

des causes « globales », Gaston Gross et ses collaborateurs distinguent des causes inchoatives (*déclencher, instaurer*), progressives (*perpétuer, maintenir*) et terminatives (*clore, éradiquer*).

Traitement de la diversité des causes. Méthodologie

Pour remédier à cette hétérogénéité, Gaston Gross et ses collaborateurs proposent une analyse détaillée du spectre argumental de chaque connecteur selon des classes sémantiques fines appelées « classes d'objets ». La multiplication des classes est indispensable pour pouvoir décrire la subordination de manière adéquate.

Les prédicats de ces classes sont décrits en extension. La description du spectre argumental de tous les connecteurs est réalisée à partir d'un corpus de grande ampleur (sont recensés 250 relateurs et 6 000 substantifs sur lesquels opèrent les prédicats de la causalité). Les compléments sont donnés par ordre de fréquence. Ce travail méthodique révèle que chaque prédicat de cause sélectionne un nombre d'opérandes qui lui est propre, les opérandes représentant des classes sémantiques spécifiques. Le verbe *provoquer* est, par exemple, compatible avec 800 substantifs et sélectionne comme arguments-objets i) des événements tels que les <accidents-incidents>, les <catastrophes naturelles>, etc., ii) des états événementiels <conflits>, <maladies>, etc., et enfin iii) des actions événementielles <attitudes et comportements>, <opérations intellectuelles>, <dire>, etc. À côté de leur combinatoire (opérateurs appropriés : verbes, adjectifs et adverbes) sont également analysées toutes les transformations possibles (thématisation, passivation, nominalisation, etc.).

Conclusion

En conclusion, la cause est une notion complexe qui ne peut pas être réduite à une définition conceptuelle unique impliquant un événement du monde des phénomènes responsable de l'existence d'un autre événement du monde des phénomènes. Cette étude, qui constitue une illustration parfaite de cette complexité, pourrait avoir des applications non seulement dans le traitement automatique des langues naturelles, mais aussi dans l'enseignement du français comme langue première et seconde. Elle pourrait aussi contribuer aux recherches en analyse du discours.

Benoît HABERT, Construire des bases de données pour le français, Tome 1. Notions, Éditions OPHRYS, 2009, 200 pages, ISBN 2-7080-1218-5.

Lu par Marie-Paule JACQUES

LIDILEM, Université Stendhal/IUFM, Université Joseph Fourier, Grenoble

Avec cet ouvrage et ses riches compléments en téléchargement gratuit, Benoît Habert continue un travail entrepris il y a plusieurs années : doter le linguiste – et plus largement le

chercheur en sciences humaines – d'outils et de ressources informatiques. Il présente ici les notions essentielles pour l'utilisation de bases de données à partir de réels cas de recherche.

Avec cette parution, Benoît Habert poursuit la démarche entreprise dans de précédents ouvrages : familiariser le linguiste – et plus largement le chercheur ou étudiant en sciences humaines – avec les possibilités offertes par les outils et ressources informatiques.

Il se focalise ici sur l'utilisation de bases de données pour recueillir, stocker, organiser, interroger les données support d'une recherche. L'une des forces de l'ouvrage est de présenter les opérations permises par l'utilisation d'un système de gestion de bases de données (SGBD) à partir de trois études de cas, chaque étude éclairant diverses manipulations en fonction de leur pertinence pour l'étude elle-même et ses objectifs. Les opérations sont ainsi ancrées dans les besoins réels d'interrogation et de sélection des données et non exposées *in abstracto*.

L'auteur n'a pas lésiné sur les moyens pour aider le lecteur à se forger une véritable compétence en matière de bases de données : le tome 1, payant, est accompagné non seulement d'un tome 2 en téléchargement gratuit, mais aussi des bases de données exemples elles-mêmes, dans des formats variés pour en permettre l'utilisation par divers gestionnaires de bases de données¹. On ne peut que louer ce souci de diversité logicielle qui offre au lecteur le choix du SGBD proprement dit. Avec les deux tomes et les exemples, un lecteur néophyte soucieux d'élargir son champ de compétences a à sa disposition les outils nécessaires.

Après une introduction qui situe clairement le propos, il ne s'agit pas de la formation à un logiciel précis, mais d'un exposé détaillé des concepts principaux des bases de données, le tome 1 est divisé en 11 chapitres. Trois études de cas, chacune courant sur plusieurs chapitres, offrent à l'auteur les illustrations nécessaires à la présentation des notions et opérations générales des SGBD.

Les quatre premiers chapitres s'appuient sur des données issues d'une enquête dans un service de prématurés, constituées aussi bien de texte (par exemple, les remarques inscrites par les infirmières sur un cahier de suivi) que de données chiffrées (par exemple l'ancienneté des infirmières dans le service, leur âge...). Le chapitre 1 « prépare le terrain » en présentant ces données avec une focalisation sur des notions utiles ultérieurement : l'association *individu* (bébé, infirmière) et *caractères* (sexe, âge, ancienneté, commentaire rédigé, etc.). Le chapitre 2 reprend ces notions sous la terminologie usuelle *enregistrement/attributs* et introduit à la notion de *table*, cruciale pour les bases de données, en faisant le détour par d'autres outils informatiques susceptibles de créer et manipuler des tables : traitement de texte et tableur. Sont ainsi mises en évidence les limites des opérations que ces outils réalisent et, par contraste, la puissance des SGBD. Les chapitres 3 et 4 débutent

1. Ce compte-rendu se restreindra à la présentation du tome 1 sans s'interdire des allusions au tome 2 (http://www.toeditions.com/Sources/Habert_Bases-de-donnees-2.htm).

l'inventaire des traitements possibles : sélection des données répondant à certaines caractéristiques, liste des combinaisons d'attributs, tris successifs, déductions de nouvelles informations par regroupements, tris, combinaisons d'opérations élémentaires qui permettent d'obtenir de nouvelles données, notamment chiffrées (par exemple, le nombre d'enregistrements pour lesquels l'attribut X présente la valeur n).

Les chapitres 5 à 7 poursuivent cette présentation d'opérations possibles, en prenant comme support une analyse de la pièce *Phèdre*. Les données sont ici le texte de la pièce, découpée en vers, mais aussi un premier niveau d'observations et de relevés effectués sur ce matériau, comme par exemple la transcription phonétique du texte, qui permet de s'affranchir des graphies et de saisir la dimension sonore, seul accès au phénomène de la rime, ou encore la position de chaque syllabe au sein d'un vers. La spécificité d'une pièce de théâtre de présenter une triple structure, dramatique, syntaxique, métrique, fournit une parfaite illustration de la notion de *relation* au sein d'une base de données. Les chapitres 6 et 7 explicitent donc les divers aspects liés à cette notion, comme les concepts de *clé primaire* et *clé étrangère*, les types de *jointures*, en n'oubliant pas de réutiliser les opérations exposées précédemment pour montrer des exemples d'exploitation des fonctions d'un SGBD et des exemples de résultats produits par de telles opérations.

Les chapitres 8 à 10 changent de perspective en s'appuyant sur un nouveau « cas » : l'étude du morphème *-esque*. Là encore, le premier des trois chapitres présente les données de façon détaillée, mais les deux chapitres suivants ne sont pas, autant que précédemment, consacrés à l'exposé de nouvelles opérations de traitement des données. Sont expliquées des combinaisons des opérations vues auparavant pour la réorganisation des données, notamment l'éclatement de la table initiale de l'étude en plusieurs tables pour améliorer la cohérence. Mais ces deux chapitres « prennent de la hauteur » en s'attaquant au niveau de la modélisation elle-même, c'est-à-dire à la détermination des entités, des attributs, des relations, en amont de leur implémentation dans une base de données. Cette implémentation, la création concrète de la base et de ses objets, ainsi que le peuplement en données de la base sont traités dans le chapitre 10. Le choix d'aborder ces questions plus abstraites après que le lecteur a touché du doigt la « matière » d'une base de données nous paraît particulièrement judicieux. Il est en effet plutôt difficile de construire un objet sans comprendre ni sa destination, ni son contour final.

Aide supplémentaire à cette construction, le chapitre 11, dernier chapitre de l'ouvrage avant la conclusion, prodigue d'utiles recommandations de « bonnes pratiques » pour la prévention des incohérences, la justesse des données, l'exploitation de la base comme maillon d'une chaîne à la fois de traitements et de chercheurs qui peuvent avec cet outil échanger des données et des résultats d'analyse.

Tout au long de l'ouvrage, l'ancrage sur des études authentiques, pour précieux qu'il soit, ne suffit pas à gommer totalement l'aridité intrinsèque du propos. Nous

voulons dire par là que la création et la manipulation de bases de données sont pour un linguiste « lambda » un sujet plutôt rébarbatif et que même le talent pédagogique d'un Benoît Habert ne suffit pas à le rendre attrayant. Il faut déjà à la fois un bagage informatique minimal et une conviction enthousiaste de l'utilité de l'outil informatique dans la recherche en sciences humaines pour digérer le sujet. L'ouvrage ne se prête guère à une lecture « de découverte » mais s'adresse à tout chercheur ou étudiant prêt à y consacrer du temps. Ce tome 1, qui recèle la quintessence des notions, gagne fortement à être utilisé en conjonction avec le tome 2, qui n'en est pas une suite ou une « application » mais un complément entremêlé. D'ailleurs, la table des matières du tome 1 fait état, pour chacun des 11 chapitres, des compléments et prolongements à trouver dans le tome 2, ainsi que de quatre chapitres additionnels. Aucun des deux tomes n'a véritablement de sens sans l'autre. Mais le chercheur ou l'étudiant qui acceptera d'ouvrir devant lui les deux tomes et les bases de données proposées afin d'en décortiquer le fonctionnement verra s'ouvrir aussi une nouvelle dimension de la recherche en linguistique et en sciences humaines, tant il est vrai que l'outil change la perception et la compréhension des données.