

Rubrique préparée par Denis Maurel

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

Gérard SABAH, Compréhension des langues et interaction, *Hermes Lavoisier*, 2006, 400 pages, ISBN 2-7462-1256-0.

Lu par **Guillaume PITEL**

LORIA, équipe Langue et Dialogue, Nancy

Cet ouvrage collectif, conçu sous la direction de Gérard Sabah, fait partie d'un traité ayant pour objectif d'écrire l'état de l'art dans plusieurs domaines, dans un « esprit d'échange disciplinaire ». Il faut noter que les quatorze chapitres ne se valent pas du point de vue de l'objectif affiché par le traité. Certains présentent en détail des travaux ou des points de vue particuliers, tandis que d'autres exposent de manière moins approfondie les domaines de recherche. Cette diversité rend l'ouvrage intéressant à la fois pour ceux qui souhaitent acquérir une connaissance d'ensemble, mais aussi pour des connaisseurs curieux de s'enrichir dans certains domaines. L'ouvrage débute par quelques indispensables généralités et rappels sur la langue et son traitement, sur les notions de sens et de signification, sur la compositionnalité et les différents niveaux d'analyse. Le reste des chapitres se compose de deux parties, la première plus théorique et généraliste, la seconde plus appliquée.

La première partie, « Les fondements », commence avec le chapitre 2, « *Le point sur l'état actuel des connaissances en traitement automatique des langues* », d'Adeline Nazarenko. Il fournira un excellent point de départ pour celles et ceux peu familiers du domaine du TAL. Après les bases allant de l'analyse morphologique jusqu'à la pragmatique, l'auteur s'intéresse aux changements qui sont apparus autour des années 1990, avec la lexicalisation des traitements et l'intérêt accru porté aux corpus. L'auteur termine en mettant l'accent sur l'enjeu actuel présenté par les ressources et leur acquisition, sur l'importance des spécificités de chaque application et des boîtes à outils pour le TAL.

Le chapitre 3, « *Modélisation en sémantique* », de Patrice Enjalbert et Bernard Victorri, aborde la question de la sémantique en suivant une vision par paliers de complexité croissante. Ils évoquent rapidement le calcul du sens et la désambiguïsation, avant de présenter en détails un modèle de la sémantique spatiale et du mouvement. Sont ainsi évoqués l'approche montagovienne, la référence, la temporalité, la structure thématique et rhétorique ainsi que la cohésion textuelle. Le chapitre s'achève sur la présentation d'un modèle de la temporalité pour le français.

Le chapitre est bien équilibré, mais on pourra regretter le localisme des références sur les modèles étudiés, qui sont souvent françaises.

Dans le chapitre 4 de Jacques Moeschler et Anne Reboul, intitulé « *Compréhension, pragmatique et argumentation* », les auteurs soulignent l'importance du contexte dans la compréhension. Ils décrivent l'importance de la sous-détermination des expressions et de l'intention informative et communicative dans le processus de compréhension. On y trouvera une revue des formalisations des différents aspects de la pragmatique, ainsi que quelques bases sur l'argumentation.

Le chapitre 5, « *Compréhension et interaction* », d'Anne Nicolle, s'intéresse plus particulièrement à la gestion de la communication entre l'humain et la machine. Les différentes architectures d'un système de TAL sont exposées, allant des architectures en tuyau aux architectures multi-agents. Un état de l'art plus spécifique au dialogue homme-machine est dressé, portant sur la structure du dialogue, les modèles BDI et les modèles de planification. Le chapitre se termine sur une étude de l'importance de l'interaction dans la compréhension et de la représentation mutuelle des agents.

Dans le chapitre 6, « *Références* », Bertrand Gaiffe expose la problématique de la résolution de la référence, en commençant par donner différentes typologies des expressions référentielles. Il dépeint par la suite le rôle de la résolution de la référence dans les différentes strates de l'analyse linguistique, en détaillant notamment la DRT et ses extensions. Enfin, il décrit l'approche des représentations mentales et des domaines de référence, d'orientation plus cognitive, et expose la complémentarité de cette théorie avec la DRT.

Le chapitre 7, « *Rhétorique et compréhension* » rédigé par Stéphane Ferrari, s'intéresse aux différents effets de styles et à leur importance dans le TAL. Les différents types de métaphores en sont l'objet principal, avec une présentation des modèles actuellement employés. On peut noter que des travaux récents, comme ceux sur le Blending, sont rapidement évoqués, après les modèles standard plus anciens. Quelques liens avec la LSA, et donc la psycholinguistique, sont aussi exposés.

La seconde partie, « *Domaines d'application de la compréhension des langues* » débute avec le chapitre 8, « *Le traitement des nouvelles formes de communication écrite* », où Jean Véronis et Emilie Guimier de Neef présentent les particularités des formes nouvelles de communication écrite, comme par exemple les échanges sur les forums, les messageries instantanées, les blogs ou encore les SMS. Le chapitre énumère les différents phénomènes propres à ces médias de communication, et en tire les conséquences pour le TAL. Ce type de recherche est assez pointu, et ne trouvera pas forcément sa place dans un plan de lecture destiné à une première approche du TAL.

Le chapitre 9, « *TAL sur corpus : histoire, acquis, défis* » de Benoît Habert, aborde dans le détail l'annotation de corpus, l'évolution et la convergence des approches par modèle (IA traditionnelle) et par observation (acquisition à partir de

données). L'auteur décrit les ressources et méthodologies actuellement établies, avant de conclure sur les défis que présentent encore cette approche, parmi lesquels : la mesure de la fiabilité, l'évaluation de ce que représente (linguistiquement) un corpus, les ressources linguistiques tirées des corpus, et enfin l'articulation entre règles et régularités.

Le chapitre **10**, « *Systèmes de questions-réponses* », de Marc El-Bèze, s'intéresse aux systèmes permettant de répondre à des questions portant sur des sujets ouverts, et recourant pour cela aux corpus. Après avoir exposé le lien de ce domaine avec le développement de ressources textuelles à large couverture, il en présente les caractéristiques et architectures typiques, avant de détailler les problèmes particuliers qui y sont rencontrés. Même si ce chapitre traite d'un sujet assez pointu, il s'agit d'un bon exemple de la manière dont des applications du TAL peuvent fournir des occasions de se confronter à des problèmes assez théoriques.

Dans le chapitre **11**, « *Les systèmes de résumé automatiques : comment assurer une continuité référentielle dans la lecture des textes* », Delphine Battistelli et Jean-Luc Minel décrivent les différents usages des résumés et la palette des solutions proposées pour automatiser le processus. Les auteurs évoquent ensuite des perspectives portant sur le traitement référentiel, primordial dans toutes les approches, la temporalité et son intégration dans les résumés. Le chapitre se termine sur une section portant sur la navigation dans les documents. Le sujet traité est un peu pointu, mais la première partie du chapitre est abordable par tous.

Le chapitre **12**, « *Dialogue et Apprentissage* », de Daniel Luzzati, aborde le sujet des systèmes de dialogue sous un angle bien particulier. On trouve une argumentation propre à l'auteur dans laquelle il oppose système exécuté et système apprenant par l'expérience (par apprentissage essai/erreur) et développe différentes propositions en se plaçant dans cette optique. Ce chapitre ne me semble pas destiné à un public peu initié au TAL, ne présentant que partiellement les autres approches.

Le chapitre **13**, « *Compréhension et multilinguisme* », écrit par Guy Lapalme et Elliott Macklovitch, est consacré en grande partie à la traduction. Après en avoir fait un historique et présenté la vision traditionnelle de la traduction (compréhension/production), les auteurs abordent la question cruciale de la nécessité (ou non) de la compréhension pour la traduction, introduisant les approches probabilistes.

En conclusion, dans « *Compréhension automatique des langues : où va-t-on ? Où pourrait-on aller ?* », Gérard Sabah fait le point sur les différences entre approches symboliques pour la compréhension profonde et approches statistiques, puis propose quelques pistes pour ces deux approches. Il présente aussi quelques éléments où l'inspiration cognitive est très présente, pour tracer une voie vers un modèle implémentable de la compréhension.

Denis CREISSELS, Syntaxe générale, une introduction typologique, tome 1, catégories et constructions, *Hermès Lavoisier*, 2006, 412 pages, ISBN 2-7462-1399-0.

Lu par **Emmanuel Schang**

Coral - Université d'Orléans

Cet ouvrage constitue le premier d'une série de deux volumes consacrés à la syntaxe générale. Tandis que le second volume est consacré à la phrase, celui-ci traite des catégories et constructions, comme l'indique le sous-titre. Il est composé d'une table des matières détaillée, d'une liste commentée des gloses et abréviations, d'une introduction suivie de vingt chapitres, de la liste des références bibliographiques, de deux index (thèmes et langues) et s'achève par le sommaire du second volume. Chaque chapitre se conclut par une notice bibliographique dont les références sont commentées.

L'introduction situe l'ouvrage parmi les travaux de D. Creissels et explicite la position de l'auteur face aux deux grands courants théoriques prétendument opposés que sont l'approche formaliste et l'approche typologique. D. Creissels adopte une position originale (et salutaire pour beaucoup de linguistes descriptivistes) en refusant l'assimilation entre « formalisme » et « générativisme » ainsi qu'entre « typologie » et « fonctionnalisme », et en défendant une approche lexicaliste et réaliste (refus des transformations syntaxiques, des éléments vides ou effacés). Il est souvent fait référence à HPSG comme exemple de théorie formelle compatible avec les analyses de l'ouvrage. Cette introduction, au ton parfois polémique¹, est la seule partie de ce livre qui soit explicitement consacrée à une discussion sur les différentes approches syntaxiques.

Le reste des chapitres se présente comme un cours de syntaxe qui commence par la notion de mot, de phrase, de constituant, pour aller progressivement vers des notions plus complexes, comme par exemple, les constructions personnelles et anti-personnelles. Le linguiste connaissant les travaux de D. Creissels retrouvera ici un nombre important de notions déjà abordées dans *Description des langues négro-africaines et théorie syntaxique*, paru en 1991, ainsi qu'une manière d'exposer les problèmes habituelle à l'auteur. Toutefois, ces notions sont souvent abordées sous un autre angle et réactualisées par des références à des travaux récents. Chaque chapitre traite d'une notion d'un point de vue typologique, en essayant donc de

¹ Page 4 : « Dans le courant chomskyen orthodoxe actuel, il ne reste pratiquement plus rien de l'objectif de formalisation affiché au départ, sinon une prolifération de symboles et de schémas susceptible d'impressionner seulement des linguistes de formation littéraire qui ignorent ce qu'est une réelle formalisation, du type pratiqué par exemple en grammaire catégorielle ou en sémantique formelle. »

dégager les points communs à la plupart des langues du monde ainsi que les limites de la variation constatée sur chaque point de syntaxe. Comme dans les ouvrages précédents, l'analyse découle de la possibilité de dégager dans toutes les langues une opposition nom/verbe, qui repose elle-même sur la possibilité d'identifier universellement le nom propre comme étant le prototype de la notion grammaticale de nom.

Cet ouvrage est rédigé de façon claire, parfois dense mais toujours agréable et les analyses sont remarquablement bien illustrées par des exemples abondants pris dans de nombreuses langues (et notamment dans celles sur lesquelles l'auteur travaille depuis longtemps : tswana, mandinka et hongrois). Le public visé par ce livre est bien entendu le linguiste (confirmé ou étudiant) s'intéressant à la description de langues peu documentées. Il trouvera ici un guide précis et complet des notions qu'il est souhaitable de connaître avant de se lancer dans l'étude de ces langues. Mais on ne peut que conseiller sa lecture à tous ceux qui travaillent dans le domaine de la linguistique formelle et à ceux qui, dans le domaine du traitement automatique, pensent qu'il n'est pas inutile de tenir compte de la diversité des langues. Grâce aux nombreux exemples, le lecteur habitué aux Grammaires de Constructions ou à HPSG trouvera ici (et peut-être plus encore dans le second volume) de quoi alimenter sa réflexion théorique.

Denis CREISSELS, Syntaxe générale, une introduction typologique, tome 2, la phrase, *Hermes Lavoisier*, 2006, 334 pages, ISBN 2-7462-1400-8.

Lu par Christophe Benzitoun

Équipe Delic - Université de Provence

L'ouvrage de D. Creissels représente la première introduction à la syntaxe générale de cette ampleur en langue française. Et malgré cela, il possède deux qualités complémentaires que l'on ne s'attendrait pas forcément à trouver dans un travail précurseur de ce type : la recherche du maximum de couverture en termes de nombre de langues et de sujets abordés et la précision de certaines descriptions, rivalisant avec des descriptions unilingues, en particulier pour le français. En effet, nous avons relevé environ une centaine de langues différentes prises comme sujet d'étude. De plus, l'ouvrage ne compte pas moins de dix-sept chapitres traitant de thèmes aussi divers que la valence verbale, le passif, la topicalisation, la focalisation, la négation, les types énonciatifs de phrases, les phrases complexes (coordination et subordination), les divers types de relatives, les constructions à montée et à contrôle, etc. Il serait évidemment trop long de décrire chaque chapitre, et deux ou trois lignes pour chacun d'eux ne permettraient pas de rendre justice au travail de l'auteur. Nous avons donc choisi de détailler les analyses proposées entre le chapitre 31 et le chapitre 35, qui concernent l'étude des phrases dites complexes.

Pour commencer, signalons que, lorsque cela est possible, l'auteur donne des exemples presque systématiquement en français afin de permettre aux lecteurs francophones d'appréhender plus facilement des configurations dans des langues qui ne lui sont pas familières. Il pose aussi des questions syntaxiques essentielles qui ne vont pas de soi même pour des langues aussi étudiées que le français et l'anglais : « À partir de quel moment peut-on considérer que les contraintes auxquelles obéissent les enchaînements d'unités phrastiques cessent d'être la pure et simple conséquence d'un principe discursif de cohérence pour devenir des règles de syntaxe ? » (p. 184). Il y répond en détaillant les diverses stratégies que les langues ont mises en place pour signaler des regroupements interphrastiques. Ainsi, en tswana, par exemple, les constructions séquentielles possèdent des formes verbales spéciales que l'on ne retrouve pas dans des phrases indépendantes.

D. Creissels s'attaque aussi au difficile problème de la subordination. Après une critique justifiée de la conception traditionnelle, il la redéfinit en termes de relation hiérarchique se manifestant par un déséquilibre des possibilités de modulation énonciative (assertion/interrogation/injonction) : modulation équivalente à celle d'une phrase simple indépendante pour la phrase matrice et bloquée pour la subordonnée. L'auteur mentionne en outre la portée de la négation ainsi que la focalisation entre *c'est* et *que* et l'ordre des constituants comme autres critères pour distinguer entre subordonnées et phrases qui ont l'apparence de ces dernières mais qui n'en sont pas. Il en tire la conclusion que les « marqueurs de subordination » sont polyfonctionnels, c'est-à-dire qu'ils ne marquent pas toujours le même type de subordination et même qu'ils peuvent ne pas marquer de subordination du tout². Ainsi, *si*, qui se traduit en bambara par *ni*, peut dans ces deux langues fonctionner tantôt comme un marqueur d'interrogative indirecte, tantôt comme un marqueur de subordonnée circonstancielle, ce qui ne peut pas être fortuit. *Si* peut même être employé en tant que connecteur discursif. Alors que l'on a pour habitude de distinguer au moins deux *si* homonymes en français, D. Creissels suggère qu'il s'agit en fait d'un seul morphème polyfonctionnel.

En ce qui concerne le terme de coordination, D. Creissels part du constat suivant : « en considérant que toute phrase complexe doit résulter, soit d'une subordination, soit d'une coordination, on se condamne à avoir une notion de coordination purement négative, et à faire du terme de coordination un de ces termes par défaut dont le succès tient uniquement à ce qu'ils entretiennent l'illusion qu'on a réussi à catégoriser ce qu'ils servent à étiqueter. » (p. 199)

Suivant la même démarche que celle utilisée pour restreindre l'acceptation de subordination, il réduit la coordination presque exclusivement à la dimension intraphrastique, le relateur étant considéré aussi comme polyfonctionnel. Les relations interphrastiques sont, selon lui, irréductibles à la seule coordination.

² La terminologie (marqueur de subordination) n'est donc pas très adaptée à ce contexte.

Quant aux relatives, la présentation est excessivement claire et l'auteur revient notamment sur le statut de « l'antécédent » (qu'il appelle « nom de domaine ») et du « pronom relatif ». Les chapitres sur les relatives sont particulièrement adaptés pour introduire auprès des étudiants les problèmes que cette catégorie pose dans une perspective typologique : « *Les véritables pronoms relatifs sont rares en dehors des langues d'Europe : dans les langues du monde, la plupart des relativiseurs qui ne sont pas de pures marques de subordination, sont des joncteurs plutôt que des pronoms relatifs.* » (p. 228)

Il propose donc de distinguer trois types de relativiseurs : les pures marques de subordination, les pronoms relatifs et les joncteurs³. Il en donne une illustration en prenant comme langues le français et l'anglais qui ne comportent apparemment que les deux premiers types. L'auteur présente par ailleurs d'autres mécanismes de relativisation, moins familiers aux linguistes travaillant sur des langues d'Europe, tels que les relatives prénominales ou marquées uniquement par la position fixe du verbe.

Toutefois, cet ouvrage n'est évidemment pas exempt de tout reproche et j'ai relevé quelques points problématiques, souvent assez mineurs, qui méritent d'être signalés. Tout d'abord, la consultation des deux tomes est à mon avis indispensable ne serait-ce que pour avoir la totalité des définitions des termes employés. Par exemple, la phrase (notion qui constitue le titre du livre) est définie seulement dans le premier tome et de nombreux renvois obligent à s'y référer. La terminologie employée est relativement dense et elle n'est parfois pas la plus judicieuse : par exemple, le terme de « subordonnée circonstancielle » est utilisé alors que l'auteur critique ceux qui l'aborde en termes de signification « circonstancielle » (p. 193).

On peut aussi regretter l'utilisation systématique d'exemples fabriqués. À ce propos, des informations de première main ont parfois été exploitées par l'auteur (recours à des informateurs). Mais il a bien évidemment aussi eu recours à des ouvrages spécialisés comme sources des données. Comme dans la plupart des études typologiques, on peut donc se poser la question de la fiabilité de ces sources, car l'on sait très bien qu'en français notamment il existe un écart significatif entre la description que proposent les grammaires et les usages. Par ailleurs, le terme de syntaxe appliqué à certaines analyses pourra sembler inapproprié à quelques lecteurs peu familiers des cadres utilisés en syntaxe générale. Mais dès lors que l'on travaille sur des langues appartenant à des familles différentes, les outils d'analyse sont obligatoirement plus souples et se rapprochent nécessairement de la sémantique. La définition de la relative par exemple s'appuie sur une caractérisation sémantique. On pourrait encore relever quelques imprécisions mineures comme le classement de

³ Je n'ai pas très bien compris à quoi correspondait cette troisième catégorie, sans doute parce que la définition est donnée dans le premier tome. Afin de rendre compte le plus fidèlement possible de mon travail de lecteur, je ne me suis pas référé au tome 1.

l'exemple « *Comme il a plu, le sol est humide* » (p. 193) dans la catégorie des subordonnées alors que la construction en *comme* a ici les propriétés des éléments exclus de ce fonctionnement (cf. p. 197).

Pour finir sur une note positive, signalons que chaque fin de chapitre est agrémentée d'une bibliographie succincte fort utile et que le livre est accompagné d'un index thématique et d'un index des langues.

Au final, cet ouvrage permet de porter un autre regard sur les langues déjà fortement décrites en posant clairement la question de l'adéquation des catégories descriptives largement utilisées en linguistique. En effet, dès lors que l'on raisonne sur un grand nombre de langues, on s'aperçoit de la difficulté qu'il y a à faire coïncider la diversité des faits avec les outils d'analyse que l'on est habitué à manipuler. Les étudiants, enseignants et chercheurs souhaitant se familiariser avec la syntaxe générale trouveront donc à la fois des réflexions à réutiliser dans la description de langues particulières et surtout une passerelle vers le monde des langues peu ou pas du tout documentées.

Brian MACWHINNEY, *The CHILDES Project* (3^e éd.). Volume I. Tools for Analyzing Talk : Transcription Format and Programs, Lawrence Erlbaum, 2000, 159 pages, ISBN 0-8058-2995-4.

Lu par **Hervé HUNKELER**

DYALANG - Université de Rouen

À l'heure actuelle, on dispose de trop peu de transcriptions orales notamment en français. Les outils pour le traitement de corpus oraux sont aussi en nombre insuffisant. Aussi face à ce déficit, le projet CHILDES défini par Brian MacWhinney présente un grand intérêt. Cet ouvrage se compose de deux grandes parties : le manuel CHAT qui décrit les principes de transcription et de codage et le manuel CLAN qui décrit les programmes de CHILDES. Ce projet qui est libre et gratuit a un site Internet (<http://childes.psy.cmu.edu/>) qui donne une mise à jour régulière de ses fonctionnalités. L'avantage de cet ouvrage est d'explicitier de façon didactique, d'une part les règles de transcription et de codage au format CHAT, d'autre part les fonctions des programmes de CLAN. D'accès facile, il s'adresse à tout chercheur en acquisition de la langue maternelle ou de langue étrangère. Signalons que le volume 2 intitulé The Database, est une banque de données de corpus oraux concernant une trentaine de langues. Le seul regret est que ce volume soit uniquement diffusé en langue anglaise.

Le livre s'ouvre sur une introduction qui présente un bref historique de l'évolution des méthodes de recherche du langage de l'enfant. Ce développement permet à tout chercheur de mesurer l'importance de l'informatisation qui, avec la naissance de CHILDES en 1981, établit un format de transcription et des programmes d'analyse facilitant ainsi le partage des données entre les chercheurs.

Cette évocation chronologique montre comment, Catherine Snow (Harvard), Brian Mac Whinney et l'équipe de Carnegie Mellon, relayés par des chercheurs du monde entier, ont fait évoluer ce projet au départ relativement artisanal, sur le plan de la transcription et des capacités d'analyse des programmes. CHILDES a l'avantage, à travers ces évolutions, de mieux prendre en compte la forme et le contenu des productions du jeune enfant ainsi que les conditions naturelles de la conversation.

Dans la dernière partie de cette introduction, nous avons trouvé originale et pratique la rubrique « *How to Use This Book* ». Celle-ci donne une méthode d'utilisation des outils de CHILDES en distinguant cinq utilisateurs potentiels : si on est un utilisateur expérimenté, elle recommande de revoir toutes les règles de transcription avant d'employer une commande spécifique ou les nouvelles caractéristiques de l'éditeur. Parmi celles-ci l'inclusion du format CA (conversation analysis), format de remplacement de CHAT, et la possibilité de transcrire directement en IPA sur une ligne dépendante. D'autres facilités sont signalées comme les possibilités de relier les fichiers transcrits (au format CHAT) aux fichiers audio et vidéo numérisés. Si l'on est un nouvel utilisateur, ce dernier devra apprendre, p. 16, les normes de présentation des fichiers au format CHAT (notamment les symboles en tête de lignes @, %, *). Puis il devra consulter le manuel CLAN, notamment le mode d'emploi sur les fonctions des programmes.

Afin de se familiariser avec le format CHAT, l'auteur encourage l'utilisateur néophyte à transcrire une petite partie de ses données dans un fichier test. Pour vérifier si le codage est correct, MacWhinney recommande d'employer régulièrement le programme CHECK. Après avoir transcrit ce petit segment de données, l'auteur recommande d'essayer tous les programmes d'analyse que l'on a prévu d'utiliser afin de vérifier si les types de résultats obtenus répondent aux besoins de son travail. Pour les nouveaux utilisateurs qui sont principalement intéressés par des données déjà contenues dans les archives CHILDES (*Database*, volume 2), ils devront apprendre les bases d'installation et de mise en route des programmes CLAN. Sinon, ils n'auront simplement qu'à utiliser l'éditeur pour ouvrir et lire les fichiers. Enfin l'auteur recommande aux universitaires désirant enseigner les outils de CHILDES à des étudiants, d'utiliser les ressources du mode d'emploi en ligne développé par Yuriko Oshima-Takane et son collègue McGill. Ces leçons fournissent, sous la forme de questions à choix multiples, de multiples possibilités de tester ses connaissances sur les règles essentielles de CHAT. Elles sont accessibles à l'adresse <http://childes.psy.cmu.edu/CHAT.html>.

La première partie de ce livre décrit les conventions et les principes de transcription du système CHAT (**C**ode for the **H**uman **A**nalysis of **T**ranscripts). Ce système de transcription permet d'éviter deux difficultés que rencontre fréquemment la personne qui transcrit : la sur-régularisation qui consiste à faire correspondre à la forme produite par l'enfant une forme adulte ou standard qui n'a pas de réelle correspondance et la sous-régularisation qui, au contraire, consiste à ne pas faire correspondre la forme produite par le sujet à la forme adulte ou standard. Il a aussi l'avantage de répondre à l'exigence de clarté (chaque symbole utilisé dans le

système renvoie à un référent définissable). De plus la variété des options du format CHAT facilite la lecture d'une transcription. Celle-ci peut s'effectuer sur plusieurs niveaux. On peut effectuer une vérification automatisée des données transcrites avec le programme CHECK. Brian MacWhinney distingue trois composantes fondamentales d'un fichier CHAT : les « Headers » — ainsi appelés parce que certains d'entre eux doivent figurer en tête du fichier — qui contiennent des informations qui restent constantes et valent pour tout le fichier. La ligne principale (niveau obligatoire) contient la transcription orthographique des formes qui correspondent aux mots produits par le locuteur. Elle n'a pas vocation à tout coder sinon elle devient illisible et difficile à analyser. Pour répondre aux besoins d'une recherche projetée, différents chercheurs peuvent ajouter à cette ligne, des « lignes secondaires » ou « lignes dépendantes » telles %pho pour les chercheurs s'intéressant à la phonologie, %spa pour ceux qui s'intéressent à la conversation. De plus, il est possible de modifier ou de transformer des systèmes de codage selon les besoins des chercheurs. Ce système se caractérise donc par sa souplesse d'adaptation et ses multiples niveaux de transcription (26 niveaux offerts).

Dans la deuxième partie de ce volume, l'auteur présente l'ensemble des programmes d'analyse de données transcrites sous le nom de **CLAN** (Computerized Language Analysis). Ce dernier comprend un éditeur (sous Windows ou Macintosh) particulièrement adapté pour travailler avec les fichiers au format CHAT. La représentation de l'éditeur à l'identique sur papier, illustré d'exemples, rend particulièrement aisé l'utilisation de ce dernier. À ce sujet, l'auteur encourage les enseignants à faire connaître les exemples proposés dans le mode d'emploi de CLAN afin d'encourager les étudiants à explorer la banque de données à la recherche de nouvelles hypothèses de travail. Ce système automatise un grand nombre de procédures de transcription ou de codage et permet la liaison des fichiers transcrits avec des fichiers son ou vidéo numérisés. Ses nombreux programmes d'analyses automatisées permettent d'effectuer des recherches lexicales, morphosyntaxiques, phonologiques ou de discours à l'intérieur de grands corpus. Chaque commande d'un programme est accompagnée d'options qui permettent d'affiner les analyses. Parmi ces programmes, l'auteur cite l'un des plus puissants et l'un des plus employés, le programme d'analyse fréquentielle **FREQ**. Ce dernier, facile à utiliser, a aussi l'avantage de servir d'initiation à l'utilisation de CLAN. À la fin du manuel CLAN, l'auteur présente une série d'exercices exemplifiés qui aident les étudiants à consolider leurs connaissances de CHAT et de CLAN.

Ce livre didactique est le premier de deux volumes qui présente les trois outils du projet CHILDES (CHAT, CLAN, the Database). Ces outils sont, pour Brian MacWhinney, particulièrement flexibles pour répondre aux besoins spécifiques des chercheurs dans le domaine de l'acquisition du langage. Ils offrent des possibilités d'analyses poussées sur de larges corpus oraux. Pour l'auteur, ce livre sera très utile aussi bien pour le néophyte que pour l'utilisateur expérimenté des outils de CHILDES. Il s'adresse particulièrement aux chercheurs et aux étudiants qui travaillent sur les transcriptions du langage du jeune enfant. Nous souhaitons que ce

livre puisse faire connaître les multiples fonctionnalités qu'offre ce projet. Il est important pour Brian MacWhinney que les outils progressent en fonction des besoins des utilisateurs. C'est la raison pour laquelle l'auteur insiste sur la nécessité de faire évoluer CHAT, les programmes et la base de données aussi longtemps que le projet CHILDES sera actif.