.018
artifact	0.016	relation	0.017
artefact	0.016	artifact	0.016
human activity	0.014	signal	0.015
human action	0.013	communication	0.015
finish	0.013	social relation	0.015
ending	0.013	material	0.014
conclusion	0.013	quantity	0.013
act	0.012	amount	0.013
cause	0.011	device	0.011
causal agent	0.011	organism	0.011
stay	0.011	living thing	0.011
instrumentality	0.011	piece	0.010
container	0.009	body part	0.009
hand tool	0.009	tissue	0.008
substance	0.008	nutriment	0.008
content	0.008	food	0.007
conveyance	0.008	meal	0.007
instrumentation	0.008	liquid	0.007
vehicle	0.006	fluid	0.007

Tableau 5.1: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'prendre' et 'donner'.

Manger		Piloter	
propriété	probabilité	propriété	probabilité
entity	0.023	whole	0.022
substance	0.023	entity	0.022
object	0.022	object	0.022
natural object	0.021	artifact	0.021
food	0.018	instrumentality	0.021
nutriment	0.018	onveyance	0.020
living thing	0.015	vehicle	0.017
foodstuff	0.015	craft	0.017
organism	0.015	aircraft	0.017
plant	0.015	heavier-than-air craft	0.015
plant part	0.015	public transport	0.015
plant organ	0.014	hand tool	0.013
vascular plant	0.014	edge too	0.013
herb	0.013	airplane	0.011
animal	0.011	aeroplane	0.011
invertebrate	0.011	drone	0.011
mollusk	0.010	pilotless aircraft	0.010
part	0.010	radio-controlled aircraft	0.009
repast	0.008	group	0.008
meal	0.008	device	0.008
whole	0.008	act	0.008
gramineous plant	0.008	activity	0.008
artifact	0.006	piece	0.008
snack food	0.006	team	0.008
baked goods	0.006	battery	0.008

Tableau 5.2: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'manger' et 'piloter'.

Achever		Demander	
propriété	probabilité	propriété	probabilité
act	0.022	entity	0.024
activity	0.021	object	0.024
line of work	0.018	physical object	0.023
occupation	0.017	abstraction	0.023
line	0.017	act	0.022
job	0.017	whole	0.021
business	0.017	whole thing	0.021
production	0.017	artifact	0.020
product	0.017	artefact	0.020
task	0.017	communication	0.020
assay	0.015	social relation	0.020
case study	0.015	human action	0.018
book	0.015	human activity	0.017
paper	0.015	relation	0.015
report	0.015	indication	0.015
prepare	0.013	activity	0.015
brochure	0.011	cause	0.015
text	0.011	state	0.014
textbook	0.011	instrumentality	0.011
reference	0.011	instrumentation	0.011
mission	0.009	event	0.011
military mission	0.006	being	0.011
operation	0.006	living thing	0.011
support	0.006	animate thing	0.011
assignment	0.006	organism	0.011

Tableau 5.3: Classement des probabilités des propriétés sémantiques de l’actant en première position des verbes ‘achever’ et ‘demander’.

Entendre		Jouer	
propriété	probabilité	propriété	probabilité
abstraction	0.024	whole	0.024
relation	0.024	entity	0.023
social relation	0.023	object	0.023
communication	0.023	artifact	0.020
auditory communication	0.019	instrumentality	0.020
speech	0.019	container	0.018
conversation	0.019	event	0.017
discussion	0.019	social event	0.017
saying	0.019	contest	0.017
expression	0.018	act	0.017
locution	0.018	activity	0.017
word	0.018	game	0.016
pronunciation	0.017	diversion	0.016
monologue	0.015	sport	0.015
module	0.013	athletic game	0.015
text	0.013	court game	0.014
rumor	0.009	board game	0.014
rumour	0.009	word game	0.013
bruit	0.009	portrayal	0.012
noise	0.009	personation	0.011
problem	0.009	enactment	0.011
issue	0.009	characterization	0.009
question	0.009	music	0.009
head	0.009	football game	0.007
difficulty	0.009	football	0.007

Tableau 5.4: Classement des probabilités des propriétés sémantiques de l'actant en première position des verbes 'entendre' et 'jouer'.

Dans une deuxième série d'expérimentations, nous avons simulé celles de (Xiaohua et al., 2010). Ces derniers ont proposé un modèle qui utilise un système d'étiquetage des rôles sémantiques pour modéliser l'utilisation des verbes dans un contexte spécifique. Le modèle a été développé pour approcher la question de l'acquisition des préférences de sélection des verbes pour les apprenants de l'anglais LS. Plusieurs chercheurs (Gui et Yang, 2002; Yi et al., 2008) mentionnent que plus de trente pour cent des erreurs en anglais LS par des apprenants chinois sont des erreurs liées aux préférences de sélection des verbes. Le modèle détecte et corrige automatiquement ce genre d'erreurs de sélection des verbes. Il utilise les caractéristiques locales, telles que les associations n-grammes avec les fonctionnalités sémantiques qui modélisent explicitement le contexte d'utilisation du verbe. L'approche combine ces deux éléments dans un classifieur linéaire généralisé et les chercheurs confirment que cette dernière a atteint des performances très élevées. La méthode est composée de trois étapes. Étant donné la saisie d'une phrase en anglais écrite par des apprenants de l'anglais LS, le système vérifie chaque verbe et génère des candidats de correction en remplaçant chaque verbe par un vecteur de caractéristiques sémantiques qui représente le contexte d'utilisation, ensuite, le candidat ayant le meilleur score est sélectionné comme résultat.

À titre d'exemple, pour la correction d'une phrase telle que 'Jane _ TV every day', le modèle de (Xiaohua et al., 2010) donne le résultat suivant :

Candidat à la correction	Score
Jane watches TV every day	10.8
Jane looks TV every day	0.8
Jane reads TV every day	0.2

De même, la nature probabiliste de notre modèle permet de prédire les meilleurs candidats qui peuvent jouer le rôle de l'objet direct le plus approprié dans des situations pareilles à celles de (Xiaohua et al., 2010). Cette caractéristique permet à notre modèle d'être utilisé comme outil d'aide à l'apprentissage en corrigeant des phrases produites par des apprenants de l'anglais LS ou du français LS.

Pour montrer cette capacité de notre modèle à prédire les objets directs appropriés en utilisant les préférences de sélection, nous avons exposé à notre modèle 500 paires d'intrants de la L1 pour l'emmener à une étape avancée d'apprentissage. Par la suite, nous avons passé un nombre de cadres (LS) différents à chaque étape d'apprentissage (10, 50, 100, 200, 300, 400 et 500). Après chaque étape d'apprentissage, nous avons exposé 186 intrants (Xiaohua et al., 2010) dont l'information du premier actant est supprimée pour calculer les probabilités des profils sémantiques et les classer selon le score le plus élevé.

La figure 5.1 montre les probabilités des profils sémantiques appropriés de l'objet direct des 182 intrants dans chaque étape d'apprentissage.

Ces résultats sont meilleurs que celles de (Xiaohua et al., 2010). Notre modèle, et à partir de 200 intrants, prédit avec un score de quatre-vingt-six pour cent les profils sémantiques appropriés de l'objet direct des 182 intrants de test. Les résultats du modèle de (Xiaohua et al., 2010) calcule le meilleur candidat avec un score de soixante-cinq pour cent en utilisant des corpus de taille énorme (LDC2005T12, verbes cibles du New York Times, OpenNLP31) et un système d'étiquetage des rôles sémantiques selon le principe de (Riedel et Meza-Ruiz, 2008).

L'avantage de notre modèle est de calculer les profils sémantiques appropriés sans utiliser un étiquetage explicite des rôles sémantiques, mais plutôt d'utiliser les propriétés sémantiques élémentaires pour construire des profils sémantiques pour chaque actant et de calculer, par la suite, le score des meilleurs profils qui peuvent être des bons candidats pour une position d'un actant donné. Dans le cas de notre modèle, et dans une première étape d'apprentissage (nombre d'intrants égal à 10), la probabilité des préférences de sélection était 0.48. Cela s'explique par le fait que les constructions de la L1 participent positivement ou négativement dans le calcul des profils sémantiques de la LS selon les semblances et les dissemblances entre les deux langues.

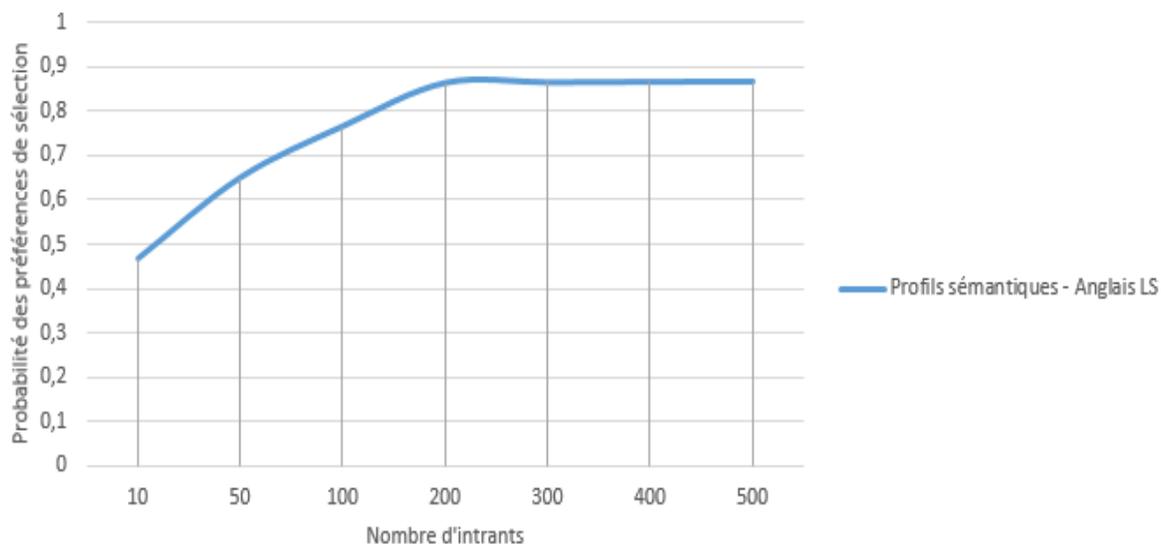


Figure 5.1: Prédiction des profils sémantiques pour 182 intrants de l'anglais LS à des étapes différentes d'apprentissage.

5.7 Discussion des résultats

Nous avons adapté le modèle computationnel pour l'acquisition des préférences de sélection en utilisant les profils sémantiques représentés par des vecteurs des propriétés sémantiques. Les rôles sémantiques et les préférences de sélection des verbes constituent des éléments principaux dans l'apprentissage d'une LS. Les expérimentations réalisées ont montré que l'acquisition des profils sémantiques et des distributions de probabilités sur les propriétés sémantiques dans le cas de la LS pour des positions des actants sont effectués grâce aux constructions générales d'un côté et des préférences de sélection d'un autre côté. Cette acquisition est fondée sur des propriétés sémantiques élémentaires et non sur l'étiquetage explicite des rôles sémantiques ou des représentations hiérarchiques complètes des classes sémantiques comme le cas de la majorité des modèles actuels (Xiaohua et al., 2010; Patrick et al, 2006; Messiant et al., 2010; Zafirain et al., 2013). Des éléments tels que l'utilisation des constructions comme représentation générale de l'information liée aux structures actanciennes, l'utilisation des préférences de sélection comme information spécifique, l'utilisation de la probabilité de couplage selon la formule d'Anderson (1991) ont permis à notre modèle d'obtenir des résultats

visiblement significatifs et le rendre cognitivement plausible et stimulant en le comparant avec les modèles actuels d'ALS. En outre, nous avons utilisé dans certaines expérimentations un ensemble d'intrants de test plus grand (182 intrants) que celui utilisé dans les expérimentations décrites dans les chapitres précédents pour comparer les résultats avec certains modèles existants. Comme nous l'avons montré, notre modèle prédit les profils sémantiques avec un score de quatre-vingt-cinq pour cent à partir de 200 intrants de LS. Ce score peut être expliqué par les intrants bruyants et incomplets à cause de la qualité des corpus linguistiques et l'ambiguïté des sens des mots de la LS qui peut conduire à des surgénéralisations.

Pour la plupart des verbes utilisés dans les expérimentations, les propriétés sémantiques générales telles que { entity, abstraction, object, whole } sont associées à des probabilités plus élevées dans les profils sémantiques des actants qui représentent l'objet direct. En ce qui concerne les verbes qui exigent des actants plus spécifiques tels que 'manger' et 'piloter', les profils sémantiques sont représentés par des propriétés sémantiques spécifiques telles que { food, nutriment, living thing, foodstuff, plant, repast, meal }. Ces profils sémantiques ont été calculés selon les fréquences des verbes et des mots utilisés dans le corpus. À titre d'exemple, pour le verbe 'piloter', les probabilités des actants liés à la propriété sémantique 'artifact' sont plus fréquentes que ceux liés à la propriété 'team'. Pour les expérimentations avec des verbes qui représentent des faux amis entre le français et l'anglais tels que 'achever' et 'demander' les profils sémantiques calculés montrent que le modèle a pris en considération l'usage des verbes et des mots et non la similitude phonologique ou morphologique entre les verbes. Ces profils sémantiques montrent que les sens des verbes 'achever' et 'demander' sont ceux de 'to complete' et 'to ask' et pas les sens de 'to achieve' et 'to demand'. Comme perspective future de cette thèse, nous allons adapter notre modèle pour prendre en considération les aspects phonologique et morphologique des verbes ce qui nous permettra de simuler l'influence des faux amis dans le cas des préférences de sélection des verbes en absence des primitives sémantiques de ces derniers. Dans le cas des verbes qui peuvent avoir des sens différents, selon le contexte d'utilisation tels que 'entendre'

(Piron, 2004) et ‘jouer’, et après plusieurs stages d’apprentissages, le modèle calcule un profil sémantique représentatif des différents sens avec des probabilités qui montrent la fréquence des phrases qui représentent chaque contexte.

5.8 Conclusion

Nous avons adapté notre modèle computationnel pour l’acquisition des préférences de sélection pour la LS à partir des intrants extraits des paires scène/énoncé. Les préférences de sélection sont représentées par des profils sémantiques sous forme de distribution de probabilité sur les propriétés sémantiques élémentaires d’un actant candidat pour une position dans une structure actancielle. En outre, notre modèle a été conçu pour pouvoir considérer les intrants de la L1 et la LS en utilisant des paramètres de lissage et la probabilité de couplage (Anderson, 1991). Ces caractéristiques constituent un avantage qui a permis d’obtenir des résultats importants en calculant les probabilités des propriétés sémantiques appropriées et prédire les profils sémantiques des actants candidats pour une position spécifique de la structure actancielle. De même, le modèle nous a permis de comparer les profils sémantiques générés dans chaque étape d’apprentissage et d’analyser les résultats selon le système conceptuel des apprenants de la L1. À titre d’exemple, la figure 5.1 montre que dans une première étape d’apprentissage (nombre d’intrants égal à 10) la probabilité des préférences de sélection était 0.48, ce qui montre que les constructions précédemment élaborées pour la L1 ont aidé pour avoir un score relativement élevé dans une première étape d’apprentissage de la LS. Cependant, une amélioration du modèle est nécessaire pour pouvoir prendre en considération les similitudes phonologiques et morphologiques entre les verbes et les mots. Cet élément permettra d’analyser l’influence des faux amis entre la L1 et la LS dans le cas des préférences de sélection des verbes. En outre, l’adaptation du modèle pour pouvoir traiter automatiquement les auxiliaires modaux tels que ‘laisser’ et ‘faire’ permet de repérer le contexte approprié du verbe (Piron, 2004). De même, la réalisation de certaines expérimentations avec des verbes ditransitifs permettra d’élucider de plus les processus d’acquisition des préférences de sélection. Un autre domaine et champ

d'expérimentation est l'application des préférences de sélection aux noms (Pustejovsky, 1995). À titre d'exemple, le gâteau préfère être mangé, mais pas bu. Cette modélisation permettra de prédire les propriétés sémantiques du verbe en utilisant celles des actants. De même, cela nous permet de comparer les préférences de sélection d'un actant avec des verbes différents. Ces aspects seront traités dans d'autres perspectives de cette thèse.

CHAPITRE VI

CONTRIBUTIONS ET PERSPECTIVES FUTURES

6.1 Contributions

L'analyse, la compréhension et la modélisation des processus cognitifs d'acquisition des langues secondes restent un problème entier. Par ailleurs, la simulation de ces processus par des modèles computationnels permet d'élucider ces mécanismes fondamentaux d'apprentissage. Pour cela, nous avons proposé un modèle computationnel d'ALS en utilisant comme champ d'expérimentation le français et l'anglais comme L1 et LS et vice versa. Ce modèle computationnel a été fondé sur une approche bayésienne non supervisée extraite d'un modèle de catégorisation humaine proposé par Anderson (1991) et qui a été adopté par d'autres modèles de L1 (Perek, 2015 ; Alishahi et Stevenson, 2010) et LS (Matusévych et al., 2016). En outre notre modèle a été fondé sur la théorie de construction appelée « Usage-Based » qui permet d'appliquer des mécanismes de catégorisation et de généralisation en utilisant uniquement les intrants observés et extraits à partir des corpus linguistiques. Ces derniers sont utilisés pour prédire des modèles syntaxiques, des structures actanciennes ou des propriétés sémantiques des prédicats et des actants. Cette modélisation computationnelle nous a permis de comprendre l'acquisition des propriétés abstraites de la langue et la façon avec laquelle un apprenant peut généraliser des associations entre la syntaxe et le sens (sémantique) des éléments d'une structure actancielle en utilisant l'unité fondamentale du langage qui est « la construction » (Fillmore, 1988 ; Goldberg, 1995). Notre modèle regroupe les

cadres en constructions. Les intrants du modèle ne sont pas étiquetés et par conséquent les constructions ne sont pas prédéfinies, mais plutôt déterminées selon la similarité des cadres observés.

Ces intrants sont des paires scène/énoncé ou la scène est représentée par une structure actancielle avec un prédicat, des arguments et des propriétés, tandis que l'énoncé est représenté par la phrase sous sa forme texte ou audio. En outre, le modèle apprend ces intrants selon les propriétés des structures actancielles et les catégorise dans des constructions. Ces constructions seront utilisées pour enraciner certaines données et calculer la probabilité des propriétés syntaxiques et sémantiques manquantes dans les cadres et par conséquent le modèle acquiert la relation sémantique entre un verbe et ses actants et les caractéristiques syntaxiques de la structure actancielle.

De même, d'autres éléments d'acquisition de LS tels que l'acquisition des rôles sémantiques et les préférences de sélection des verbes ont été étudiés dans cette thèse. Ces éléments représentent un domaine très riche pour la modélisation des relations sémantiques et syntaxiques entre un prédicat et ces actants.

D'un autre côté, l'adoption de la théorie « Usage-Based » nous a permis de confirmer le rôle du volume et de la qualité des intrants dans l'ALS. Comme nous l'avons mentionné dans notre premier chapitre, deux théories principales se concurrencent sur la nature de l'acquisition de la langue naturelle. Les théories nativistes confirment l'existence d'un système inné, chez l'être humain, responsable de l'apprentissage de la langue. Ce système est composé de symboles, de règles et de relations qui sont responsables des processus d'acquisition de la langue (L1, LS, etc.). Cette théorie affirme que les principes généraux et les éléments communs de la langue sont déjà encodés dans le cerveau de l'être humain sous forme d'une Grammaire Universelle et que la relation entre les structures syntaxiques et la sémantique des différents composants de la langue est assurée par ce dispositif inné et les apprenants n'ont pas besoin de les apprendre (Chomsky 1965, 1968; Grimshaw, 1990). Pour leur part, les théories de construction

affirment qu'on apprend et acquiert à force de l'activité langagière elle-même (Fillmore, Kay et O'Connor, 1988; Goldberg, 1995). Les expérimentations réalisées dans les chapitres III, IV, et V ont montré que la capacité de notre modèle à généraliser de nouvelles constructions avec le temps, et sans étiquetage explicite des propriétés, n'est pas un simple ajustement des intrants par rapport aux règles établies, mais toute une élaboration de nouvelles structures abstraites et de nouveaux modèles. Par conséquent, nous pouvons confirmer que ce ne sont pas les règles linguistiques et les structures abstraites de la langue qui sont innées et codées dans le cerveau, mais plutôt la capacité de catégorisation et de généralisation à partir des intrants. Ce qui est supporté par d'autres recherches telles que (Ellis 2013; MacWhinney, 1987).

Nous allons montrer nos contributions au niveau du modèle computationnel et des expérimentations dans les sections suivantes.

6.1.1 Modèle computationnel

Nous avons adopté d'un modèle de Anderson (1991) en utilisant certains principes tels que le paramètre de lissage qui représente le nombre total de valeurs qu'une propriété sémantique ou syntaxique ou qu'une unité lexicale peut avoir. En outre, pour aligner l'intégralité des distributions de probabilités à une proportion commune, nous avons ajusté chaque probabilité sur l'ensemble des constructions ou des unités lexicales, selon le type d'expérimentation. Nous avons utilisé le principe de lissage additif (Manning et al., 2008; Jurafsky et Martin, 2008), en ajoutant une valeur constante à chaque occurrence d'événement pour éviter d'avoir un événement avec une probabilité nulle. Puisque le modèle va progresser vers l'utilisation des cadres dans leur contexte convenable, les cadres de la LS seront favorisés dans le cas d'ALS. Par conséquent, nous avons adopté la probabilité de couplage $c = 0.5$ pour ces derniers. De même, nous avons utilisé la probabilité de couplage indiquée par Anderson (1991) $c = 0.3$ dans le cas des cadres de la L1 pour orienter le modèle vers l'élaboration des constructions de la LS comme dans le cas d'un apprentissage réel de la LS. Ceci a pour avantage de donner au

modèle la capacité de catégoriser deux ensembles d'intrants de la langue comme L1 et L2 ce qui simule des situations réelles d'apprentissage.

La représentation des constructions avec des associations probabilistes entre les propriétés syntaxiques et sémantiques d'une structure actancielle permet de concevoir deux éléments principaux qui sont la nature incertaine de la langue naturelle et l'ALS en s'appuyant seulement sur l'utilisation de la langue elle-même « Usage-Based ». D'un autre côté, et en utilisant le même principe, une distribution de probabilités pour repérer le meilleur profil sémantique a été adoptée comme représentation des rôles sémantiques et des préférences de sélection des verbes. Cette représentation a permis d'utiliser et de confirmer le rôle des informations générales groupées dans les constructions et des informations spécifiques représentées par les préférences de sélection, dans la prédiction des profils sémantiques, des modèles syntaxiques et des autres propriétés des verbes et des actants. Cela nous a permis de confirmer l'interférence entre les connaissances spécifiques associées au verbe et les connaissances générales associées à la construction dans la généralisation des connaissances sur des nouveaux verbes dans le cas d'ALS, ce qui appuie d'autres recherches telles que celle de (Ellis, 2013).

6.1.2 Processus d'ALS

Des processus cognitifs d'apprentissage tels que la catégorisation, la généralisation et la prédiction ont été étudiés dans ce projet de recherche. Nous avons adopté la notion des constructions qui sont considérées comme des structures qui regroupent des verbes qui ont une similarité élevée des propriétés syntaxiques et sémantiques. Cette similarité est calculée à partir d'un modèle bayésien en utilisant les valeurs de ces propriétés. En outre, le modèle utilise des prédictions probabilistes pour calculer les valeurs possibles des propriétés syntaxiques et sémantiques et pour regrouper les cadres. Cette caractéristique le rend capable de simuler l'apprentissage d'ALS et donne la possibilité de prédire les composants manquants dans des tâches similaires à la production ou la compréhension de la langue.

Nous avons montré l'influence de certains facteurs tels que l'âge de début d'apprentissage, le nombre d'intrants et le quotient LS/L1 dans l'ALS. Les résultats des expérimentations ont permis de confirmer le rôle de l'âge de début d'ALS. Cependant, le nombre d'intrants et le quotient LS/L1 sont les éléments les plus importants dans cette acquisition. D'un autre côté, les résultats des simulations ont démontré l'émergence des constructions abstraites des structures actanciennes à partir de l'utilisation des verbes. Les calculs bayésiens du modèle ont permis de tenir compte de deux types de généralisation des règles et modèles qui sont la généralisation appropriée et la surgénéralisation dans le cas d'ALS. De nombreux principes cognitifs ont été pris en considération par notre modèle tels que les notions de la preuve positive, les principes d'analyse universelle, l'enracinement des cadres, la concurrence entre constructions, la correction à partir des exemples et l'apprentissage à base d'utilisation. Le modèle a démontré que l'intégration de ces notions dans les processus de généralisation et de correction des surgénéralisations peut expliquer le comportement d'apprentissage des apprenants de la LS.

Pour le cas d'acquisition des rôles sémantiques de la LS, nous avons illustré le rôle des profils sémantiques, des positions des actants, le cas grammatical de chaque actant et les primitives sémantiques des verbes dans cette acquisition. De même, nous avons montré le rôle de la généralisation des profils sémantiques et la façon avec laquelle émergent les propriétés sémantiques vers des propriétés plus générales proportionnellement avec le nombre d'intrants du modèle. En effet, les associations probabilistes entre ces propriétés et les positions syntaxiques des actants permettent au modèle d'apprendre des conceptions générales des rôles sémantiques sans avoir besoin d'un étiquetage explicite de ces derniers. La prédiction des rôles sémantiques appropriés dépend aussi des constructions de la L1 et de la similitude entre les propriétés des verbes de la L1 et de la LS. Dans le cas des cadres similaires, l'acquisition de ces rôles est plus rapide et le degré de certitude est élevé. La généralisation des profils sémantiques a permis d'apprendre les rôles sémantiques des verbes qui n'ont pas d'équivalents exacts dans la L1, tels que les verbes psychologiques dont les actants possèdent des cas grammaticaux différents dans

la L1 et la LS. Cependant, cette généralisation doit être accompagnée d'autres éléments tels que les primitives sémantiques du verbe.

En ce qui concerne l'acquisition des préférences de sélection des verbes de la LS, notre modèle calcule les probabilités des propriétés sémantiques appropriées et prédit les profils sémantiques des actants candidats pour une position spécifique de la structure actancielle. De même, le modèle nous a permis de comparer les profils sémantiques générés dans chaque étape d'apprentissage et d'analyser les résultats selon le système conceptuel des apprenants de la L1. À titre d'exemple, les résultats ont montré que les constructions précédemment élaborées pour la L1 ont aidé à obtenir un score relativement élevé dans une première étape d'apprentissage de la LS.

6.1.3 Dimensions cognitives et informatiques

Principalement, notre modèle est fondé sur la théorie « Usage-Based » qui affirme que les apprenants d'une langue acquièrent les différents aspects de cette dernière en force de l'activité langagière elle-même (Tomasello, 2003; Behrens, 2009; Eskildsen, 2008). Cette théorie oppose aux théories nativistes de l'acquisition de la langue et qui confirment l'existence d'un système inné, chez l'être humain, responsable de l'apprentissage de la langue naturelle. Ce dernier est composé de symboles, de règles et de relations qui sont responsables des processus d'apprentissage de langue naturelle et d'acquisition des langues secondes. Ce système inné est fondé sur la théorie de la Grammaire Universelle et considère la langue comme un nombre infini de phrases générées par un ensemble limité de règles et de symboles (Chomsky 1965, 1968; Grimshaw, 1990). D'un autre côté, le modèle est fondé sur l'existence d'un système conceptuel mature chez les apprenants (adultes et moins jeunes) d'une LS. Ce dernier est considéré comme l'élément le plus important qui distingue l'acquisition de la LS de l'apprentissage des langues maternelles (Lakoff, 1994). Ce système est représenté, dans notre modèle, par l'ensemble de constructions de la L1. En outre, les processus cognitifs de compréhension et de production sont simulés par des tâches de prédictions, en calculant la probabilité des profils sémantiques, des propriétés sémantiques et de

caractéristiques syntaxiques. Ces processus sont des tâches qui s'effectuent progressivement et ils sont fondés sur la catégorisation des cadres, l'élaboration des constructions et la généralisation des informations. Ces derniers permettent l'enracinement de certaines connaissances proportionnellement avec le nombre d'intrants, la qualité de données (des intrants bruyants et incomplets) et la fréquence des verbes et des actants dans le corpus. Par conséquent, l'information au niveau des constructions de la LS, pour les structures actanciennes, les préférences de sélection des verbes et les rôles sémantiques est mis à jour avec chaque nouvel intrant. De nombreux concepts cognitifs ont été pris en considération par notre modèle tels que la preuve positive, les principes d'analyse universelle, l'enracinement des cadres, la concurrence entre constructions et l'apprentissage à base d'utilisation. Cependant, le concept de la preuve négative n'est pas disponible d'une façon angulaire dans le cas d'ALS. Ce dernier est peut-être étudié dans le cadre d'un apprentissage formel et pas dans celui d'ALS à base d'utilisation. D'un autre côté, l'ALS peut être très rapide à ses débuts vu que l'apprenant connaît déjà des congénères ou des vrais amis et des modèles syntaxiques qui sont partagés entre la L1 et la LS. De même, l'application des modèles syntaxiques inappropriés de la L1 sur la LS et l'utilisation des faux amis peuvent provoquer des erreurs de surgénéralisation. Cependant, nos expérimentations dans les chapitres précédents ont montré que ces erreurs` de surgénéralisation peuvent être corrigées en recevant plus d'intrants de la LS.

Un autre élément cognitif qui a été abordé dans cette étude est l'évolution des profils sémantiques d'un ensemble de propriétés plus spécifiques à un ensemble plus général sous forme de distribution de probabilités des propriétés sémantiques élémentaires. Ces profils vont être utilisés pour prédire l'actant le plus approprié à une position spécifique de la structure actancielle. En outre, les profils peuvent être utilisés pour déterminer les préférences de sélection pour un verbe donné, ou pour la désambiguïsation du sens de ce dernier. En résumé, la dimension cognitive de notre modèle a été montrée par plusieurs théories linguistiques et cognitives qui décrivent l'ALS par des apprenants dans des situations réelles. Les résultats de nos expérimentations concordent parfaitement avec

ces théories cognitives et ils ont montré le rôle des intrants de la LS dans l'acquisition de cette dernière et ils ont fourni des explications pour certains processus et modèles observés d'ALS.

En ce qui concerne la modélisation informatique ou computationnelle des différents processus d'acquisition de LS tels que la compréhension, la production et l'apprentissage des sens des verbes, nous avons opté pour un modèle bayésien d'apprentissage qui répond de plus à la nature incertaine de la langue naturelle. Cette dernière est fondée sur des informations inexacts, imprécises, incomplètes et non mesurables. Le modèle a été fondé sur des algorithmes d'apprentissage non supervisé des structures actanciennes, des profils sémantiques et des préférences de sélection des verbes qui peuvent grouper des cadres dans des constructions et calculer les probabilités des propriétés sémantiques et syntaxiques manquantes. Pour l'élaboration de notre modèle, nous avons choisi le langage de programmation C# et la technologie .NET avec un environnement de Visual studio qui offrent un ensemble de mécanismes simples et des possibilités sophistiquées pour traiter les problèmes de prédiction, tels que la plateforme Infer.NET et l'outil WebBayes.Net. De même, des bases de données et des outils tels que WordNet, WOLF et TalkBank ont été adoptés pour extraire et élaborer les intrants du modèle.

D'un autre côté, notre modèle pourra être utilisé comme outil de support à l'apprentissage pour l'acquisition de l'anglais/français LS. En outre, il pourra être utilisé pour générer des exemples d'apprentissage, corriger des résultats des apprenants de la LS, compléter certains autres exemples bruyants ou incomplets ou comparer les résultats des apprenants réels avec ceux du modèle.

6.2 Intrants du modèle et évolution des constructions

6.2.1 Corpus linguistiques et intrants du modèle

Un autre avantage de notre démarche est représenté par l'utilisation d'un ensemble élargi de verbes en anglais (24) et en français (32) ce qui permet d'avoir plus de types

d'intrants qui peuvent aider à compléter différents genres d'expérimentations, tel que mentionné dans les chapitres précédents. En revanche, cela a demandé beaucoup de traitement automatique et manuel pour utiliser les données d'un corpus au format texte brut et les transformer en forme d'intrants convenables. En ce qui concerne la qualité des données, nous avons choisi des corpus dédiés à l'ALS à partir de la base de données TalkBank pour les intrants en anglais et français parlé de Kate Beeching (Beeching, 2013) pour les intrants en français. Le corpus PAROLE (Parallel Oral Foreign Language) a été conçu et compilé pour étudier les caractéristiques des différents niveaux de compétence de la LS. La deuxième partie de nos intrants en anglais a été extraite à partir du corpus BELC (Barcelona English Language Corpus). BELC est un projet qui examine les effets de l'âge sur l'acquisition de l'anglais LS. Tandis que le corpus du français parlé de Kate Beeching contient les transcriptions d'une centaine d'interviews de différentes longueurs avec des participants dont la langue maternelle est le français. Ces corpus ont été créés dans un contexte d'ALS et le nombre de verbes sélectionnés ont permis d'obtenir des intrants très adéquats à l'ALS et, par conséquent, des résultats très significatifs qui ont aidé à expliquer l'effet des différents facteurs qui peuvent influencer cette acquisition. Ces éléments fondamentaux tels que la qualité des corpus choisis, le nombre de verbes utilisés, l'adaptation du modèle de Anderson (1991) à l'ALS en utilisant le principe de lissage additif (Manning et al., 2008; Jurafsky et Martin, 2008) et la probabilité de couplage sont des éléments qui manquent aux modèles computationnels existants d'ALS et qui ont permis à notre modèle d'avoir des résultats qui ont pu expliquer l'influence des intrants dans l'ALS.

En comparant avec des modèles existants d'ALS qui ont utilisé un nombre de verbes qui ne dépasse pas six pour chaque langue et qui adopte des corpus tels que SALSA (Burchardt et al., 2006), PropBank, TIGER (Brants et al., 2004), Treebank (Palmer et al., 2005), nos intrants étaient beaucoup plus appropriés aux modèles d'ALS. Des corpus tels que SALSA et PropBank ne contiennent pas le type de langage que les enfants ou les apprenants de LS reçoivent et ils ont été adoptés seulement parce qu'ils contiennent les annotations nécessaires. À titre d'exemple, TIGER comprend 40.000 phrases de

journaux annotés syntaxiquement. Le type de langage qui est différent de celui des apprenants adultes et enfants moins jeunes de LS, ne favorise pas la simulation d'ALS par ces modèles. L'ALS est fondée principalement sur des intrants bruyants ce qui permet la comparaison du modèle avec un réel apprentissage. Cependant, les intrants adoptés par ces modèles ne permettent pas d'expliquer certains processus cognitifs tels que l'influence de l'étape d'apprentissage de l'ALS ou celui de l'âge de début d'acquisition, d'une façon efficace et significative.

6.2.2 Évolution des constructions

Notre modèle applique le principe de construction de Goldberg (1995) et infère des associations probabilistes entre le modèle syntaxique et les propriétés sémantiques d'un verbe et de ces actants. Les constructions regroupent des cadres qui ont une similarité élevée de leurs propriétés syntaxiques et sémantiques. Cette similarité est calculée à partir d'un modèle bayésien en utilisant les valeurs de ces propriétés. Les cadres peuvent être regroupés sous des constructions selon leurs similarités. Le modèle calcule la similarité d'un nouveau cadre par rapport aux constructions existantes; ensuite il le catégorise avec la construction la plus similaire. Si la similarité avec toutes les constructions ne dépasse pas le seuil prédéfini, une nouvelle construction sera créée pour ce cadre. Dans le calcul de similarité, nous prenons en considération toutes les propriétés sémantiques et syntaxiques du cadre d'un côté et de l'ensemble des cadres qui forment la construction d'un autre côté. De ce fait, l'ensemble de construction est dynamique et chaque nouvel intrant permet d'avoir une nouvelle version de cet ensemble en enracinant une information au niveau d'une construction donnée ou en élaborant une nouvelle construction.

D'un autre côté, le modèle calcule et par la suite utilise des prédictions probabilistes des valeurs possibles des variables syntaxiques et sémantiques, dans les différents niveaux d'apprentissage. Cette caractéristique, permet de simuler l'apprentissage de la LS selon le nombre d'intrants offerts dans chaque étape d'apprentissage, ce qui permet d'analyser l'information probabiliste sur l'effet de l'usage de la langue sur son apprentissage.

Nous avons adopté une approche bayésienne non supervisée extraite d'un modèle de catégorisation humaine proposé par Anderson (1991) et qui a été adopté par d'autres modèles de L1 (Perek, 2015 ; Alishahi et Stevenson, 2010) et LS (Matusevych et al., 2016). Ce modèle regroupe les cadres en constructions. Les intrants du modèle ne sont pas étiquetés et par conséquent les constructions ne sont pas prédéfinies, mais plutôt déterminées selon la similarité des cadres observés. Nous appliquons la formule de Bayes et l'argument du maximum en parcourant toutes les constructions pour trouver la construction la plus similaire. Dans le cas d'ALS, deux ensembles de constructions vont être parcourus par le modèle, SL1 et SLS. SL1 est plus stable durant les simulations d'ALS tandis que le SLS est dynamique et s'incrémente avec le temps. Dans notre modèle, le SLS est l'équivalent d'un système de conceptualisation. À une étape avancée d'apprentissage, le modèle s'oriente automatiquement vers l'ensemble SLS.

6.2.3 Système conceptuel mature et ensemble de constructions de la L1

Le système conceptuel mature chez les apprenants adultes contient les représentations sémantiques de nature symbolique qui peuvent être liées à la perception, au raisonnement et au langage (Lakoff, 1994).

Comme nous l'avons montré dans la section précédente, chaque nouvel intrant traité par le modèle permet d'avoir une nouvelle version de l'ensemble de constructions de la L1 ou de la LS. De cet effet, ces ensembles de constructions sont de nature dynamique soit avec l'enracinement de certaines données, la correction de certaines surgénéralisation ou la création de nouvelles constructions.

D'un autre côté, comme nous l'avons vu durant nos expérimentations, l'ensemble de constructions de la L1 joue le rôle d'un système conceptuel mature chez les apprenants de la LS. À titre d'exemple, durant les expérimentations des sections 4.6 et 5.6 nous avons remarqué que lorsqu'il s'agit des données semblables entre la L1 et la LS, l'acquisition des compétences de la LS se réalise dès les premiers stades d'apprentissage

et avec un nombre d'intrants égal seulement à 10. Ces résultats concordent avec ceux de (Rappoport et al., 2005).

6.2.4 Généralisation et surgénéralisation

Un nouvel intrant aura plus de possibilités d'être attiré par une construction très enracinée qu'une construction qui contient peu de données. De même, les constructions qui contiennent des données enracinées et qui attirent plus de prédicats offrent des candidats sous forme de propriétés sémantiques, de profils sémantiques ou de modèles syntaxiques qui peuvent remplacer des informations manquantes dans une structure actancielle donnée. Ce calcul des composants manquants à partir des constructions ou des préférences de sélections des verbes en utilisant les associations probabilistes entre les propriétés sémantiques et syntaxiques est une forme de généralisation des connaissances générales et spécifiques du modèle.

Techniquement, les paramètres de lissage et la probabilité de couplage (Anderson, 1991) et l'ajustement de chaque probabilité sur l'ensemble des unités lexicales, des profils sémantiques ou des constructions ont permis d'obtenir la valeur la plus appropriée de la propriété linguistique manquante ce qui est considéré comme une généralisation des données de ces éléments.

D'un autre côté, les résultats de nos expérimentations de la section 3.6.3 ont montré l'aspect de la surgénéralisation ou l'application d'une règle grammaticale dans des situations non appropriées. Plusieurs autres études affirment que les apprenants débutants et intermédiaires de la LS sont fortement influencés par leur L1 (Juffs 2000, White 2003). Cependant, ces derniers finissent par adopter les règles de la LS, avec quelques intrants (preuves positives) supplémentaires. À titre d'exemple, en anglais, il existe souvent deux formes que les verbes ditransitifs de transfert de possession peuvent prendre:

- John gave/sent a book to Mary.
- John gave/sent Mary a book.

Dans la section 3.6.3 nous avons imité l'acquisition du français LS par des apprenants anglophones. Notre modèle a montré qu'au début d'apprentissage, ces derniers font une surgénéralisation avec des verbes tels que 'donner' et 'envoyer' en les utilisant dans des constructions à double objet comme le cas de leur L1. Dans les premières étapes d'ALS, des phrases similaires à "Adam a donné Ève un livre." ont été générées avec une probabilité proche de 0.5. Par la suite, et avec plus d'intrants de la LS, cette probabilité a été diminuée à 0. En outre, le principe de la « preuve positive » permet aux apprenants de corriger les erreurs de surgénéralisation, toutefois, il préserve la capacité de ces derniers à générer des composants syntaxiques et sémantiques dans des situations de production et de compréhension pour des nouveaux verbes et modèles.

6.2.5 Profils sémantiques des actants et des verbes

Les simulations présentées dans les chapitres IV et V ont démontré la capacité du modèle et celle de l'apprenant de la LS à prédire des rôles sémantiques et à généraliser des profils sémantiques des actants en passant d'un ensemble de propriétés spécifiques à un ensemble plus général dans un stade avancé d'apprentissage. Les constructions de la LS qui ont des propriétés en commun avec celles de la L1 jouent un rôle important dans l'apprentissage de différents types de verbes de la LS tels que les verbes transitifs et psychologiques. À titre d'exemple, le profil sémantique d'un actant qui occupe la première position dans une construction transitive d'un verbe psychologique a les propriétés d'une entité animée et le modèle prédit son rôle sémantique (Expérient ou Stimulus) dès la première étape d'apprentissage de la LS en se fondant sur les primitives sémantiques du verbe qui sont les mêmes dans la L1.

Pour les verbes qui ont des propriétés différentes de la L1, tel qu'un actant qui peut être datif dans la L1 et nominatif dans la LS, le modèle commence avec un score très faible dans les premières étapes d'apprentissage et qui s'améliore avec le nombre d'intrants de la LS. D'un autre côté, les résultats confirment que d'autres facteurs sont pris en considération dans l'acquisition des rôles sémantiques avec les profils sémantiques des actants. Les primitives sémantiques des verbes et les positions de ces derniers jouent un

rôle primordial dans la prédiction des rôles sémantiques adéquats par les apprenants de la LS. Par conséquent, le modèle prédit aussi les rôles sémantiques des nouveaux actants en calculant leurs propriétés à partir des éléments précédemment mentionnés.

Dans le chapitre V, les expérimentations ont montré que l'acquisition des profils sémantiques et des distributions de probabilités sur les propriétés sémantiques dans le cas de la LS pour des positions des actants sont effectués grâce aux constructions générales d'un côté et des préférences de sélection d'un autre côté. Cette acquisition est fondée sur des propriétés sémantiques élémentaires et pas sur l'étiquetage explicite des rôles sémantiques ou des représentations hiérarchiques complètes des classes sémantiques. Dans ces expérimentations, le modèle a prédit les profils sémantiques avec un score de quatre-vingt-cinq pour cent à partir de 200 intrants de LS. Ce score peut être expliqué par les intrants bruyants et incomplets à cause de la qualité des corpus linguistiques et l'ambiguïté des sens des mots de la LS qui peut conduire à des surgénéralisations.

Pour la plupart des verbes utilisés dans les expérimentations, les propriétés sémantiques générales telles que { entity, abstraction, object, whole } sont associées à des probabilités plus élevées dans les profils sémantiques des actants qui représentent l'objet direct. Les profils sémantiques des verbes qui exigent des actants plus spécifiques tels que 'manger' et 'piloter' sont représentés par des propriétés sémantiques plus spécifiques telles que { food, nutriment, living thing, foodstuff, plant, repast, meal }. Ces profils sémantiques ont été calculés selon les fréquences des verbes et des mots utilisés dans le corpus. Dans le cas des expérimentations avec des verbes qui représentent des faux amis entre le français et l'anglais tels que 'achever' et 'demander' les profils sémantiques calculés montrent que le modèle a pris en considération l'usage des verbes et des mots et pas la similitude phonologique ou morphologique entre les verbes.

6.3 Futures perspectives

6.3.1 Acquisition de la morphologie grammaticale d'une LS

L'acquisition de la morphologie grammaticale d'une LS est un facteur principal du développement de cette dernière. Les mots ont des marques morphologiques telles que la personne, le temps, l'aspect, le mode et la voix pour le verbe et le genre et le nombre pour le nom. D'autres mots tels que les adjectifs qualificatifs et certains déterminants et pronoms portent principalement les marques morphologiques du nom.

Dans le cas d'ALS, nous pouvons réaliser une étude comparative sur l'émergence des formes verbales en français et en anglais. Comparer le développement de la morphologie verbale entre une langue romane et langue germanique chez les apprenants de la LS peut élucider certains aspects d'ALS qui n'ont pas été couverts dans cette thèse tels que l'acquisition du genre et du nombre. L'utilisation d'une approche probabiliste pour montrer les ressemblances et les différences dans les étapes d'évolution de l'une des deux langues comme LS en ayant l'autre langue comme L1 pourra aider à illustrer les facteurs principaux et les processus cognitifs élémentaires impliqués dans l'acquisition de ces aspects morphologiques d'ALS. Plusieurs approches et méthodes peuvent être adoptées dans cette future étude telles que l'approche fonctionnaliste de l'acquisition de la grammaire, la description des systèmes verbaux de chaque langue et l'analyse des facteurs qui peuvent influencer l'émergence des morphologies verbales. En utilisant chaque langue comme LS et L1 dans des expérimentations différentes, nous pouvons analyser l'évolution du système verbal selon les calculs probabilistes du modèle et les comparer avec les productions des apprenants de LS des autres études telles que (Bassano et al., 2001; Hawkins et al., 2006).

Certains facteurs peuvent jouer un rôle primordial dans le développement des morphologies verbales de LS tels que la complexité cognitive, les éléments linguistiques liés aux désinences intégrées aux mots des langues flexionnelles. Ces dernières sont des langues dont les mots changent de forme morphologique et phonologique, selon certains critères grammaticaux, dans la phrase.

Le concept de complexité cognitive a été étudié par plusieurs recherches récentes en montrant l'évolution de l'apprentissage des verbes par rapport aux noms dans l'acquisition de la langue. Certaines études telles que (Tomasello, 1995) ont montré que le développement des noms est plus rapide par rapport à celui des verbes à cause de la complexité de ces derniers qui ont de fortes variations selon le contexte de la phrase. D'un autre côté, la complexité sémantique est proportionnelle à la morphologie des verbes, car les catégories plus complexes sont les catégories sémantiquement marquées. À titre d'exemple, l'acquisition de la négation, du pluriel et du passé simple d'une LS est généralement plus complexe que celle de l'affirmation, du singulier et du présent simple (Dressler et al., 1987).

En déterminant les catégories les plus marquées entre le français et l'anglais (Personnes grammaticales, genre, nombre, etc.) nous pouvons établir des hypothèses et prédire le comportement de notre modèle selon ces dernières.

En outre, les ressemblances et les différences entre les organisations linguistiques des différentes langues tels que la complexité morphologique, la complexité syntaxique ou le rapport psycholinguistique peuvent être un champ d'expérimentation pour cette perspective.

6.3.2 Le rôle de la motivation dans l'ALS

Beaucoup de tâches et décisions liées à nos comportements sont sélectionnées à partir d'un vaste ensemble de possibilités. En outre, notre état de motivation joue un rôle primordial dans la sélection d'une tâche ou une décision qui peut générer un résultat positif ou négatif, selon le cas (Saggar et al., 2007). En ce qui concerne la LS, plusieurs aspects et modèles de motivation affectent l'acquisition de cette dernière. Généralement, les recherches liées à l'ALS se focalisent notamment sur la L1 de l'apprenant et la qualité et le nombre d'intrants de la LS plutôt que la motivation de l'apprenant lui-même. D'un autre côté, diverses théories de la motivation en LS (Dörnyei, 2001, 2009; William et Burden, 1997) ont exploré l'impact des facteurs externes tels que le

renforcement, la tâche, le contrôle de l'action et la dimension culturelle dans l'ALS. Comme nous l'avons déjà montré dans le chapitre III, certaines études (Gardner et al., 1959) ont confirmé que le rôle de la motivation est plus important que celui de l'âge de début d'ALS et qu'il n'est jamais trop tard pour apprendre une LS. De même, des facteurs tels que la curiosité ou le voyage peuvent aider énormément à acquérir une langue dans un cadre non formel.

De même, la motivation peut-être intégrative ou instrumentale. Un apprenant de la LS peut avoir besoin d'acquérir une LS pour s'intégrer à une nouvelle société pour des raisons culturelles ou économiques. Par conséquent, le degré de volonté a un ascendant très important sur la maîtrise d'une langue seconde.

En ce qui concerne la modélisation computationnelle de la motivation, plusieurs modèles (Saggar et al., 2007, Merrick, 2014) ont été conçus pour élucider l'influence de la motivation dans le développement de compétences dans des domaines comme l'apprentissage de langues et de jeux. Dans le cas de notre modèle, le défi sera la conception de la motivation dans un cadre d'ALS fondé seulement sur la pratique de cette dernière dans un environnement favorable et non sur un apprentissage formel. Ces modèles computationnels existants utilisent des techniques différentes, pour la conception de la motivation, telles que le perceptron de (Rosenblatt, 1958), les réseaux neuronaux et les classificateurs linéaires qui effectuent des classifications binaires qui aident à ce type de modélisation.

La plupart des modèles computationnels existants de la motivation prennent en considération des facteurs internes et externes tels que 'le professeur', 'la classe', 'les outils' et 'l'état émotionnel de l'apprenant'. Dans notre cas, nous allons nous focaliser sur les éléments internes tels que les émotions au lieu de considérer des éléments externes qui sont plus adéquats à un contexte d'apprentissage formel (professeur, classe, preuve négative, outil, etc.).

6.3.3 Influence des verbes très fréquents dans l'ALS

Plusieurs recherches (Diependaele et al., 2013; Ellis, 2014; Koirala, 2015) ont étudié le facteur de la distribution de fréquence des mots/verbes par rapport à l'acquisition des profils sémantiques des verbes et des actants. Ces études ont montré que la distribution de la fréquence des verbes dans l'ensemble d'intrants de la LS influence l'ALS, plus spécifiquement dans l'élaboration de ses constructions. À titre d'exemple, les résultats de (Koirala, 2015) ont montré une corrélation entre la fréquence des mots en anglais et leur difficulté d'acquisition par des apprenants espagnols et portugais. La plupart des apprenants ont considéré les mots moins fréquents comme des mots difficiles à apprendre. De même, les mots très fréquents ont été évalués comme des mots faciles à apprendre indépendamment de leurs complexités structurelles ou phonétiques. À partir de ces résultats (Koirala, 2015) confirme que la fréquence des mots et plus spécifiquement des verbes d'une langue peut influencer l'acquisition des autres verbes de cette dernière.

En ce qui concerne les constructions de la LS, nos résultats des chapitres précédents ont été liés aux corpus précédemment mentionnés, sans modification des fréquences des verbes ou des autres mots. Cependant, d'autres recherches telles que (Year et al., 2009; Nakamura, 2012; McDonough et al., 2012) ont étudié l'effet de la répartition équilibrée ou déséquilibrée des verbes dans l'ensemble des intrants. Cette distribution des verbes est souvent différente entre l'ensemble d'intrants de la LS et de la L1, ce qui peut être un champ d'expérimentation très important pour notre modèle probabiliste, pour cerner l'influence de ce facteur en considérant les verbes les plus fréquents de chaque langue. En outre, nous allons montrer l'effet de cette distribution dans l'élaboration des constructions de la LS et la L1. Les modèles précédemment mentionnés n'ont pas pris en considération les différences entre la L1 et la LS et leurs résultats ont montré que le niveau de la CL des apprenants et des modèles avec un ensemble équilibré des intrants est plus élevé par rapport à un ensemble déséquilibré de ces derniers. Ces résultats peuvent être confirmés avec notre modèle bayésien en considérant d'autres éléments

plus fins de l'ALS tels que les préférences de sélection des verbes, les profils sémantiques et les autres différences entre les deux langues.

6.3.4 Désambiguïsation de sens dans le cas d'ALS

Dans le cas d'ALS certaines propriétés sémantiques et syntaxiques fondamentales de la structure actancielle et sa compatibilité avec la scène jouent un rôle principal dans l'acquisition de la LS. À titre d'exemple, dans le cas des faux amis entre le français et l'anglais, un apprenant francophone de l'anglais LS peut entendre l'énoncé "I demanded an explanation." avec une scène qui semble appropriée. Ce dernier peut confondre le verbe 'to demand' avec le verbe 'demander' de sa L1. Cette confusion conduit bien aisément à une catégorisation du cadre du verbe 'to demand' dans des constructions avec des caractéristiques syntaxiques et sémantiques similaires à celles de 'demander'. L'élimination de ce genre de généralisation nécessite une désambiguïsation de sens. Notre modèle peut adopter des mesures statistiques de cooccurrence (Brousseau-Villeneuve et al., 2010) pour construire des modèles de contexte qui permettent au modèle computationnel de faire la différence dans ce genre de situation d'ambiguïté. Ces modèles de contexte peuvent aider à discriminer le sens non approprié. Ils permettent de calculer le poids de la relation entre un mot et les autres mots de son contexte en utilisant sa distance dans l'énoncé et dans l'ensemble d'intrants. D'autres méthodes ont été adoptées pour la désambiguïsation de sens telles que les méthodes fondées sur un dictionnaire et qui utilisent l'information des ressources lexicales; les méthodes d'apprentissage supervisé et celles de l'apprentissage automatique non supervisé.

D'un autre côté, la désambiguïsation lexicale est un problème non trivial dans le traitement de la langue naturelle et de l'ontologie informatique. En outre, le concept est utilisé dans plusieurs domaines tels que l'analyse du discours et l'amélioration de la recherche de l'information. Cependant, la désambiguïsation de sens n'a pas été utilisée dans le cas de la modélisation de l'ALS malgré que cette désambiguïsation peut aider à expliquer certains processus cognitifs liés à l'ALS.

BIBLIOGRAPHIE

- Adone, D. (ed.) (2003). *Recent development in Creole Studies*. VII/234 S. Tübingen: Niemeyer.
- Akhtar, N. (1999). Acquiring basic word order: Evidence for data-driven learning of syntactic structure. *Journal of Child Language*, 26, 339–356.
- Alishahi, A., & Stevenson, S. (2010). Learning general properties of semantic roles from usage data: a computational model. *Language and Cognitive Processes*, 25:1, 50-93.
- Alishahi, A., & Stevenson, S. (2008). A computational model for early argument structure acquisition. *Cognitive Science*, 32:5, 789-834
- Anders B., Love H., and Pierre N. (2009). Multilingual semantic role labeling. In *Proceedings of The Thirteenth Conference on Computational Natural Language Learning (CoNLL)*, Boulder, Colorado, pp. 43--48
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3):409–429.
- Andrew, I., D. (2003). *Most Honourable Remembrance: The Life and Work of Thomas Bayes*. New York : Springer-Verlag
- Andrews, M., & Vigliocco, G. (2010). The Hidden Markov Topic Model: A Probabilistic Model of Semantic Representation, DOI: 10.1111/j.1756-8765.2009.01074.x
- Andrews, M., & Vinson, D. (2008). Inferring a Probabilistic Model of Semantic Memory from Word Association Norms. In: (Proceedings) *Proceedings of the 30th Meeting of the Cognitive Science Society*
- Anita T. (2014). *Effet d’amorçage structurel en français langue seconde : une étude de corpus longitudinale* Centre de Langues et Littérature, Université de Lund, Box 201, 221 00 Lund, Suède
- Ayoun, D. (2005). Verb movement in the L2 acquisition of English by adult native speakers of French. In Susan Foster-Cohen, Maria del Pilar Garcia Mayo, and Jasone Cenoz (eds.), *EUROSLA Yearbook Volume 5*, 35-76. Amsterdam: John Benjamins.

- Bae, H.S., M.C. L'Homme and G. Lapalme (2008). "Semantic Roles in Multilingual Terminological Descriptions: Application to French and Korean Contexts", *Multilingual and Comparative Perspectives in Specialized Language Resources. Proceedings of the Workshop. Language Resources and Evaluation, LREC 2008, Marrakech, Morocco.*
- Basili, V., R., Selby, R. & Hutchens, D. 1986. *Experimentation in Software Engineering.* IEEE Transactions on Software Engineering.
- Bassano, D., I. Maillachon, S. Kampfler 8: Dressler, WU. 2001. *L'acquisition de la morphologie verbale en français et en allemand autrichien :II. L'épreuve des faits. Enfance 2 : 117-148.*
- Becker, A., & Naïm, P. (1999). *Les réseaux bayésiens. Modèles graphiques de connaissances.* Paris : Editions Eyrolles
- Behrens, H. (2009). *Usage-based and emergentist approaches to language acquisition*", in: *Linguistics 47, H. 2, S. 383-41*
- Beeching, K. (2013) *A parallel corpus approach to investigating semantic change.* In: Aijmer, K. and Altenberg, B., eds. (2013) *Advances in corpus-based contrastive linguistics. Studies in honour of Stig Johansson.* Amsterdam: John Benjamins, pp. 103-125. ISBN
- Benati, A. (2013). *The input processing theory in second language acquisition.* In M. P. G. Mayo, M. J. G. Mangado & M. M. Adrián (Eds.), *Contemporary approaches to second language acquisition* (pp. 93-110). Amsterdam: John Benjamins Publishing Company.
- Bencini, G., & Goldberg, A. (2000). *The Contribution of Argument Structure Constructions to Sentence Meaning.* *Journal of Memory and Language, 43, 640-651*
- Benson, C. (2002). *Transfer/Cross-linguistic influence.* *ELTJournal, 56 (1), 68-70.*
- Bernard C. (2005). *COGNITIQUE : Science et pratique des relations à la machine à penser.* Paris, France : L'Harmattan.
- Bernard V. & Catherine F. (1996) *La polysémie - construction dynamique du sens.* Hermes, pp.131, 1996.
- Blumenthal-Dramé, A. (2012). *Entrenchment in usage-based theories: what corpus data do and do not reveal about the mind.* Berlin etc.: Walter de Gruyter.

- Bowerman, M. (1988). The 'no negative evidence' problem: How do children avoid constructing an overly general grammar? In J. Hawkins (Ed.), *Explaining language universals* (pp. 73-101). Oxford: Basil Blackwell
- Brants, S., Dipper, S., Eisenberg, P., Hansen-Schirra, S., König, E., Lezius, W., Uszkoreit, H. (2004). TIGER: Linguistic interpretation of a German corpus. *Research on Language and Computation*, 2(4), 597–620.
- Brosseau-Villeneuve B. , Nie J. and Kando N. (2010). Towards an optimal weighting of context words based on distance. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 107–115, Beijing, China, August 2010. Coling 2010 Organizing Committee.
- Bruhn-Garavito J. and Montrul S. (1996) Verb-movement and clitic placement in French and Spanish as a second language. *Proceedings of the 20th Annual Boston University Conference on Language Development*, 123-34. Somerville, MA: Cascadilla Press.
- Burchardt, A., Erk, K., Frank, A., Kowalski, A., Pado, S., & Pinkal, M. (2006). The SALSA corpus: a German corpus resource for lexical semantics. In *Proc. of LREC-2006*.
- Cabrera, M., & Zubizarreta, M.L. (2005). Overgeneralization of causatives and transfer in L2 Spanish and L2 English. In D. Eddington (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp.15–30). Somerville, MA: Cascadilla Proceedings Project.
- Carlucci, L. Case, J. (2013). On the Necessity of U-Shaped Learning, Article first published online: 17 JAN 2013, Volume 5, Issue 1, pages 56–88, January 2013
- Muñoz C. (ed.), (2006) *Age and the Rate of Foreign Language Learning*. Clevedon: Multilingual Matters.
- Chang, N. (2004). Putting meaning into grammar learning. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 17–24.
- Chaudron, C. (1988). *Second language classrooms: Research on teaching and learning*. Cambridge: Cambridge University Press.
- Chomsky, N. (1968). *Le Langage et la pensée*, Paris : Payot, coll.
- Chomsky, N. (1981). *Lectures on government and binding*. Berlin and New York: Mouton de Gruyter.

- Chomsky, N. (1991). Some Notes on Economy of Derivation and Representation, in R. Frieden, ed., *Principles and Parameters of Comparative Grammar*, MIT Press
- Ciaramita, M. & Johnson, M. (2000). Explaining away ambiguity: learning verb selectional preference with Bayesian networks. In *Proceedings of the 18th International Conference on Computational Linguistics (COLING 2000)*.
- Clark, E. V. (2009). *First Language Acquisition* (2nd edn). Cambridge: Cambridge University Press.
- Clark, S. & Weir, D. (2002). Class-based probability estimation using a semantic hierarchy. *Computational Linguistics*, 28(2):187–206.
- Corder, S.P. (1975). Error Analysis, Interlanguage and Second Language Acquisition. *Language Teaching and Linguistics: Abstracts* 8: 201-218.
- Cuervo, M. C. (2007). "Double Objects in Spanish as a Second Language: Acquisition of Morphosyntax and Semantics". *SSLA* 29. 583-615.
- Das, D. (2010). *Structured Probabilistic Models for Natural Language Semantics*, Language Technologies Institute, School of Computer Science, Carnegie Mellon University, USA
- De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics*, 28(03), 411–424. doi:10.1017/S0142716407070221
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first and second language word recognition: A lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66, 843-863.
- Dekydspotter, L. Rex, A. S. & Rachel, T. (1999/2000) The Interpretation of Quantification at a Distance in English-French Interlanguage: Domain-Specificity and Second Language Acquisition. *Language Acquisition*, 8, 265-320.
- Dressler, W. U., Mayerthaler, W., Panagl, O. & Wurzel, W. U. (1987). *Leitmotifs in Natural Morphology*. Amsterdam: Benjamins.
- Dittmar N. 1990. La sémantique en construction. *Langage et société* 50-51. 39-66.
- Dittmar N. 1984. Semantic Features of Pidginized Learner Varieties of German, in : R. W. Andersen (éd.) *Second Languages* : 243 -270..

- Dominey, P. F. & Inui, T. (2004). A developmental model of syntax acquisition in the construction grammar framework with cross-linguistic validation in English and Japanese. In *Proceedings of the First Workshop on Psycho-computational Models of Human Language Acquisition*, pages 33–40.
- Dörnyei, Z. (2001). *Teaching and Researching Motivation*. Don Mills, ON: Pearson Education Ltd.
- Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.
- Dowty, D. (1989). On the Semantic Content of the Notion “Thematic Role”. In: *Properties, Types, and Meanings*. Vol. II. Ed. by G. Chierchia, B. H. Partee, and R. Turner. Dordrecht: Kluwer. 69-130.
- Draper, D. (2004). *Bayesian Modeling, Inference and Prediction* Retrieved Jan 9, 2011
Workshop on Psycho-computational Models of Human Language Acquisition, pages 33–40.
- Ellis, N. C. & Wulff, S. (2014). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd Edn), chapter 6.. NY: Routledge
- Ellis, N. C. (2013). Second language acquisition. In *Oxford Handbook of Construction Grammar* (pp. 365-378), G. Trousdale & T. Hoffmann (Eds.). Oxford: Oxford University Press
- Ellis, N. C. and Sagarra, N. (2010). The Bounds of Adult Language Acquisition: Blocking and Learned Attention. *Studies in Second Language Acquisition*, 32, 553-580.
- Ellis, N. C. & Cadierno, T. (2009). Constructing a second language. Introduction to the Special Section. *Annual Review of Cognitive Linguistics*, 7, 111-139.
- Ellis, N. C. (2006a). Cognitive perspectives on SLA: The Associative-Cognitive CREED. *AILA Review*, 19, 100-121. 175
- Ellis, N. C. (2006b). Selective Attention and Transfer Phenomena in L2 Acquisition: Contingency, Cue Competition, Salience, Interference, Overshadowing, Blocking, and Perceptual Learning. *Applied Linguistics*, 27(2), 164-194.
- Ellis, N., C. (2002). Frequency effects in language processing—a review with implications for theories of implicit and explicit language acquisition, *Studies in Second Language Acquisition* 24/2: 143–88.

- Erk, K., & Pado, S. (2008). A Structured Vector Space Model for Word Meaning in Context, University of Texas at Austin, USA
- Erk, K. (2007). A simple, similarity-based model for selectional preferences. In Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, pages 216–223, Prague, Czech Republic
- Eskildsen W., S. (2008) Constructing another Language Usage-Based Linguistics in Second Language Acquisition, Oxford Journals, Humanities Applied Linguistics, Volume30, Issue3Pp. 335-357
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 1017-1063.
- Fillmore, C., Kay, P., & O'Connor, M., K. (1988). Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language*, 64:501–538
- Fisher, C. (1996). Structural limits on verb mapping: The role of analogy in children's interpretations of sentences. *Cognitive Psychology*, 31(1):41–81.
- Fleischman, M., & Roy, D. (2005). Intentional context in situated language learning. In Ninth Conference on Computational Natural Language Learning. Stroudsburg : MIT Media Lab
- Fodor, J., Katz, J. (Eds.). (1964). The structure of language. In, Readings in the philosophy of language. Englewood Cliffs, N. J.: Prentice-Hall, 1964. Pp. 479-518
- French, R., M. (1998). A simple recurrent network model of bilingual memory. In Proceedings of the Twentieth Annual Cognitive Science Society Conference, pages 368–373
- Gaëlle D. (2011). Introduction a une approche instrumentee de la synonymie, Crisco Cahiers de recherche linguistique diffusés par le CRISCO depuis 2000. ISSN 2116-6242
- Gazzaniga, M., S. (1995). *The Cognitive Neurosciences* (Ed.) Cambridge : MIT Press.
- Geeraerts, D., & Cuyckens, H., (2007) *The Oxford Handbook of Cognitive Linguistics*. Oxford University Press US.
- George A. M. (1995). WordNet: A Lexical Database for English. *Communications of the ACM* Vol. 38, No. 11: 39-41.
- Gildea, D. (2002). Probabilistic models of verb- argument structure. In Proc. of COLING.

- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1:135–176.
- Gleitman, L. and Gillette, J. (1995). The role of syntax in verb learning. In *Handbook of child language*. Oxford: Blackwell.
- Goldberg, A., E. (1995). *Constructions: a construction grammar approach to argument structure*, The University of Chicago Press.
- Goldberg, A., E. (2006). *Constructions at Work: The nature of generalization in language*. Oxford University Press
- Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika*, 40, 237–264.
- Gardner, R. C., & Lambert, W. E. (1959). Motivational variables in second language acquisition. *Canadian Journal of Psychology*, 13(4), 266–272. doi:10.1037/h0083787
- Griffiths, T., L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, 114, 211-244
- Gruber, J. (1965). *Studies in Lexical Relations*. MIT Dissertation. Bloomington: Indiana University Linguistics Club. [Reprint 1970].
- Harley, H. (2006). *English words: A linguistic introduction*. Malden, MA: Blackwell Publishers.
- Hart, B., & Todd, R., R. (1995). Meaningful differences in the everyday experience of young American children". Baltimore, MD: Brookes.
- Hawkins, R., & Lozano, C. (2006). Second Language acquisition of phonology, morphology, and syntax. In *Encyclopedia of language & linguistics* (pp. 67–74).
- Hilton, H. E. (2008) *Le corpus PAROLE : architecture du corpus et conventions de transcription*. BilingBank dans TalkBank (Pittsburgh : Carnegie Mellon University), <http://talkbank.org/>.
- Hendriks, H. & Demagny, A-C. (2008). The expression of caused motion in French by English Learners. International conference on The acquisition of French as a second language. Southampton.
- Hendriks, H., Hickmann, H. and Demagny, A.C. (2008) How adult English learners of French express caused motion: A comparison with English and French natives. *Acquisition et Interaction en Langue Etrangere* 27, 15-41.

- Hickmann, M., & Henriette H. (2010). Typological constraints on the acquisition of spatial language in French and English. *Cognitive Linguistics* 21 (2): 189-215
- Hendriks, H., Hickmann, M. & Demagny, A.-C. (2008). How adult English learners of French express caused motion : a comparison with English and French natives. In *Acquisition et Interaction en Langue Étrangère*, 27, 15-41. <http://aile.revues.org/3993>
- Jackendoff, R. (1972) *Semantic interpretation in generative grammar*. Cambridge, MA: The MIT Press.
- Jackendoff, R. (1983). *Semantics and cognition*. Cambridge, MA: MIT Press.
- Jackendoff, R. (1990). *Semantic structures*. Cambridge : MIT Press
- Jackendoff, R. (2002). *Foundations of language—Brain, meaning, grammar, evolution*. New York, NY: Oxford University Press.
- Juffs, Alan. (2000). An Overview of the Second Language Acquisition of Links between Verb Semantics and Morpho-Syntax, ed. John Archibald, *Second language acquisition and linguistic theory*, 187-227, Oxford:Blackwell.
- Jurafsky D. Martin J. H. (2008). *Speech and Language Processing* (2nd ed.). Prentice Hall. p. 132. ISBN 978-0-13-187321-6.
- Joo, H. (2003). Second language learnability and the acquisition of the argument structure of English locative verbs by Korean speakers, *Second Language Research*; 19; 305
- Kail, M. (1991). Acquisition des langues premières et secondes dans une perspective interlangues, *Revue Française de Pédagogie*, 96, 67-78.
- Kailuweit, R. (2003). Classes de prédication, macrorôles et corrélation (linking) dans la RRG, *Cahiers du CRISCO* 13 : 21-36.
- Kako, E. (2006). Thematic role properties of subjects and objects. *Cognition*, 101:1–42.
- Kasper, S. (2008). A comparison of thematic role theories. Unveröffentlichte Magisterarbeit. Philipps-Universität Marburg. Marburg.
- Kerrie E. E. et John N. W. (2008). First language polysemy affects second language meaning interpretation: evidence for activation of first language concepts during second language reading, *Second Language Research*, Vol. 24, No. 2 (April 2008), pp. 167-187
- Khaled, K., et Hossein, N. (2013). First language transfer in second language writing: An examination of current research, *Iranian Journal of Language Teaching Research*

- Kibrik, A. (1997). Beyond Subject and Object: towards a Comprehensive Relational Typology". *Linguistic Typology* 1: 279-346.
- Koirala, C. (2015). The word frequency effect on second language vocabulary learning. *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy*
- Kullback, S. & Leibler, R.A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79–86.
- Ladevie-Roche P. J. (1853), *Éléments de logique, de psychologie, de théodicée et de morale, rédigés d'après les nouveaux programmes du baccalauréat ès-lettres et du baccalauréat ès-sciences.* - Bordeaux : Chaumas-Gayet, VIII-479 p.
- Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What Categories. Reveal about the Mind.* Chicago: The University of Chicago Press. Pp. xvii +. 614.
- Lakoff, G. (1994). What is a Conceptual System?, in: Willis F. Overton & David Stuart Palermo eds., *The Nature and Ontogenesis of Meaning*
- Landauer, T., K., Foltz, P., W., & Laham, D. (1989). Introduction to Latent Semantic Analysis., *Discourse Processes*, 25, 259-284
- Landauer, T., K., & Dumais, S., T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2), 211-240.
- Langacker, R., W. (2000). *A Dynamic Usage-Based Model* in M. Barlow and S. Kemmer (eds): *Usage-Based Models of Language.* Stanford: CSLI.
- Lawrence, R., R. (1989). A tutorial on Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE* 77 (2): 257–286. doi:10.1109/5.18626
- Lefebvre, C, Jourdan, C. & White, L., (Eds.). (2006). *L2 Acquisition and Creole Genesis:Dialogues.* Amsterdam: Benjamins.
- Levin, B. (2004). *Verbs and constructions: Where next?* Retrieved February 2009 from <http://www.stanford.edu/~bclevin/pubs.html>.
- Levin, B & Rappaport, M. (2005). *Argument realization.* Hovav. Cambridge: Cambridge University Press, Pp. 286. ISBN 0521663768
- Long, M. H. (1996). The role of linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413-468). San Diego: Academic Press.

- Li, H. & Abe, N. (1998). Generalizing case frames using a thesaurus and the MDL principle. *Computational Linguistics*, 24(2):217–244.
- Li, P., & Farkas, I. (2002). A self-organizing connectionist model of bilingual processing. In Heredia, R. and Altarriba, J., editors, *Bilingual sentence processing*. Elsevier.
- Light, M. & Greiff, W. (2002). Statistical models for the induction and use of selectional preferences. *Cognitive Science*, 26(3):269–281.
- Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, 24, 187-219.
- Ludington, J. D. (2013). Cross-situational learning of foreign vocabulary reveals a possible noun bias in adulthood. *Theory and Practice in Language Studies* 3 (4), 545
- Lumsden, J. S. (1999). Language acquisition and creolization. In M. DeGraff (Ed.), *Language Creation and Language Change: Creolization, Diachrony, and Development*, 129-158. Cambridge, MA: the MIT Press.
- MacWhinney, B. (2015). *Multidimensional SLA in S. Eskildsen & T. Cadierno Usage-based perspectives on second language learning*. New York: Oxford University Press. pp. 22–45.
- MacWhinney, B. (2004). Parameters or cues? *Bilingualism: Language and Cognition*, 7, 35-36.
- MacWhinney, B. (1987). The Competition Model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition*. Hillsdale, NJ: Lawrence Erlbaum.
- Maling, J. 1990. "The hierarchical assignment of grammatical case". Paper presented at the Symposium on Icelandic Linguistics, Reykjavik, June 13th, 1990.
- Manning, C. D., Raghavan P. & Schütze M. (2008). *Introduction to Information Retrieval*. Cambridge University Press, p. 260.
- Marchman, V. A., Fernald, A., & Hurtado, N. (2010). How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish–English bilinguals. *Journal of Child Language*, 37(4), 817–840. doi:10.1017/S0305000909990055
- Matusevych, Yevgen, Alishahi, Afra, & Backus, Albert (2016). The impact of first and second language exposure on learning second language constructions. *Bilingualism: Language and Cognition*, FirstView, 1-22.

- Matusevych, Yevgen, Alishahi, Afra, & Backus, Albert (2015). Distributional determinants of learning argument structure constructions in first and second language. In D.C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 1547-1552).: Austin, TX: Cognitive Science Society.
- Matusevych, Y., Alishahi, A., & Backus, A.M. (2014). Isolating second language learning factors in a computational study of bilingual construction acquisition. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 988-993). Austin, TX: Cognitive Science Society.
- Matusevych, Y., Alishahi, A., & Backus, A. (2013). Computational simulations of second language construction learning. In V. Demberg & R. Levy (Eds.), *Proceedings of the fourth annual workshop on cognitive modeling and computational linguistics* (pp. 47-56). Stroudsburg: Association for Computational Linguistics.
- Mazurkewich, Irene (1984). The Acquisition of the Dative Alternation by Second Language Learners and Linguistic Theory. *Language Learning* 34(1):91-109.
- McDonough, K., & Trofimovich, P. (2013). Learning a novel pattern through balanced and skewed input. *Bilingualism: Language and Cognition*, 16, 654-662. doi:10.1017/S1366728912000557.
- McDonough, K., & Nekrasova-Becker, T. (2012). Comparing the effect of skewed and balanced input on English as a foreign language learners' comprehension of the double-object dative construction. *Applied Psycholinguistics, FirstView*, 1–24.
- McDonough, K. (2006). Responses to recasts: Repetitions, primed production, and linguistic development *A Mackey Language Learning* 56 (4), 693-720
- McRae, K., Ferretti, T. R., & Amyote, L. (1997). Thematic roles as verb-specific concepts. *Language and Cognitive Processes*, 12(2/3), 137-176.
- Medin, D., & Ross, B., M., A. (2005). *Cognitive Psychology*, 4th ed. John Wiley & Sons.
- Merrick, KE. (2014). Evolution of Intrinsic Motives in a Multi-Player Common Pool Resource Game. Florida, presented at IEEE Symposium on Computational Intelligence for Human-like Intelligence, Florida, 09 - 14 December 2014

- Messiant C., Gábor K., Poibeau T. et al. (2010). Acquisition de connaissances lexicales à partir de corpus : la sous-catégorisation verbale en français. *Traitement automatique des langues*, 51(1), 65–96.
- Meunier, J. G. (1989). Artificial intelligence and the theory of Signs, *Semiotica*, September, p. 43 – 63.
- Meunier, J. G. (1999) . Représentation en Sciences Cognitives, *RSSI*, vol 19, no 2-3 p 83-105.
- Michael S. (2005). *PRACTICAL. ENGLISH. USAGE*. Third Edition. OXFORD
- Mitchell, J., & Lapata., M. (2008). Vector-based models of semantic composition. In *Proceedings of ACL*, 236–244
- Monner, D., Vatz, K., Morini, G., Hwang, S.-O., & DeKeyser, R. M. (2013). A neural network model of the effects of entrenchment and memory development on grammatical gender learning. *Bilingualism: Language and Cognition*, 16, 246–265.
- Munoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *IRAL*, 49(2), 113–133.
- Nakamura, D. (2012). Input skewedness, consistency, and order of frequent verbs in frequency-driven second language construction learning. *IRAL*, 50(1), 1–37.
- Palmer, M., Gildea, D., & Kingsbury, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*, 31(1), 71–106.
- Pardo, T., A., S., & Nunes, M., G., V., (2005). A Statistical Generative Model for Unsupervised Learning of Verb Argument Structures, (NILC), CP 668 – ICMC-USP, 13.560-970 São Carlos, SP, Brasil
- Patrick, Y. and Timothy, B. (2006). Verb Sense Disambiguation Using Selectional Preferences Extracted with a State-of-the-art Semantic Role Labeler, *Proceedings of the 2006 Australasian Language Technology Workshop (ALTW2006)*, pages 139–148.
- Pavlov, I. P. 1927. *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex*. London: Routledge and Kegan Paul.
- Perek, F. (2015). *Argument structure in usage-based construction grammar: experimental and corpus-based perspectives*, University of Basel. John Benjamins Publishing Company. 2015. x, 246 pp.

- Peters, J., M. (1990). A Cognitive Computational Model of Risk Hypothesis. Generation, *Journal of Accounting Research*, pp. 83103
- Pollock, J.-Y. (1989). Verb Movement, Universal Grammar, and the Structure of IP. *Linguistic Inquiry* 20, 365- 424. MIT Press.
- Pinker, S. (1984). *Language Learnability and Language Development* . Cambridge, MA: Harvard University Press.
- Pinker, S. (1989). *Learnability and cognition: the acquisition of argument structure*. Cambridge : MIT Press.
- Piron S. (2004). Contraintes syntaxiques et préférences sélectionnelles du verbe entendre. in PURNELLE G., C. FAIRON et A. DISTER (dir.) *Le poids des mots. Actes des 7ièmes Journées d'Analyse statistique des Données Textuelles (JADT)*, 10-12 mars 2004, Presses Universitaires de Louvain, Louvain-la-Neuve, Belgique, vol. 2, pp. 885-896.
- Pereira, F. (2000). *Formal grammar and information theory: Together again?* London: Royal Society
- Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence twoyear-olds' bilingual proficiency. *Child Development*, 82(6), 1834–1849. doi:10.1111/j.1467-8624.2011.01660.x
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Quine, W. (1960). *Word and object*". Cambridge : MIT Press
- Radimsky J. (2012). Actants, arguments et rôles sémantiques : combien de niveaux d'analyse?, In: Tomaszewicz T., Vetulani G., *L'apport linguistique et culturel français à l'Europe : du passé aux défis de l'avenir*, Lask, Leksem, p. 97-103.
- Randall M. (2007). *Memory, Psychology and Second Language Learning*, Amsterdam: Benjamins
- Rappoport, A., & Sheinman, V. (2005). A second language acquisition model using example generalization and concept categories. In *Proceedings of the Second Workshop on Psychocomputational Models of Human Language Acquisition*, pages 45–52.
- Resnik, P. (1993). *Selection and information: A class-based approach to lexical relationships*. PhD thesis, University of Pennsylvania.

- Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization. *Cognition*, 61:127–199.
- Richards, J. (1972). *Social Factors, Interlanguage, and Language Learning*. *Language Learning*, 22: 149-188.
- Rosch, E. H. (1973), "Natural categories, *Cognitive Psychology*" 4, 328 – 350
- Salton, G. and McGill, M.J. (1983) *Introduction to Modern Information Retrieval*. McGraw-Hill Book Co., New York.
- Saggar, M., Markman AB, Maddox WT, Miikkulainen R. (2007). A Computational Model of the Motivation-learning Interface. *Proc. of the Cog. Sci. Soc.*
- Sagot B. & Fišer D. (2008). Construction d'un WordNet libre du français à partir de ressources multilingues. In *TALN 2008*, Avignon, France.
- Saville-Troike, M. (2006). *Introducing Second Language Acquisition*. Cambridge: CUP. EN476-576.
- Seibert H. A. (2012). How second language learners process argument structure: The effects of first language and individual differences, *THE PENNSYLVANIA STATE UNIVERSITY*, 2012, 228 pages; 3647795
- Selinker, L., Swain, M., and Dumas, G. (1975). The Interlanguage Hypothesis Extended to Children. *Language Learning* Vol. 25, No. 1, pp. 139-152.
- Séaghdha., Ó., D. (2010). Latent variable models of selectional preference. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*. Uppsala, Sweden.
- Sikos, J., Versley, Y., and Frank, A. (2016). Implicit Semantic Roles in a Multilingual Setting. *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics (* SEM 2016)*, Berlin, Germany, 11-12 August, pp. 45--54, *SEM.
- Shayan, S., & Gershkoff, S., L. (2007). How do children learn Agent and Patient roles? In *Poster presentation at SRCD Biennial Meeting Boston, Massachusetts*
- Gui, S. & Yang, H. (2002). *Chinese Learner English Corpus*. Shanghai Foreign Languages Education Press, Shanghai, China.
- Siskind, J., M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61:39–91.

- Souza, R. A. (2011). Argument structure in L2 acquisition: Language transfer revisited in a semantics and syntax perspective. *Ilha do Desterro*, 60, 153–187.
- Spiegelhalter, D., J., Dawid, P., Lauritzen, S., L., & Cowell R., G., (1993). Bayesian analysis in expert systems”, *Statistical Science*, 8 :219–283
- Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Springer-Verlag, Wien and New York
- Talmy, L. (1985). Lexicalization patterns: semantic structure in lexical forms. In Timothy Shopen (ed.), *Language Typology and Syntactic Descriptions*, Vol. 3:Grammatical Categories and the Lexicon. CUP. Cambridge. 57-149.
- Talmy, L. 2000. *Toward a Cognitive Semantics*. Cambridge, MA: MITPress
- Theakston, A. L. (2004). The role of entrenchment in children’s and adults’ performance on grammaticality judgment tasks. *Cognitive Development*, 19:15–34.
- Thomas E. Payne. Résumé des rôles sémantiques et des relations grammaticales, 19 octobre 2007
- Thothathiri M, Snedeker J. (2011). The role of thematic roles in sentence processing: evidence from structural priming in young children. Unpublished manuscript, Harvard University.
- Tomasello, M. (1995) Pragmatic contexts for early verb learning. In M. Tomasello & W. Merriman (eds.), *Beyond Names for Things: Young Children's Acquisition of Verbs*. Hillsdale, NJ: Lawrence Erlbaum
- Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition*, 74, 209–253.
- Tomasello, M. (2003). *Constructing a Language: a Usage Based Theory of Language Acquisition*. Cambridge : Harvard University Press.
- Von Wangenheim C. G., Bortolon A., Wangenheim V., (2001). A hybrid approach for the management of FAQ Document in Latin languages. In:icabr 2001- international conference of case-based reasoning, 2001. Vancouver. International conference of case-based reasoning 2001.
- Weber, P., & Jouffe, L. (2003). Reliability modelling with dynamic Bayesian networks. In *Proceedings of 5th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*

- Wellens, P., Loetzsch, M., & Steels, L. (2008). Flexible word meaning in embodied agents. *Connection Science*, 20(2-3):173--191.
- Werker I.E. Tees R.C. (1984), Cross-language speech perception : Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49-63.
- White, Lydia (1987). Markedness and Second Language Acquisition: The Question of Transfer. *Studies in Second Language Acquisition* 9:261-286.
- White, L. (1990). Implications of learnability theories for second language learning and teaching, In M. A. K. Halliday, J. Gibbons, & H. Nicholas (Eds.), *Learning, Keeping and Using Language* (pp. 271-286). Amsterdam, John Benjamins.
- White, Lydia (1991). Argument Structure in Second Language Acquisition. *Journal of French Language Studies* 1: 189-207.
- White, L., C. Brown, J. Bruhn-Garavito, D. Chen, M. Hirakawa, S. Montrul (1999). "Psych verbs in second language acquisition." In E. Klein and G. Martohardjono (eds.) *The Development of Second Language Grammars: a Generative Approach*, pp. 171-196. Amsterdam: John Benjamins.
- White, Lydia (2003). *Second Language Acquisition and Universal Grammar*. Cambridge University Press.
- Zubizarreta, María Luisa and Eunjeong Oh (2004). "The lexicon-Syntax Interface: The Case of Motion Verbs". University of Southern California, ms.
- Williams, M., Burden, R. (1997). *Psychology for Language Teachers*, Cambridge University Press.
- Williams, E. (1981). Argument Structure and Morphology. *Linguistic Review* 1. 81-114.
- Witten, I. H., & Eibe F. (2005). *Data Mining: Practical machine learning tools and techniques*. San Francisco : Morgan Kaufmann
- Xiaohua L., Bo H., Kuan L., Stephan H. S., and Ming Z. (2010). SRL-based verb selection for ESL, In *EMNLP 2010*, 1068–1076, Cambridge, MA, USA.
- Year J., Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal*, 93, 399–417.
- Yi, X., Gao, J., & Dolan, W. B. (2008). A Web-based English Proofing System for English as a Second Language Users. *Proc. of International Joint Conference on Natural Language Processing*, pages 619-624

Yu, C., Ballard, D., H., & Aslin, R. N. (2005). The role of embodied intention in early lexical acquisition. *Cognitive Science*, 29:961–1005.

Zapirain, B., Agirre, E., Màrquez, L., & Surdeanu, M. (2013). Selectional preferences for semantic role classification. *Computational Linguistics*, 39(3), 631-663.

APPENDICE A

CORPUS ANGLAIS

A.1 Lexicon

#-----
#----- VERBS -----
#-----

WORD: go
CATG: <predicate>
FREQ: 404

SEMR: GO(\$1)
FIDX: 1
FREQ: 43
FTRS: act,move
ARGS: \$1 is <agent>
COND: \$1 is <animate>
PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,going>
PATT: \$1 <word>

SEMR: GO(\$1, \$2)
FIDX: 2
FREQ: 62
FTRS: act,move
ARGS: \$1 is <agent>
ARGS: \$2 is <destination>
COND: \$1 is <animate>
COND: \$2 is <location>
PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
PROL: \$2 is <location,path,destination>
PATT: \$1 <word> \$2

SEMR: GO(\$1, \$2)
FIDX: 3
FREQ: 197
FTRS: act,move
ARGS: \$1 is <agent>
ARGS: \$2 is <destination>
COND: \$1 is <animate>

COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> to \$2

SEMR: GO(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 3
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,going>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> on \$2 \$3

SEMR: GO(\$1, \$2)
 FIDX: 5
 FREQ: 11
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> in \$2

SEMR: GO(\$1, \$2)
 FIDX: 6
 FREQ: 3
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> into \$2

SEMR: GO(\$1,\$2)
 FIDX: 7
 FREQ: 44
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <purpose>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>

PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <end,goal,usefulness,utility>
 PATT: \$1 <word> to \$2

SEMR: GO(\$1, \$2)
 FIDX: 8
 FREQ: 10
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> at \$2

SEMR: GO(\$1,\$2)
 FIDX: 9
 FREQ: 21
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PATT: \$1 <word> with \$2

SEMR: GO(\$1,\$2)
 FIDX: 10
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <ind-exist,affected,motion,direction,going>
 PATT: \$1 <word> through \$2

ENDW

#-----

WORD: eat
 CATG: <predicate>
 FREQ: 267

SEMR: EAT(\$1,\$2)
 FIDX: 1
 FREQ: 109
 FTRS: act,consume

ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <edible>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,consuming,eating>
 PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>
 PATT: \$1 <word> \$2

SEMR: EAT(\$1,\$2)
 FIDX: 2
 FREQ: 46
 FTRS: act,consume
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <meal>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,consuming,eating>
 PROL: \$2 is <non-ind-exist,affected,change,change-matter,activity>
 PATT: \$1 <word> \$2

SEMR: EAT(\$1)
 FIDX: 3
 FREQ: 82
 FTRS: act,consume
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,consuming,eating>
 PATT: \$1 <word>

ENDW

#-----

WORD: see
 CATG: <predicate>
 FREQ: 197

SEMR: SEE(\$1,\$2)
 FIDX: 1
 FREQ: 166
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <thing>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2

SEMR: SEE(\$1,\$2)

FIDX: 2
 FREQ: 15
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <predicate>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> that \$2

ENDW

#-----

WORD: like
 CATG: <predicate>
 FREQ: 179

SEMR: LIKE(\$1,\$2)
 FIDX: 1
 FREQ: 126
 FTRS: appreciate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,emotion,liking>
 PROL: \$2 is <ind-exist,appreciable>
 PATT: \$1 <word> \$2

SEMR: LIKE(\$1,\$2)
 FIDX: 2
 FREQ: 5
 FTRS: appreciate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,emotion,liking>
 PROL: \$2 is <ind-exist,appreciable>
 PATT: \$1 <word> that \$2

ENDW

#-----

WORD: play
 CATG: <predicate>
 FREQ: 113

SEMR: PLAY(\$1,\$2)

FIDX: 1
 FREQ: 81
 FTRS: act,playfully
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <game>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <non-ind-exist,theme,activity,change,change-formal>
 PATT: \$1 <word> \$2

SEMR: PLAY(\$1)
 FIDX: 2
 FREQ: 13
 FTRS: act,playfully
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PATT: \$1 <word>

SEMR: PLAY(\$1,\$2)
 FIDX: 3
 FREQ: 5
 FTRS: act,playfully
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <ind-exist,change,change-location,affected,motion,manner,tool>
 PATT: \$1 <word> with \$2

SEMR: PLAY(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 2
 FTRS: act,playfully
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <game>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <non-ind-exist,theme,activity,change,change-formal>
 PROL: \$3 is <volitional,sentient,animate,ind-exist,playing>
 PATT: \$1 <word> \$2 with \$3

SEMR: PLAY(\$1,\$2)
 FIDX: 5
 FREQ: 6
 FTRS: act,playfully
 ARGS: \$1 is <agent>

ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> in \$2

SEMR: PLAY(\$1,\$2)
 FIDX: 6
 FREQ: 1
 FTRS: act,playfully
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <non-ind-exist,theme,activity,change,change-formal>
 PATT: \$1 <word> on \$2

ENDW

#-----

WORD: think
 CATG: <predicate>
 FREQ: 105

SEMR: THINK(\$1,\$2)
 FIDX: 1
 FREQ: 61
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> \$2

SEMR: THINK(\$1)
 FIDX: 2
 FREQ: 10
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PATT: \$1 <word> so

SEMR: THINK(\$1,\$2)
 FIDX: 3
 FREQ: 22
 FTRS: act,cogitate

ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> that \$2

SEMR: THINK(\$1,\$2)
 FIDX: 4
 FREQ: 2
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> about \$2

ENDW

#-----

WORD: want
 CATG: <predicate>
 FREQ: 97

SEMR: WANT(\$1,\$2)
 FIDX: 1
 FREQ: 25
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PATT: \$1 <word> \$2

SEMR: WANT(\$1,\$2)
 FIDX: 2
 FREQ: 57
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PATT: \$1 <word> to \$2

SEMR: WANT(\$1,\$2)
 FIDX: 3
 FREQ: 5
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PATT: \$1 <word> that \$2

ENDW

#-----

WORD: look
 CATG: <predicate>
 FREQ: 93

SEMR: LOOK(\$1,\$2)
 FIDX: 1
 FREQ: 11
 FTRS: act,perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,looking>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> at \$2

SEMR: LOOK(\$1,\$2)
 FIDX: 2
 FREQ: 14
 FTRS: perceive
 ARGS: \$1 is <stimulus>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <affecting,perceivable,ind-exist>
 PROL: \$2 is <state,quality>
 PATT: \$1 <word> \$2

SEMR: LOOK(\$1,\$2)
 FIDX: 3
 FREQ: 37
 FTRS: act,perceive
 ARGS: \$1 is <experiencer>

ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,looking>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2

SEMR: LOOK(\$1)
 FIDX: 4
 FREQ: 11
 FTRS: act,perceive
 ARGS: \$1 is <experiencer>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,looking>
 PATT: \$1 <word>

SEMR: LOOK(\$1,\$2)
 FIDX: 5
 FREQ: 8
 FTRS: act,perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,looking>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> in \$2

SEMR: LOOK(\$1,\$2)
 FIDX: 6
 FREQ: 5
 FTRS: act,perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,looking>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> to \$2

ENDW

#-----

WORD: study
 CATG: <predicate>
 FREQ: 81

SEMR: STUDY(\$1,\$2)

FIDX: 1
 FREQ: 13
 FTRS: act, learn
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2

SEMR: STUDY(\$1)
 FIDX: 2
 FREQ: 60
 FTRS: act,learn
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PATT: \$1 <word>

SEMR: STUDY(\$1,\$2)
 FIDX: 3
 FREQ: 3
 FTRS: act,learn
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PROL: \$2 is <location,destination>
 PATT: \$1 <word> in \$2

SEMR: STUDY(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: act, learn
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PROL: \$2 is <location,destination>
 PATT: \$1 <word> at \$2

ENDW

#-----

WORD: get
 CATG: <predicate>
 FREQ: 79

SEMR: GET(\$1,\$2)
 FIDX: 1
 FREQ: 59
 FTRS: act,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,possessioning,ind-exist,getting>
 PROL: \$2 is <change,change-possession,ind-exist>
 PATT: \$1 <word> \$2

SEMR: GET(\$1,\$2,\$3)
 FIDX: 2
 FREQ: 6
 FTRS: act,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <src-pred>
 PROL: \$1 is <volitional,sentient,possessioning,ind-exist,getting>
 PROL: \$2 is <change,change-possession,ind-exist>
 PROL: \$3 is <source,animate,ind-exist,affected,non-possessioning,onset>
 PATT: \$1 <word> \$2 \$3

SEMR: GET(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 3
 FTRS: cause,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <for-pred>
 PROL: \$1 is <volitional,sentient,ind-exist,getting>
 PROL: \$2 is <affected,change,change-possession,ind-exist>
 PROL: \$3 is <animate,ind-exist,affected,possessioning,getting,beneficiary>
 PATT: \$1 <word> \$2 \$3

SEMR: GET(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 3
 FTRS: cause,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>

ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,ind-exist,getting>
 PROL: \$2 is <affected,change,change-possession,ind-exist>
 PROL: \$3 is <animate,ind-exist,affected,possessing,getting,beneficiary>
 PATT: \$1 <word> \$3 \$2

SEMR: GET(\$1,\$2)
 FIDX: 5
 FREQ: 3
 FTRS: changestate
 ARGS: \$1 is <theme>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <adjective>
 PROL: \$1 is <sentient,animate,ind-exist,change,change-emotional,affected,becoming>
 PROL: \$2 is <state>
 PATT: \$1 <word> \$2

ENDW

#-----

WORD: say
 CATG: <predicate>
 FREQ: 75

SEMR: SAY(\$1,\$2)
 FIDX: 1
 FREQ: 53
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,pronounceable,speakable>
 PATT: \$1 <word> \$2

SEMR: SAY(\$1,\$2)
 FIDX: 2
 FREQ: 14
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,pronounceable,speakable>

PATT: \$1 <word> that \$2

ENDW

#-----

WORD: come

CATG: <predicate>

FREQ: 70

SEMR: COME(\$1)

FIDX: 1

FREQ: 39

FTRS: act,move

ARGS: \$1 is <agent>

COND: \$1 is <animate>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PATT: \$1 <word>

SEMR: COME(\$1,\$2)

FIDX: 2

FREQ: 19

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>

PATT: \$1 <word> \$2

SEMR: COME(\$1,\$2)

FIDX: 3

FREQ: 2

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <source>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,source>

PATT: \$1 <word> from \$2

SEMR: COME(\$1,\$2)

FIDX: 4

FREQ: 1

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> in \$2

SEMR: COME(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 ARGS: \$3 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PROL: \$3 is <location,path,source>
 PATT: \$1 <word> \$2 from \$3

SEMR: COME(\$1,\$2)
 FIDX: 6
 FREQ: 3
 FTRS: act,move,descend
 ARGS: \$1 is <agent>
 ARGS: \$2 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,source>
 PATT: \$1 <word> up \$2

ENDW

#-----

WORD: watch
 CATG: <predicate>
 FREQ: 65

SEMR: WATCH(\$1,\$2)
 FIDX: 1
 FREQ: 57
 FTRS: look,observe
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,watching>
 PROL: \$2 is <ind-exist,perceptible>
 PATT: \$1 <word> \$2

SEMR: WATCH(\$1)

FIDX: 2
 FREQ: 1
 FTRS: attend,beware
 ARGS: \$1 is <experiencer>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,watching>
 PATT: \$1 <word> out

ENDW

#-----

WORD: try
 CATG: <predicate>
 FREQ: 54

SEMR: TRY(\$1,\$2)
 FIDX: 1
 FREQ: 52
 FTRS: act,attempt
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,trying>
 PROL: \$2 is <ind-exist,affected>
 PATT: \$1 <word> to \$2

ENDW

#-----

WORD: give
 CATG: <predicate>
 FREQ: 53

SEMR: GIVE(\$1,\$2,\$3)
 FIDX: 1
 FREQ: 31
 FTRS: cause,possession
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,affecting,cause,cause-possession,giving>
 PROL: \$2 is <animate,ind-exist,affected,possessing,given,beneficiary>
 PROL: \$3 is <affected,change,change-possession,ind-exist>
 PATT: \$1 <word> \$2 \$3

SEMR: GIVE(\$1,\$2,\$3)

FIDX: 2
 FREQ: 13
 FTRS: cause,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,affecting,cause,cause-possess,giving>
 PROL: \$2 is <affected,change,change-possession,ind-exist>
 PROL: \$3 is <animate,ind-exist,affected,possessing,given,beneficiary>
 PATT: \$1 <word> \$2 to \$3

ENDW

#-----

WORD: fall
 CATG: <predicate>
 FREQ: 53

SEMR: FALL(\$1)
 FIDX: 1
 FREQ: 50
 FTRS: move,fall
 ARGS: \$1 is <theme>
 COND: \$1 is <thing>
 PROL: \$1 is <ind-exist,change-location,motion,direction,vertical,falling>
 PATT: \$1 <word>

SEMR: FALL(\$1,\$2)
 FIDX: 2
 FREQ: 2
 FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,vertical,falling>
 PATT: \$1 <word> \$2

ENDW

#-----

WORD: drink
 CATG: <predicate>
 FREQ: 52

SEMR: DRINK(\$1,\$2)

FIDX: 1
 FREQ: 23
 FTRS: act,absorb
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <potable>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,absorbing,drinking>
 PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>
 PATT: \$1 <word> \$2

SEMR: DRINK(\$1)
 FIDX: 2
 FREQ: 25
 FTRS: act,absorb
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,absorbing,drinking>
 PATT: \$1 <word>

ENDW

#-----

WORD: prepare
 CATG: <predicate>
 FREQ: 48

SEMR: PREPARE(\$1,\$2)
 FIDX: 1
 FREQ: 45
 FTRS: act,produce
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <producing>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,producing,preparing>
 PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>
 PATT: \$1 <word> \$2

SEMR: PREPARE(\$1)
 FIDX: 2
 FREQ: 2
 FTRS: act,produce
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,producing,preparing>
 PATT: \$1 <word>

ENDW

#-----

WORD: arrive
 CATG: <predicate>
 FREQ: 44

SEMR: ARRIVE(\$1,\$2)
 FIDX: 1
 FREQ: 16
 FTRS: act,appear
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,appearing,arriving>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> at \$2

SEMR: ARRIVE(\$1,\$2)
 FIDX: 2
 FREQ: 9
 FTRS: act,appear
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,appearing,arriving>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> to \$2

SEMR: ARRIVE(\$1,\$2)
 FIDX: 3
 FREQ: 7
 FTRS: act,appear
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,appearing,arriving>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> in \$2

SEMR: ARRIVE(\$1)
 FIDX: 4
 FREQ: 10
 FTRS: act,appear
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>

PROL: \$1 is <volitional,sentient,animate,ind-exist,appearing,arriving>
 PATT: \$1 <word>

ENDW

#-----

WORD: listen
 CATG: <predicate>
 FREQ: 38

SEMR: LISTEN(\$1,\$2)
 FIDX: 1
 FREQ: 5
 FTRS: act,attend
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,listening,attending>
 PROL: \$2 is <ind-exist,listenable>
 PATT: \$1 <word> \$2

SEMR: LISTEN(\$1,\$2)
 FIDX: 2
 FREQ: 30
 FTRS: act,attend
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,listening,attending>
 PROL: \$2 is <ind-exist,listenable>
 PATT: \$1 <word> to \$2

SEMR: LISTEN(\$1)
 FIDX: 3
 FREQ: 1
 FTRS: act,attend
 ARGS: \$1 is <experiencer>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,listening,attending>
 PATT: \$1 <word>

ENDW

#-----

WORD: make
 CATG: <predicate>

FREQ: 36

SEMR: MAKE(\$1,\$2)

FIDX: 1

FREQ: 24

FTRS: cause,become,create

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,creating,making>

PROL: \$2 is <change,change-matter,affected,stationary,non-ind-exist,become,made>

PATT: \$1 <word> \$2

SEMR: MAKE(\$1,\$2,\$3)

FIDX: 2

FREQ: 1

FTRS: cause,changestate

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <state>

COND: \$1 is <thing>

COND: \$2 is <thing>

COND: \$3 is <adjective>

PROL: \$1 is <affecting,ind-exist,cause,cause-formal,making1>

PROL: \$2 is <ind-exist,change,change-formal,affected,becoming,made1>

PROL: \$3 is <state>

PATT: \$1 <word> \$2 \$3

SEMR: MAKE(\$1,\$2,\$3)

FIDX: 3

FREQ: 1

FTRS: cause,become,create

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <source>

COND: \$1 is <animate>

COND: \$2 is <concrete>

COND: \$3 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,creating,making>

PROL: \$2 is <change,change-matter,affected,stationary,non-ind-exist,become,made>

PROL: \$3 is <material,source>

PATT: \$1 <word> \$2 with \$3

SEMR: MAKE(\$1,\$2,\$3)

FIDX: 4

FREQ: 3

FTRS: cause,changestate

ARGS: \$1 is <agent>

ARGS: \$2 is <experiencer>

ARGS: \$3 is <state>
 COND: \$1 is <thing>
 COND: \$2 is <animate>
 COND: \$3 is <adjective>
 PROL: \$1 is <affecting,ind-exist,cause,cause-formal,making1>
 PROL: \$2 is <ind-exist,animate,sentient,change,change-formal,affected,becoming,made1>
 PROL: \$3 is <state>
 PATT: \$1 <word> \$2 \$3

SEMR: MAKE(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 2
 FTRS: cause,become,create
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,creating,making>
 PROL: \$2 is <change,change-matter,affected,stationary,non-ind-exist,become,made>
 PROL: \$3 is <animate,ind-exist,affected,possessing,getting,beneficiary>
 PATT: \$1 <word> \$3 \$2

SEMR: MAKE(\$1,\$2,\$3)
 FIDX: 6
 FREQ: 1
 FTRS: cause,become,create
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,creating,making>
 PROL: \$2 is <change,change-matter,affected,stationary,non-ind-exist,become,made>
 PROL: \$3 is <animate,ind-exist,affected,possessing,getting,beneficiary>
 PATT: \$1 <word> \$2 for \$3

ENDW

#-----

WORD: put
 CATG: <predicate>
 FREQ: 31

SEMR: PUT(\$1,\$2)
 FIDX: 1
 FREQ: 6

FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PATT: \$1 <word> \$2

SEMR: PUT(\$1,\$2,\$3)
 FIDX: 2
 FREQ: 1
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PROL: \$3 is <location,destination,surface>
 PATT: \$1 <word> \$2 \$3

SEMR: PUT(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 17
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PROL: \$3 is <location,destination,surface>
 PATT: \$1 <word> \$2 in \$3

SEMR: PUT(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 1
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PROL: \$3 is <location,destination,surface>

PATT: \$1 <word> \$2 on \$3

ENDW

#-----

WORD: take

CATG: <predicate>

FREQ: 28

SEMR: TAKE(\$1,\$2,\$3)

FIDX: 1

FREQ: 1

FTRS: cause,move

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <source>

COND: \$1 is <animate>

COND: \$2 is <concrete>

COND: \$3 is <location>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>

PROL: \$2 is <ind-exist,motion,change,change-location>

PROL: \$3 is <location,path,destination>

PATT: \$1 <word> \$2 from \$3

SEMR: TAKE(\$1,\$2)

FIDX: 2

FREQ: 18

FTRS: cause,possess

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-possess,animate,ind-exist>

PROL: \$2 is <ind-exist,change,change-possession>

PATT: \$1 <word> \$2

SEMR: TAKE(\$1,\$2,\$3)

FIDX: 3

FREQ: 2

FTRS: cause,move

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <destination>

COND: \$1 is <animate>

COND: \$2 is <concrete>

COND: \$3 is <location>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>

PROL: \$2 is <ind-exist,motion,change,change-location>

PROL: \$3 is <location,path,destination>

PATT: \$1 <word> \$2 to \$3

SEMR: TAKE(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 1
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>
 PROL: \$2 is <ind-exist,motion,change,change-location>
 PROL: \$3 is <animate,ind-exist,affected,possessing,getting,beneficiary>
 PATT: \$1 <word> \$2 to \$3

SEMR: TAKE(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 1
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>
 PROL: \$2 is <ind-exist,motion,change,change-location>
 PROL: \$3 is <location,path,source>
 PATT: \$1 <word> \$2 out of \$3

SEMR: TAKE(\$1,\$2,\$3)
 FIDX: 6
 FREQ: 2
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>
 PROL: \$2 is <ind-exist,motion,change,change-location>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> \$2 in \$3

ENDW

#-----
 #----- Example of WORDS -----
 #-----

WORD: catch

CATG: <noun-category>
FREQ: 0
SEMR: catch
ENDW

#-----
WORD: enter
CATG: <noun-category>
FREQ: 0
SEMR: enter
ENDW

#-----
WORD: do
CATG: <noun-category>
FREQ: 0
SEMR: do
ENDW

#-----
WORD: move
CATG: <noun-category>
FREQ: 0
SEMR: move
ENDW

#-----
WORD: lift
CATG: <noun-category>
FREQ: 0
SEMR: lift
ENDW

#-----
WORD: pass
CATG: <noun-category>
FREQ: 0
SEMR: pass
ENDW

#-----
WORD: ball
CATG: <noun-category>
FREQ: 0
SEMR: ball
ENDW

#-----
WORD: spoon
CATG: <noun-category>
FREQ: 0

SEMR: spoon
ENDW

#-----
WORD: tramp
CATG: <noun-category>
FREQ: 0
SEMR: tramp
ENDW

#-----
WORD: pen
CATG: <noun-category>
FREQ: 0
SEMR: pen
ENDW
#-----

A.2 Extrait de 50 entrées de paires scène/énoncé générées aléatoirement par l'outil automatique

0-----
 SENTENCE: you go to basement
 SEM_REP: GO-1[act,move](you<agent>,basement<destination>)

1-----
 SENTENCE: you come out
 SEM_REP: COME-1[act,move](you<agent>,out<destination>)

2-----
 SENTENCE: she watch movie
 SEM_REP: WATCH-1[look,observe](she<experiencer>,movie<stimulus>)

3-----
 SENTENCE: you eat dinner
 SEM_REP: EAT-1[act,consume](you<agent>,dinner<theme>)

4-----
 SENTENCE: they say that something
 SEM_REP: SAY-1[(they<agent>,something<theme>)]

5-----
 SENTENCE: they go to page
 SEM_REP: GO-1[act,move](they<agent>,page<destination>)

6-----
 SENTENCE: you get money
 SEM_REP: GET-1[act,possess](you<agent>,money<theme>)

7-----
 SENTENCE: we give this to her
 SEM_REP: GIVE-1[cause,possess](we<agent>,this<theme>,her<beneficiary>)

8-----
 SENTENCE: dad listen to her
 SEM_REP: LISTEN-1[(dad<experiencer>,her<stimulus>)]

9-----
 SENTENCE: you come here
 SEM_REP: COME-1[(you<agent>,here<destination>)]

10-----
 SENTENCE: eve eat supper
 SEM_REP: EAT-1[act,consume](eve<agent>,supper<theme>)

11-----
 SENTENCE: you prepare breakfast
 SEM_REP: PREPARE-1[act,produce](you<agent>,breakfast<theme>)

12-----
 SENTENCE: eve come
 SEM_REP: COME-1[act,move](eve<agent>)

13-----
 SENTENCE: eve like movie
 SEM_REP: LIKE-1[(eve<experiencer>,movie<stimulus>)]

14-----
 SENTENCE: i see thing
 SEM_REP: SEE-1[perceive](i<experiencer>,thing<stimulus>)

15-----
 SENTENCE: mary watch map
 SEM_REP: WATCH-1[look,observe](mary<experiencer>,map<stimulus>)

16-----

SENTENCE: you get doll
 SEM_REP: GET-1[act,possess](you<agent>,doll<theme>)
 17-----
 SENTENCE: you see this
 SEM_REP: SEE-1[perceive](you<experiencer>,this<stimulus>)
 18-----
 SENTENCE: it make you big
 SEM_REP: MAKE-1[cause,changestate](it<agent>,you<experiencer>,big<state>)
 19-----
 SENTENCE: you get quiet
 SEM_REP: GET-1[changestate](you<theme>,quiet<state>)
 20-----
 SENTENCE: eve like god
 SEM_REP: LIKE-1[appreciate](eve<experiencer>,god<stimulus>)
 21-----
 SENTENCE: i arrive in clock
 SEM_REP: ARRIVE-1[act,appear](i<agent>,clock<destination>)
 22-----
 SENTENCE: it go in there
 SEM_REP: GO-1[act,move](it<agent>,there<destination>)
 23-----
 SENTENCE: i listen to you
 SEM_REP: LISTEN-1[(i<experiencer>,you<stimulus>)]
 24-----
 SENTENCE: mary eat one
 SEM_REP: EAT-1[act,consume](mary<agent>,one<theme>)
 25-----
 SENTENCE: i see mary
 SEM_REP: SEE-1[perceive](i<experiencer>,mary<stimulus>)
 26-----
 SENTENCE: i try to pass
 SEM_REP: TRY-1[act,attempt](i<agent>,pass<theme>)
 27-----
 SENTENCE: i give me this
 SEM_REP: GIVE-1[cause,possess](i<agent>,me<beneficiary>,this<theme>)
 28-----
 SENTENCE: you watch television
 SEM_REP: WATCH-1[look,observe](you<experiencer>,television<stimulus>)
 29-----
 SENTENCE: she go to basement
 SEM_REP: GO-1[act,move](she<agent>,basement<destination>)
 30-----
 SENTENCE: you see her
 SEM_REP: SEE-1[perceive](you<experiencer>,her<stimulus>)
 31-----
 SENTENCE: he prepare lunch
 SEM_REP: PREPARE-1[act,produce](he<agent>,lunch<theme>)
 32-----
 SENTENCE: you like music
 SEM_REP: LIKE-1[appreciate](you<experiencer>,music<stimulus>)
 33-----
 SENTENCE: she say that

SEM_REP: SAY-1[(she<agent>,that<theme>)]
 34-----
 SENTENCE: dad listen voice
 SEM_REP: LISTEN-1[act,attend](dad<experiencer>,voice<stimulus>)
 35-----
 SENTENCE: you drink pope
 SEM_REP: DRINK-1[act,absorb](you<agent>,pope<theme>)
 36-----
 SENTENCE: you arrive at home
 SEM_REP: ARRIVE-1[act,appear](you<agent>,home<destination>)
 37-----
 SENTENCE: you go on this bus
 SEM_REP: GO-1[act,move](you<agent>,this<theme>,bus<destination>)
 38-----
 SENTENCE: dave play her
 SEM_REP: PLAY-1[act,playfully](dave<agent>,her<theme>)
 39-----
 SENTENCE: i see carpenter
 SEM_REP: SEE-1[perceive](i<experiencer>,carpenter<stimulus>)
 40-----
 SENTENCE: she come in home
 SEM_REP: COME-1[act,move](she<agent>,home<destination>)
 41-----
 SENTENCE: you eat bread
 SEM_REP: EAT-1[act,consume](you<agent>,bread<theme>)
 42-----
 SENTENCE: i give it to mary
 SEM_REP: GIVE-1[cause,possess](i<agent>,it<theme>,mary<beneficiary>)
 43-----
 SENTENCE: she study spanish
 SEM_REP: STUDY-1[act](she<agent>,spanish<theme>)
 44-----
 SENTENCE: she eat ice_cream
 SEM_REP: EAT-1[(she<agent>,ice_cream<theme>)]
 45-----
 SENTENCE: you look at that
 SEM_REP: LOOK-1[act,perceive](you<experiencer>,that<stimulus>)
 46-----
 SENTENCE: you put it in top
 SEM_REP: PUT-1[cause,move](you<agent>,it<theme>,top<destination>)
 47-----
 SENTENCE: they say that
 SEM_REP: SAY-1[tell,affirm](they<agent>,that<theme>)
 48-----
 SENTENCE: they watch map
 SEM_REP: WATCH-1[(they<experiencer>,map<stimulus>)]
 49-----
 SENTENCE: you take it in kitchen
 SEM_REP: TAKE-1[cause,move](you<agent>,it<theme>,kitchen<beneficiary>)

APPENDICE B

CORPUS FRANCAIS

B.1 Lexicon

```
#-----
#----- VERBS -----
#-----
```

WORD: faire
 CATG: <predicate>
 FREQ: 712

SEMR: FAIRE(\$1,\$2)
 FIDX: 1
 FREQ: 358
 FTRS: cause,become,create
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,creating,making>
 PROL: \$2 is <change,change-matter,affected,stationary,non-ind-exist,become,made>
 PATT: \$1 <word> \$2

SEMR: FAIRE(\$1,\$2)
 FIDX: 2
 FREQ: 108
 FTRS: act,playfully
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <non-ind-exist,theme,activity,change,change-formal>
 PATT: \$1 <word> \$2

SEMR: FAIRE(\$1,\$2)
 FIDX: 3
 FREQ: 42
 FTRS: act,become,learn

ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2

SEMR: FAIRE(\$1,\$2)
 FIDX: 4
 FREQ: 10
 FTRS: cause,become,create
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,creating,making>
 PROL: \$2 is <change,change-matter,affected,stationary,non-ind-exist,become,made>
 PATT: \$1 <word> ce_qu_il \$2

SEMR: FAIRE(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 32
 FTRS: cause,become,create
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,creating,making>
 PROL: \$2 is <animate,ind-exist,affected,possessing,getting,beneficiary>
 PROL: \$3 is <change,change-matter,affected,stationary,non-ind-exist,become,made>
 PATT: \$1 \$2 <word> \$3

SEMR: FAIRE(\$1,\$2)
 FIDX: 6
 FREQ: 2
 FTRS: cause,changestate
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <affecting,ind-exist,cause,cause-formal,making1>
 PROL: \$2 is <ind-exist,change,change-location,affected,motion,manner,tool>
 PATT: \$1 <word> à \$2

SEMR: FAIRE(\$1,\$2)
 FIDX: 7
 FREQ: 1

FTRS: cause,changestate
 ARGS: \$1 is <agent>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$3 is <animate>
 PROL: \$1 is <affecting,ind-exist,cause,cause-formal,making1>
 PROL: \$3 is <animate,ind-exist,affected,possessing,getting,beneficiary>
 PATT: \$1 <word> pour \$2

ENDW

#-----

WORD: aller
 CATG: <predicate>
 FREQ: 363

SEMR: ALLER(\$1, \$2)
 FIDX: 1
 FREQ: 96
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> à \$2

SEMR: ALLER(\$1, \$2)
 FIDX: 2
 FREQ: 15
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> aux \$2

SEMR: ALLER(\$1, \$2)
 FIDX: 3
 FREQ: 11
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> en \$2

SEMR: ALLER(\$1, \$2)
 FIDX: 4
 FREQ: 12
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> dans \$2

SEMR: ALLER(\$1, \$2)
 FIDX: 5
 FREQ: 22
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <purpose>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <end,goal,usefulness,utility>
 PATT: \$1 <word> \$2

SEMR: ALLER(\$1,\$2)
 FIDX: 6
 FREQ: 3
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PATT: \$1 <word> avec \$2

SEMR: ALLER(\$1,\$2,\$3)
 FIDX: 7
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,going>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> \$2 dans \$3

SEMR: ALLER(\$1, \$2)
 FIDX: 8
 FREQ: 2
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> sur \$2

SEMR: ALLER(\$1, \$2)
 FIDX: 9
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> jusqu_à \$2

SEMR: ALLER(\$1,\$2)
 FIDX: 10
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <purpose>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <end,goal,usefulness,utility>
 PATT: \$1 <word> pour \$2

SEMR: ALLER(\$1,\$2)
 FIDX: 11
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <ind-exist,affected,motion,direction,going>
 PATT: \$1 <word> par \$2

SEMR: ALLER(\$1,\$2,\$3)
 FIDX: 12
 FREQ: 1
 FTRS: act,move

ARGS: \$1 is <agent>
 ARGS: \$2 is <source>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <location,path,source>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> d_une \$2 à \$3

SEMR: ALLER(\$1, \$2)
 FIDX: 13
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> au \$2

SEMR: ALLER(\$1, \$2, \$3)
 FIDX: 14
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 ARGS: \$3 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <location,path,destination>
 PROL: \$3 is <location,path,source>
 PATT: \$1 <word> à \$2 par \$3

ENDW

#-----

WORD: dire
 CATG: <predicate>
 FREQ: 305

SEMR: DIRE(\$1,\$2)
 FIDX: 1
 FREQ: 133
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>

COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,pronounceable,speakable>
 PATT: \$1 <word> \$2

SEMR: DIRE(\$1)
 FIDX: 2
 FREQ: 7
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PATT: \$1 <word>

SEMR: DIRE(\$1, \$2)
 FIDX: 3
 FREQ: 4
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,pronounceable,speakable>
 PATT: \$1 <word> en \$2

SEMR: DIRE(\$1, \$2)
 FIDX: 4
 FREQ: 5
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <animate,ind-exist,affected,beneficiary>
 PATT: \$1 <word> à \$2

SEMR: DIRE(\$1, \$2)
 FIDX: 5
 FREQ: 7
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <animate,ind-exist,affected,beneficiary>
 PATT: \$1 <word> aux \$2

SEMR: DIRE(\$1, \$2)

FIDX: 6
 FREQ: 1
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,affected>
 PATT: \$1 <word> devant \$2

SEMR: DIRE(\$1,\$2)
 FIDX: 7
 FREQ: 1
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <non-ind-exist>
 PATT: \$1 <word> depuis \$2

SEMR: DIRE(\$1,\$2,\$3)
 FIDX: 8
 FREQ: 1
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <animate,ind-exist,affected,possessing>
 PROL: \$3 is <ind-exist,affected>
 PATT: \$1 <word> que \$2 a \$3

ENDW

#-----

WORD: voir
 CATG: <predicate>
 FREQ: 208

SEMR: VOIR(\$1,\$2)
 FIDX: 1
 FREQ: 129
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>

COND: \$1 is <animate>
 COND: \$2 is <thing>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2

SEMR: VOIR(\$1,\$2)
 FIDX: 2
 FREQ: 9
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <sentient,animate,ind-exist,perceivable>
 PATT: \$1 \$2 <word>

SEMR: VOIR(\$1,\$2)
 FIDX: 3
 FREQ: 5
 FTRS: perceive
 ARGS: \$1 is <agent>
 ARGS: \$2 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <ind-exist,visitable>
 PATT: \$1 <word> \$2

SEMR: VOIR(\$1,\$2)
 FIDX: 4
 FREQ: 2
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <thing>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> sur \$2

SEMR: VOIR(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 1
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 ARGS: \$3 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <thing>

PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <sentient,animate,ind-exist,perceivable>
 PROL: \$3 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2 dans \$3

SEMR: VOIR(\$1,\$2,\$3)
 FIDX: 6
 FREQ: 1
 FTRS: perceive
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 ARGS: \$3 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <thing>
 COND: \$3 is <animate>
 PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>
 PROL: \$2 is <ind-exist,perceivable>
 PROL: \$3 is <sentient,animate,ind-exist,perceivable>
 PATT: \$1 <word> \$2 de \$3

ENDW

#-----

WORD: vouloir
 CATG: <predicate>
 FREQ: 188

SEMR: VOULOIR(\$1,\$2)
 FIDX: 1
 FREQ: 136
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PATT: \$1 <word> \$2

SEMR: VOULOIR(\$1,\$2)
 FIDX: 2
 FREQ: 3
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PATT: \$1 <word> être \$2

SEMR: VOULOIR(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 5
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PROL: \$3 is <advisable,ind-exist>
 PATT: \$1 <word> \$2 \$3

SEMR: VOULOIR(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 1
 FTRS: demand,desire
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,missing,ind-exist,wanting>
 PROL: \$2 is <advisable,ind-exist>
 PROL: \$3 is <advisable,ind-exist>
 PATT: \$1 <word> \$2 à \$3

ENDW

#-----

WORD: savoir
 CATG: <predicate>
 FREQ: 167

SEMR: SAVOIR(\$1)
 FIDX: 1
 FREQ: 5
 FTRS: comprehend,discern
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,distinguish,ind-exist,cognizing>
 PATT: \$1 <word>

SEMR: SAVOIR(\$1,\$2)
 FIDX: 2
 FREQ: 111
 FTRS: comprehend,discern
 ARGS: \$1 is <agent>

ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,distinguish,ind-exist,cognizing>
 PROL: \$2 is <learnable,ind-exist>
 PATT: \$1 <word> \$2

SEMR: SAVOIR(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 2
 FTRS: comprehend,discern
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,distinguish,ind-exist,cognizing>
 PROL: \$2 is <learnable,ind-exist>
 PROL: \$3 is <learnable,ind-exist>
 PATT: \$1 <word> \$2 \$3

SEMR: SAVOIR(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 1
 FTRS: comprehend,discern
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,distinguish,ind-exist,cognizing>
 PROL: \$2 is <learnable,ind-exist>
 PROL: \$3 is <volitional,sentient,ind-exist>
 PATT: \$1 <word> \$2 avec \$3

SEMR: SAVOIR(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 1
 FTRS: comprehend,discern
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,distinguish,ind-exist,cognizing>
 PROL: \$2 is <learnable,ind-exist>
 PROL: \$3 is <volitional,sentient,ind-exist>
 PATT: \$1 <word> \$2 de \$3

ENDW

#-----

WORD: travailler
 CATG: <predicate>
 FREQ: 182

SEMR: TRAVAILLER(\$1)
 FIDX: 1
 FREQ: 16
 FTRS: act,product
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PATT: \$1 <word>

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 2
 FREQ: 50
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <location,ind-exist>
 PATT: \$1 <word> à \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 3
 FREQ: 37
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <location,ind-exist>
 PATT: \$1 <word> dans \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 4
 FREQ: 25
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <workable,ind-exist>
 PATT: \$1 <word> \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 5
 FREQ: 12
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <workable,ind-exist>
 PATT: \$1 <word> sur \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 6
 FREQ: 9
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <workable,ind-exist>
 PATT: \$1 <word> avec \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 7
 FREQ: 2
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <location,ind-exist>
 PATT: \$1 <word> en \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 8
 FREQ: 2
 FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <workable,ind-exist>
 PATT: \$1 <word> comme \$2

SEMR: TRAVAILLER(\$1,\$2)
 FIDX: 9
 FREQ: 3

FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <workable,ind-exist>
 PATT: \$1 <word> pour \$2

SEMR: TRAVAILLER(\$1,\$2,\$3)
 FIDX: 10
 FREQ: 1

FTRS: act,product
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,accomplish,ind-exist,acting>
 PROL: \$2 is <location,ind-exist>
 PROL: \$3 is <volitional,sentient,ind-exist>
 PATT: \$1 <word> sur \$2 pour \$3

ENDW

#-----

WORD: venir
 CATG: <predicate>
 FREQ: 138

SEMR: VENIR(\$1)
 FIDX: 1
 FREQ: 10
 FTRS: act,move
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PATT: \$1 <word>

SEMR: VENIR(\$1,\$2)
 FIDX: 2
 FREQ: 19
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>

PATT: \$1 <word> \$2

SEMR: VENIR(\$1,\$2)

FIDX: 3

FREQ: 11

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <state>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <state,quality>

PATT: \$1 <word> \$2

SEMR: VENIR(\$1,\$2)

FIDX: 4

FREQ: 6

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <co-agent>

COND: \$1 is <animate>

COND: \$2 is <animate>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <ind-exist,visitable>

PATT: \$1 <word> avec \$2

SEMR: VENIR(\$1,\$2)

FIDX: 5

FREQ: 28

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>

PATT: \$1 <word> à \$2

SEMR: VENIR(\$1,\$2,\$3)

FIDX: 6

FREQ: 1

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

ARGS: \$3 is <theme>

COND: \$1 is <animate>

COND: \$2 is <location>

COND: \$3 is <concrete>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>

PROL: \$3 is <ind-exist,change,change-location,affected,motion,manner,tool>

PATT: \$1 <word> à \$2 en \$3

SEMR: VENIR(\$1,\$2)

FIDX: 7

FREQ: 7

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>

PATT: \$1 <word> en \$2

SEMR: VENIR(\$1,\$2)

FIDX: 8

FREQ: 14

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <state>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <state,quality>

PATT: \$1 <word> en \$2

SEMR: VENIR(\$1,\$2)

FIDX: 9

FREQ: 20

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>

PATT: \$1 <word> de \$2

SEMR: VENIR(\$1,\$2)

FIDX: 10

FREQ: 11

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <location,path,destination>

PATT: \$1 <word> dans \$2

ENDW

#-----

WORD: penser
 CATG: <predicate>
 FREQ: 113

SEMR: PENSER(\$1,\$2)
 FIDX: 1
 FREQ: 6
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> \$2

SEMR: PENSER(\$1,\$2)
 FIDX: 2
 FREQ: 15
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> que c'est \$2

SEMR: PENSER(\$1,\$2)
 FIDX: 3
 FREQ: 2
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> qu'il y a \$2

SEMR: PENSER(\$1,\$2)
 FIDX: 4
 FREQ: 2
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>

PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> à \$2

SEMR: PENSER(\$1,\$2)
 FIDX: 5
 FREQ: 1
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> que ça \$2

ENDW

#-----

WORD: trouver
 CATG: <predicate>
 FREQ: 108

SEMR: TROUVER(\$1,\$2)
 FIDX: 1
 FREQ: 47
 FTRS: act,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,possessing,ind-exist,getting>
 PROL: \$2 is <change,change-possession,ind-exist>
 PATT: \$1 <word> \$2

SEMR: TROUVER(\$1,\$2)
 FIDX: 2
 FREQ: 1
 FTRS: act,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,possessing,ind-exist,getting>
 PROL: \$2 is <change,change-possession,ind-exist>
 PATT: \$1 <word> à \$2

SEMR: TROUVER(\$1,\$2)
 FIDX: 3
 FREQ: 1
 FTRS: act,possess
 ARGS: \$1 is <agent>

ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,possessing,ind-exist,getting>
 PROL: \$2 is <change,change-possession,ind-exist>
 PATT: \$1 <word> en \$2

SEMR: TROUVER(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> dans \$2

SEMR: TROUVER(\$1,\$2)
 FIDX: 5
 FREQ: 1
 FTRS: changestate
 ARGS: \$1 is <theme>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <adjective>
 PROL: \$1 is <sentient,animate,ind-exist,change,change-emotional,affected,becoming>
 PROL: \$2 is <state>
 PATT: \$1 <word> dans \$2

SEMR: TROUVER(\$1,\$2,\$3)
 FIDX: 6
 FREQ: 1
 FTRS: act,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,sentient,possessing,ind-exist,getting>
 PROL: \$2 is <change,change-possession,ind-exist>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> du \$2 en \$3

ENDW

#-----

WORD: prendre
 CATG: <predicate>

FREQ: 101

SEMR: PRENDRE(\$1,\$2)

FIDX: 1

FREQ: 40

FTRS: cause,possess

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-possess,animate,ind-exist>

PROL: \$2 is <ind-exist,change,change-possession>

PATT: \$1 <word> \$2

SEMR: PRENDRE(\$1,\$2)

FIDX: 2

FREQ: 2

FTRS: cause,possess

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <potable>

PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,absorbing,drinking>

PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>

PATT: \$1 <word> \$2

SEMR: PRENDRE(\$1,\$2,\$3)

FIDX: 3

FREQ: 4

FTRS: cause,move

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

COND: \$3 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>

PROL: \$2 is <ind-exist,change,change-possession>

PROL: \$3 is <ind-exist,change,change-possession>

PATT: \$1 <word> \$2 de \$3

SEMR: PRENDRE(\$1,\$2)

FIDX: 4

FREQ: 2

FTRS: cause,possess

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-possess,animate,ind-exist>

PROL: \$2 is <ind-exist,change,change-possession>

PATT: \$1 <word> en \$2

SEMR: PRENDRE(\$1,\$2,\$3)

FIDX: 5

FREQ: 1

FTRS: cause,move

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <beneficiary>

COND: \$1 is <animate>

COND: \$2 is <concrete>

COND: \$3 is <animate>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>

PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>

PROL: \$3 is <non-ind-exist,affected,change,change-matter,activity>

PATT: \$1 <word> \$2 aux \$3

SEMR: PRENDRE(\$1,\$2,\$3)

FIDX: 6

FREQ: 1

FTRS: cause,move

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

ARGS: \$3 is <beneficiary>

COND: \$1 is <animate>

COND: \$2 is <concrete>

COND: \$3 is <animate>

PROL: \$1 is <volitional,sentient,affecting,cause,cause-location,animate,ind-exist>

PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>

PROL: \$3 is <non-ind-exist,affected,change,change-matter,activity>

PATT: \$1 <word> \$2 pour le \$3

ENDW

#-----

WORD: partir

CATG: <predicate>

FREQ: 101

SEMR: PARTIR(\$1)

FIDX: 1

FREQ: 1

FTRS: act,move

ARGS: \$1 is <them>

COND: \$1 is <concrete>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,going>

PATT: \$1 <word>

SEMR: PARTIR(\$1, \$2)

FIDX: 2

FREQ: 2

FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <purpose>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <end,goal,usefulness,utility>
 PATT: \$1 <word> au \$2

SEMR: PARTIR(\$1, \$2)
 FIDX: 3
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> à \$2

SEMR: PARTIR(\$1, \$2)
 FIDX: 4
 FREQ: 2
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,source>
 PATT: \$1 <word> de \$2

SEMR: PARTIR(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 1
 FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> en \$2 chez \$3

SEMR: PARTIR(\$1, \$2)
 FIDX: 6

FREQ: 2
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> en \$2

SEMR: PARTIR(\$1,\$2,\$3)
 FIDX: 7
 FREQ: 1
 FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,going>
 PROL: \$3 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PATT: \$1 <word> en \$2 avec \$3

ENDW

#-----

WORD: aimer
 CATG: <predicate>
 FREQ: 90

SEMR: AIMER(\$1,\$2)
 FIDX: 1
 FREQ: 24
 FTRS: appreciate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,emotion,liking>
 PROL: \$2 is <ind-exist,appreciable>
 PATT: \$1 <word> \$2

SEMR: AIMER(\$1,\$2,\$3)
 FIDX: 2
 FREQ: 1
 FTRS: appreciate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 ARGS: \$2 is <stimulus>

COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,emotion,liking>
 PROL: \$2 is <ind-exist,appreciable>
 PROL: \$2 is <ind-exist,appreciable>
 PATT: \$1 <word> que \$2 \$3

SEMR: AIMER(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 1
 FTRS: appreciate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,emotion,liking>
 PROL: \$2 is <ind-exist,appreciable>
 PROL: \$2 is <ind-exist,appreciable>
 PATT: \$1 <word> \$2 dans \$3

SEMR: AIMER(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: appreciate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,emotion,liking>
 PROL: \$2 is <ind-exist,appreciable>
 PATT: \$1 \$2 <word>

ENDW

#-----

WORD: arriver
 CATG: <predicate>
 FREQ: 91

SEMR: ARRIVER(\$1)
 FIDX: 1
 FREQ: 5
 FTRS: act,move
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PATT: \$1 <word>

SEMR: ARRIVER(\$1,\$2)
 FIDX: 2
 FREQ: 17
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> \$2

SEMR: ARRIVER(\$1,\$2)
 FIDX: 3
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <source>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,source>
 PATT: \$1 <word> du \$2

SEMR: ARRIVER(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> sur \$2

SEMR: ARRIVER(\$1,\$2)
 FIDX: 5
 FREQ: 2
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> dans \$2

SEMR: ARRIVER(\$1,\$2)
 FIDX: 6
 FREQ: 4
 FTRS: act,move,descend

ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <ind-exist,affected,motion,direction,coming>
 PATT: \$1 <word> avec \$2

ENDW

#-----

WORD: passer
 CATG: <predicate>
 FREQ: 89

SEMR: PASSER(\$1)
 FIDX: 1
 FREQ: 9
 FTRS: act,move
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,moving>
 PATT: \$1 <word>

SEMR: PASSER(\$1, \$2)
 FIDX: 2
 FREQ: 34
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <purpose>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <end,goal,usefulness,utility>
 PATT: \$1 <word> \$2

SEMR: PASSER(\$1, \$2)
 FIDX: 3
 FREQ: 10
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,moving>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> à \$2

SEMR: PASSER(\$1,\$2)
 FIDX: 4

FREQ: 5
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,moving>
 PATT: \$1 <word> sous \$2

SEMR: PASSER(\$1,\$2)
 FIDX: 5
 FREQ: 4
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,moving>
 PATT: \$1 <word> en \$2

SEMR: PASSER(\$1,\$2,\$3)
 FIDX: 6
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,moving>
 PROL: \$3 is <state,quality>
 PATT: \$1 <word> \$2 en \$3

SEMR: PASSER(\$1,\$2)
 FIDX: 7
 FREQ: 4
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,moving>
 PATT: \$1 <word> par \$2

SEMR: PASSER(\$1,\$2)
 FIDX: 8
 FREQ: 2

FTRS: spend
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause>
 PROL: \$2 is <ind-exist,change-location,affected,motion>
 PROL: \$3 is <volitional,affecting,animate,ind-exist,cause>
 PATT: \$1 <word> \$2 avec \$3

SEMR: PASSER(\$1,\$2)
 FIDX: 9
 FREQ: 1
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,direction,moving>
 PATT: \$1 <word> dans \$2

ENDW

#-----

WORD: croire
 CATG: <predicate>
 FREQ: 82

SEMR: CROIRE(\$1)
 FIDX: 1
 FREQ: 2
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PATT: \$1 <word>

SEMR: CROIRE(\$1,\$2)
 FIDX: 2
 FREQ: 14
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> que c'est \$2

SEMR: CROIRE(\$1,\$2)
 FIDX: 3
 FREQ: 2
 FTRS: act,cogitate
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,thinking>
 PROL: \$2 is <ind-exist,cogitable>
 PATT: \$1 <word> qu'il y a \$2

SEMR: CROIRE(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: act,trust
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,mind,reasoning>
 PROL: \$2 is <ind-exist,trustable>
 PATT: \$1 <word> pour \$2

ENDW

#-----

WORD: donner
 CATG: <predicate>
 FREQ: 79

SEMR: DONNER(\$1,\$2,\$3)
 FIDX: 1
 FREQ: 31
 FTRS: cause,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,affecting,cause,cause-possess,giving>
 PROL: \$2 is <animate,ind-exist,affected,possessing,given,beneficiary>
 PROL: \$3 is <affected,change,change-possession,ind-exist>
 PATT: \$1 \$2 <word> \$3

SEMR: DONNER(\$1,\$2,\$3)
 FIDX: 2
 FREQ: 2
 FTRS: cause,possess

ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,affecting,cause,cause-possess,giving>
 PROL: \$2 is <affected,change,change-possession,ind-exist>
 PROL: \$3 is <animate,ind-exist,affected,possessing,given,beneficiary>
 PATT: \$1 <word> \$2 aux \$3

SEMR: DONNER(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 1
 FTRS: cause,possess
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <beneficiary>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,affecting,cause,cause-possess,giving>
 PROL: \$2 is <affected,change,change-possession,ind-exist>
 PROL: \$3 is <animate,ind-exist,affected,possessing,given,beneficiary>
 PATT: \$1 <word> \$2 à \$3

ENDW

#-----

WORD: mettre
 CATG: <predicate>
 FREQ: 77

SEMR: METTRE(\$1,\$2)
 FIDX: 1
 FREQ: 8
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PATT: \$1 <word> \$2

SEMR: METTRE(\$1,\$2,\$3)
 FIDX: 2
 FREQ: 4
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>

ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <animate,ind-exist,affected,putting,beneficiary>
 PROL: \$3 is <ind-exist,change-location,affected,motion,stationary,putting>
 PATT: \$1 \$2 <word> en \$3

SEMR: METTRE(\$1,\$2,\$3)
 FIDX: 3
 FREQ: 3
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <beneficiary>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <animate,ind-exist,affected,putting,beneficiary>
 PROL: \$3 is <ind-exist,change-location,affected,motion,stationary,putting>
 PATT: \$1 \$2 <word> dans \$3

SEMR: METTRE(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 2
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PROL: \$3 is <location,destination,surface>
 PATT: \$1 <word> \$2 à \$3

SEMR: METTRE(\$1,\$2,\$3)
 FIDX: 5
 FREQ: 1
 FTRS: cause,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,change-location,affected,motion,stationary,putting>
 PROL: \$3 is <location,destination,surface>

PATT: \$1 <word> \$2 dans \$3

ENDW

#-----

WORD: parler

CATG: <predicate>

FREQ: 76

SEMR: PARLER(\$1,\$2)

FIDX: 1

FREQ: 14

FTRS: tell,affirm

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>

PROL: \$2 is <ind-exist,pronounceable,speakable>

PATT: \$1 <word> \$2

SEMR: PARLER(\$1,\$2)

FIDX: 2

FREQ: 26

FTRS: tell,affirm

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>

PROL: \$2 is <ind-exist,pronounceable,speakable>

PATT: \$1 <word> de \$2

SEMR: PARLER(\$1,\$2)

FIDX: 3

FREQ: 17

FTRS: tell,affirm

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>

PROL: \$2 is <ind-exist,pronounceable,speakable>

PATT: \$1 <word> des \$2

SEMR: PARLER(\$1,\$2)

FIDX: 4

FREQ: 4

FTRS: tell,affirm

ARGS: \$1 is <agent>

ARGS: \$2 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <volitional,affecting,animate,ind-exist,cause,saying>
 PATT: \$1 <word> à \$2

SEMR: PARLER(\$1,\$2)
 FIDX: 5
 FREQ: 1
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,pronounceable,speakable>
 PATT: \$1 <word> sur \$2

SEMR: PARLER(\$1,\$2)
 FIDX: 6
 FREQ: 2
 FTRS: tell,affirm
 ARGS: \$1 is <agent>
 ARGS: \$2 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <volitional,affecting,animate,ind-exist,cause,saying>
 PATT: \$1 <word> avec \$2

ENDW

#-----

WORD: appeler
 CATG: <predicate>
 FREQ: 76

SEMR: APPELER(\$1,\$2)
 FIDX: 1
 FREQ: 42
 FTRS: label,name
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,markable,noticeable>
 PATT: \$1 <word> \$2

SEMR: APPELER(\$1,\$2)
 FIDX: 2
 FREQ: 24
 FTRS: label,name
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,markable,noticeable>
 PATT: \$1 se <word> \$2

SEMR: APPELER(\$1,\$2)
 FIDX: 3
 FREQ: 3
 FTRS: label,name
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,markable,noticeable>
 PATT: \$1 me <word> \$2

SEMR: APPELER(\$1,\$2)
 FIDX: 4
 FREQ: 4
 FTRS: appeal,request
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,saying>
 PROL: \$2 is <ind-exist,helper,benefactor>
 PATT: \$1 <word> \$2

ENDW

#-----

WORD: rester
 CATG: <predicate>
 FREQ: 67

SEMR: RESTER(\$1)
 FIDX: 1
 FREQ: 2
 FTRS: be,remain
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>

PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PATT: \$1 <word>

SEMR: RESTER(\$1,\$2)
 FIDX: 2
 FREQ: 34
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <state,quality>
 PATT: \$1 <word> \$2

SEMR: RESTER(\$1,\$2)
 FIDX: 3
 FREQ: 8
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> à \$2

SEMR: RESTER(\$1,\$2)
 FIDX: 4
 FREQ: 6
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> dans \$2

SEMR: RESTER(\$1,\$2)
 FIDX: 5
 FREQ: 3
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> au \$2

SEMR: RESTER(\$1,\$2)

FIDX: 6
 FREQ: 1
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> en \$2

SEMR: RESTER(\$1,\$2)
 FIDX: 7
 FREQ: 2
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> sur \$2

SEMR: RESTER(\$1,\$2)
 FIDX: 8
 FREQ: 2
 FTRS: be,remain
 ARGS: \$1 is <agent>
 ARGS: \$2 is <co-agent>
 COND: \$1 is <animate>
 COND: \$2 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,staying>
 PROL: \$2 is <volitional,affecting,animate,ind-exist,cause,staying>
 PATT: \$1 <word> avec \$2

ENDW

#-----

WORD: apprendre
 CATG: <predicate>
 FREQ: 44

SEMR: APPRENDRE(\$1,\$2)
 FIDX: 1
 FREQ: 15
 FTRS: act,learn
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>

PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> \$2

SEMR: APPRENDRE(\$1,\$2)
 FIDX: 2
 FREQ: 18
 FTRS: act,learn
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PROL: \$2 is <ind-exist,perceivable>
 PATT: \$1 <word> à \$2

SEMR: APPRENDRE(\$1,\$2)
 FIDX: 3
 FREQ: 1
 FTRS: act,learn
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,studying>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> dans \$2

ENDW

#-----

WORD: manger
 CATG: <predicate>
 FREQ: 37

SEMR: MANGER(\$1)
 FIDX: 1
 FREQ: 1
 FTRS: act,consume
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,consuming,eating>
 PATT: \$1 <word>

SEMR: MANGER(\$1,\$2)
 FIDX: 2
 FREQ: 7
 FTRS: act,consume
 ARGS: \$1 is <agent>

ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <edible>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,consuming,eating>
 PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>
 PATT: \$1 <word> \$2

SEMR: MANGER(\$1,\$2)
 FIDX: 3
 FREQ: 6
 FTRS: act,consume
 ARGS: \$1 is <agent>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,consuming,eating>
 PROL: \$2 is <state,quality>
 PATT: \$1 <word> \$2

SEMR: MANGER(\$1,\$2)
 FIDX: 4
 FREQ: 3
 FTRS: act,consume
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,consuming,eating>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> dans \$2

SEMR: MANGER(\$1,\$2)
 FIDX: 5
 FREQ: 3
 FTRS: act,consume
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-matter,consuming,eating>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> à \$2

SEMR: MANGER(\$1,\$2)
 FIDX: 6
 FREQ: 2
 FTRS: act,consume
 ARGS: \$1 is <agent>

ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,consuming,eating>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> au \$2

SEMR: MANGER(\$1,\$2)
 FIDX: 7
 FREQ: 1
 FTRS: act,consume
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,consuming,eating>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> en \$2

ENDW

#-----

WORD: essayer
 CATG: <predicate>
 FREQ: 35

SEMR: ESSAYER(\$1)
 FIDX: 1
 FREQ: 1
 FTRS: act,attempt
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,trying>
 PATT: \$1 <word>

SEMR: ESSAYER(\$1,\$2)
 FIDX: 2
 FREQ: 14
 FTRS: act,attempt
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,trying>
 PROL: \$2 is <ind-exist,affected>
 PATT: \$1 <word> de \$2

SEMR: ESSAYER(\$1,\$2,\$3)

FIDX: 3
 FREQ: 11
 FTRS: act,attempt
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,trying>
 PROL: \$2 is <ind-exist,affected>
 PROL: \$3 is <ind-exist,affected>
 PATT: \$1 <word> \$2

SEMR: ESSAYER(\$1,\$2,\$3)
 FIDX: 4
 FREQ: 1
 FTRS: act,attempt
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is destination>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,trying>
 PROL: \$2 is <ind-exist,affected>
 PROL: \$3 is <location,path,destination>
 PATT: \$1 <word> \$2 en \$2

ENDW

#-----

WORD: revenir
 CATG: <predicate>
 FREQ: 35

SEMR: REVENIR(\$1)
 FIDX: 1
 FREQ: 4
 FTRS: act,move
 ARGS: \$1 is <agent>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PATT: \$1 <word>

SEMR: REVENIR(\$1,\$2)
 FIDX: 2
 FREQ: 9
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$3 is <state>

COND: \$1 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,motion,direction,coming>
 PROL: \$3 is <state,quality>
 PATT: \$1 <word> \$2

SEMR: REVENIR(\$1,\$2)
 FIDX: 3
 FREQ: 4
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> \$2

SEMR: REVENIR(\$1,\$2)
 FIDX: 4
 FREQ: 2
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> en \$2

SEMR: REVENIR(\$1,\$2)
 FIDX: 5
 FREQ: 9
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$3 is <state>
 COND: \$1 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,motion,direction,coming>
 PROL: \$3 is <state,quality>
 PATT: \$1 <word> à \$2

SEMR: REVENIR(\$1,\$2)
 FIDX: 6
 FREQ: 9
 FTRS: act,move
 ARGS: \$1 is <agent>
 ARGS: \$3 is <state>
 COND: \$1 is <animate>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,motion,direction,coming>
 PROL: \$3 is <state,quality>

PATT: \$1 <word> aux \$2

SEMR: REVENIR(\$1,\$2)

FIDX: 7

FREQ: 9

FTRS: act,move

ARGS: \$1 is <agent>

ARGS: \$2 is <time>

COND: \$1 is <animate>

COND: \$2 is <hour>

PROL: \$1 is <volitional,sentient,animate,ind-exist,change-location,motion,direction,coming>

PROL: \$2 is <hour,period,time>

PATT: \$1 <word> à \$2

ENDW

#-----

WORD: regarder

CATG: <predicate>

FREQ: 34

SEMR: REGARDER(\$1,\$2)

FIDX: 1

FREQ: 21

FTRS: perceive

ARGS: \$1 is <experiencer>

ARGS: \$2 is <stimulus>

COND: \$1 is <animate>

COND: \$2 is <thing>

PROL: \$1 is <sentient,animate,ind-exist,visual,seeing>

PROL: \$2 is <ind-exist,perceivable>

PATT: \$1 <word> \$2

SEMR: REGARDER(\$1,\$2)

FIDX: 2

FREQ: 5

FTRS: look,observe

ARGS: \$1 is <experiencer>

ARGS: \$2 is <stimulus>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,watching>

PROL: \$2 is <ind-exist,perceptible>

PATT: \$1 <word> \$2

SEMR: REGARDER(\$1,\$2,\$3)

FIDX: 3

FREQ: 1

FTRS: look,observe

ARGS: \$1 is <experiencer>

ARGS: \$2 is <stimulus>

ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,visual,watching>
 PROL: \$2 is <ind-exist,perceptible>
 PROL: \$3 is <ind-exist,tool>
 PATT: \$1 <word> \$2 à \$3

SEMR: REGARDER(\$1,\$2)
 FIDX: 4
 FREQ: 2
 FTRS: examine,explore
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>
 PROL: \$2 is <ind-exist>
 PATT: \$1 <word> ce-que \$2

SEMR: REGARDER(\$1,\$2)
 FIDX: 5
 FREQ: 2
 FTRS: examine,explore
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>
 PROL: \$2 is <ind-exist>
 PATT: \$1 <word> sur \$2

SEMR: REGARDER(\$1,\$2)
 FIDX: 6
 FREQ: 1
 FTRS: attend,see
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,looking,ind-exist,visiting>
 PROL: \$2 is <ind-exist>
 PATT: \$1 <word> \$2

ENDW

#-----

WORD: chercher
 CATG: <predicate>

FREQ: 32

SEMR: CHERCHER(\$1,\$2)

FIDX: 1

FREQ: 22

FTRS: examine,explore

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>

PROL: \$2 is <ind-exist,searchable>

PATT: \$1 <word> \$2

SEMR: CHERCHER(\$1,\$2)

FIDX: 2

FREQ: 1

FTRS: examine,explore

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>

PROL: \$2 is <ind-exist>

PATT: \$1 <word> sur \$2

SEMR: CHERCHER(\$1,\$2)

FIDX: 3

FREQ: 4

FTRS: examine,explore

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>

PROL: \$2 is <ind-exist>

PATT: \$1 <word> dans \$2

SEMR: CHERCHER(\$1,\$2)

FIDX: 4

FREQ: 3

FTRS: examine,explore

ARGS: \$1 is <agent>

ARGS: \$2 is <destination>

COND: \$1 is <animate>

COND: \$2 is <location>

PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>

PROL: \$2 is <location,path,destination>

PATT: \$1 <word> \$2

SEMR: CHERCHER(\$1,\$2)

FIDX: 5

FREQ: 1
 FTRS: examine,explore
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 ARGS: \$3 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 COND: \$3 is <concrete>
 PROL: \$1 is <volitional,sentient,searching,ind-exist,investigating>
 PROL: \$2 is <ind-exist,searchable>
 PROL: \$3 is <ind-exist,tool>
 PATT: \$1 <word> \$2 avec \$3

ENDW

#-----

WORD: visiter
 CATG: <predicate>
 FREQ: 20

SEMR: VISITER(\$1,\$2)

FIDX: 1

FREQ: 18

FTRS: attend,see

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,looking,ind-exist,visiting>

PROL: \$2 is <ind-exist>

PATT: \$1 <word> \$2

ENDW

#-----

WORD: boire
 CATG: <predicate>
 FREQ: 17

SEMR: BOIRE(\$1,\$2)

FIDX: 1

FREQ: 3

FTRS: act,absorb

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <potable>

PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,absorbing,drinking>

PROL: \$2 is <ind-exist,affected,change,change-matter,stationary>
 PATT: \$1 <word> \$2

SEMR: BOIRE(\$1,\$2)

FIDX: 2

FREQ: 2

FTRS: act,absorb

ARGS: \$1 is <agent>

ARGS: \$2 is <state>

COND: \$1 is <animate>

COND: \$2 is <concrete>

PROL: \$1 is <volitional,sentient,affecting,animate,ind-exist,cause,cause-
 matter,absorbing,drinking>

PROL: \$2 is <state,quality>

PATT: \$1 <word> \$2

ENDW

#-----

WORD: jouer

CATG: <predicate>

FREQ: 16

SEMR: JOUER(\$1,\$2)

FIDX: 1

FREQ: 4

FTRS: act,playfully

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <game>

PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>

PROL: \$2 is <non-ind-exist,theme,activity,change,change-formal>

PATT: \$1 <word> \$2

SEMR: JOUER(\$1,\$2)

FIDX: 2

FREQ: 3

FTRS: act,playfully

ARGS: \$1 is <agent>

ARGS: \$2 is <theme>

COND: \$1 is <animate>

COND: \$2 is <game>

PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>

PROL: \$2 is <non-ind-exist,theme,activity,change,change-formal>

PATT: \$1 <word> à \$2

SEMR: JOUER(\$1,\$2)

FIDX: 3

FREQ: 4

FTRS: act,play

ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <ind-exist,affected,motion,drama,show>
 PATT: \$1 <word> \$2

SEMR: JOUER(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: act,play
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,playing>
 PROL: \$2 is <ind-exist,affected,motion,music>
 PATT: \$1 <word> de \$2

ENDW

#-----

WORD: tomber
 CATG: <predicate>
 FREQ: 12

SEMR: TOMBER(\$1)
 FIDX: 1
 FREQ: 1
 FTRS: move,fall
 ARGS: \$1 is <theme>
 COND: \$1 is <thing>
 PROL: \$1 is <ind-exist,change-location,motion,direction,vertical,falling>
 PATT: \$1 <word>

SEMR: TOMBER(\$1,\$2)
 FIDX: 2
 FREQ: 2
 FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> \$2

SEMR: TOMBER(\$1,\$2)
 FIDX: 3
 FREQ: 1

FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> sur \$2

SEMR: TOMBER(\$1,\$2)
 FIDX: 4
 FREQ: 1
 FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> à \$2

SEMR: TOMBER(\$1,\$2)
 FIDX: 5
 FREQ: 1
 FTRS: move
 ARGS: \$1 is <agent>
 ARGS: \$2 is <destination>
 COND: \$1 is <animate>
 COND: \$2 is <location>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <location,path,destination>
 PATT: \$1 <word> par \$2

SEMR: TOMBER(\$1,\$2)
 FIDX: 6
 FREQ: 3
 FTRS: act,play
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,affected>
 PATT: \$1 <word> dans \$2

SEMR: TOMBER(\$1,\$2)
 FIDX: 7
 FREQ: 1
 FTRS: act,play
 ARGS: \$1 is <agent>
 ARGS: \$2 is <theme>
 COND: \$1 is <animate>

COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <ind-exist,affected>
 PATT: \$1 <word> sur \$2

SEMR: TOMBER(\$1,\$2)
 FIDX: 8
 FREQ: 2
 FTRS: act,play
 ARGS: \$1 is <agent>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,affecting,animate,ind-exist,cause,cause-location>
 PROL: \$2 is <state,quality>
 PATT: \$1 <word> en \$2

ENDW

#-----

WORD: écouter
 CATG: <predicate>
 FREQ: 12

SEMR: ECOUTER(\$1,\$2)
 FIDX: 1
 FREQ: 5
 FTRS: act,attend
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <stimulus>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,listening,attending>
 PROL: \$2 is <ind-exist,listenable>
 PATT: \$1 <word> \$2

SEMR: ECOUTER(\$1,\$2)
 FIDX: 2
 FREQ: 1
 FTRS: act,attend
 ARGS: \$1 is <experiencer>
 ARGS: \$2 is <state>
 COND: \$1 is <animate>
 COND: \$2 is <concrete>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,listening,attending>
 PROL: \$2 is <state,quality>
 PATT: \$1 <word> \$2

SEMR: ECOUTER(\$1)
 FIDX: 3
 FREQ: 1

FTRS: act,attend
 ARGS: \$1 is <experiencer>
 COND: \$1 is <animate>
 PROL: \$1 is <volitional,sentient,animate,ind-exist,listening,attending>
 PATT: \$1 <word>

ENDW

#-----
 #----- Example of WORDS -----
 #-----

WORD: il
 CATG: <noun-category>
 FREQ: 0
 SEMR: il
 ENDW

#-----
 WORD: on
 CATG: <noun-category>
 FREQ: 0
 SEMR: on
 ENDW

#-----
 WORD: je
 CATG: <noun-category>
 FREQ: 0
 SEMR: je
 ENDW

#-----
 WORD: tu
 CATG: <noun-category>
 FREQ: 0
 SEMR: tu
 ENDW

#-----
 WORD: ils
 CATG: <noun-category>
 FREQ: 0
 SEMR: ils
 ENDW

#-----
 WORD: tourEiffel
 CATG: <noun-category>
 FREQ: 0

SEMR: tourEiffel
ENDW

#-----

WORD: image
CATG: <noun-category>
FREQ: 0
SEMR: image
ENDW

#-----

WORD: images
CATG: <noun-category>
FREQ: 0
SEMR: images
ENDW

#-----

WORD: caisse
CATG: <noun-category>
FREQ: 0
SEMR: caisse
ENDW

#-----

WORD: shampoing
CATG: <noun-category>
FREQ: 0
SEMR: shampoing
ENDW

#-----

WORD: film
CATG: <noun-category>
FREQ: 0
SEMR: film
ENDW

#-----

WORD: jeux
CATG: <noun-category>
FREQ: 0
SEMR: jeux
ENDW

#-----