
Notes de lecture

Rubrique préparée par Denis Maurel

Université François Rabelais Tours, LI (Laboratoire d'informatique)

Afra ALISHAHI, Computational Modeling of Human Language Acquisition, Morgan & Claypool publishers, 2011, 93 pages, ISBN 978-1-6084-5339-9.

Lu par **Pascal AMSILI**

Université Paris-Diderot, LLF

L'acquisition des langues intéresse les chercheurs depuis des siècles et cette question a été explorée avec toutes sortes de méthodes, mais c'est seulement très récemment que la modélisation informatique est venue enrichir l'inventaire des méthodes disponibles. C'est un mouvement assez naturel, si l'on s'avise que la recherche en TAL fait appel de façon de plus en plus massive à des méthodes d'apprentissage automatique. Il faut aussi ajouter que des données expérimentales et des corpus pertinents ont été rendus accessibles récemment. Cette situation rend la parution du livre d'Afra Alishahi très utile, aussi bien pour les chercheurs venus du TAL qui s'intéressent à l'application de leurs méthodes à la question de l'acquisition des langues, qu'aux linguistes et psycholinguistes qui veulent savoir ce que l'informatique peut apporter à leur domaine de recherche.

Ce livre est le onzième d'une série récemment publiée chez Morgan & Claypool, sous la direction de Graeme Hirst, titrée « *Synthesis Lectures on Human Language Technologies* ». Le concept est de fournir des textes relativement courts qui font un survol récent et informé de différentes questions relevant des technologies linguistiques, au sens le plus large possible. Il réalise parfaitement les objectifs de la série et propose un survol synthétique, et relativement complet, de la question.

L'ouvrage est organisé en six chapitres qui peuvent être regroupés en trois parties distinctes. La première partie – premier et deuxième chapitres – présente la problématique et les principales méthodologies du domaine. Dans le premier chapitre, on montre quelques-uns des grands débats qui ont animé le domaine de l'acquisition des langues, avec la question de la modularité du langage et celle de l'opposition entre la vue innéiste du langage, célèbre pour l'argument de la pauvreté du stimulus proposé par Chomsky en 1965, et la vue « *usage-based* », qui explore sans *a priori* les possibilités d'induire une compétence linguistique à partir des données typiquement accessibles aux enfants pendant leur apprentissage. Quelques pointeurs sur les résultats mathématiques pertinents pour ce débat sont redonnés plus loin dans l'ouvrage (p. 45-46). Cette présentation rapide du contexte ne sera sans doute pas très utile aux chercheurs ayant déjà une formation en linguistique, mais

elle a le mérite de replacer ces recherches dans leur contexte, qui reste un contexte très débattu aujourd'hui. Le chapitre 1 présente aussi des arguments en faveur de l'ajout de méthodes de modélisation informatique à l'inventaire des méthodes de la psycholinguistique. Le chapitre 2 présente la problématique de la modélisation computationnelle et discute en particulier deux aspects importants qui sont d'une part la question de la plausibilité cognitive des modèles que l'on produit (question rarement pertinente en TAL, ce qui rend beaucoup de modèles venus du TAL inappropriés pour les sciences cognitives), et, d'autre part, la question de l'évaluation, sachant que là encore les habitudes venues du TAL doivent être adaptées, puisque l'on ne dispose en général pas de « *gold standard* » pour les tâches modélisées.

La deuxième partie regroupe les trois chapitres suivants, qui sont chacun consacrés à un domaine particulier : l'apprentissage du vocabulaire (chapitre 3), l'apprentissage de la morphologie et de la syntaxe (chapitre 4) et, enfin, l'apprentissage de la relation syntaxe-sémantique, tout du moins en ce qui concerne les structures argumentales et les rôles thématiques (chapitre 5). Enfin, la troisième partie, représentée par le chapitre 6, conclut l'ouvrage avec quelques remarques conclusives.

Chacun des chapitres de cette deuxième partie est organisé de manière similaire, avec une division en sous-domaines (par exemple la morphologie, les parties du discours et la grammaire pour le chapitre 4), et chaque fois une présentation de la problématique, avec une revue assez exhaustive des modèles récents qui ont pu être proposés et une « étude de cas », censée présenter de façon plus approfondie un modèle particulier.

La structure de l'ouvrage, et son contenu, en font un excellent outil pour tout chercheur qui s'intéresse à la modélisation de l'acquisition des langues, qu'il vienne du champ du TAL ou de la psychologie cognitive, et qui va trouver ici un texte bref, dans lequel se trouvent listées la plupart des références pertinentes dans le domaine. De plus, l'autonomie des différentes parties contribue à faire de ce livre un texte de référence facile à consulter et qui pourra même être utile aux chercheurs déjà engagés dans le champ du TAL ou autres, qui trouveront un inventaire de références utiles.

Cependant, le projet de l'ouvrage conduit à une limite qui peut engendrer une certaine frustration : le niveau de détail des discussions et des explications est souvent insuffisant en soi et rend nécessaire la consultation des sources directes, mais ce serait malvenu d'en faire le reproche à l'auteure compte tenu du projet et des contraintes de la collection.

On peut cependant exprimer quelques regrets, qui n'enlèvent rien à l'intérêt de l'ouvrage. Tout d'abord, les études de cas ne sont pas la partie la plus réussie de l'ouvrage. On comprend le but, qui était de donner un peu de chair à un texte qui est souvent assez abstrait et comprend surtout des inventaires. Mais le niveau de détail des études de cas est souvent insuffisant pour qu'elles puissent vraiment être

exploitées (on suppose dans ces études une compétence accrue du lecteur). Par ailleurs, leur choix semble assez arbitraire, car il ne s'agit pas toujours d'un modèle particulièrement influent, ou représentatif. Enfin, il semble que l'importance donnée aux approches neuromimétiques et, en corrélation, la faible représentation des approches bayésiennes, ne reflètent pas la réalité du domaine aujourd'hui.

Pour des raisons évidentes de place, l'auteure est conduite à une certaine simplification des positions quand il s'agit de rendre compte des quelques grands débats qui traversent (quelquefois très violemment) la communauté. Ainsi l'opposition entre ce qu'on appelle le « *semantic bootstrapping* » et les approches dites « *usage-based* » est résumée en quelques lignes d'une façon forcément un peu réductrice. Mais il faut créditer l'auteure de sa neutralité assumée, qui la conduit à présenter dans tous les cas les réussites et les manques des approches en présence.

Sur un plan différent, il faut noter que tous les domaines de l'acquisition ne sont pas couverts. En particulier, les domaines de la phonologie, de la phonétique, et de la pragmatique, ne sont pas du tout évoqués, alors que dans ces domaines aussi, la convergence entre les psycholinguistes et les informaticiens s'est réalisée (en particulier dans le domaine de la phonologie).

D'autres domaines pertinents ont été négligés, là encore sans doute pour des raisons de place. D'abord, on ne mentionne que l'acquisition de la première langue, alors que de nombreux travaux existent sur l'apprentissage d'une langue seconde (en linguistique, en psycholinguistique et en modélisation). Il faudrait ajouter le domaine de l'acquisition pathologique, qui est certes un domaine plus délicat, mais où, là encore, une convergence est ébauchée entre les psycholinguistes et les informaticiens. Enfin, il serait intéressant d'ajouter à ce panorama les domaines où la psycholinguistique et la modélisation trouvent des points de rencontre, mais qui ne concernent pas directement l'acquisition. Par exemple, les études psycholinguistiques sur le temps de traitement (par exemple en syntaxe) font intervenir une notion de complexité qui peut être approchée par des modèles informatiques venant de la communauté du « *parsing* » (probabiliste).

Mais on s'éloigne clairement du domaine couvert par cet ouvrage de synthèse, dont la lecture doit être absolument recommandée, malgré les quelques remarques qui précèdent, à tout chercheur ou étudiant intéressé par l'application des techniques du TAL à des investigations psycholinguistiques.

Ismail BISKRI, Adel JEBALI, Traitement automatique des langues naturelles : de l'analyse à l'application, Hermès-Lavoisier, 2011, 256 pages, ISBN 978-2-7462-3138-7.

Lu par **Elisa Lavagnino**

Centre de recherche CeRTeM – Université de Gênes