

Table des matières

Résumé.....	i
Abstract.....	iii
Table des matières.....	v
Liste des tableaux.....	ix
Liste des figures.....	xi
Liste des équations.....	xiii
Liste des sigles.....	xiv
Remerciements.....	xv
Chapitre 1 Introduction.....	1
1.1 Le renouvellement de l'instrumentation en évaluation des apprentissages.....	2
1.2 La cartographie conceptuelle et l'évaluation des apprentissages.....	3
1.3 Objectif et structure de la thèse.....	5
Chapitre 2 Cadre conceptuel.....	10
2.1 Un bref historique au sujet de la cartographie conceptuelle.....	10
2.2 Les fondements théoriques de la cartographie conceptuelle.....	11
2.3 Les définitions de la carte conceptuelle.....	13
2.4 Les deux principales composantes d'une carte conceptuelle : les concepts et les propositions.....	17
2.5 Les autres composantes d'une carte conceptuelle.....	21
2.6 Synthèse.....	23
Chapitre 3 Recension des écrits.....	24
3.1 Considérations méthodologiques de la recension des écrits.....	25
3.2 Les caractéristiques des études.....	29
3.3 Les objets d'évaluation.....	32
3.4 Les modalités de formation.....	34
3.5 Les modalités de passation.....	36
3.6 Les modalités de correction et de notation.....	43
3.7 Appréciation des propriétés psychométriques d'une tâche de d'évaluation.....	54
3.8 Synthèse.....	78

3.9 Un bref rappel au sujet des problématiques ciblées dans les articles de la thèse.....	81
Chapitre 4 Le cadre méthodologique de la collecte de données	83
4.1 Plan et contraintes de la recherche.....	83
4.2 Les participants	86
4.3 Le matériel informatique.....	89
4.4 L'instrumentation.....	89
4.5 La procédure de collecte de données	92
4.6 Quelques précisions au sujet de la collecte de données de la deuxième étude	98
4.7 La synthèse des données	100
4.8 La relation entre la collecte de données et les trois articles	101
Les articles de recherche	102
Chapitre 5 Étude des démarches déployées par des élèves de quatrième secondaire en science et technologie lors de la production d'une carte conceptuelle sur support papier ou à l'ordinateur	103
5.1 Introduction.....	104
5.2 Cadre théorique.....	106
5.3 Objectif de la recherche	113
5.4 Méthodologie	114
5.5 Analyse des résultats.....	123
5.6 Discussion.....	136
5.7 Conclusion	140
Chapitre 6 Production semi-automatisée d'une carte conceptuelle en science et technologie	142
6.1 Introduction.....	142
6.2 Cadre théorique.....	143
6.3 Contexte	151
6.4 Objectif de la recherche	152
6.5 Démarche de production d'une carte conceptuelle.....	152
6.6 Méthodologie	158
6.7 Résultats.....	162
6.8 Discussion.....	170
6.9 Conclusion	172

Chapitre 7 Modélisation des scores issus de cartes conceptuelles à l'aide de modèles de Rasch	174
7.1 Introduction.....	175
7.2 Cadre théorique.....	176
7.3 Objectif de la recherche	183
7.4 Méthodologie	183
7.5 Résultats.....	191
7.6 Discussion et Conclusion.....	207
Chapitre 8 Discussion et Conclusion	211
8.1 Retour sur le premier article	211
8.2 Retour sur le deuxième article	212
8.3 Retour sur le troisième article.....	213
8.4 Les limites de la recherche.....	215
8.5 Les retombées et les recommandations pour des recherches potentielles	217
8.6 Réflexion quant au contexte d'évaluation des apprentissages.....	220
8.7 Réflexion au sujet du processus de validation	225
8.8 Conclusion générale.....	226
Bibliographie.....	228
Annexe A. Formulaire de consentement de l'enseignant	247
Annexe B. Formulaire de consentement des élèves.....	250
Annexe C. Lettre d'information aux parents	253
Annexe D. Description de la tâche d'entraînement	254
Annexe E. Description de la tâche principale.....	255
Annexe F. Présentation des consignes au sujet de la cartographie conceptuelle.....	256
Annexe G. Protocole de verbalisation de la pensée.....	258
Annexe H. Lexique initial de la verbalisation	262
Annexe I. Lexique initial des stratégies.....	266
Annexe J. Lexique final des stratégies.....	274
Annexe K. Certificat d'approbation éthique.....	286
Annexe L. Illustrations des difficultés rencontrées par les participants lors de la production d'une carte conceptuelle	288

Annexe M. Représentation des démarches mobilisées par les participants lors des séances de verbalisation de la pensée	301
Annexe N. Concepts et synonymes acceptés lors de la production des cartes des références	306

Liste des tableaux

Tableau I. Quelques exemples de relations sémantiques lexicales.....	19
Tableau II. Composition graphique et textuelle de diverses composantes de la carte conceptuelle	22
Tableau III. Termes considérés lors de la recherche de textes dans les bases de données	25
Tableau IV. Description des facteurs d'exclusion lors de la récupération des textes.....	26
Tableau V. Description des facteurs d'exclusion lors de l'analyse du contenu des textes répertoriés	27
Tableau VI. Distribution des objets d'évaluation selon les objectifs de recherche	33
Tableau VII. Distribution de la durée de la formation dans les études recensées	35
Tableau VIII. Distribution des variantes de tâches dans les études recensées.....	39
Tableau IX. Distribution des supports de passation dans les études recensées	42
Tableau X. Description des différentes unités d'analyse répertoriées	45
Tableau XI. Exemples d'échelles de notation de la qualité des propositions.....	47
Tableau XII. Distribution des types de preuves de fidélité et méthodes répertoriées.....	59
Tableau XIII. Coefficients de généralisabilité des six modèles de notation étudiés dans McClure et coll., 1999.....	61
Tableau XIV. Distribution des études selon trois catégories de preuves et sept types de variables externes.	69
Tableau XV. Répartition des participants entre la modalité traditionnelle et la modalité informatisée lors des deux tâches de cartographie conceptuelle	88
Tableau XVI. Synthèse des données collectées en fonction des objectifs de recherche	100
Tableau XVII. Support de production des cartes conceptuelles des huit participants lors de la tâche d'entraînement et de la tâche principale.....	115
Tableau XVIII. Lexique des actions	119
Tableau XIX. Lexique des stratégies	121
Tableau XX. Lexique des difficultés et contraintes.....	122
Tableau XXI. Caractéristiques des cartes produites par les participants	124
Tableau XXII. Fréquence des actions.....	129

Tableau XXIII. Distribution des stratégies de planification mobilisées par les huit participants	130
Tableau XXIV. Distribution des stratégies de mise en relation mobilisées par les huit participants.....	132
Tableau XXV. Distribution des stratégies de révision mobilisée par les huit participants.....	133
Tableau XXVI. Distribution des difficultés et des contraintes rencontrées par les huit participants.....	134
Tableau XXVII. Adaptation des patrons lexico-syntaxiques de Hearst (1992).....	156
Tableau XXVIII. Exemple de catégorisation de mots de liaison.....	157
Tableau XXIX. Description des corpus des deux mises à l'essai.....	159
Tableau XXX. Taux de restitution et taux de précision de l'extraction des propositions des deux mises à l'essai	163
Tableau XXXI. Fréquence des erreurs d'extraction dans les deux corpus	164
Tableau XXXII. Pertinence et fréquence d'apparition des propositions distinctes dans le corpus et dans les inventaires de propositions extraites par EPTAO	166
Tableau XXXIII. Paires de concepts ajoutées à la carte de référence initiale	187
Tableau XXXIV. Statistiques d'items de la première modélisation.....	192
Tableau XXXV. Indices de corrélations statistiquement significatifs entre les résidus standardisés.....	198
Tableau XXXVI. Coefficients de saturation des six premières composantes de l'ACP des résidus standardisés de la première modélisation	200
Tableau XXXVII. Statistiques d'items de la quatrième modélisation.....	205
Tableau XXXVIII. Coefficients de saturation de l'ACP des résidus standardisés de la quatrième modélisation.....	207
Tableau XXXIX. Concepts et synonymes acceptés pour le corpus sur l'effet de serre	306
Tableau XL. Concepts et synonymes acceptés pour le corpus sur les relations trophiques ...	307

Liste des figures

Figure 1. Carte conceptuelle sur l'univers merveilleux d'Harry Potter.....	4
Figure 2. Exemple d'une carte conceptuelle au sujet de la notion de « concept »	14
Figure 3. Exemples d'une représentation semi-formelle et d'une représentation formelle de la proposition « Jean attrape une balle »	20
Figure 4. Distribution des années de publication des études recensées selon les objectifs de recherche.....	30
Figure 5. Distribution de la taille d'échantillon des études selon les objectifs de recherche ...	31
Figure 6. Exemple d'une carte conceptuelle.....	105
Figure 7. Exemple d'une unité de l'analyse des stratégies	117
Figure 8. Démarche du participant 3.....	125
Figure 9. Démarche du participant 7.....	126
Figure 10. Démarche du participant 4.....	127
Figure 11. Extrait du participant 4 illustrant un écart entre la pensée et le contenu de la carte	136
Figure 12. Exemple de relations unaires, binaires et tertiaires	145
Figure 13. Arbre de dépendance syntaxique de la phrase.....	147
Figure 14. Démarche de production de la carte, fonctionnement d'EPTAO et modèle de données	153
Figure 15. Carte conceptuelle de référence sur le thème « relations trophiques ».....	168
Figure 16. Carte produite sur la base des propositions extraites par EPTAO.....	169
Figure 17. Exemple simple d'une carte conceptuelle	177
Figure 18. Carte conceptuelle sur le thème « Relations trophiques » tiré des trois manuels scolaires.....	186
Figure 19. Distribution des estimations de niveaux d'habileté et des paramètres de difficulté des items pour la première modélisation.....	195
Figure 20. Diagramme des éboulis de l'analyse en composantes principales des résidus.	199
Figure 21. Distribution des estimations de niveaux d'habileté et des paramètres de difficulté des items pour la quatrième modélisation.....	203
Figure 22. Démarche du participant 1.....	301

Figure 23. Démarche du participant 2.....	302
Figure 24. Démarche du participant 5.....	303
Figure 25. Démarche du participant 6.....	304
Figure 26. Démarche du participant 8.....	305

Liste des équations

(6.1) Taux de restitution.....	160
(6.2) Taux de précision	160
(7.1) Modèle à crédit partiel.....	182
(7.2) Seuils critiques de la statistique <i>Outfit</i>	190

Liste des sigles

CRESST : *Center for Research on Evaluation, Standards and Student Testing*

EPTAO : Extraction de propositions à partir de textes et assistée par ordinateur

PFNet : *Pathfinder network*

TALN : Traitement automatique du langage naturel

Remerciements

La réalisation d'un projet de doctorat est une démarche unique, une bataille de longue haleine, qui marque notre vie et celles de plusieurs gens autour de nous. Cette démarche nous transforme au fur et à mesure que nous progressons, à chaque difficulté et défi que nous rencontrons, à chaque étape que nous franchissons.

Cette démarche nous transforme sur le plan intellectuel : nous rencontrons des gens brillants et passionnés de recherche, qui partagent des intérêts communs et avec qui ont peu échangé. Je peux penser à mes collègues de longue date, Angel Manuel Arias de Los Santos, Dan Thanh Duong Thi, Pierre Gignac, Gilles Leclerc, Sébastien Béland, que pour en nommer que quelques-uns.

Nous rencontrons aussi des mentors dévoués et prêts à nous aider, à partager leurs savoirs sans compter les heures, et ce malgré leurs nombreuses obligations. Je remercie celles qui m'ont suivi depuis le début, les professeures Nathalie Loye et Micheline Joanne Durand, ainsi que ceux qui se sont joints au cours de ma démarche doctorale, dont le professeur Éric Dionne, pour leur appui et toutes les occasions d'apprendre.

Au cours d'études aux cycles supérieurs nous sommes liés, presque par le sang, à une direction de recherche. Ma collaboration avec le professeur Jean-Guy Blais a commencé lorsque j'étais étudiant au baccalauréat. Il était alors directeur du GRIÉMÉtic et cherchait des étudiants de premier cycle qui aimeraient entreprendre des projets de recherche au cours de l'été. Cette expérience a été parmi les plus enrichissantes : j'ai eu la chance de me familiariser avec la recherche en mesure et évaluation des apprentissages, à mon rythme et dans un contexte détendu, et de découvrir mes propres intérêts de recherche. C'est alors que j'ai attrapé la piqûre.

Ça m'a pris quelques années pour réaliser pleinement la chance que j'ai eue de me retrouver à travailler avec le professeur Jean-Guy Blais. Sa réputation est solide dans le domaine et sa générosité est indéniable. Tout au long de mon parcours aux études supérieures, il m'a guidé habilement dans mon projet de recherche doctorale, mais il a aussi su me proposer d'autres défis et me mettre en contact avec plusieurs chercheurs du domaine. Il est aussi d'une extrême patience. Nombre de fois que je ne lui ai pas donné de nouvelles pendant plusieurs semaines, ou même quelques mois, parce que je m'étais égaré par curiosité dans tel ou tel aspect de ma

recherche. Le volet de ma recherche doctorale sur la production semi-automatisée d'une carte conceptuelle est d'ailleurs le fruit d'un de ces égarements.

Enfin, cette démarche de recherche nous transforme sur le plan personnel. Un projet de doctorat ne se réalise pas en quelques minutes sur le coin d'une table. Nous y consacrons des heures à n'en plus finir — de jour, de soir et de fin de semaine —, ce qui exige que nous fassions des sacrifices, mais aussi, que d'autres personnes autour de nous en fassent. Je pense au soutien indéfectible de mes amis et de ma famille, qui malgré mes absences répétées, sont toujours là pour moi. Je pense surtout au soutien de Magalie, qui m'a accompagné tout au long de ma démarche.

Bonne lecture,

Maxim Morin

Chapitre 1 Introduction

Étymologiquement, le mot « évaluer » tire ses origines de deux termes de l'ancien français : « é- » qui désigne « extraire » et « value » qui signifie « valeur ». Au sens moderne, ce mot renvoie à donner son appréciation sur quelque chose. [Antidote 9, 2018]

L'évaluation est un processus mental qui s'opère dans une multitude de contextes différents chez l'être humain. Il est parfois naturel et spontané, par exemple lorsqu'un individu juge un plat qui lui est servi au restaurant, mais il est souvent le résultat d'une pratique instrumentée, qui répond à des critères explicites et à une méthodologie réfléchie. Lors d'une épreuve de bosses en ski acrobatique, par exemple, cinq à sept membres d'un jury se partagent la tâche d'apprécier la performance des athlètes. Trois à cinq juges évaluent la qualité de la descente des bosses, deux autres accordent un résultat à la qualité des sauts et un dernier juge chronomètre le temps requis pour la descente. Et selon les normes en vigueur, ces trois critères contribuent respectivement à 60 %, 20 % et 20 % du résultat final.

En éducation, on parle d'évaluation des apprentissages pour désigner le processus formel ou informel qui consiste à collecter un ensemble de données sur l'apprentissage réalisé par l'élève et à examiner le degré d'adéquation entre ces données et un ensemble de critères de qualité en vue de prendre une décision (De Ketele, 1989). Au contraire de l'appréciation d'une performance en ski de bosses, l'évaluation des apprentissages concerne des attributs qui sont généralement considérés comme n'étant pas directement observables (par ex., la résolution de problèmes en sciences ou la compétence à écrire) et par conséquent, qui doivent être inférés à partir d'observations. Ce processus inférentiel a comme rôle de produire du sens à partir des données collectées (De Ketele et Roegiers, 2015; Roegiers, 2010) et est essentiel pour assurer la conformité entre l'objet de l'évaluation, les observations et l'interprétation des données, qui est de fait un enjeu fondamental de l'évaluation des apprentissages (Biggs, 1996; Dionne, 2005; Pellegrino, Chudowsky et Glaser, 2001).

1.1 Le renouvellement de l'instrumentation en évaluation des apprentissages

Depuis plusieurs années, les systèmes éducatifs sont en perpétuels changements ce qui exige que les pratiques de l'évaluation des apprentissages se renouvellent de manière continue (Alberts, 2009; Baird, Andrich, Hopfenbeck et Stobart, 2017; Mislevy, Almond et Lukas, 2003; Roegiers, 2010; Wiggins, 1993). Non seulement il s'avère important de déterminer les connaissances, habiletés et compétences qui satisferont aux attentes de la société de demain, mais il faut aussi cerner les sources de données et les outils de collecte de données nécessaires pour y arriver (Bennett, 2014; Mislevy et coll., 2003).

Il est généralement attendu que les technologies de l'information et de la communication jouent un rôle de premier plan dans l'évolution des pratiques évaluatives, mais dans les faits les applications concrètes se font lentement (Blais, 2011, 2009). À un point tel qu'il n'est pas exagéré d'affirmer que l'évaluation des apprentissages n'a pas encore réussi à se diversifier tant dans sa nature que dans son support, et ce surtout en salle de classe. De nos jours, ce retard en fait d'intégration des technologies pourrait même être jugé comme une menace à la validité des interprétations des résultats d'évaluation, dans la mesure où les dispositifs utilisés en situation d'évaluation ne sont pas conformes à ceux auxquels font appel les enseignants et les élèves en situation d'apprentissage. Les logiciels de traitement de texte en sont de bons exemples. Même s'ils sont de plus en plus utilisés, que ce soit pour la réalisation de travaux en classe ou hors classe, ils sont souvent oubliés ou interdits lorsque vient le temps d'évaluer la compétence à écrire, notamment dans des épreuves certificatives à grande échelle comme l'épreuve unique de français de la cinquième année du secondaire au Québec.

La technologie s'avère une des solutions prometteuses pour renouveler plusieurs facettes de la démarche d'évaluation, que ce soit pour faciliter la passation de tâches plus complexes et représentatives des compétences à développer (Mislevy, Steinberg, Almond et Lukas, 2006), pour automatiser plusieurs processus, de la saisie de données à la communication des résultats, pour assurer la collecte et la gestion de données transversales et longitudinales ou pour la constitution de banques de données riches et détaillées (Bartram, 2006; Blais, 2009; Hambleton, 2006; Parshall, Harnes, Davey et Pashley, 2010; Sireci et Zenisky, 2006). Cela dit,

l'implantation de nouveaux dispositifs ne peut se faire sans se soucier de leurs qualités techniques. L'instrument d'évaluation doit en effet permettre de collecter des données pertinentes, nécessaires et suffisantes pour soutenir les inférences ciblées par la tâche, et ce relativement à l'objet d'évaluation. Dans le jargon de la mesure en éducation, ces qualités techniques renvoient aux notions de fidélité et de validité. Ces concepts sont définis plus en détail lors de la recension des écrits au troisième chapitre.

1.2 La cartographie conceptuelle et l'évaluation des apprentissages

Certes moins connue que d'autres dispositifs d'évaluation, comme la question à choix multiple, la cartographie conceptuelle est une technique de représentation des connaissances qui attire de plus en plus l'attention lorsqu'il s'agit d'évaluer des apprentissages. Même si pour certains elle est considérée comme un dispositif d'évaluation « novateur » (Parshall et coll., 2010), la cartographie conceptuelle n'est pas tout à fait nouvelle. Comme il sera expliqué au prochain chapitre, la définition la plus courante de cette technique a été mise au point dans les années 1970 lors des travaux de recherche de Joseph D. Novak à l'Université Cornell. Ces outils graphiques connaissent pourtant un regain d'intérêt depuis plusieurs années, vraisemblablement en raison d'un accès accru à des logiciels gratuits et conviviaux pour représenter des connaissances (Marchand et d'Ivernois, 2004; Novak et Canas, 2008; Rochette et coll., 2010). Ces logiciels offrent plusieurs solutions et fonctionnalités pour produire, stocker, échanger et analyser des cartes conceptuelles.

Une carte conceptuelle est un moyen pour l'individu de représenter et d'organiser un ensemble de connaissances sous forme d'un réseau composé de nœuds, de liens et de mots-liens (voir La Figure 1). Les nœuds comportent des termes (par ex., Objets magiques ou Horcruxes) qui représentent des concepts et sont unis entre eux par des expressions qui définissent les relations entre les concepts (par ex., « Objets magiques *comme* Horcruxes »).

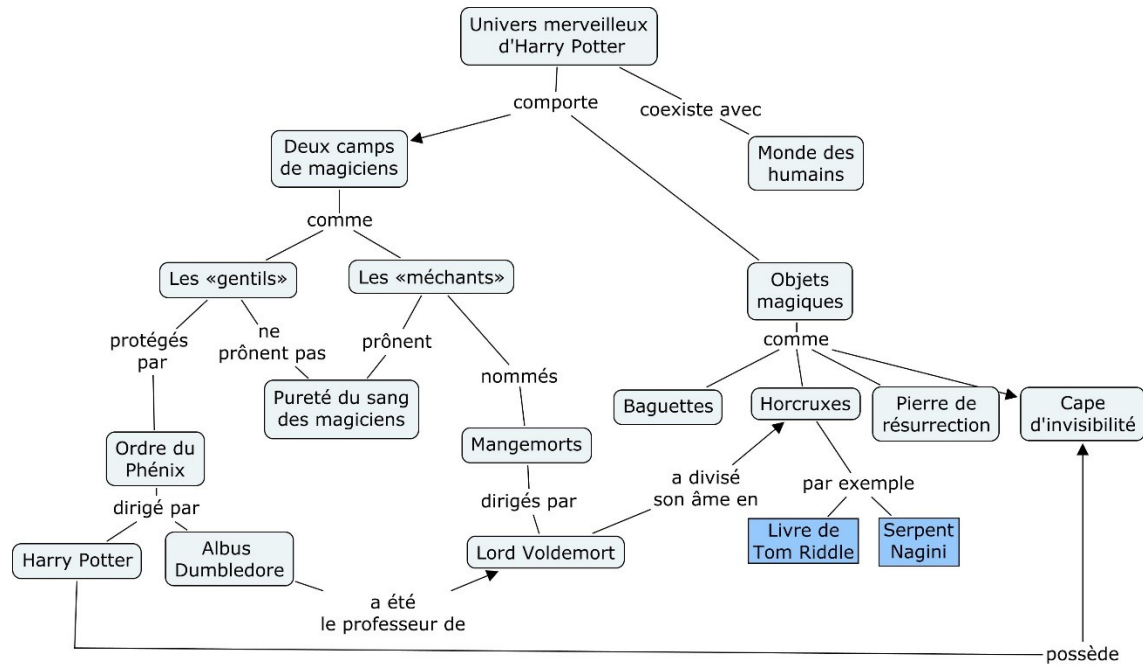


Figure 1. Carte conceptuelle sur l'univers merveilleux d'Harry Potter.

Plusieurs raisons semblent expliquer l'intérêt pour la cartographie conceptuelle en éducation. Cette technique de représentation des connaissances aurait d'abord comme avantage de s'accorder avec les conceptions modernes de l'apprentissage. En psychologie cognitive, l'organisation des connaissances est de fait d'ores et déjà reconnue comme étant une caractéristique fondamentale de l'apprentissage (Ausubel, 2000; Ausubel, Novak et Hanesian, 1978; Collins et Quillian, 1969; Novak et Gowin, 1984; Rumelhart et Ortony, 1977; Tardif, 1992). Cette caractéristique permettrait entre autres d'expliquer certaines différences entre le raisonnement des novices et des experts dans un domaine de connaissances (Baxter, Elder et Glaser, 1996).

Cette technique aurait également comme avantage de s'harmoniser avec plusieurs sphères de l'action pédagogique (Daley et Torre, 2010; Marchand et d'Ivernois, 2004; Novak et Canas, 2008). Même si elle est principalement exploitée comme outil d'apprentissage (Jonassen, 2006; Kinchin, De-Leij et Hay, 2005), elle peut aussi servir de dispositif de collecte de données en vue de l'évaluation des apprentissages (Ruiz-Primo, Schultz, et coll., 2001; Scallon, 2004;

Yin et Shavelson, 2008; Yin, Vanides, Ruiz-Primo, Ayala et Shavelson, 2005) ou de technique de planification de l'enseignement (Edmondson, 1995).

Les savoirs à l'égard de la carte conceptuelle comme dispositif d'évaluation des apprentissages demeurent pourtant limités et épars et les chercheurs ne s'entendent pas encore sur les meilleures pratiques d'évaluation qui y sont associées. Ces désaccords concernent principalement deux axes : la passation des tâches de cartographie conceptuelle et la correction et la notation des productions.

D'une part, la construction d'une carte conceptuelle étant une tâche multiforme, il existe un nombre quasi infini de modalités pour produire une carte conceptuelle. Ces modalités ne sont pourtant pas toujours équivalentes entre elles (Brandstädter, Harms et Großschedl, 2012; Ruiz-Primo, Schultz, et coll., 2001; Ruiz-Primo, Shavelson, Li et Schultz, 2001; Yin et coll., 2005). D'autre part, il existe encore plusieurs doutes au sujet des manières de traiter les données collectées au moyen de cartes conceptuelles (Besterfield-Sacre, Gerchak, Lyons, Shuman et Wolfe, 2004). Au fil des années et des travaux de recherche, les modalités de correction et de notation de cartes conceptuelles se sont multipliées et il est de plus en plus évident que ces modèles ne permettent pas d'évaluer les mêmes aspects de l'apprentissage et qu'ils ne possèdent pas les mêmes propriétés psychométriques (McClure, Sonak et Suen, 1999; Ruiz-Primo, Schultz, et coll., 2001; Ruiz-Primo, Shavelson, et coll., 2001; Srinivasan, McElvany, Shay, Shavelson et West, 2008; Yin et coll., 2005). Les différences fondamentales entre ces modèles ne sont pourtant pas claires et il reste encore beaucoup de travail à accomplir pour mieux comprendre les tenants et aboutissants de la cartographie conceptuelle dans le domaine de l'évaluation des apprentissages.

1.3 Objectif et structure de la thèse

À la lumière de la recension des écrits présentée dans cette thèse (voir le troisième chapitre), il est évident que toutes les possibilités en matière d'évaluation des apprentissages à l'aide de cartes conceptuelles n'ont pas été explorées à ce jour. La présente recherche a ainsi pour objectif d'explorer certaines avenues qui n'ont pas encore été abordées dans les études recensées sur le sujet.

Cette thèse adopte le format de thèse par articles et s'intéresse en particulier à trois volets de la cartographie conceptuelle lorsque celle-ci est utilisée comme dispositif de collecte de données en vue de l'évaluation des apprentissages. Même s'ils sont présentés dans des articles distincts, les trois volets de ce projet sont le fruit d'une seule collecte de données auprès d'élèves du secondaire en science et technologie et s'inscrivent dans un seul et même processus de validation de la carte conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages.

Objectif général :

Cette recherche explore ainsi certaines modalités de passation, de correction et de notation de la carte conceptuelle dans le but de collecter de nouvelles preuves empiriques et ainsi de soutenir le processus de validation de la carte conceptuelle lorsqu'elle est utilisée comme dispositif de collecte de données en vue de l'évaluation des apprentissages.

Au départ, le projet de thèse avait comme objectifs spécifiques de mieux comprendre le processus de réponses des élèves lorsque ceux-ci construisent leur carte conceptuelle à l'ordinateur avec un logiciel dédié à cet effet ou de façon traditionnelle avec un crayon sur un support papier (le premier article) et d'examiner les propriétés psychométriques d'une tâche de cartographie conceptuelle, en particulier la structure interne des scores, à l'aide de modèles de la famille de modèles de Rasch (le troisième article). Or, il a fallu considérer différentes options pour produire une carte de référence en vue de corriger les productions des élèves au cours du projet et, étant donné certains doutes au sujet des méthodes courantes, il a été décidé d'explorer les techniques permettant d'extraire une telle carte à partir d'un corpus de textes. Il s'est donc greffé au projet original une autre étude concernant la production semi-automatisée d'une carte conceptuelle à partir de documents textuels (le deuxième article).

En plus des trois articles, cette thèse comporte quatre chapitres : l'introduction, la présentation d'un cadre conceptuel, une recension des écrits et une discussion générale. Le deuxième chapitre retrace les origines de la cartographie conceptuelle et les définitions de cette technique de représentation des connaissances, dans le but de mieux situer la définition

considérée dans cette thèse. Le troisième chapitre présente une recension systématique des écrits sur les usages et modalités de la cartographie conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages. Cette recension fait ressortir la diversité des modalités examinées dans les études antérieures et met en évidence plusieurs dimensions des tâches de cartographie conceptuelle qui n'ont pas encore été étudiées ou l'ont été de manière incomplète. Les trois articles de la thèse ont ainsi pour objectif d'approfondir la réflexion vis-à-vis de certaines de ces zones d'ombre. Le quatrième chapitre décrit en profondeur la méthodologie de la collecte de données qui, pour des raisons de limite de mots, n'est que survolée dans les articles de recherche. Par comparaison aux articles, ce chapitre donne plus de détails en ce qui concerne la sélection des participants, l'élaboration des instruments de collecte de données et le déroulement de la collecte de données. En contrepartie, il ne couvre pas certaines considérations méthodologiques propres à chacun des articles, comme les méthodes d'analyse des données.

Le premier article (cinquième chapitre) concerne les supports de production d'une carte conceptuelle. Avec l'émergence et la popularisation de logiciels spécialisés, il y a de plus en plus de chercheurs (et vraisemblablement de praticiens) qui adoptent un support électronique plutôt qu'un support papier (Anohina-Naumeca et Grundspenkis, 2010; Anohina, Graudina et Grundspenkis, 2008; Anohina et Grundspenkis, 2007; Grundspenkis, 2011; Ifenthaler, 2010b; Liu, 2010). Or la production d'une carte ne sollicite pas nécessairement les mêmes démarches selon qu'elle est produite avec un support ou avec l'autre. À notre connaissance, aucune étude ne s'est penchée à ce jour sur les démarches et les stratégies sollicitées lors de la production d'une carte conceptuelle en modalité papier-crayon et en modalité informatisée. Ce premier article a pour objectif de combler ce manque en décrivant et en documentant ces processus.

Objectif spécifique du premier article

Documenter les démarches de cartographie conceptuelle, et ce dans deux modalités, à savoir lorsque la carte est produite à l'ordinateur ou lorsqu'elle est produite de manière manuscrite.

Comme évoqué précédemment, le deuxième article (sixième chapitre) est le fruit d'une réflexion concernant les modalités de correction de cartes conceptuelles. Comme il sera vu plus en détail dans la recension des écrits, plusieurs stratégies pour l'évaluation recourent, d'une façon ou d'une autre, à une carte de référence pour corriger et noter les cartes produites par des individus. Le second article propose et compare deux méthodes pour élaborer une carte de référence à partir d'une analyse de contenu de documents textuels. La première technique consiste à extraire de manière manuelle les concepts et les relations entre ceux-ci à partir d'une analyse syntaxique et sémantique du contenu de la phrase. La deuxième technique s'apparente à la première, sauf qu'elle exploite les techniques de traitement automatique du langage naturel (TALN), d'exploration de données et d'extraction d'information afin d'automatiser certains processus nécessaires à la production de cartes à partir de documents textuels. L'article décrit la démarche de conception d'un programme informatique permettant la production semi-automatisée d'une carte conceptuelle, ainsi que la mise à l'essai de cette démarche dans le contexte du programme de formation du Québec en science et technologie au secondaire.

Objectif spécifique du deuxième article

Concevoir et mettre à l'essai une stratégie d'extraction de propositions assistée par ordinateur en vue de la production d'une carte conceptuelle et dans le but éventuel de créer une carte de référence pouvant servir d'outil de correction des cartes produites par des élèves.

Le troisième article (septième chapitre) porte sur la notation de cartes conceptuelles. Comme il sera présenté au troisième chapitre, les modèles de notation recensés sont tous, à de rares exceptions (par ex., Liu, 1994), basés sur des scores observés, c'est-à-dire sur la somme pondérée des notes attribuées à des composantes d'une carte conceptuelle. Or, les limites de telles approches sont bien documentées dans le domaine de la mesure en éducation (Bertrand et Blais, 2004; De Ayala, 2009). L'article explore comment l'application du modèle de Rasch pour données dichotomiques (Rasch, 1960) et du modèle à crédit partiel (Masters, 1982), deux modèles de mesure de la famille des modèles de Rasch, peuvent être utiles pour mieux comprendre la structure interne des scores issus de cartes conceptuelles.

Objectif spécifique du troisième article

Illustrer les avantages d'appliquer la modélisation de Rasch, et plus particulièrement le modèle pour données dichotomiques et le modèle à crédit partiel, pour apprécier les propriétés psychométriques de tâches de cartographie conceptuelle.

Le dernier chapitre présente la discussion générale et la conclusion de cette recherche. Il fait une synthèse des résultats des trois articles et met en lumière la contribution scientifique de chacun d'entre eux. Il revient également sur les limites et les retombées de la recherche, de même qu'il suggère de possibles études complémentaires. La discussion générale présente également une réflexion quant au contexte d'évaluation particulier de cette recherche et revient sur le processus de validation de la carte conceptuelle comme dispositif de collecte de données. Le prochain chapitre présente d'ailleurs ce qui est entendu par la notion de carte conceptuelle dans le cadre de cette recherche.

Chapitre 2 Cadre conceptuel

Ce chapitre définit les principaux concepts en jeu dans cette recherche. Il relate d'abord les origines et les fondements théoriques de la carte conceptuelle, puis il présente les différentes définitions de cet outil de représentation des connaissances et de ses composantes.

2.1 Un bref historique au sujet de la cartographie conceptuelle

Il faut remonter aux travaux d'un groupe de recherche de l'Université Cornell, dirigé par Joseph D. Novak, pour retracer les débuts de la cartographie conceptuelle. Durant les années 1970, ce groupe a lancé une étude longitudinale ayant pour objectif de suivre et de comprendre l'évolution de conceptions en sciences d'élèves de la première année du primaire à la dernière année du secondaire. Pendant les douze années de leur étude, ils ont réalisé une quantité importante d'entretiens semi-structurés au cours desquels ils ont sondé les idées conceptuelles et propositionnelles des élèves¹. L'équipe de recherche ayant été rapidement inondée par la masse de données récoltées, elle a développé la cartographie conceptuelle comme solution de rechange pour schématiser et synthétiser les données contenues dans les transcriptions d'entrevues (Novak, 1990a, 2003; Novak et Musonda, 1991).

Après quelques années d'exploration des nombreux usages de la carte conceptuelle, et ayant constaté son potentiel en éducation, Novak et Gowin (1984) ont cristallisé leur pensée et leurs expériences dans plusieurs chapitres d'un ouvrage, intitulé *Learning How To Learn*. Ils y expliquent les fondements théoriques de cet outil et ils y dressent le panorama de ses nombreux usages en éducation. Comme il sera présenté au troisième chapitre, cet ouvrage comporte d'ailleurs ce qui est le premier système de notation connu des cartes conceptuelles en vue de l'évaluation des apprentissages.

¹ Au total, ils ont effectué neuf cycles d'entretiens au cours desquels le nombre d'élèves diminuait. Lors du premier cycle, ils ont interviewé 239 élèves et lors du dernier cycle, 55 élèves.

2.2 Les fondements théoriques de la cartographie conceptuelle

Les origines et fondements de la cartographie conceptuelle sont multiples. Dans la forme promue par Novak, du moins, la cartographie conceptuelle s'inspire des principes de la psychologie constructiviste et de la psychologie cognitive. Cet outil d'extériorisation et d'organisation des connaissances a en effet pour but d'engager l'apprenant dans un processus actif de construction du sens du monde qui l'entoure (Mintzes, Wandersee et Novak, 1998, 2001, 2005; Novak, 2002; Novak et Canas, 2008; Novak et Gowin, 1984). Cela étant dit, cet outil tire avant tout ses origines des modèles de représentation propositionnelle des connaissances de la psychologie cognitive.

En effet, Novak et Gowin (1984) ont puisé une grande partie de leur inspiration dans la théorie cognitive d'Ausubel (Ausubel, 1968, 2000; Ausubel et coll., 1978). Au début des années 1960, ce spécialiste de la psychologie cognitive s'est notamment intéressé à l'opposition entre l'apprentissage signifiant et l'apprentissage machinal (*rote learning*). D'un côté, l'apprentissage est dit signifiant lorsque de nouvelles connaissances sont intégrées à la structure cognitive et que ces nouvelles connaissances (concepts) permettent de préciser la signification des concepts existants (principe de différenciation progressive des concepts) ou de reconnaître de nouvelles relations entre les concepts déjà existants (principe de réconciliation intégrative). De l'autre, l'apprentissage est dit machinal lorsqu'il correspond à une simple mémorisation d'informations sans en comprendre le sens ou sans les relier à d'autres informations déjà emmagasinées dans la mémoire. Selon Novak et Gowin (1984), la cartographie conceptuelle est un dispositif pédagogique bien adapté à l'apprentissage signifiant, parce qu'elle sollicite naturellement des mécanismes d'apprentissage propositionnel et d'organisation hiérarchique des connaissances chez les personnes qui construisent la carte.

Même si plusieurs articles de recherche au sujet de la cartographie conceptuelle font encore appel à la théorie cognitive d'Ausubel, cette théorie a de manière générale été quelque peu éclipsée par les théories du traitement de l'information développées durant les mêmes années (Driscoll, 2005), et dont font partie les modèles du réseau sémantique et ceux de la théorie des schémas. Dans les travaux concernant la cartographie conceptuelle, il n'est néanmoins pas rare de trouver des références aux modèles de réseaux sémantiques, comme ceux

de Collins et Quillian (1969). Un réseau sémantique est un schéma complexe et multidimensionnel, composé de centaines voire de milliers de nœuds interconnectés entre eux par des relations. Les nœuds renferment des mots uniques et les liens sont composés de mots de liaison qui désignent la nature de la relation (« catégorie de », « attributs de », etc.) entre les nœuds. Les nœuds peuvent être organisés de manière hiérarchique, mais cela n'est pas toujours nécessaire, pour la bonne raison qu'un réseau ou un sous-ensemble d'éléments contenus dans un réseau peuvent être réorganisés et illustrés de nombreuses façons. En recherche, les modèles de réseaux sémantiques se veulent un moyen de simuler et d'étudier certaines caractéristiques de la mémoire sémantique (Collins et Loftus, 1975; Collins et Quillian, 1969), mais ils sont aussi des précurseurs dans les systèmes de traitement automatique du langage naturel (Quillian, 1969).

Les modèles de représentation propositionnelle des connaissances se sont aussi développés sous l'impulsion de la théorie des schémas (Rumelhart, 1980; Rumelhart et Norman, 1978; Rumelhart et Ortony, 1977). Le schéma est une structure d'information qui peut être symbolisée sous forme de paquets de connaissances et il évoque les concepts génériques (abstraits) tels qu'ils sont enregistrés dans la mémoire. Ces structures peuvent à la fois représenter des objets, des choses, des situations, des manifestations, des actions ou des séquences de manifestations ou d'actions (Rumelhart, 1980). Le schéma du concept de table, par exemple, peut comporter autant des informations relatives aux caractéristiques de cet objet, comme le fait d'avoir une surface plane et une ou des pattes, que des informations en lien avec ces fonctions, comme le fait de pouvoir y déposer d'autres objets. La représentation des connaissances en schémas a ainsi plusieurs similitudes avec les modèles propositionnels. Dans un système comme dans l'autre, seules les informations générales et abstraites sont encodées dans la mémoire et celles-ci deviennent ensuite disponibles pour traiter les nouvelles informations transmises par les sens, pour catégoriser les expériences vécues, pour remémorer des informations, pour comprendre des situations particulières ou pour résoudre des problèmes (Smith, 1989). Et comme pour les propositions, les nouvelles informations sont susceptibles d'augmenter ou de modifier les schémas existants.

Que ce soit la théorie cognitive d'Ausubel, les modèles de réseaux sémantiques ou les modèles de la théorie de schémas, ces modèles ont en commun qu'ils accordent une place

prépondérante à la représentation propositionnelle des connaissances, c'est-à-dire à la relation entre des concepts (ou des schémas), dans les mécanismes de traitement de l'information. Comme il sera discuté dans la prochaine section, les concepts et les propositions sont les matériaux de base de la carte conceptuelle.

2.3 Les définitions de la carte conceptuelle

Il est assez évident dans les écrits scientifiques et pédagogiques qu'il n'existe pas une définition, mais plusieurs définitions de la carte conceptuelle. Ces écrits se distinguent par rapport à leur conception générale de ce qu'est une carte conceptuelle, mais aussi par rapport à la définition de ses composantes et des principes de construction de celle-ci. Et qui plus est, dans plusieurs travaux de recherche, il n'est pas rare de constater un écart entre la définition qui est prise sur le plan théorique et celle qui est véritablement mise en application. Cela se produit lorsque les chercheurs adoptent un sens précis de la carte conceptuelle ou lorsqu'ils font référence à une définition particulière, par exemple à celle de Novak, mais qu'eux-mêmes ne respectent pas entièrement les principes énoncés ou sous-entendus.

Les définitions de la carte conceptuelle peuvent être situées sur un continuum, des plus ouvertes aux plus strictes. À un extrême, il y a ces chercheurs qui utilisent les termes « carte conceptuelle » ou « cartographie conceptuelle » de manière très libre. Pour eux, les nœuds peuvent renfermer n'importe quel type d'information (des mots, des phrases, des paragraphes et des images) et il n'y a pas nécessairement de contrainte sur la formation des arcs, c'est-à-dire sur la formation des objets qui reflètent les relations entre les nœuds (voir la Figure 2). Les arcs peuvent être représentés par des traits unidirectionnels ou bidirectionnels, accompagnés de mots de liaison ou non.

En général, ces définitions, dites ouvertes, ne permettent pas de cerner les types de connaissances qui sont représentés au moyen de ces schémas ni de différencier la carte conceptuelle d'autres variantes d'organiseurs graphiques², comme la carte mentale (Buzan et Buzan, 1995), la carte de connaissances (Bahr et Dansereau, 2001; Bahr et Dansereau, 2005;

² Des définitions plus complètes de ces variantes sont offertes dans Cañas et coll. (2003) et dans Morin et Paquette (2014)

Chmielewski et Dansereau, 1998; Chung, Baker et Cheak, 2002; Chung, O'Neil et Herl, 1999), le réseau sémantique (Fisher, 1990; Jonassen, 2006) et la carte cognitive (Eden, 1988, 1992).

À l'autre opposée sur le continuum des définitions de la carte conceptuelle se situe la définition suggérée par Novak, parfois appelée l'approche « novakienne » qui, elle, est beaucoup plus restrictive. En plus de définir ce qui doit être contenu dans les nœuds et les arcs, cette approche dicte un mode d'organisation des connaissances et précise le rôle de l'individu (le « cartographe ») dans la construction de sa carte. Selon Novak et Gowin (1984), une carte conceptuelle est un réseau de nœuds et d'arcs qui sert essentiellement à « représenter les relations significatives entre des concepts sous la forme de propositions (p. 15) ». Un nœud est une figure géométrique dans laquelle est inscrit un terme, généralement un nom ou un groupe nominal, qui désigne un concept. L'arc est formé d'un trait muni d'une tête de flèche et sur lequel est apposée une expression, généralement un verbe, qui précise la relation entre deux concepts. Les concepts et les relations s'assemblent ainsi en énoncés de sens nommés des propositions. Une proposition comprend généralement deux concepts et une relation, mais elle comporte parfois plus d'éléments. La Figure 2 illustre une carte conceptuelle réduite au sujet de la notion de « concept ».

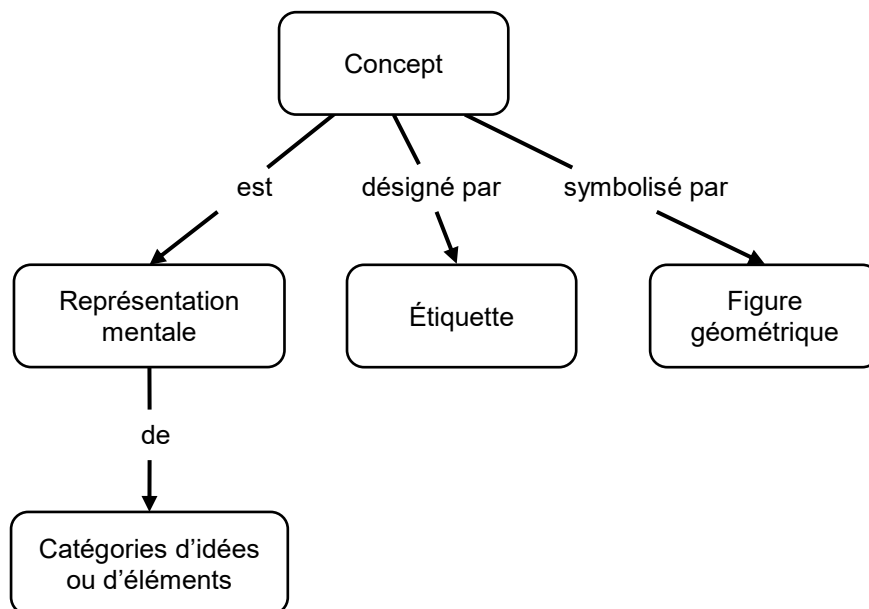


Figure 2. Exemple d'une carte conceptuelle au sujet de la notion de « concept »

La définition de Novak de la cartographie conceptuelle ne se limite toutefois pas aux caractéristiques formulées ci-dessus. Elle repose aussi sur un bon nombre de principes concernant l'élaboration de ces schémas de connaissances. Cañas et coll. (2003) résument bien ces principes :

1. Le « cartographe » doit énoncer lui-même les mots ou les expressions de liaison entre les concepts ;
2. La carte conceptuelle doit être élaborée autour d'une question spécifique (en anglais, *focus question*) ;
3. Les nœuds doivent être des mots ou de courtes expressions désignant des concepts ;
4. L'organisation des concepts doit être semi-hiérarchique ;
5. La carte devrait être composée d'exemples et de relations transversales ;
6. La carte devrait généralement comporter entre 15 et 25 concepts.

Les fondements de ces principes sont multiples. Ils s'appuient à la fois sur quelques notions théoriques, comme les principes de la psychologie constructiviste (conditions 1 et 2), les principes de la théorie d'Ausubel (conditions 3 et 4) ou les mécanismes d'apprentissage qui y sont associés (conditions 5), et à la fois sur des considérations d'ordre pédagogique résultant des nombreuses années d'expérimentation de la cartographie conceptuelle auprès d'élèves et d'étudiants de tout horizon.

Ces considérations étant plus d'ordre pratique, il n'est donc pas sans grande surprise de constater qu'elles sont souvent adaptées pour s'harmoniser avec le contexte et les conditions de réalisation de la tâche d'apprentissage ou d'évaluation (voir la recension des écrits, dans la seconde partie de ce chapitre). À cet égard, même si plusieurs chercheurs ou pédagogues font référence aux écrits de Novak dans leurs travaux, ils n'emploient pas toujours sa définition telle quelle.

Dans le but d'éviter les malentendus, certains chercheurs préfèrent donc utiliser des termes distincts pour nommer les nombreuses variantes et adaptations de cette dernière. Dans une recension des écrits, Cañas et coll. (2003) emploient les termes « réseau sémantique » et «

cartes de connaissances » pour désigner deux variantes de cartes non hiérarchiques : la première est construite librement alors que la deuxième est produite à partir d'une liste préétablie de concepts et d'expressions de liaison. Dans sa thèse, Vikaros (2017) aime mieux l'expression « graphe sémantique » pour désigner ce que Cañas et coll. (2003) nomment une carte de connaissances. L'emploi de termes distincts pour désigner chaque variante d'organisations graphiques des connaissances est selon toute vraisemblance un problème en constante évolution et, après tout, les expressions proposées ci-dessus sont généralement méconnues et rarement employées.

La présente recherche adopte donc une définition de la carte conceptuelle un peu plus souple que celle de Novak, mais qui demeure cohérente avec celles qui se retrouvent dans plusieurs travaux de recherche (Cañas et coll., 2003; Jonassen, 2000; Jonassen, 2006; Ruiz-Primo, Schultz, et coll., 2001; Ruiz-Primo et Shavelson, 1996; Ruiz-Primo, Shavelson, et coll., 2001). Dans cette recherche, une carte conceptuelle est simplement un schéma bidimensionnel composé de nœuds et d'arcs qui représentent respectivement des concepts et les relations entre ceux-ci. Les relations sont symbolisées par des traits munis d'une flèche et sont accompagnées de mots de liaison pour désigner le sens de la relation.

Et au contraire de l'approche de Novak, pour les raisons évoquées ci-dessus, il apparaît profitable de séparer la définition de la carte conceptuelle des modalités de production d'une carte conceptuelle. Fondamentalement, il n'y a pas lieu d'imposer un mode d'organisation du contenu de la carte, car cette organisation est largement tributaire du domaine des connaissances et des modalités de production de la carte conceptuelle. La prochaine section approfondira la définition des principales composantes d'une carte conceptuelle, le concept et la proposition, ainsi que celle de quelques autres composantes secondaires. Le troisième chapitre présentera diverses modalités de production de ces schémas de connaissances dans le contexte de l'évaluation des apprentissages.

2.4 Les deux principales composantes d'une carte conceptuelle : les concepts et les propositions

Une carte conceptuelle est composée de deux unités d'informations fondamentales : le concept et la proposition. Dans une perspective cognitiviste, un concept est défini comme une catégorisation d'objets, d'évènements, de situations ou de propriétés possédant des attributs communs (Ausubel, 1968, 2000; Ausubel et coll., 1978). Parfois il peut désigner une entité plutôt concrète (le concept d'un chien), parfois une entité plutôt abstraite (le concept de chaleur). Quelle que soit la nature de l'entité, il est primordial de différencier le concept et le vocable qui le désigne, autrement dit le signifié et le signifiant (Astolfi, Darot, Ginsburger-Vogel et Toussaint, 2008; Fisher, 1990). Un même vocable peut avoir une signification différente d'un individu à l'autre. Par exemple, la notion de chaleur ne désigne vraisemblablement pas la même chose chez un jeune enfant et chez un ingénieur spécialisé en énergie thermique. Le vocable prend seulement son sens lorsqu'il est mis en relation avec d'autres énoncés.

Dans les modèles de représentations propositionnelles des connaissances, les concepts s'assemblent entre eux pour former des unités de connaissances, nommées les propositions. La notion de proposition n'est toutefois pas unique au domaine de la cartographie conceptuelle ; elle est aussi employée dans plusieurs domaines connexes tels que la psychologie cognitive, la logique, la grammaire et le domaine de la représentation des connaissances en informatique. Même si les définitions de la proposition issues de ces domaines partagent plusieurs points en commun qui méritent d'être soulignés, elles ont également leurs particularités.

En psychologie cognitive, une proposition correspond au plus court énoncé auquel peut être attribué une valeur de vérité et se manifeste sous la forme de relations entre des concepts ou de relations entre un concept et une propriété (Bernstein, Penner, Clarke-Stewart et Roy, 2003). Le caractère de vérité sous-jacent à une proposition implique donc que celle-ci est vérifiable. Par exemple, la proposition « la Terre *tourne autour du* Soleil » représente la vision héliocentrique du système solaire et correspond à la notion scientifiquement juste, tandis que la proposition inverse « le Soleil tourne autour de la Terre » renvoie à la vision géocentrique et constitue évidemment une conception erronée du système solaire. En psychologie cognitive, la

proposition peut aussi concerner la relation entre une entité et un attribut, comme « la porte *est* ouverte ». Cet énoncé est également vérifiable.

La définition de la proposition en psychologie cognitive découle elle-même de la définition généralement admise en logique classique. Dans ce domaine de la philosophie, une proposition est une « assertion fondamentale jugée trop élémentaire pour être qualifiée de théorème. Sa caractéristique essentielle est d'être susceptible de recevoir une valeur de vérité » (Larousse en ligne, s.d.). En logique, une phrase comme « Paris est la capitale du Canada » sera décortiquée en plusieurs propositions élémentaires comme « Paris est une capitale » ou « du Canada », et la valeur de vérité de chacune de ces propositions sera évaluée tour à tour. Malgré les rapprochements entre les définitions en psychologie cognitive et en logique, l'analyse en logique mène à des propositions plus fines (par exemple, « du Canada ») que celles qui sont étudiées dans les modèles propositionnels de la connaissance.

La notion de proposition prend aussi un sens particulier en grammaire³. Dans ce domaine, une proposition est un constituant syntaxique de la phrase construit autour d'un verbe. La plupart du temps, elle est composée d'un sujet et d'un verbe ; d'un sujet, d'un verbe d'attribution et d'un attribut ; ou encore d'un sujet, d'un verbe et d'un complément. Tantôt ces constructions syntaxiques s'accordent avec la définition d'une proposition en psychologie cognitive, tantôt elles ne s'y conforment pas.

Toujours en linguistique, il y a également un rapprochement évident à faire entre la notion de relation sémantique et celle de proposition. Une relation sémantique établit une relation de sens entre des termes ou des concepts. Le Tableau I dresse d'ailleurs une liste des relations sémantiques lexicales qui sont couramment rencontrées. À cette liste s'ajoutent d'autres types de relations sémantiques comme les relations de causalité, les relations de finalité, etc.

³ La notion de proposition est issue de la grammaire traditionnelle. La grammaire nouvelle préfère la notion de phrase, au sens élargi.

Tableau I. Quelques exemples de relations sémantiques lexicales

<i>Catégorie/Type de relation</i>	<i>Description</i>	<i>Exemple</i>
Hiérarchique		
Hyperonymie	Relation entre un terme et un terme de classe inférieure	Les <i>félins</i> comprennent les lions.
Hyponymie	Rapport d'inclusion entre un terme et un terme de classe supérieure	Les <i>lions</i> font partie des félins.
Holonymie	Relation entre un terme désignant un ensemble et un terme désignant une partie de l'ensemble. Relation entre le tout et une partie.	La <i>maison</i> a une porte.
Méronymie	Relation entre un terme désignant une partie et un terme désignant le tout.	La <i>porte</i> d'une maison
Équivalence/opposition		
Synonymie	Relation entre deux termes ayant un sens semblable ou très rapproché de l'un et de l'autre	Automobile et voiture
Antonymie	Relation d'opposition entre deux termes qui sont le contraire de l'un et de l'autre.	Grand et petit

Enfin, la notion de proposition, ou plus précisément la notion de relation entre des concepts est employée dans le domaine de la représentation des connaissances en informatique. Dans ce domaine, il est relativement courant d'utiliser des structures de représentation des connaissances, comme les graphes conceptuels (Sowa, 2000). Ces structures, qui sont parfois à la base des systèmes de raisonnement intelligent, sont toutefois beaucoup plus formelles que celle de la cartographie conceptuelle. Prenons, par exemple la proposition « Jean attrape la balle » telle qu'illustrée à la Figure 3. Dans une carte conceptuelle, cette proposition serait représentée par deux nœuds (Jean et Balle) et un mot de liaison (attrape), mais elle serait décortiquée bien différemment dans un graphe conceptuel. Le verbe « Attraper » serait lui aussi un nœud et il serait en relation avec les deux autres nœuds, « Jean » et « balle ». Plus précisément, le nœud « Jean » constituerait l'agent associé au verbe et le nœud « Balle » serait l'objet de ce verbe.

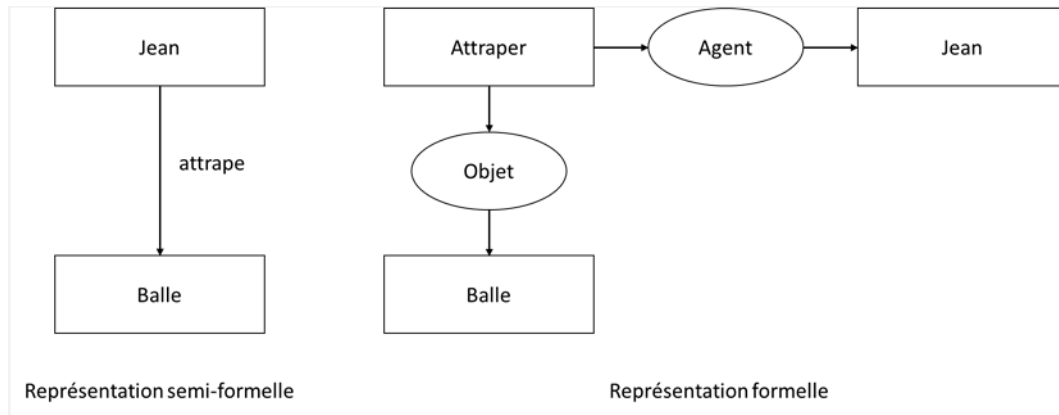


Figure 3. Exemples d’une représentation semi-formelle et d’une représentation formelle de la proposition « Jean attrape une balle »

Dans le contexte de la cartographie conceptuelle, la définition d’une proposition rejoint un peu chacune des définitions spécifiques présentées ci-dessus, mais elle s’en distingue également. Dans leur ouvrage, Novak et Gowin (1984) définissent une proposition comme étant « deux énoncés de concepts ou plus reliés entre eux par des mots pour former une unité sémantique (p. 15) ». Cette définition est déjà plus restrictive que celle en psychologie cognitive ou en grammaire, car elle se restreint aux cas où les nœuds représentent des concepts. Il est en effet très rare qu’un nœud désigne une propriété ou un attribut dans une carte conceptuelle. Une proposition telle que « le chien *est* poilu » n’est pas signifiante dans ce type de représentation des connaissances.

Et naturellement, la carte conceptuelle s’accorde mieux avec la relation binaire entre deux concepts. Les relations unaires comme « Jean pleure » sont tout simplement interdites parce qu’elles ne comportent pas deux nœuds et les relations ternaires sont quant à elle bien souvent reformulées en deux relations binaires. Ainsi, la proposition « Jean donne un livre à Johanne » serait exprimée sous forme de deux relations : « Jean donne un livre » et « livre à Johanne ». Cette reconstruction d’une relation ternaire en deux relations binaires n’est pourtant pas sans risque, car elle peut mener à la perte de sens, ce qui peut avoir des conséquences notamment lors de la correction et de la notation de cartes conceptuelles.

Les propositions d'une carte conceptuelle peuvent aussi représenter divers types de connaissances. Dans une note théorique, Safayeni, Derbentseva et Cañas (2005) expliquent qu'il existe principalement deux catégories de propositions : celles dites statiques et celles dites dynamiques. Les premières décrivent l'état d'une chose ou de la relation entre deux concepts et s'opposent aux secondes qui, elles, définissent en quoi le changement d'état d'un concept fait varier l'état d'un autre concept.

Ces deux types de propositions se divisent eux-mêmes en sous-catégories. Toujours selon Safayeni et ses collaborateurs, les propositions de type statique peuvent désigner des relations d'inclusion (Lion *fait partie* des Félines), des relations d'appartenance commune (Lion et Chat *font partie* des Félines) ou encore des relations d'intersection (Lions *ont des griffes comme* les Chats), tandis que les propositions de type dynamique peuvent désigner des relations de cause à effet (la durée d'un trajet *dépend de* la vitesse moyenne et de la distance parcourue) ou des relations corrélationnelles (les individus qui terminent des études postsecondaires *ont des salaires plus élevés*).

Même si la notion de « proposition » est bien établie dans le contexte de la cartographie conceptuelle, il serait hasardeux d'ignorer les définitions des domaines connexes. La carte conceptuelle est un support, composé de contraintes graphiques et de contraintes linguistiques, servant à extérioriser des connaissances emmagasinées dans la mémoire. Les différentes définitions présentées ci-dessus permettent donc de mieux situer la définition de la cartographie conceptuelle et quelques-unes de ses limites pour représenter des connaissances.

2.5 Les autres composantes d'une carte conceptuelle

En dernière analyse, les concepts et les propositions sont certainement les deux composantes fondamentales de la carte conceptuelle, mais au fil des années, un vocabulaire assez riche a été élaboré pour décrire d'autres composantes de la carte conceptuelle. Le Tableau II décrit les caractéristiques de la composition graphique et textuelle de chacun de ces constituants.

Tableau II. Composition graphique et textuelle de diverses composantes de la carte conceptuelle

<i>Éléments</i>	<i>Composition graphique</i>	<i>Composition textuelle</i>
Concept	Une forme géométrique (généralement un rectangle ou un ovale) nommée un nœud.	Un libellé sous forme de mots ou de symboles
Lien direct	Un tracé de lignes généralement unidirectionnel entre deux concepts rapprochés (avec une tête de flèche)	Une étiquette (mots de liaison) caractérise la relation entre les concepts
Lien transversal	Un tracé de lignes généralement unidirectionnel entre deux concepts éloignés (avec une tête de flèche)	Une étiquette (mots de liaison) caractérise la relation entre les concepts
Exemple	Une forme géométrique différente de celle des concepts	Un libellé ou une image représentant un exemple d'un concept.
Proposition	Un ensemble de nœuds et de liens	Une unité sémantique généralement composée d'une triade concept – relation - concept
Embranchement	Deux tracés de lignes ou plus qui proviennent d'un même concept.	Tout concept général qui se scinde en deux propositions ou plus formé de concepts spécifiques.
Niveau hiérarchique	Une organisation du haut vers le bas de nœud. Le nœud supérieur se divise en plusieurs nœuds inférieurs.	Le concept superordonné est général, alors que les concepts subordonnés sont plus spécifiques. Les concepts spécifiques forment un regroupement logique.

2.6 Synthèse

Dans ce chapitre, la cartographie conceptuelle a été présentée comme étant une technique servant à représenter et à organiser un ensemble de connaissances sous forme d'un schéma bidimensionnel. Ce chapitre a permis de définir clairement ce qui est entendu par une carte conceptuelle dans cette thèse et de faire ressortir les différences entre cette définition et d'autres. Bien entendu, définir la carte conceptuelle n'est qu'un premier pas pour comprendre l'objet de cette recherche. Pour bien appréhender le rôle de la carte conceptuelle lorsque celle-ci est utilisée comme dispositif de collecte de données en vue de l'évaluation des apprentissages, il faut également porter une attention particulière aux modalités de passation, de correction et de notation de tâches de cartographie conceptuelle, ce qui sera traité dans le prochain chapitre.

Chapitre 3 Recension des écrits

La cartographie conceptuelle est souvent décrite comme une technique permettant de rendre compte de dimensions de la connaissance qui ne sont pas évaluées au moyen d'instruments d'évaluation traditionnels (Ruiz-Primo et Shavelson, 1996). Il n'est pas clair toutefois ce à quoi renvoie ce quelque chose d'autre. Tout comme il n'est pas clair comment faut-il s'y prendre pour l'évaluer ? Ce chapitre tente d'apporter quelques éclairages à ces égards en examinant ce qui a été fait dans les études antérieures.

Quelques recensions ont déjà abordé la contribution de la carte conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages. Cela dit, la plupart d'entre elles (Baharom, 2012; Hay, Kinchin et Lygo-Baker, 2008; Kinchin, 2001; Novak et Cañas, 2006; Pudelko, Young, Vincent-Lamarre et Charlin, 2012; Venkatesan, Bharathi Priya et Rajarajan, 2017; Yue, Zhang, Zhang et Jin, 2017) sont très générales et ne s'intéressent pas seulement à la fonction d'évaluation des apprentissages. Elles brossent plutôt un portrait très large des usages et fonctions de la cartographie conceptuelle en éducation.

Les recensions dédiées à la fonction d'évaluation des apprentissages sont plus rares et le contenu de celles-ci est très varié. Parfois elles s'intéressent à plusieurs domaines de connaissances, parfois elles se limitent à un seul domaine de connaissances. Par exemple, Mintzes et coll. (2001) ont limité leur recension au domaine de la biologie et (Keppens et Hay, 2008), à celui des sciences informatiques. Certaines recensions ont également été consacrées à une dimension bien précise, comme la fonction d'évaluation formative (Hartmeyer, Stevenson et Bentsen, 2017; Krabbe, 2014), les modèles de notation de cartes conceptuelles (Anohina et Grundspenkis, 2009) ou encore la facilité d'utilisation (*usability*) des programmes informatiques servant à construire des cartes conceptuelles (Weinerth, Koenig, Brunner et Martin, 2014).

Encore à ce jour, la recension de Ruiz-Primo et Shavelson (1996) est probablement la plus exhaustive à l'égard des utilisations de la carte conceptuelle en évaluation des apprentissages. Ces chercheurs y décrivent de manière systématique les modalités de passation et de notation de 21 études, de même qu'ils présentent les preuves de fidélité de scores et de validité mentionnées dans ces publications.

Ce chapitre s’inspire en grande partie de l’organisation du contenu de la recension Ruiz-Primo et Shavelson (1996). Dans un premier temps, il fait état des usages et des modalités de la cartographie conceptuelle comme dispositif de collecte de données en vue de l’évaluation des apprentissages (sections 3.3 à 3.6) et dans un deuxième temps, il brosse le portrait des études qui ont examiné les propriétés psychométriques de tâches de cartographie conceptuelle (section 3.7).

3.1 Considérations méthodologiques de la recension des écrits

Une recension systématique des écrits⁴ a été réalisée dans le but d’assurer la plus grande représentation possible des études concernant la cartographie conceptuelle. Pour ce faire, une recherche par mots-clés a été effectuée dans trois champs (le titre, les sujets et le résumé) de trois bases de données (PsychINFO, Scopus et ERIC) pour répertorier tous les textes en anglais qui concernent à la fois le domaine de l’évaluation des apprentissages et celui de la cartographie conceptuelle (voir les termes recherchés dans le Tableau III). La collecte de données comprend tous les textes publiés avant janvier 2018.

Tableau III. Termes considérés lors de la recherche de textes dans les bases de données

<i>Concept</i>	<i>Termes</i>
Évaluation des apprentissages	<i>assessment, assess, measurement, measure, evaluation, evaluate</i>
Carte conceptuelle	<i>concept map, concept maps, concept mapping, knowledge map, knowledge maps, knowledge mapping</i>

⁴ Par recension systématique des écrits, il est entendu que la collecte de textes a été faite de manière systématique en examinant tous les documents repérés dans les bases de données considérées, et ce dans le but de brosse le portrait des études portant sur les usages de la carte conceptuelle comme outil de collecte de données en vue de l’évaluation des apprentissages. La méthode suivie lors de cette recension ne suit pas de normes particulières, comme la norme PRISMA (Moher, Liberati, Tetzlaff et Altman, 2009), même si elle partage plusieurs principes de base.

Par souci de contrôle de qualité, seuls les documents révisés par des pairs, c'est-à-dire les rapports de recherche avec comité de lecture, les articles publiés dans les actes de conférence, les articles publiés dans les revues scientifiques et les thèses ont été retenus. Au total 2481 références ont été répertoriées dans les trois bases de données, soit 581 références dans ProQuest, 515 dans PsychINFO et 1385 dans Scopus. Les résultats de recherche des trois bases de données ont été mis en commun et les doublons ont été éliminés. Ce sont finalement 1984 références distinctes qui ont été traitées par la suite.

L'étape suivante a consisté à colliger les textes. Certains textes n'ont plus être collectés pour l'une des quatre raisons énumérées dans le Tableau IV. Puis s'en sont suivis une première lecture des résumés et un travail de filtrage des documents restants (N = 1885).

Tableau IV. Description des facteurs d'exclusion lors de la récupération des textes

<i>Facteur d'exclusion</i>	<i>Description</i>	<i>Fréquence</i>
Non disponible	Soit le texte est introuvable, soit il existe uniquement en version papier et celle-ci n'est pas disponible.	49
Accès payant	La revue scientifique requiert un accès payant qui n'est couvert pas dans les catalogues de l'Université de Montréal	36
Langue étrangère	Le texte est rédigé dans une langue autre que le français ou l'anglais.	13
Très mauvaise qualité du document	La copie disponible du texte est illisible.	1

La stratégie de recherche documentaire de départ étant très ouverte, un grand nombre de textes a été répertorié dans les bases de données, dont plusieurs qui ont fini par ne pas être retenus pour cette recension. Le Tableau V dresse la liste des facteurs d'exclusion à la suite de l'analyse de contenu des résumés ou des textes. En fin de compte, ce sont seulement 7,6 % des textes répertoriés (150 sur 1984) au départ qui ont fini par faire l'objet des analyses plus détaillées.

Comme il est possible de le constater, la majorité des textes a été exclue parce qu'ils n'étaient pas d'intérêt pour le présent projet, et ce même si les champs de recherche renfermaient les mots-clés. Les termes associés à l'évaluation et à la cartographie conceptuelle sont en effet assez généraux et polysémiques, de telle sorte que plusieurs textes d'autres domaines de connaissances ont été identifiés par les moteurs de recherche. Ces textes ont été exclus de la liste finale à la suite d'une première lecture de leurs résumés. Le Tableau V dresse l'inventaire des autres raisons pour lesquelles les textes ont été écartés des analyses.

Tableau V. Description des facteurs d'exclusion lors de l'analyse du contenu des textes répertoriés

<i>Facteur d'exclusion</i>	<i>Description</i>	<i>Fréquence</i>
Texte non pertinent	Même si les champs sondés lors de la recherche contiennent au moins une référence aux termes liés à l'évaluation et une référence aux termes associés à la cartographie conceptuelle, le texte ne concerne pas le domaine de connaissances de la recension. Par exemple, il peut faire référence à la méthode de recherche de « concept mapping » de Trochim (1989).	1434
Définition non conforme	La définition de la carte conceptuelle n'est pas conforme à celle présentée au deuxième chapitre ou le chercheur ne définit pas clairement ce qui est entendu par une carte conceptuelle.	85
Aucune étude empirique	Le document ne présente pas d'étude empirique.	70
Informations manquantes	Plusieurs informations relatives aux considérations théoriques ou méthodologiques sont absentes du texte.	57
Modalités de passation non conformes	La méthode de collecte de données n'exige pas que la personne évaluée produise ou complète une carte conceptuelle ou la carte est produite en équipe.	41

<i>Facteur d'exclusion</i>	<i>Description</i>	<i>Fréquence</i>
Automatisation de la production d'une carte conceptuelle ⁵	Le texte traite de méthode pour générer une carte conceptuelle de manière semi-automatisée ou automatisée, à partir de sources de données variées (réponses aux tests, textes produits par les répondants, textes issus d'ouvrage de référence). Il ne traite pas nécessairement de l'évaluation des apprentissages.	19
Absence de modalité de notation	Les chercheurs procèdent à un traitement strictement qualitatif du contenu des cartes, souvent sous la forme d'une analyse de contenu. L'étude n'a pas directement comme objectif d'apprécier le niveau d'apprentissage des personnes évaluées.	17
Duplicata	Le texte est quasi identique à un autre document ou il présente la même étude empirique.	12

Les 150 textes retenus ont donc été analysés selon plusieurs dimensions, plusieurs trouvant leur source d'inspiration dans le cadre conceptuel de Ruiz-Primo et Shavelson (1996). Ces derniers ont étudié quatre dimensions des tâches de cartographie conceptuelle : les modalités de passation, les modalités de notation, les estimations de la fidélité des scores et les preuves de validité. Ce chapitre aborde ces quatre dimensions en plus d'en ajouter deux nouvelles : les objets d'évaluation et les modalités de formation. Toutes ses dimensions se divisent elles-mêmes en sous-dimensions, qui sont définies dans leur section respective.

D'autres dimensions et sous-dimensions ont été considérées au cours du traitement des données, mais faute d'information dans les textes, elles ont été abandonnées. L'une de ces dimensions est d'ailleurs la fonction de l'évaluation. Même si certains chercheurs mentionnent que la carte conceptuelle peut servir de dispositif d'évaluation formative (par ex., Berlanga, van Rosmalen, Boshuizen et Sloep, 2012; Cathcart, 2016; Conlon, 2006; Joseph, Conradsson,

⁵ La recherche documentaire ne visait pas directement cette thématique qui, il faut le rappeler, concerne le deuxième article de cette thèse. Cela dit, ces textes ont été consultés et certains d'entre eux ont été considérés lors de l'écriture de cet article. Les autres textes n'ont pas été retenus parce qu'ils ne répondaient pas aux critères choisis dans cet article.

Nilsson Wikmar et Rowe, 2017) ou de dispositif d'évaluation sommative (par ex., Rice, Ryan et Samson, 1998), c'est assez rare qu'ils situent leurs travaux de recherche par rapport à l'une ou l'autre des fonctions d'évaluation. Cela s'explique probablement par le fait que les données collectées au moyen de tâche de cartographie conceptuelle servent essentiellement aux fins de la recherche et qu'en fin de compte, le ou les enseignants n'exploitent pas les données en salle de classe.

3.2 Les caractéristiques des études

Les analyses plus approfondies de cette recension systématique ont donc été effectuées sur un corpus contenant 150 textes, et ce pour un total de 155 études distinctes⁶. Lors d'une première analyse globale, les textes repérés ont pu être regroupés en deux grandes catégories d'objectifs de recherche. D'un côté, il y a les études qui se sont servies de la carte conceptuelle pour tout simplement évaluer les apprentissages d'individus. La majorité des textes (71 %, N=106) tombent dans cette première catégorie. De l'autre, il y a les études qui ont examiné les propriétés psychométriques de différentes modalités et tâches de cartographie conceptuelle en vue de l'évaluation des apprentissages.

Ces études sont réparties sur près de quatre décennies et comme l'illustre la Figure 4, les utilisations de la carte conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages gagnent en popularité depuis les premières publications de Novak et Gowin (1983, 1984). En revanche, le nombre d'études qui visent à apprécier les propriétés psychométriques des tâches de cartographie conceptuelle demeure assez stable au cours de la même période de temps (moins de dix études par période de cinq ans).

⁶ Certaines publications présentent les travaux de plusieurs études distinctes à la fois. Un article comporte trois études et trois articles renferment deux articles chacun.

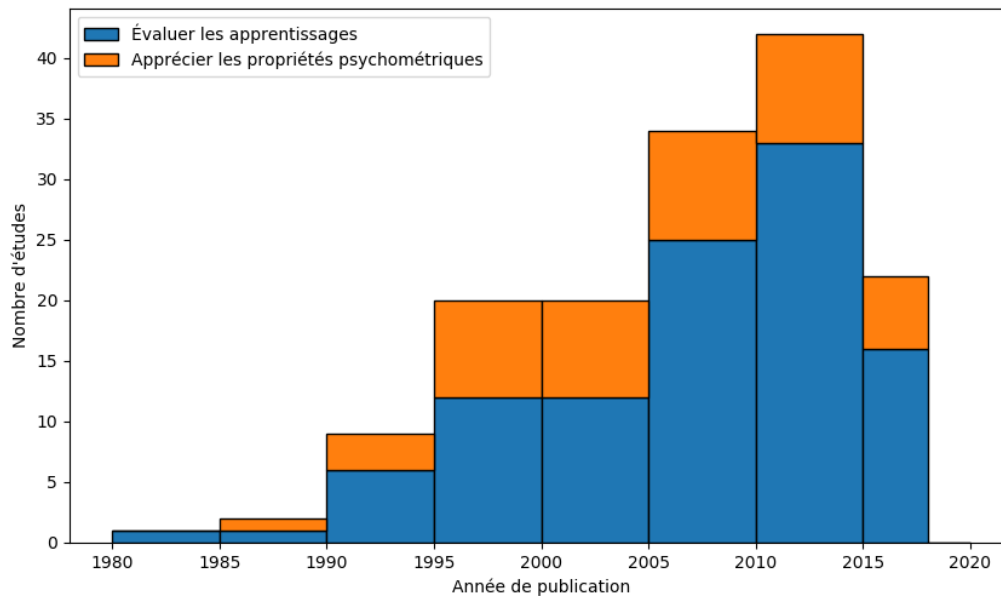


Figure 4. Distribution des années de publication des études recensées selon les objectifs de recherche.

Afin de mieux comprendre les conditions expérimentales des études recensées, ces dernières ont également été analysées et codifiées selon le domaine de connaissances ciblé par la ou les tâches d'évaluation, selon le niveau de scolarité et selon la taille d'échantillon des répondants. À l'origine, la cartographie conceptuelle a été développée dans le domaine de l'apprentissage en sciences. Il n'est donc pas surprenant de constater que la majorité des tâches ($N = 81$, 52 %) concernent une discipline scientifique (physique, biologie, chimie, etc.) ou une discipline connexe comme les sciences de la santé ($N = 20$, 13 %), le génie ($N = 17$, 11 %) ou les mathématiques ($N = 8$, 5 %). En revanche, seulement sept études (5 %) ont proposé des tâches dans le domaine des sciences sociales et les études restantes ($N = 22$, 14 %) ont présenté des tâches dans des domaines variés, comme l'éducation, la technologie de l'information, l'art, la gestion ou encore la santé et la sécurité au travail.

Qui plus est, les chercheurs ont majoritairement préféré réaliser leurs projets de recherche auprès d'étudiants de niveau postsecondaire ($N = 98$, 63 %). Autrement, le quart des études ont été effectuées auprès d'étudiants de niveau secondaire ($N = 39$, 25 %) et à peu près

5 % d'entre elles se sont déroulées auprès d'élèves du primaire. Quatre études (3 %) ont aussi été effectuées auprès d'étudiants de niveaux variés. Les autres études (4 %) se sont faites à l'extérieur du système scolaire, soit en formation professionnelle en entreprise, soit auprès d'enseignants ou d'experts de contenu, soit avec des patients.

Enfin, comme il peut être observé à la Figure 5, un grand nombre d'études (37 %) est basé sur des échantillons de moins de 50 répondants et la majorité (70 %) sous la barre de 100 répondants. Et ce constat est relativement le même pour les études qui avaient pour objectif d'évaluer les apprentissages à l'aide de la carte conceptuelle (en bleu) et celles qui cherchaient à apprécier les propriétés psychométriques de tâches de cartographie conceptuelle (en orange). De si petits échantillons ont bien entendu pour effet de restreindre la mise en place des plans de recherche complexe et la généralisation des résultats.

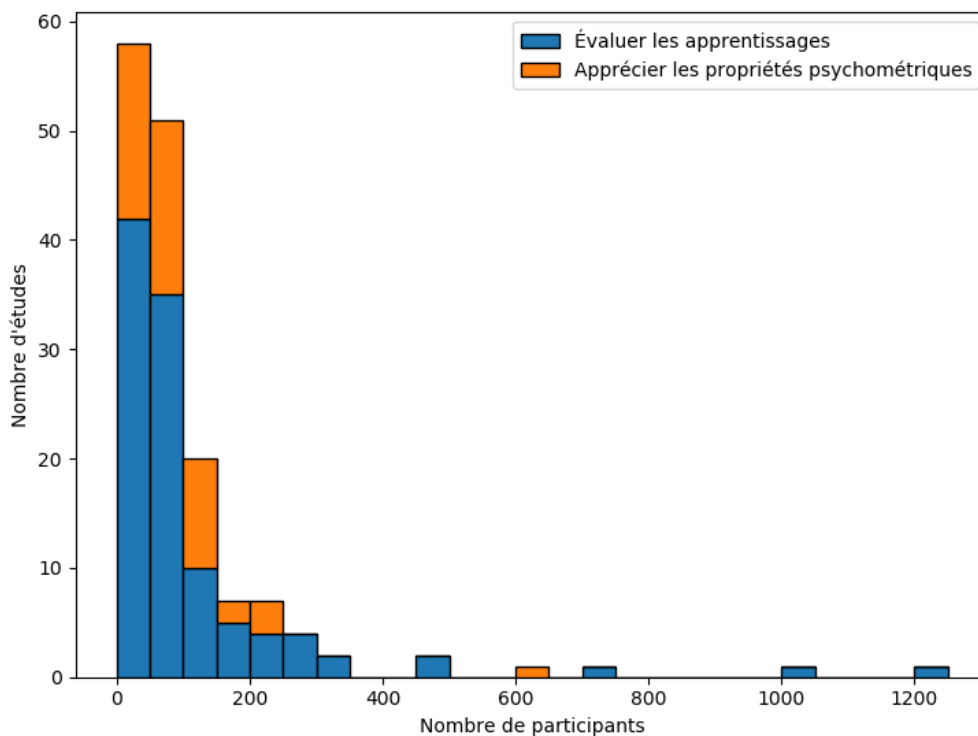


Figure 5. Distribution de la taille d'échantillon des études selon les objectifs de recherche

3.3 Les objets d'évaluation

Avant de s'intéresser aux modalités d'évaluation en soi, il est important de définir, ou au moins de nommer, ce qui est hypothétiquement évalué à l'aide de tâches de cartographie conceptuelle. Une erreur assez fréquente, parfois le fruit d'un simple abus de langage, consiste à confondre l'évaluation de cartes conceptuelles et l'évaluation des apprentissages au moyen de cartes conceptuelles. Cette recherche s'intéresse au deuxième objet d'évaluation. Dans une note théorique intitulée « *How good is my concept map? Am I a good Cmapper?* » (en français, « À quel point ma carte conceptuelle est-elle bonne ? Suis-je un bon cartographe ? »), Cañas, Novak et Reiska (2015) expliquent qu'une « excellente carte conceptuelle » se démarque par la qualité de son contenu, la qualité de sa structure et la qualité globale de la carte. Il n'est pas clair toutefois si ces trois dimensions de la carte conceptuelle sont des manifestations suffisantes et nécessaires de l'apprentissage, comme ces auteurs le prétendent.

Dans les études recensées, les chercheurs empruntent une panoplie d'expressions et de termes pour désigner ce qu'ils prétendent évaluer au moyen de leurs tâches de cartographie conceptuelle. Par souci de clarté, les vocables répertoriés ont été regroupés en six grandes catégories, qui couvrent divers aspects de la connaissance (voir le Tableau VI). Comme en font foi les vocables énumérés dans ce tableau, il n'y a, en toute apparence, aucun consensus à l'égard de l'objet d'évaluation d'une tâche de cartographie conceptuelle.

Tableau VI. Distribution des objets d'évaluation selon les objectifs de recherche

Catégorie	Objets de l'évaluation	Objectif de la recherche		Total
		Évaluer les apprentissages (N = 107)	Apprécier les propriétés psychométriques (N = 48)	
Compréhension	Compréhension ; Compréhension conceptuelle ; Compréhension connectée ; Compréhension de textes ; Compréhension des concepts ; Compréhension générale ; Compréhension signifiante	30 %	23 %	28 %
Structure des connaissances	Complexité de la structure ; Connexion entre concepts ; Degré de connexion des connaissances ; Interrelation entre les concepts ; Intégration des connaissances ; Représentation et organisation des connaissances ; Structure cognitive ; Structure des connaissances	23 %	33 %	26 %
Rendement	Rendement scolaire	13 %	23 %	16 %
Connaissances	Acquisition de connaissances ; Connaissances ; Connaissances antérieures ; Connaissances conceptuelles ; Connaissances du contenu ; Connaissances déclaratives ; Connaissances extériorisées	14 %	8 %	12 %
Changement	Changement de la compréhension ; Évolution des connaissances, Changement conceptuel	5 %	4 %	5 %
Autres	Apprentissage signifiant ; Changement conceptuel ; Conceptions ; Niveau cognitif ; Pensée ; Pensée critique ; Pensée systémique ; Qualité de l'apprentissage ; Propositions limitées ou inappropriées	7 %	4 %	6 %
Non indiqué		7 %	4 %	6 %

Ce qui est peut-être étonnant dans cette analyse est l'écart qui existe entre le vocable désignant les objets d'évaluation et certaines caractéristiques des études. Lors de l'analyse de contenu des textes, toutes les études ont été classifiées en deux types d'interprétation : celles dont les interprétations sont basées sur une évaluation ponctuelle des apprentissages et celles dont les interprétations reposent sur la progression des apprentissages. Parmi les 107 études qui ont pour objectif d'évaluer les apprentissages, près de la moitié (48 %) font passer la même tâche à deux occasions ou plus et s'intéressent par le fait même à la progression des apprentissages. En comparaison, seulement 8 % des études concernant les propriétés psychométriques adoptent une approche similaire.

Ce portrait met en évidence un constat bien intéressant. Il semble que les chercheurs emploient assez souvent la carte conceptuelle pour évaluer la progression des apprentissages, mais ils ne le reconnaissent pas comme tel. Il y a pourtant une réflexion à faire entre l'objet d'évaluation, le type d'interprétation proposé et les modalités d'évaluation. Même s'il existe possiblement plusieurs dérives de leur approche, Novak et Gowin (1984) sont assez clairs lorsqu'ils définissent leur modèle de notation d'une carte conceptuelle : celui-ci est plus adapté à une évaluation de la progression des apprentissages qu'à une évaluation ponctuelle.

3.4 Les modalités de formation

Malgré qu'il soit largement reconnu que la formation à la cartographie conceptuelle est un incontournable pour assurer la qualité des productions des individus (Horton et coll., 1993), il faut bien se rendre compte que les bonnes pratiques dans ce domaine sont très variables et très peu documentées, en comparaison aux modalités de notation, par exemple. Parmi les 150 textes recensés, 33 (22 %) n'ont même pas fait mention des modalités de formation dans leur texte et parmi les études restantes, les informations sont parfois sommaires et morcelées et les modalités sont hétéroclites.

Les retombées de la formation dépendent en partie du temps consacré à l'acquisition de connaissances de base au sujet de la cartographie conceptuelle et au développement de certaines habiletés liées à cette technique de représentation des connaissances. Or, la durée de la phase de formation est très variable d'une étude à l'autre. La formation la plus courte répertoriée dure huit minutes (Herl, O'Neil, Chung et Schacter, 1999), alors que d'autres modalités de formation

se sont étalées sur plusieurs séances de cours, durant lesquelles les apprenants ont réalisé différentes activités de cartographie conceptuelle (Bell, 2018; Liu et Hinchey, 1996; Novak, Gowin et Johansen, 1983; Rice et coll., 1998). Cela dit, comme l'indique le Tableau VII, près du tiers (32 %) des modalités de formation recensées ont duré moins de trente minutes, et environ 59 % ont pris moins d'une heure, ce qui équivaut à peu près à une période de classe. Il ne faut pas oublier que la majorité des études recensées ont été faites auprès d'étudiants postsecondaires, ce qui explique peut-être que la formation a été relativement courte. Évidemment, il va de soi que celle-ci doit être adaptée à l'âge des individus (Novak et Gowin, 1984). La durée de la formation est fort probablement aussi tributaire des contenus traités et des activités pédagogiques proposées.

Tableau VII. Distribution de la durée de la formation dans les études recensées

Durée de la formation	Pourcentage (N = 60*)
30 minutes ou moins	32 %
Entre 30 et 60 minutes	27 %
Entre 60 et 120 minutes	12 %
Entre 120 et 180 minutes	8 %
Plus de 180 minutes	5 %
Plusieurs séances	17 %

* Seulement 60 des 155 études ont fourni des informations au sujet de la durée de la formation

Les activités de formation peuvent être séparées en activités d'ordre théorique et en activités d'ordre pratique. D'une part, un peu plus des deux tiers des études (N = 102, 68 %) ont offert une présentation magistrale lors de laquelle une personne, soit un chercheur ou un enseignant, a fait un court exposé aux participants afin de leur expliquer les principes de la cartographie conceptuelle et de leur donner des consignes relatives à la construction d'une carte. Dans certaines de ces études, le présentateur a aussi consacré un peu de temps à des éléments de contenu plus spécifiques, comme la formation au logiciel permettant de tracer la carte (N = 17, 11 %). D'autre part, à peu près la moitié des études (N = 75, 48 %) ont aussi proposé des activités préparatoires, permettant aux individus de s'exercer à construire une carte conceptuelle et à développer les habiletés spécifiques à cette technique de représentation des connaissances.

Ces activités peuvent se faire selon plusieurs modalités. En s'inspirant de la thèse de Bartels (1995), Bell (2018) a proposé, par exemple, un plan de formation réparti en sept séances qui intègre une grande variété d'activités préparatoires. Son plan comprend une activité de cartographie conceptuelle réalisée en grand groupe (enseignant et élèves), des activités en dyades et des activités individuelles. Son objectif était ainsi de varier les activités et d'amener tous les élèves à maîtriser peu à peu cette technique de représentation des connaissances.

Quelques rares études ont également présenté un protocole d'évaluation de la qualité de la formation. Plus précisément, six d'entre elles ont analysé de manière informelle les cartes produites par les étudiants lors des activités préparatoires et sept d'entre elles l'ont fait de manière plus formelle, en vérifiant, par exemple, que les liens étaient effectivement étiquetés, que chaque carte contenait au moins une proposition juste et valide ou qu'il y avait au moins un niveau hiérarchique dans la carte de chaque répondant. Dans chacune de ces études, les chercheurs ont conclu que la formation a rempli sa mission.

Même si la formation préalable à la cartographie conceptuelle est parfois négligée, le choix des modalités de formation peut avoir un effet véritable sur les productions des individus évalués. Il y a déjà plusieurs années, Anderson et Huang (1989) ont montré que des élèves qui reçoivent une formation « détaillée » composée d'un exposé magistral et d'exercices préparatoires obtiennent de meilleurs résultats lors de tâche de cartographie conceptuelle que des élèves qui reçoivent une formation « brève », composée seulement d'un exposé magistral. Et Yin et coll. (2005) ont peut-être raison d'affirmer que leur modèle de formation de 20 minutes est suffisant pour amener les répondants à respecter les règles de construction d'une carte conceptuelle, mais une telle formation ne donne pas nécessairement à ces individus l'occasion de développer certaines habiletés spécifiques à cette technique de représentation des connaissances, comme dans les modalités de formation échelonnées sur plusieurs séances. En toute logique, la formation doit être adaptée aux besoins et au contexte de la situation d'évaluation des apprentissages.

3.5 Les modalités de passation

La cartographie conceptuelle étant une activité relativement multiforme, il existe un nombre quasi infini de modalités de passation et conséquemment, il est difficile, voire

impossible, de toutes les énumérer. Dans leur recension des écrits, Ruiz-Primo et Shavelson (1996) décortiquent une tâche de cartographie conceptuelle en deux dimensions, à savoir la tâche comme telle et le format de réponse, qui sont elles-mêmes divisés en sous-dimensions. Les études répertoriées dans la présente recension ont été analysées systématiquement à l'égard de ces dimensions et de leurs sous-dimensions respectives.

3.5.1 La tâche

La première dimension, la tâche, est décrite en matière d'exigences, de contraintes et de structure du contenu. Les exigences d'une tâche correspondent tout simplement à ce qui est demandé à l'individu. Dans leur recension, non seulement Ruiz-Primo et Shavelson (1996) considèrent les tâches lors desquelles la personne évaluée doit produire ou compléter une carte conceptuelle, mais ils tiennent compte aussi de tâches d'évaluation apparentées, comme des tâches d'organisation (ou de mise en ordre d'importance) de cartons sur lesquelles sont inscrits des énoncés de concepts, des tâches d'appréciation du degré de connexité de plusieurs paires de concepts ou des tâches de production écrite (Ruiz-Primo et Shavelson, 1996). Au contraire de Ruiz-Primo et Shavelson, la présente recherche écarte un bon nombre de ces modalités et ne prend en considération que les tâches lors desquelles l'individu interagit directement avec la carte conceptuelle.

La deuxième sous-dimension, c'est-à-dire les contraintes de la tâche, correspond aux degrés de liberté (ou de restriction) qu'ont les répondants lors de la réalisation de la tâche. Le type de tâche promue par Novak et Gowin (1984), qui consiste à produire une carte à partir de zéro, est en toute fin pratique la modalité connue la plus ouverte. Cette liberté conduit généralement à des productions plus diversifiées, ce qui est intéressant dans certaines situations, mais ce qui a aussi un prix. Dans un contexte d'évaluation des apprentissages, il est parfois préférable de restreindre les possibilités d'une tâche, notamment en donnant certaines consignes ou en fournissant certains éléments aux répondants, comme des énoncés des concepts ou des expressions de liaison, de sorte à uniformiser un peu plus leurs productions et ainsi à faciliter la correction (Ruiz-Primo, Schultz, et coll., 2001).

Le Tableau VIII donne le portrait des types de tâches, des exigences et des contraintes préconisées dans les études recensées. La grande majorité des tâches, soit environ 90 % d'entre

elles, ont requis de produire une carte. Ces tâches se répartissent à peu près en parts égales entre la modalité ouverte recommandée par Novak et Gowin (1984), c'est-à-dire la production d'une carte à partir de zéro, et une modalité un peu plus restrictive qui consiste à produire une carte conceptuelle à partir d'une liste de concepts. Quelques études (N=20, 12 %) ont également demandé de produire une carte conceptuelle à partir d'une liste de concepts et de mots de liaison. En revanche, les tâches du type « compléter une carte » sont plutôt rares. Le Tableau VIII fait aussi état de deux autres modalités qui s'apparentent aux tâches du type « compléter les trous ». Plutôt que de remplir des composantes laissées vides, les répondants dans ces modalités doivent soit corriger les erreurs préalablement identifiées par l'évaluateur soit trouver les erreurs et les corriger (Correia, Cabral et Aguiar, 2016).

Tableau VIII. Distribution des variantes de tâches dans les études recensées

Variante de tâche	Dispositif d'évaluation des apprentissages	Apprécier les propriétés psychométriques	Total (N=173*)
Production d'une carte	94 %	83 %	90 %
à partir d'une liste de concepts	41 %	41 %	41 %
à partir de zéro	46 %	20 %	36 %
à partir d'une liste de concepts et d'une liste de mots de liaison	6 %	20 %	12 %
à partir d'une structure fixe	0 %	2 %	1 %
à partir d'une liste de concepts et d'une structure fixe	1 %	0 %	1 %
Compléter une carte	2 %	14 %	6 %
Compléter les nœuds d'une carte à partir d'une liste de concepts	1 %	11 %	5 %
Compléter les arcs d'une carte à partir d'une liste de mots de liaison	1 %	3 %	2 %
Autres	0 %	3 %	1 %
Corriger les erreurs d'une carte	0 %	2 %	1 %
Trouver les erreurs et les corriger	0 %	2 %	1 %
Non indiqué	4 %	0 %	2 %

* Treize expérimentations ont étudié plus d'une modalité de passation.

Le Tableau VI compare également la distribution des modalités de passation entre les deux objectifs ciblés par la présente recherche. Étant naturellement plus exploratoires, les études au sujet des propriétés psychométriques ont mis à l'essai une plus grande variété de tâches de cartographie conceptuelle et certaines de ces études se sont penchées sur des modalités de passation qui sont très peu utilisées dans la pratique. En contrepartie, la production d'une carte à partir de zéro, qui est en fait la modalité la plus populaire pour évaluer les apprentissages des individus, ne semble pas avoir reçu autant d'attention de la part des chercheurs.

La dernière sous-dimension de la tâche considérée par Ruiz-Primo et Shavelson (1996), soit la structure du contenu, concerne l'interaction entre les contenus disciplinaires et les deux dimensions précédentes. Cette interaction peut prendre plusieurs formes, mais la hiérarchisation des connaissances et des concepts est une caractéristique d'une tâche qui incarne bien ce qui est entendu par la structure du contenu. Comme il a été mentionné au deuxième chapitre, Novak (Cañas et Novak, 2008; Novak et Gowin, 1984) prône l'organisation (semi-) hiérarchique du contenu d'une carte conceptuelle. Or, selon plusieurs (Ruiz-Primo, Schultz, et coll., 2001; Ruiz-Primo et Shavelson, 1996; Ruiz-Primo, Shavelson, et coll., 2001; Yin et coll., 2005), il n'y aurait aucune raison de demander à tout coup aux individus d'organiser hiérarchiquement leurs concepts. Certains domaines de connaissances, par exemple des branches de la biologie s'y prêtent peut-être, mais d'autres domaines ne s'y accommodent pas. Qui plus est, il est parfois très difficile d'identifier les niveaux hiérarchiques d'une carte lors de la correction de cartes produites par des individus (Watson, Pelkey, Noyes et Rodgers, 2016; West, Pomeroy, Park, Gerstenberger et Sandoval, 2000).

3.5.2 Le format de réponse

La deuxième composante du cadre conceptuel de Ruiz-Primo et Shavelson (1996) concerne le format de réponse et se divise en trois sous-dimensions. La première de ces sous-dimensions, le « mode de réponse », désigne le support avec lequel l'individu manifeste ses connaissances. Dans la présente recherche, deux supports sont pris en considération : le support papier et le support informatisé. Ruiz-Primo et Shavelson (1996) ont considéré quant à eux un troisième mode de réponse qui ne répond pas aux conditions de cette thèse : la production orale. Dans ce mode de réponse, une autre personne, souvent un chercheur, a comme responsabilité

de tracer une carte à partir des connaissances verbalisées par l'individu lors de la réalisation de la tâche. Il n'est donc pas certain que la carte finale, tracée par une autre personne, reflète exactement l'organisation des connaissances de la personne qui fait l'objet de l'évaluation. C'est pourquoi ce mode de réponse a été écarté dans cette recension.

Même si le support de passation est une dimension importante d'une tâche de cartographie conceptuelle, il n'est pas toujours évident d'établir lequel a été privilégié dans les études antérieures. En fait, le support de passation n'est précisé que dans 58 % des études. Comme le Tableau IX l'indique, les cartes conceptuelles ont été tracées de manière traditionnelle sur support papier la majorité du temps, et ce quel que soit l'objectif de la recherche. Dans quelques rares études, les participants ont également eu la liberté de choisir entre le support informatisé ou le support papier. La tendance observée au Tableau IX est vraisemblablement appelée à évoluer : les études recensées s'étalent sur une période de plus de trente ans et les chercheurs n'avaient pas nécessairement accès aux mêmes logiciels et ressources informatiques que celles auxquelles ils pourraient avoir accès maintenant et dans les années à venir.

La variété des supports de passation soulève toutefois des enjeux différents selon le contexte d'évaluation des apprentissages. Le choix d'un ou plusieurs supports suscite d'abord des enjeux en matière de formation préalable à la cartographie conceptuelle. La personne qui construit une carte conceptuelle doit évidemment se familiariser avec les outils et ressources mis à sa disposition (Weinerth et coll., 2014). Dans certaines situations, le support de passation soulève aussi des enjeux relativement à la comparabilité des notes. En contexte d'évaluation à grande échelle, par exemple, il n'est peut-être pas possible d'offrir les mêmes supports et les mêmes conditions de passation à toutes les personnes et dès lors, il faut s'assurer que le support de passation n'avantage ni ne désavantage pas certains groupes de candidats à l'examen (Raikes et Harding, 2003). À ce jour, l'étude de Brandstädter et coll. (2012) est la seule qui a comparé les notes d'élèves qui ont subi une tâche de cartographie conceptuelle à l'ordinateur ou à la main. Ils en sont arrivés à la conclusion que les supports ne sont pas équivalents : les participants ont obtenu des notes plus élevées lorsqu'ils construisent leur carte à l'ordinateur.

Tableau IX. Distribution des supports de passation dans les études recensées

Support	Apprécier les propriétés psychométriques	Dispositif d'évaluation des apprentissages	Total
Papier-crayon	74 %	65 %	69 %
Informatique	26 %	31 %	29 %
Papier-crayon/Informatique	0 %	5 %	3 %

La seconde dimension du format de réponse correspond à la personne qui trace la carte conceptuelle, ce que Ruiz-Primo et Shavelson (1996) nomment le « cartographe ». Dans les articles analysés dans leur recension, la carte peut être produite directement par la personne évaluée, tout comme elle peut aussi être tracée par une autre personne ou par un groupe de personnes. Au risque de se répéter et conformément aux conditions énoncées précédemment, seule la première situation est prise en considération dans cette recherche. La personne évaluée doit rendre une carte conceptuelle à la fin de la tâche.

Les caractéristiques du format de réponse constituent la troisième et dernière dimension du format de réponse. Cette dimension correspond aux spécificités de la tâche et des ressources mises à la disposition des individus. Ces caractéristiques varient donc d'une étude à l'autre et il est très difficile de les répertorier. Quelques exemples peuvent pourtant en illustrer la variété. L'utilisation d'étiquettes amovibles, communément appelées des Post-its™, lors de la production d'une carte conceptuelle sur support papier est un premier exemple de caractéristiques du format de réponse. Ces étiquettes permettent de déplacer plus facilement les nœuds de la carte (Cañas et Novak, 2008), ce qui est pratique autant lors de la phase d'élaboration d'une carte que lors de sa révision. L'utilisation de feuilles optiques dans les modalités du type « compléter les trous » est un deuxième exemple (Schau et Mattern, 1997; Schau, Mattern, Zeilik, Teague et Weber, 2001). Dans cette approche, il est attendu que les répondants retranscrivent leur réponse sur la feuille optique, c'est-à-dire qu'ils noircissent les numéros correspondant aux énoncés du concept qu'ils ont inscrits dans chaque trou, avant que les tâches soient corrigées et notées.

En somme, cette analyse pointue des modalités de passation, en se basant sur le cadre conceptuel de Ruiz-Primo et Shavelson (1996), a mis en lumière la diversité des manières de collecter des données à partir de cartes conceptuelles. Même s'il existe une certaine cohérence pour ce qui est des exigences générales de la tâche, il y a plusieurs facteurs et consignes spécifiques de la tâche qui sont différents d'une étude à l'autre. Nul doute, la variété des modalités de passation amène à se poser des questions au sujet de la comparabilité de celles-ci. Ces questions seront approfondies à la section 3.8.

3.6 Les modalités de correction et de notation

Alors qu'il semble y avoir un certain consensus au sujet des modalités de passation les plus populaires, il n'y a vraisemblablement pas de modèles de notation⁷ qui font l'unanimité dans les études antérieures. Dans les faits, il n'est pas rare de constater que les chercheurs traitent et analysent les cartes produites par leurs répondants à la lumière de deux modèles de notation ou plus. Comme von der Heide (2015) l'explique, les modèles de notation capteraient différents aspects de l'apprentissage et l'application de plusieurs modèles de notation permettrait d'éliminer ou de réduire les biais inhérents à chacun d'entre eux. De même, les chercheurs ont tendance à créer leur propre modèle de notation ou à adapter un modèle existant. Ils vont rarement faire appel à un modèle de notation d'une étude précédente : seulement 29 modèles ont été repris dans deux études ou plus. Au total, ce sont donc 218 modèles de notation différents qui ont été repérés et utilisés parmi les 155 études analysées.

Compte tenu du grand nombre de modèles de notation inventoriés, il est particulièrement compliqué de les catégoriser et de les comparer. Dans leur recension des écrits, Ruiz-Primo et Shavelson (1996) proposent un cadre conceptuel très général dans lequel ils font la distinction entre trois catégories de modèles : ceux qui reposent sur la notation des composantes d'une carte conceptuelle, ceux qui comparent la représentation des connaissances d'un individu avec une carte de référence et ceux qui combinent les deux approches précédentes. Ce cadre, trop simple,

⁷ L'expression « modèle de notation » est prise au sens large ici ; elle s'assimile au concept de « modèle d'interprétation ».

ne permet pourtant pas de différencier toutes les manières de traiter et d'interpréter les productions des individus dans les publications recensées.

Cette section analyse plutôt les divers modèles recensés selon cinq volets : 1) les critères d'évaluation, les indicateurs observables et les échelles de notation ; 2) la combinaison des notes partielles ; 3) la transformation des notes brutes ; 4) la modélisation des scores ; et 5) l'emploi de la carte de référence. Mais avant de se lancer dans cette analyse, un exemple est fort utile afin d'introduire certaines des idées qui seront discutées ci-dessous.

Dans leur ouvrage fondateur, *Learning how to learn*, Novak et Gowin (1984) proposent un modèle de notation qui illustre bien les caractéristiques fondamentales de plusieurs autres modèles de notation. Avec leur modèle, parfois nommé le modèle de notation structurale, le correcteur doit accorder un point à chaque relation valide, cinq points à chaque niveau hiérarchique valide, dix points à chaque relation transversale valide et un point à chaque exemple de concept valide. Les points accordés à chaque élément sont additionnés pour former des notes partielles correspondant à chacun des quatre indicateurs, puis ils sont combinés pour former une note totale. Dans plusieurs études, les résultats sont d'ailleurs analysés à ces deux niveaux, à savoir au niveau des notes partielles et au niveau de la note totale.

Le modèle de notation de Novak et Gowin est de loin le modèle le plus répandu dans les écrits analysés dans cette recension et il est celui qui a été adapté le plus souvent. Compte tenu de l'importance des travaux de Novak à l'égard de la cartographie conceptuelle, il n'est pas surprenant que plusieurs modèles s'inspirent de cette approche. Les sections suivantes reprendront d'ailleurs cet exemple à quelques reprises.

3.6.1 Les critères d'évaluation, les indicateurs observables et les échelles de notation

Une planification rigoureuse de l'évaluation des apprentissages exige d'abord de définir l'objet de l'évaluation, d'établir des critères d'évaluation et des indicateurs observables qui reflètent ce qui est évalué, et d'élaborer ou d'utiliser des dispositifs appropriés pour collecter les données et pour les interpréter (Durand et Chouinard, 2012). En règle générale, ce souci de cohérence entre les objectifs et les moyens est un principe bien établi en évaluation des

apprentissages (Pellegrino et coll., 2001; Wilson et Sloane, 2000). Cela dit, ce principe n'est pas toujours appliqué dans les études concernant la cartographie conceptuelle : les critères d'évaluation sont rarement énoncés et la plupart du temps, les modèles de notation sont formulés directement à partir des indicateurs et des échelles de notation. Le modèle de Novak et Gowin (1984) est une exception à la règle : ces derniers sont parmi les rares chercheurs à avoir justifié le choix des indicateurs sur une base théorique, en l'occurrence sur les principes de l'apprentissage significatif⁸ (Ausubel, 2000; Ausubel et coll., 1978).

Comme dans le modèle de notation donné en exemple, les indicateurs et les échelles de notation renvoient fréquemment à des composantes de la carte, ces dernières devenant en quelque sorte les unités d'analyse de la carte. Le modèle de notation structurelle, par exemple, est ainsi composé de quatre types d'unités : les relations, les niveaux hiérarchiques, les relations transversales et les exemples. Le Tableau X brosse le portrait des diverses unités d'analyse répertoriées dans les études antérieures. Ces unités sont présentées en ordre croissant de degré de granularité. Le nœud, le lien et les mots-liens sont les trois unités les plus élémentaires. Les autres unités reflètent soit une combinaison de ces unités élémentaires, soit ils représentent l'ensemble de la carte.

Tableau X. Description des différentes unités d'analyse répertoriées

No	Unité d'analyse	Description/Exemple d'indicateurs	Nombre d'indicateurs distincts
1	Nœud/Concept	Une figure géométrique renfermant un énoncé qui désigne un concept	26
2	Exemple	Une figure géométrique renfermant un énoncé qui désigne un exemple d'un concept. La figure peut être de couleur différente de celle d'un concept.	6
3	Lien	Un trait entre deux concepts	7

⁸ Comme il a été vu à la section 2.1, les fondements de la cartographie conceptuelle tels que promus par Novak et Gowin (1984) sont inspirés de la définition de l'apprentissage significatif d'Ausubel (Ausubel, 2000; Ausubel et coll., 1978).

No	Unité d'analyse	Description/Exemple d'indicateurs	Nombre d'indicateurs distincts
4	Proposition/relation	Une unité sémantique élémentaire composée d'au moins deux concepts et d'une relation. Dans plusieurs modèles, la proposition est définie par la relation qui unit deux concepts.	78
5	Relation transversale	Une relation entre deux concepts éloignés ou entre deux concepts provenant d'un même niveau hiérarchique	9
7	Embranchement	Un nœud qui a deux liens sortants ou plus	3
8	Niveau hiérarchique	Un ensemble de nœuds pour lequel un nœud désignant un concept superordonné se scinde en deux ou plusieurs nœuds représentant des concepts subordonnés. Ces nœuds sont souvent organisés du haut vers le bas.	8
9	Liens émanant d'un concept	Les liens émanant d'un concept	11
10	Idée/Conception	Parcelles de concepts et de propositions Par ex., Présence de conceptions erronées	8
11	Réseau (inspiré de la théorie des graphes)	L'ensemble du réseau d'arcs et de nœuds. Par ex., Diamètre du réseau Par ex., Nombre de cycles du réseau.	10
12	Organisation générale de la carte	Analyse globale visuelle de la structure de la carte Par ex., Typologie de structure de Kinchin (2000) : la structure en chardon, la structure en chaîne, la structure en réseau intégré.	17
13	Holistique/Ensemble de la carte	Critères d'évaluation et indicateurs variés	24
14	Autre	Références ; ressources ; processus	3

Les définitions des indicateurs répertoriées dans les études sont nombreuses et variées. Comme le dénote la colonne de droite du Tableau X, il peut en effet y avoir plusieurs variantes d'indicateurs pour chaque type d'unité d'analyse. Pour ce qui est des liens, par exemple, les chercheurs peuvent s'intéresser au nombre de liens présents dans une carte (Walker, Cordray,

King et Fries, 2005), au nombre de liens valides dans une carte (Molaison, Taylor, Erickson et Connell, 2009), au nombre de liens de la carte de l'individu qui sont aussi présents dans une carte de référence (Ifenthaler, 2010a), etc.

Ces définitions se distinguent à plusieurs égards. Les indicateurs peuvent tantôt concerner la quantité d'éléments contenus dans la carte, tantôt la qualité de ces éléments. Cela dit, bon nombre des indicateurs recensés (95 des 210 indicateurs distincts, 45 %) ont comme caractéristique qu'ils requièrent à la fois d'apprécier la qualité du contenu et de dénombrer des composantes. Les indicateurs concernant les relations et les propositions en sont de bons exemples. Plusieurs indicateurs consistent à évaluer la justesse ou la pertinence des relations une à une, puis à sommer les points à accorder à chacune d'entre elles.

La définition d'un indicateur dépend également du support de notation employé pour apprécier la carte produite par un individu. Les échelles de notation sont alors des dispositifs très courus pour soutenir le jugement du correcteur. Parmi les publications analysées, il n'y a pas moins de 28 échelles différentes pour la notation de la qualité des propositions. Ces échelles se distinguent selon le format de l'échelle (numérique ou descriptif), selon le type de descripteurs et selon le nombre d'échelons. Le Tableau XI donne trois exemples parmi tant d'autres de ces échelles. Tout comme les critères d'évaluation et des indicateurs, les échelles de notation constituent elles aussi une autre source de variation entre les études répertoriées.

Tableau XI. Exemples d'échelles de notation de la qualité des propositions

Format d'échelle	Échelons	Auteurs
Numérique	0, 1, 2, 3	(Austin et Shore, 1995)
Qualitatif (court)	Proposition incorrecte, Proposition partiellement correcte Proposition correcte	(Conlon, 2006)

Format d'échelle	Échelons	Auteurs
Numérique/descriptif	0 : proposition incorrecte ou scientifiquement non pertinente, 1 : proposition partiellement incorrecte, 2 : proposition correcte, mais scientifiquement « mince », 3 : proposition scientifiquement correcte et formulée en termes scientifiques.	(Yin et coll., 2005)

La diversité des critères, des indicateurs et des échelles de notation est une chose, mais il y a au moins autant de manières de les combiner entre eux pour former un modèle de notation. En fait, dans 74 % des études recensées, les chercheurs synthétisent les informations récoltées au moyen des indicateurs sous forme d'un résultat global, souvent présenté sous forme d'une note aux élèves. Les modèles d'interprétation recensés étant fortement axés sur une approche quantitative, ce sont principalement ces modèles qui sont discutés dans ce qui suit. Les approches essentiellement qualitatives sont en effet assez rares. Le lecteur peut consulter Alhashem et Al-Jafar (2015); Besterfield-Sacre et coll. (2004); Hay (2007); Hay et Kinchin (2006); Kinchin, Hay et Adams (2000); Roth et Bernhardt (2016); Shallcross (2016).

3.6.2 La combinaison des notes partielles

Le degré de complexité d'un modèle de notation croit rapidement selon la nature et le nombre de critères et d'indicateurs qui le compose. Exactement 50 % des modèles de notation répertoriés (N = 109) sont composés de plus d'un indicateur et certains vont jusqu'à huit. Les chercheurs proposent évidemment divers mécanismes pour combiner les notes partielles attribuées aux indicateurs. Trois stratégies générales ont été observées pour établir un score final.

La première stratégie, et vraisemblablement la plus simple, consiste à calculer la somme des notes individuelles, c'est-à-dire à faire la somme des notes décernées à chaque composante de la carte. À plusieurs occasions, les évaluateurs préfèrent pourtant accorder différentes pondérations à chaque indicateur ou composante. Par exemple, Rye et Rubba (2002) expliquent dans leur article qu'ils accordent plus d'importance à la justesse des relations qu'à la pertinence

des concepts étant donné que ce premier indicateur est une meilleure manifestation de la compréhension. La seconde stratégie consiste donc à appliquer diverses pondérations aux notes partielles de chaque indicateur. Dans leur modèle, Rye et Rubba décident donc d'accorder deux fois plus de points aux relations qu'aux concepts. La troisième stratégie est encore plus pointue : la pondération est appliquée indépendamment à chaque composante. Dans le modèle nommé « Importance/Qualité » (Srinivasan et coll., 2008), par exemple, les correcteurs attribuent un facteur d'importance (sur une échelle de 0 à 2) à chaque paire de concepts possible, puis ils donnent leur appréciation à l'égard de la qualité de chaque relation formulée par les répondants (sur une échelle en six points). La note attribuée à chaque relation équivaut ainsi au produit du facteur d'importance et de la note de la justesse de la proposition et la note totale est la somme de chaque note individuelle pondérée. Des variantes de cette stratégie ont aussi été mises à profit par d'autres chercheurs (Anderson et Huang, 1989; Greene, Lubin, Slater et Walden, 2013; Gregoriades, Pampaka et Michail, 2009; Shymansky et coll., 1993).

3.6.3 La transformation de la note brute

Les méthodes de calculs de notes basées sur la somme (ou la somme pondérée) de notes partielles sont peut-être faciles à mettre en application, mais l'interprétation des résultats qui en découle n'est pas toujours intuitive. Dans les variantes de tâches ouvertes pour lesquelles il n'y a pas de limite quant au nombre de composantes, car ces notes s'expriment sur des échelles de valeurs qui varient de zéro à l'infini, et en l'absence de valeur maximale, il est difficile de se prononcer sur le niveau de performance d'un individu. En effet, que veut dire une note de 34 dans ce contexte ?

Une stratégie courante pour remédier à cet obstacle à l'interprétation des scores consiste alors à transformer les notes sur une échelle plus commune, comme une échelle de valeurs se situant entre 0 ou 1 ou une échelle en pourcentages. Quelques méthodes pour effectuer de telles transformations ont été repérées dans les études.

Les deux premières méthodes divisent la note décernée à la carte par une note jugée maximale. Plus précisément, la première d'entre elles correspond au quotient entre la note obtenue par l'individu et la note qu'il aurait obtenue si toutes les composantes de sa carte avaient été exactes. Cette méthode est parfois nommée le score de qualité (Abd-El-Khalick et

BouJaoude, 1997; Burrows et Mooring, 2015; Francisco, Nakhleh, Nurrenbern et Miller, 2002). La deuxième méthode consiste plutôt à diviser la note finale décernée à la carte par la note attribuée à une carte de référence. Cette méthode est communément appelée le score de convergence (Abd-El-Khalick et BouJaoude, 1997; Ruiz-Primo, Schultz, et coll., 2001; Schwendimann, 2013; Wills, 2008). Malgré certaines similitudes, ces deux modèles de notation renvoient pourtant à des interprétations bien différentes des apprentissages réalisés et ne sont pas à l'abri de certaines dérives. Par exemple, un individu peut facilement obtenir une note parfaite (de 1) avec le premier modèle de notation en produisant une carte qui ne contient qu'une seule paire de concepts. Ou encore un individu peut dépasser la note maximale théorique (de 1) avec le modèle du score de convergence lorsque sa carte dépasse les attentes (Novak et coll., 1983; Ruiz-Primo, Schultz, et coll., 2001).

Une autre famille de méthode pour transformer les scores consiste à comparer directement le contenu des deux cartes et à produire des scores de similarité. Dans ces méthodes, le score final correspond donc à la proportion d'éléments (d'ordinaire les concepts et les relations) concordants entre les deux cartes et se situe véritablement sur une échelle variant de 0 à 1. Plus la valeur de l'indice de similarité est élevée, plus la carte de l'individu est similaire à la carte de référence. Parmi les méthodes repérées, c'est celle de Goldsmith qui a été citée et reprise le plus souvent dans les études recensées (N = 5). Cette méthode consiste à examiner et à comparer le voisinage de chaque concept entre deux cartes, généralement la carte de l'individu et une carte de référence, puis à déterminer le degré de correspondance entre ces deux cartes. Fait intéressant : les valeurs observées de l'indice de Goldsmith sont en général faibles ou très faibles dans les études analysées. Par exemple, les valeurs oscillent entre 0,04 et 0,49 (M = 0,22, É.-T. = 0,12) dans l'étude de Michael (1995) et entre 0,00 et 0,42 (M = 0,15, É.-T. = 0,08) dans celle de Jablokow, Defranco, Richmond, Piovoso et Bilén (2015). Il semble en effet que cet indice soit très sensible aux moindres différences entre deux cartes, ce qui fait en sorte que les valeurs observées soient concentrées dans la partie inférieure de l'échelle et que la variance de celles-ci est plutôt faible. En conséquence, ce modèle n'est peut-être pas le plus approprié pour discriminer les performances des individus.

3.6.4 La modélisation des scores

Quoique peu évoquée dans les études concernant la notation des cartes conceptuelles, la modélisation des scores est une approche bien connue dans le domaine de l'évaluation des apprentissages, notamment dans l'évaluation à grande échelle, et fort utile pour donner un cadre à l'interprétation des données. Un modèle de mesure est un modèle mathématique qui met en relation une série d'observations (souvent des notes aux items) et un score qui reflète le rendement de l'individu à la tâche d'évaluation. Il dicte les inférences qui peuvent être tirées à partir des données récoltées et les conditions dans lesquelles ces inférences sont valides.

Trois familles de modèles de mesure sont couramment utilisées et appliquées en éducation : les modèles de la théorie classique des tests (Lord et Novick, 1968; Novick, 1966), les modèles de la théorie de la généralisabilité (Brennan, 2001; Cardinet, Johnson et Pini, 2010; Cronbach, 1972) et les modèles de la théorie de la réponse à l'item (Birnbaum, 1968; Lord et Novick, 1968; Rasch, 1960). Les modèles des deux premières familles sont apparentés et s'intéressent particulièrement à quantifier l'erreur de mesure tandis que les modèles de la théorie de la réponse à l'item portent une attention un peu plus particulière à la relation entre les observations, les caractéristiques des items (par ex, leur difficulté) et les caractéristiques des personnes (par ex., leur habileté). Une description complète de ces modèles de mesure dépasse largement le cadre de cette recension. Le lecteur est invité à consulter les références ci-dessus ou encore à se référer à des ouvrages de nature plus didactique, comme ceux de Crocker et Algina (1986), de Bertrand et Blais (2004) ou de Bond et Fox (2007) pour se familiariser avec les concepts fondamentaux des modèles de mesure en éducation.

Les applications de la théorie de la généralisabilité et de la théorie de la réponse à l'item sont assez rares dans les études répertoriées. La plupart des modèles de notation des cartes conceptuelles sont basés sur des méthodes de calcul des scores observés et conséquemment partagent plusieurs points en commun avec les modèles de la théorie classique des tests. Les chercheurs dans ces études n'accordent toutefois pas autant d'attention à la quantification de l'erreur de mesure, comme c'est le cas dans la théorie classique des tests. Ils se satisfont de calculer un score final. Autrement, il n'y a que onze études parmi les 150 publications recensées qui font appel à la théorie de la généralisabilité et une seule s'appuie sur la théorie de la réponse à l'item. En raison de leurs particularités et de leur rôle en ce qui concerne l'appréciation des

propriétés psychométriques des tâches de cartographie conceptuelle, ces études seront discutées plus longuement à la section 3.7.

3.6.5 L'utilisation de la carte de référence

Un tour d'horizon des modalités de notation des cartes conceptuelles ne serait pas complet sans une discussion plus approfondie au sujet de la carte de référence, un dispositif assez courant dans les études répertoriées (N = 66, 43 %) et utilisé de nombreuses façons. Lors de l'élaboration des tâches, par exemple, la carte de référence peut servir à mettre au point des tâches du type « compléter les trous » ou à déterminer les listes de concepts et de mots de liaison. Lors de la correction, elle peut être utilisée comme outil d'appoint ou comme référent absolu.

Même si la carte de référence est un outil assez populaire, le recours à ce dispositif soulève des questions importantes, surtout en ce qui a trait aux méthodes pour l'élaborer et la valider. Une carte de référence représente une structure « idéale » des connaissances. Son utilisation présuppose ainsi quelques conditions telles que l'existence d'un certain consensus à l'égard de l'organisation des contenus dans le domaine de connaissances ciblé et l'existence d'un moyen de construire une représentation acceptable de ces contenus (Hatami, Farrokhnia et Hassanzadeh, 2016).

L'approche la plus fréquente pour élaborer une carte de référence consiste à consulter des « experts de contenu » et à leur demander de coconstruire cette représentation idéale. Les variantes répertoriées de cette approche sont pourtant très nombreuses. Parmi les 50 études recensées⁹, la carte de référence a été élaborée à 17 occasions par un seul individu et 33 fois par plusieurs personnes. Ces cartes ont été construites principalement par des enseignants (N = 18), mais elles ont aussi été construites par des spécialistes de contenu (des chercheurs ou des professeurs-chercheurs dans le domaine, N = 12). Les auteurs des articles sont aussi intervenus à plusieurs occasions dans l'élaboration de la carte de référence, parfois seuls (N = 7), parfois en collaboration avec des enseignants (N = 6) ou avec des spécialistes du contenu (N = 7).

⁹ Au total, 66 études ont mentionné avoir utilisé une carte de référence, seize d'entre elles n'ont pas précisé qui l'a élaborée.

Les protocoles d'élaboration des cartes de référence sont aussi très variables selon les études. La carte peut tout simplement être tracée par un seul expert de contenu dans des conditions similaires à celles qui sont offertes aux répondants (Borrego, Newswander, McNair, McGinnis et Paretti, 2009). À d'autres occasions, les personnes consultées construisent individuellement leur propre carte conceptuelle, puis ces cartes sont mises en commun pour obtenir une carte finale. Cette démarche d'agrégation des cartes peut être réalisée par les individus qui ont tracé les cartes individuelles (Dennis, 1999; Mahler, Großschedl et Harms, 2017; Wallace et Mintzes, 1990), comme elle peut être effectuée par le chercheur (Harris, 2008; Jimenez Snelson, 2011). Dans certaines situations, il arrive aussi que la carte soit révisée par des individus qui ne sont pas intervenus directement dans l'élaboration de celle-ci. Dans l'étude de Harris (2008), par exemple, la chercheuse fait la synthèse des cartes produites par dix assistants d'enseignement et chargés de cours, pour ensuite présenter cette carte à un expert de contenu dans le domaine évalué.

Bien sûr l'élaboration d'une carte de référence est régie par des contraintes pratiques. Un enseignant seul qui désire évaluer les apprentissages des élèves dans sa salle de classe peut se satisfaire de construire la carte lui-même. D'autres conditions et d'autres contextes d'évaluation nécessitent de mettre en œuvre une méthodologie plus robuste. Dans une étude au sujet de la structure des connaissances, Acton, Johnson et Goldsmith (1994) ont d'ailleurs mis en évidence que les représentations des connaissances d'experts de contenu sont idiosyncrasiques, c'est-à-dire qu'elles varient grandement d'un individu à l'autre. Pour cette raison, ils recommandent d'élaborer une structure de référence à partir des représentations de plusieurs experts en conservant dans la structure finale les éléments qui font l'unanimité au sein du groupe d'experts, ou du moins les éléments pour lesquels une majorité des experts s'entendent, et en éliminant les éléments pour lesquels les experts ne s'entendent pas.

La méthode d'élaboration d'une carte de référence à l'aide des représentations d'experts de contenu est certes répandue, mais elle a ses limites. Sur le plan méthodologique, elle dépend entre autres de l'échantillonnage des experts de contenu et de la capacité de ces derniers à s'entendre sur l'organisation des connaissances. Une méthode quelque peu différente pour construire une carte de référence consiste à l'extraire de documents textuels. À la différence des humains, les textes ont cette qualité qu'ils sont plus stables et qu'ils demeurent toujours

accessibles. Parmi les publications recensées, deux études (Abd-El-Khalick et BouJaoude, 1997; Hay, Kehoe, et coll., 2008) ont proposé des procédures pour extraire manuellement une carte conceptuelle à la suite de l'analyse de contenu de textes. Une autre démarche prometteuse consiste à extraire ces informations automatiquement à l'aide de techniques de traitement automatique du langage naturel. Certaines des méthodes répertoriées dans les 19 articles exclus à la section 3.1 s'apparentent d'ailleurs à ces approches, sauf que la plupart d'entre elles sont basées sur des corpus de données quelque peu différents (comme des données issues d'examens scolaires). Le deuxième article (voir le sixième chapitre) de cette thèse explore un peu plus ce processus d'extraction d'une carte conceptuelle, en proposant une démarche semi-automatisée et en comparant les résultats de cette démarche avec ceux d'un processus d'extraction manuelle.

3.7 Appréciation des propriétés psychométriques d'une tâche de d'évaluation

Comme il a été mentionné à la section 3.2, 44 des 150 textes scientifiques consultés lors de cette recension ont été classifiés comme ayant pour objectif d'apprécier les propriétés psychométriques des tâches de cartographie conceptuelle. Cette section va approfondir l'analyse de ces textes. Elle vise à tracer le portrait des diverses preuves de fidélité des scores et de validité qui sont collectées pour appuyer les interprétations et les utilisations des scores.

3.7.1 Quelques notions au sujet de la fidélité ou de la précision des scores

En mesure et en évaluation en éducation, la fidélité des scores renvoie au degré de confiance qu'une personne peut avoir dans la précision des scores d'un test (American Educational Research Association, American Psychology Association et National Council on Measurement in Education, 2014; Haertel, 2006; Laurier, Tousignant et Morissette, 2005). Malgré un consensus autour de sa définition, il existe dans la pratique plusieurs méthodes et statistiques pour estimer la fidélité des scores, selon le contexte et les conditions de la collecte de données. Parmi les méthodes courantes se trouvent celles qui permettent d'estimer la stabilité des résultats dans le temps, l'équivalence des tâches, la cohérence interne (Bertrand et Blais, 2004).

Dans sa définition classique, la stabilité d'un test, qui est aussi appelé la fidélité « test-retest » en anglais, correspond à la corrélation des scores calculée lorsqu'un « test est administré deux fois et que chaque répétition est considérée comme une mesure parallèle » (Bertrand et Blais, 2004, p. 49). L'équivalence de tests renvoie quant à elle à la corrélation des scores obtenue lorsque « deux formes d'un test considérées parallèles sont administrées au même groupe d'individus » (Bertrand et Blais, 2004, p. 50). Le coefficient d'équivalence porte aussi le nom de fidélité des versions parallèles d'un test (en anglais, parallel-forms reliability). Enfin, la cohérence interne est une méthode d'estimation de la fidélité qui vise à déterminer le degré de cohésion entre les items d'un instrument. Le coefficient alpha de Cronbach est probablement la méthode d'estimation la plus répandue.

Un autre volet important de la fidélité des scores dans le domaine de l'éducation concerne l'accord (ou la fidélité) entre les correcteurs. Dans certaines épreuves à enjeux élevés composées d'items à réponse construite, il est d'usage de demander à deux individus ou plus de corriger chaque réponse, afin d'assurer l'exactitude de la note qui lui est attribuée. Il existe plusieurs méthodes et statistiques pour estimer le degré d'accord entre deux ou plusieurs observateurs : parmi les plus connues se trouvent le pourcentage d'accord parfait et la statistique Kappa de Cohen (Cohen, 1960). La théorie de la généralisabilité, qui a été brièvement introduite précédemment, peut aussi être d'une grande utilité pour estimer la fidélité interjuges, surtout lorsqu'il est nécessaire de quantifier et de comparer la contribution de plusieurs facettes d'instrumentation à l'erreur de mesure (Crocker et Algina, 1986).

3.7.2 Quelques notions au sujet de la validité

Même si la fidélité et la validité sont souvent présentées de façon distincte, ces deux notions sont intimement liées entre elles. Si le concept de fidélité fait appel à la précision ou à la suffisance des résultats, la validité correspond traditionnellement à la justesse, à la signification et à l'utilité des inférences produites à partir des données collectées au moyen des instruments d'évaluation des apprentissages. Il n'est pas rare d'entendre dire que la validité renvoie au degré pour lequel « le test mesure bien ce qu'il prétend mesurer ».

Ce regard quelque peu simpliste et intuitif du concept de validité ne peut toutefois en appréhender le sens profond. Il suffit de suivre l'évolution historique de ce concept (Kane, 2001,

2006, 2013; Sireci, 2009) pour constater qu'il n'y a pas d'unanimité sur la manière de le définir ni de l'opérationnaliser. Cela dit et malgré certaines oppositions sur le plan théorique, un courant de pensée dominant s'est cristallisé au cours des dernières décennies, ce courant étant largement influencé par les textes fondateurs de Messick (1989) sur la validité et de Kane (2001, 2006, 2013) sur la validation. En particulier, Messick (1989, p. 13) a défini la validité comme étant :

un jugement intégré du degré avec lequel des preuves empiriques et des arguments théoriques soutiennent la justesse et la pertinence des interprétations et des actions basées sur le score d'un test ou d'autres instruments d'évaluation¹⁰.

Kane (2001, 2006) a approfondi cette conception en expliquant que la démarche de validation s'appuie sur des arguments interprétatifs, qui consistent à déterminer dès le départ les interprétations et les utilisations de la démarche d'évaluation, et sur des arguments de validité, qui cherchent à en évaluer la plausibilité, la cohérence et la pertinence en s'appuyant sur des éléments de preuve théoriques et empiriques. La validation est dans la circonstance une démarche permettant de former un argumentaire en vue de soutenir les interprétations, les actions et les décisions basées sur les scores de l'épreuve. Plusieurs stratégies peuvent être mises en œuvre pour construire cet argumentaire et recueillir ainsi des preuves de validité.

La plus récente version des *Standards of Educational and Psychological Testing* (AERA/APA/NCME 2014), un document élaboré conjointement par l'American Educational Research Association, l'American Psychological Association et le National Council on Measurement in Education dans le but « de promouvoir des pratiques de testing réfléchies et de fournir un socle pour apprécier la qualité de ces pratiques (p. 1) », décrit cinq sources de preuves de validité.

Les preuves fondées sur le contenu sont le résultat d'une analyse (conceptuelle ou empirique) de l'adéquation entre le contenu d'une épreuve, le domaine de connaissances évalué et le construit évalué par cette épreuve. Parfois ces preuves sont explicites et reposent sur une série d'expérimentations visant à déterminer l'adéquation entre le contenu de la tâche ou de

¹⁰ Traduction libre de : *Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.*

l'épreuve et ce qui est évalué, parfois elles sont implicites et se trouvent dans les informations fournies par les concepteurs au sujet de la démarche d'élaboration de l'instrument d'évaluation.

Les preuves fondées sur les processus de réponses cherchent à vérifier l'adéquation entre les processus de pensée suscitée lors de la tâche, la performance observée chez l'individu et le construit évalué. Ces preuves servent entre autres à établir dans quelle mesure les processus de réponses concordent avec les clés de correction, les scores attribués à la réponse et ultimement les interprétations de ces scores. Elles permettent aussi de s'assurer que le format, le support et le contenu de l'épreuve sont conformes aux attentes initiales.

Les preuves fondées sur la structure interne concernent la relation entre les variables observables d'une épreuve, par exemple les notes attribuées aux items, et les variables dépendantes qui reflètent le ou les construits ciblés. La modélisation des données à partir de modèles à variables latentes, comme l'analyse factorielle, est une approche statistique très populaire pour récolter des preuves de ce type.

Comme son nom le dit, les preuves de validité fondées sur la relation avec d'autres variables concernent la relation entre les résultats ou les scores décernés à une tâche ou à une épreuve d'évaluation et les valeurs d'autres variables. La variable externe peut être le score à une épreuve jugée similaire (mesure dite convergente), à une épreuve considérée différente (mesure dite discriminante) ou à un indicateur du niveau de performance (mesure dite « critère » ou prédictive) comme la performance dans un domaine professionnel. La variable externe peut aussi refléter un facteur exogène à la tâche d'évaluation qui devrait induire un effet sur les scores observés. Les études expérimentales ou quasi expérimentales qui tentent de vérifier à quel degré un test ou une tâche d'évaluation est sensible à l'effet de l'enseignement ou de l'apprentissage en sont de bons exemples.

Le dernier type de preuves a pour objet d'examiner les conséquences de l'utilisation du « testing » et vise à mettre en évidence les conséquences positives ou négatives, anticipées ou inattendues d'une épreuve, afin que les utilisateurs de cette épreuve puissent apprécier de manière éclairée les utilisations et les interprétations qu'ils font des scores au test. Même s'il s'avère judicieux de porter une attention particulière aux dérives potentielles de l'utilisation d'un test, il n'y a pourtant pas de consensus à savoir si ces considérations doivent faire partie de

la définition de la validité ou non (Markus et Borsboom, 2013). Il est d'autant plus difficile de prévoir les conséquences d'une épreuve lorsque cette dernière n'est pas mise en œuvre dans un contexte réel, comme dans les études recensées dans ce texte. Dans la pratique, les données récoltées pour cette source de preuves sont bien souvent collectées après quelques passations de l'épreuve.

Malgré que les sources de preuves aient été présentées séparément ci-dessus, la validation n'est ni une démarche de cumul arithmétique de preuves ni une démarche machinale. Non seulement il est essentiel de recueillir des preuves de validité suffisantes et nécessaires, mais il faut aussi porter un jugement sur l'importance à accorder à chacune d'entre elles. Il s'agit alors d'une responsabilité partagée entre les concepteurs et les utilisateurs de l'épreuve. Une épreuve peut bien être élaborée selon les règles de l'art, mais il suffit que ses utilisateurs perçoivent une seule conséquence de l'épreuve comme étant trop négative pour que celle-ci soit mise de côté, ou modifiée en profondeur. Et comme la mise en œuvre et les utilisations des résultats de l'évaluation peuvent changer dans le temps, la démarche de validation n'est jamais définitive. Il est important de renouveler périodiquement cette démarche afin de s'assurer que les interprétations des scores s'accordent toujours avec les intentions de la démarche d'évaluation.

Bref, il n'existe pas de méthodologie unique ni d'approche idéale pour « valider » les interprétations des scores. Les démarches de validation, comme les démarches d'estimation de la précision des scores, dépendent de l'instrument d'évaluation et du contexte de la recherche. Les sections suivantes tracent le portrait des preuves de fidélité des scores et de validité dans les études recensées.

3.7.3 Les preuves de fidélité dans les études

Près des deux tiers (N = 29, 66 %) des études analysées font état d'au moins une preuve de fidélité des scores. Le Tableau XII présente la répartition des types de preuves et des méthodes d'estimation qui y sont associées. Manifestement, les chercheurs ont accordé beaucoup plus d'attention à apprécier l'accord interjuges et beaucoup moins aux autres méthodes d'estimation de la fidélité des scores. La plupart des études recensées ayant mis à contribution des variantes de tâches relativement ouvertes, il n'est donc pas surprenant que

celles-ci aient vérifié l'effet potentiel des correcteurs sur la précision des scores. En revanche, le nombre restreint d'études qui ont apprécié la fidélité des scores sous l'angle de la stabilité des scores, de l'équivalence des tâches, de la cohérence interne ou de la théorie de la généralisabilité traduit certains manques et lacunes dans les écrits scientifiques. Les différentes méthodes d'estimation de la fidélité des scores n'étant pas interchangeables, il reste donc encore beaucoup de travail à accomplir pour examiner certaines des propriétés psychométriques des tâches de cartographie conceptuelle. Les prochaines sous-sections passent en revue chacun des types de preuves de fidélité des scores répertoriés.

Tableau XII. Distribution des types de preuves de fidélité et méthodes répertoriées

Type de preuves	Méthodes répertoriées	Pourcentage*
Accord interjuges	Coefficient de corrélation	43 %
	Coefficient de corrélation intraclasse	
	Coefficient de généralisabilité (modèle Individu X Correcteur)	
	Coefficient de généralisabilité (modèle Individu X Correcteur X Occasion de correction)	
	Coefficient alpha de Cronbach	
	Différentes versions du coefficient Kappa de Cohen	
	Pourcentage d'accord	
Cohérence interne	Coefficient alpha de Cronbach	11 %
Généralisabilité	Modèles de la généralisabilité variés	9 %
Stabilité	Coefficient de corrélation	7 %
Équivalence	Statistiques descriptives	5 %
Accord intrajuge	Coefficient de corrélation intraclasse	2 %
Accord interexperts	Coefficient alpha de Cronbach	2 %
Fidélité (TRI)	Pourcentage d'accord	
	Coefficient de fidélité marginale	2 %

* N = 44 études

3.7.3.1 L'accord interjuges

Malgré le bon nombre d'études qui se sont penchées sur le degré d'accord interjuges de tâches de cartographie conceptuelle, il est assez compliqué de faire la synthèse de leurs résultats parce que chacune d'entre elles a ses particularités : les statistiques d'estimation du degré d'accord interjuges ne sont pas les mêmes, les modalités de passation et de correction sont différentes, les méthodes de sélection et de formation des correcteurs ne sont pas uniformes, etc.

Cela dit, les conclusions des chercheurs sont assez similaires d'une étude à l'autre : la plupart d'entre eux rapportent des indices de fidélité de l'accord interjuges qui sont près ou qui dépassent les seuils considérés comme minimalement acceptables¹¹ et en général, les auteurs des études concluent que les correcteurs contribuent de manière négligeable à l'erreur de mesure (Cetin, Guler et Sarica, 2016; Ruiz-Primo, Schultz, et coll., 2001; Srinivasan et coll., 2008). Il n'est malheureusement pas évident de statuer sur les conditions (par ex., des modalités de formation des correcteurs) qui ont mené à ces résultats, faute d'information dans plusieurs articles. Dans quelques rares publications, les chercheurs expliquent avoir formé les correcteurs et avoir réalisé un exercice de calibrage (Borrego et coll., 2009; Harris, 2008; Millard et coll., 2006).

Deux études se démarquent pourtant des autres, en observant des degrés d'accord interjuges plutôt faibles entre les correcteurs. Dans une première étude, Kessler (1995) a examiné la corrélation entre les notes attribuées par deux correcteurs à un ensemble de cartes d'élèves, et ce pour deux tâches de cartographie conceptuelle (prétest et post-test). Les productions ont été corrigées au moyen du modèle de notation structurelle de Novak. Les niveaux de fidélité interjuges des notes sont respectivement de 0,61 et de 0,40 pour la tâche prétest et post-test, et le niveau de fidélité des notes partielles obtenues correspondant à chaque indicateur oscille entre 0,21 et 0,68. Le chercheur ne formule pas d'hypothèse ou ne donne pas de raisons pour expliquer ces résultats.

¹¹ Une valeur de coefficient de généralisabilité qui dépasse 0,80 (Brennan, 2001), une valeur de corrélation ou de corrélation intraclasse dépassant 0,70, une valeur du coefficient Kappa de Cohen ou de statistiques dérivées qui dépasse 0,60 (Landis et Koch, 1977).

Dans la deuxième étude, McClure et coll. (1999) ont comparé le degré d'accord interjuges de six modèles de notation à l'aide de la théorie de la généralisabilité. Les valeurs des coefficients de généralisabilité de chaque modèle sont présentées au Tableau XIII. Aucune d'entre elles ne dépasse le seuil minimal acceptable de 0,80 (Brennan, 2001). Les auteurs de l'étude avancent deux facteurs pour expliquer les faibles niveaux d'accord observés : premièrement, le manque de clarté et d'organisation dans les procédures de correction comme dans le modèle d'appréciation globale de la carte ; deuxièmement l'imposition d'une trop grande charge cognitive sur la mémoire de travail des correcteurs lorsque le modèle de notation est trop complexe, comme dans le modèle de notation structurelle. Selon eux, ces deux facteurs expliqueraient pourquoi le modèle de notation de la justesse des propositions a obtenu le meilleur degré de fidélité interjuges. Cela n'explique pourtant pas les niveaux faibles d'accord interjuges en général.

Tableau XIII. Coefficients de généralisabilité des six modèles de notation étudiés dans McClure et coll., 1999

No	Description	Coefficient de généralisabilité
1	Échelle d'appréciation globale de 0 à 10	0,67
2	+ carte de référence en appui	0,36
3	Notation de la justesse des propositions (échelle de notation de 0 à 3 pour chaque relation)	0,51
4	+ carte de référence en appui	0,76
5	Notation structurelle de Novak et Gowin	0,41
6	+ carte de référence en appui	0,23

Dans le même ordre d'idées, quelques études ont également comparé la fidélité interjuges entre divers modèles de notation. Les résultats de ces études sont en général unanimes. Toutes choses étant égales par ailleurs, il semble que les notes obtenues au moyen des modèles de notation de la qualité des relations sont plus précises (fidèles) que celles obtenues à partir de la notation structurelle de Novak et Gowin (McClure et coll., 1999; West, Park, Pomeroy et Sandoval, 2002; Wunnasri, Pailai, Hayashi et Hirashima, 2017) et des méthodes de notation

basée sur une échelle d'appréciation globale (Ingec, 2009; McClure et coll., 1999; Vikaros, 2017).

Deux autres études (Brandstädter et coll., 2012; Yin et coll., 2005) ont également comparé l'accord interjuges entre deux modalités de passation courantes, la tâche du type « produire une carte à partir d'une liste de concepts » et la tâche du type « produire une carte à partir d'une liste de concepts et de mots de liaison ». Malgré quelques différences du point de vue méthodologique, les deux études en sont arrivées à des résultats très similaires. Yin et coll. (2005) ont obtenu des estimations de la fidélité interjuges (méthode non précisée) de 0,81 pour la première tâche et de 0,92 pour la seconde tâche, alors que Brandstädter et coll. (2012) ont obtenu des estimations du coefficient intraclass respectivement de 0,92 et de 0,97 pour ces deux mêmes variantes. En toute logique, ces résultats semblent donc indiquer que l'ajout de restrictions lors de la production d'une carte conceptuelle, comme l'imposition d'une liste d'expressions de liaison, est un facteur qui contribue à accroître l'accord entre les correcteurs.

3.7.3.2 La cohérence interne

Alors que les méthodes basées sur la cohérence interne sont très courantes pour estimer la fidélité des scores d'épreuves traditionnelles, celles-ci sont rarement exploitées lors de l'appréciation de tâche de cartographie conceptuelle. L'application de ces méthodes semble en fait limitée à un ensemble restreint de modalités de passation et de notation, ce qui s'explique probablement du fait que plusieurs méthodes et statistiques pour estimer la cohérence interne, comme le coefficient alpha, reposent sur la notion d'items et que cette dernière n'a pas toujours d'équivalent dans le contexte de la cartographie conceptuelle.

Cela dit, certaines modalités remplissent les conditions nécessaires pour estimer la cohérence interne des scores. Parmi les textes analysés, trois études (Hatami et coll., 2016; Ruiz-Primo, Schultz, et coll., 2001; Schau et coll., 2001) ont réussi à estimer la cohérence interne de tâches du type « compléter la carte » en considérant les nœuds ou les arcs vides comme des items d'un test. Les valeurs du coefficient alpha varient entre 0,71 et 0,91 dans ces trois études. Ces résultats doivent cependant être nuancés, car ils dépendent de certaines caractéristiques des tâches. Ruiz-Primo, Schultz, et coll. (2001) ont observé, par exemple, une valeur moyenne du coefficient alpha plus élevée pour une tâche du type « compléter les arcs vides » (0,85) que pour

une tâche du type « compléter les nœuds vides » (0,71). Schau et coll. (2001) ont obtenu quant à eux une valeur beaucoup plus élevée (0,91) en combinant les données de quatre tâches du type « compléter les nœuds ». Au total, ce sont 38 concepts qui ont été retirés parmi les 105 nœuds composant les quatre tâches dans cette dernière étude. En comparaison, Ruiz-Primo, Schultz, et coll. (2001) ont retiré seulement douze mots de liaison ou douze énoncés de concepts d'une carte maîtresse composée de 38 relations et de 30 nœuds.

Sous certaines conditions, il est aussi possible d'adapter le modèle de notation d'une tâche du type « produire une carte » dans le but d'estimer la cohérence interne des scores. Cela requiert toutefois d'établir quelles observations agiront à titre d'items d'un test. Coleman (1995) et Taricani et Clariana (2006) se sont servis, par exemple, d'une carte de référence pour établir quelles paires de concepts vont représenter les items. Les modèles de notation et les résultats de ces deux études sont pourtant très différents. Dans la première étude, les correcteurs ont décerné un point chaque fois que la relation entre les deux concepts ciblés était juste et les chercheurs ont estimé la cohérence interne entre 0,31 et 0,38. Dans l'étude de Taricani et Clariana (2006), les correcteurs ont plutôt attribué un point chaque fois qu'un lien a été tracé entre les deux concepts, et ce peu importe la justesse ou la pertinence des mots de liaison. Les chercheurs ont obtenu des valeurs du coefficient alpha beaucoup plus élevées, variant entre 0,76 et 0,86, ce qui s'explique peut-être par le fait que les scores ne dépendent pas du jugement des correcteurs, comme dans la première étude.

3.7.3.3 La généralisabilité

La théorie de la généralisabilité est une approche de modélisation très avantageuse pour estimer la précision des scores, notamment parce qu'elle s'adapte à une multitude de plans de recherche (Brennan, 2001; Cardinet et coll., 2010). Elle peut servir à estimer la fidélité interjuges, comme dans certaines des études discutées ci-dessus, mais elle peut aussi être utile pour quantifier la contribution relative ou absolue de différents facteurs à l'erreur de mesure. Elle peut également se substituer aux méthodes classiques d'estimation de la stabilité des scores et d'équivalence des tâches. Quatre articles répertoriés modélisent les scores des cartes conceptuelles à l'aide de l'appareillage de cette théorie de la mesure en éducation.

Dans le premier article, Srinivasan et coll. (2008) comparent la généralisabilité des scores obtenus par quatre modèles de notation : 1) la notation de Novak et Gowin ; 2) la notation de la justesse des propositions ; 3) un modèle de notation qui tient compte à la fois de la justesse et de la pertinence (importance) des propositions ; 4) un modèle hybride qui reprend le troisième modèle et qui y inclut la notation des relations transversales. Chaque étudiant en médecine ayant participé à leur étude a réalisé deux tâches (une au sujet de l'asthme, l'autre concernant le diabète), et ce à deux occasions pour chacune d'entre elles. En analysant séparément les notes décernées aux cartes des étudiants pour chaque modèle de notation, les chercheurs en sont arrivés à la conclusion que les notes obtenues par les trois modèles basés sur la qualité des relations (0,76 – 0,78) sont nettement plus fidèles que les notes attribuées au moyen du modèle de notation de Novak et Gowin (0,24). Cela dit, il est important de constater que les correcteurs ont très peu contribué à l'erreur de mesure, et ce pour les quatre modèles. C'est surtout l'interaction entre les répondants et les occasions de mesure (50,1 % de la variance totale) et l'interaction entre les répondants, les domaines de connaissances et les occasions (25,2 %) qui ont contribué à l'erreur totale lorsque le modèle de notation de Novak et Gowin a été utilisé. Pour ce qui est des trois autres modèles, c'est avant tout la composante de variance correspondant à l'interaction triple entre les répondants, les domaines de connaissances et les occasions qui explique la variance totale (entre 20,2 % et 23,3 %). En principe, les chercheurs s'attendaient à des résultats stables entre les deux occasions de mesure parce que leurs répondants n'avaient pas accès à du nouveau contenu d'apprentissage. Les résultats de l'étude de généralisabilité laissent plutôt croire que le modèle de notation de Novak et Gowin est sensible à des facteurs exogènes et potentiellement indésirables qui affecteraient la stabilité de l'opération d'évaluation des apprentissages. Plus de travaux de recherche sont nécessaires pour identifier ces facteurs.

Dans la seconde étude, Yin et Shavelson (2008) ont étudié les propriétés psychométriques d'un modèle de notation qui consiste à noter seulement un sous-ensemble de propositions, nommé l'ensemble des propositions critiques, et ils ont mis en application ce modèle de notation pour deux variantes de tâches. Lors de leurs analyses, ils ont entre autres choses examiné à quel degré l'occasion de la mesure et l'échantillonnage des propositions critiques ont une influence sur la généralisabilité des scores. Ils ont également fait des

projections (une étude de décision) et estimé combien de propositions critiques seraient nécessaires pour atteindre un seuil de généralisabilité jugé acceptable. Leurs conclusions sont assez claires, et ce pour les deux modalités de passation : il faut plus que seize propositions critiques pour noter de manière précise les cartes produites par des élèves. Dans les faits, ils établissent qu'il aurait fallu noter au moins 18 propositions critiques dans leur tâche du type « produire une carte à partir d'une liste de concepts et de mots de liaison » et au moins 30 propositions critiques dans leur tâche du type « produire une carte à partir d'une liste de concepts » afin d'atteindre une valeur du coefficient de généralisabilité considérée acceptable, soit au-delà de 0,80.

Dans la troisième étude, Cetin et coll. (2016) ont étudié la généralisabilité des notes obtenues sur une échelle globale de 0 à 10 après avoir demandé à un groupe d'élèves de septième année de réaliser quatre tâches et à trois enseignants de corriger chaque production. Leur étude de généralisabilité montre que l'interaction entre l'étudiant et la tâche est le facteur qui contribue le plus à l'erreur de mesure (34,8 %), suivi de la facette de la tâche (13,5 %). Comparativement, les composantes de variances associées à l'univers des correcteurs sont quant à elles négligeables (moins de 4,0 %). Ces résultats ne sont pas tant surprenants, car les tâches de performance sont réputées être sensibles à l'échantillonnage des sujets de la tâche (Shavelson, Baxter et Gao, 1993).

Dans la dernière étude, Kassab et coll. (2016) ont examiné la généralisabilité de notes attribuées aux cartes conceptuelles produites par 116 étudiants en médecine, en considérant particulièrement deux facettes d'instrumentation : le domaine de connaissances et les correcteurs. À la différence des études précédentes, ils ont observé une contribution plus importante de la facette des correcteurs (10 %¹²) et de l'interaction entre les étudiants et les correcteurs (25 %) à la variance totale de scores. Les auteurs de cette recherche ont toutefois proposé un modèle de notation quelque peu différent, fondée sur l'appréciation globale de cinq

¹² Dans le texte original, les résultats sont présentés sous forme de proportions et sont arrondis à la deuxième décimale.

dimensions¹³ des cartes plutôt que sur la notation des composantes, ce qui explique peut-être pourquoi les résultats ne s'accordent pas avec ceux d'autres études qui se sont intéressées à l'accord interjuges.

3.7.3.4 La stabilité et l'équivalence

Comme il a été vu précédemment, les articles qui relatent des preuves de stabilité ou d'équivalence des notes sont rares dans les études recensées. Ces preuves de fidélité exigent en effet de faire passer la même tâche ou des tâches similaires à plusieurs occasions, ce qui n'est pas toujours opportun lorsque la recherche est réalisée en contexte naturel de salle de classe.

Trois études ont examiné la stabilité des notes. Dans une première étude, Michael (1995) a examiné la stabilité des résultats d'une tâche de type « produire une carte à partir d'une liste de concepts », et ce par trois groupes distincts. Il a observé des valeurs de corrélation élevées pour les deux groupes ayant reçu une formation à la cartographie conceptuelle (0,78 et 0,92), tandis qu'il a observé une valeur de corrélation plus modérée pour le groupe n'ayant pas reçu de formation (0,44). Les deux autres études (Ifenthaler, 2010b; Yin et coll., 2005) ont quant à elles observé des valeurs de coefficients de stabilité au-delà de 0,80, ce qui concorde avec les deux groupes de répondants qui ont reçu une formation dans l'étude précédente.

Deux autres recherches se sont penchées sur l'équivalence de modalités de passation. Dans la pratique courante en psychométrie, l'examen de l'équivalence des tâches consiste à vérifier à quel degré, deux versions d'une même épreuve, par exemple deux épreuves composées d'items à choix multiple tirés de la même banque d'items, sont équivalentes. La notion d'équivalence des tâches prend une signification quelque peu particulière dans les études concernant la cartographie conceptuelle : elle concerne l'équivalence de variantes de tâches. Dans la première étude, Ruiz-Primo, Schultz, et coll. (2001) ont montré que les notes obtenues au moyen de l'une des trois variantes de tâches étudiées, soit une tâche du type « produire une carte à partir d'une liste de concepts », une tâche du type « compléter les nœuds » et une tâche

¹³ Les dimensions sont la « validité » du choix des concepts, l'arrangement hiérarchique des concepts, l'intégration entre les concepts, l'adéquation entre le contenu de la carte et le contexte de la tâche, ainsi que le degré de créativité de l'étudiant. Elles ont été notées selon une échelle de notation (1 = pauvre à 5 = excellent).

du type « compléter les arcs vides », ne sont pas équivalentes. En effet, les valeurs de corrélation observées entre les notes attribuées à chacune de ces tâches sont modérées (entre 0,44 et 0,53), ce qui suggère que ces modalités évaluent des aspects quelque peu différents de l'apprentissage. De même, les chercheurs ont remarqué que les résultats sont, toutes choses étant égales par ailleurs, plus faibles dans la modalité plus ouverte.

En adoptant une méthodologie similaire, Yin et coll. (2005) en sont arrivés à des conclusions semblables, mais cette fois en comparant les notes entre une tâche du type « produire une carte à partir d'une liste de concepts » et celles d'une tâche du type « produire une carte conceptuelle à partir d'une liste de concepts et de mots de liaison ».

3.7.3.5 L'accord entre les experts

Comme il a été discuté tout au long de ce chapitre, mais surtout à la section 3.6.5, plusieurs modalités s'appuient sur une carte de référence pour décerner des notes aux cartes produites par des individus, cette carte de référence étant souvent la représentation d'un ou plusieurs experts de contenu. L'élaboration de la carte de référence est un travail exigeant, surtout lorsqu'il est nécessaire d'obtenir un consensus entre plusieurs experts.

Plutôt que de leur demander d'élaborer une seule carte de référence, Herl et coll. (1999) ont préféré demander à quatre experts de contenu d'élaborer individuellement leur carte et ils ont comparé les cartes des élèves sondés à chacune de ces cartes. Ils ont ainsi obtenu quatre scores de similarité pour chaque carte produite par les élèves. Puis, afin de vérifier le degré de précision des scores, ils ont analysé le degré d'accord entre chaque dyade de correcteurs, et ce pour chaque paire de concepts se trouvant dans les cartes des élèves. Ils ont obtenu des degrés d'accord entre 79 % et 90 %, ce qu'ils considèrent comme étant de très bons résultats. Comme il s'agit d'une des rares études à implanter une telle modalité de notation, ils recommandent toutefois fortement d'examiner les propriétés de cette approche dans d'autres conditions et contextes.

3.7.3.6 La fidélité marginale selon la TRI

À la section 3.7.3.2, il a été expliqué que, pour estimer la cohérence interne des scores, il est nécessaire de définir clairement ce qui compose les items d'une tâche de cartographie

conceptuelle. Les quatre études discutées dans cette section adoptent une approche relativement similaire : en utilisant une carte de référence, ils déterminent un sous-ensemble de concepts, de liens ou de relations qui agissent en qualité d'items d'un test. Dans cette approche, une partie des connaissances représentées par les répondants est tout simplement ignorée lors de la correction, surtout lorsque les répondants produisent eux-mêmes leur carte conceptuelle.

Dans un article de conférence, Liu (1994) a proposé une autre méthode pour transformer le contenu d'une carte en items d'un test. Pour y arriver, il a d'abord corrigé les productions des élèves conformément au modèle de notation de Novak et Gowin, puis il a converti les notes partielles de chaque indicateur en valeurs catégorielles échelonnées entre 0 et 9 sur la base de la distribution empirique des notes partielles. Par exemple, il a divisé le nombre de relations par quatre et arrondi cette valeur à l'entier le plus près, étant donné que le nombre maximal observé de relations a été de 39 dans les cartes collectées.

Ces transformations lui ont alors servi à modéliser les scores avec le modèle gradué pour données ordonnées de Samejima (1969) et par le fait même, à estimer la fidélité des scores dans le cadre de la théorie de la réponse à l'item. Dans son article, il emploie plus précisément deux stratégies pour estimer la fidélité des scores : d'abord en estimant la fidélité marginale, ensuite en calculant la moyenne des erreurs types obtenues pour chaque niveau d'habileté. Il a obtenu ainsi des valeurs respectivement de 0,78 et de 0,85 pour ces deux méthodes, ce qui le mène à conclure que son approche est relativement fidèle et qu'elle répond au minimum aux besoins de l'évaluation en salle de classe.

3.7.4 Les preuves de validité dans les études

L'analyse de contenu des 44 publications retenues a aussi permis de retracer les méthodes et les résultats en ce qui a trait à quatre des cinq sources de preuves de validité. Un peu comme il a été observé avec les méthodes d'estimation de la fidélité, les études ont nettement privilégié une source de preuves par rapport aux autres: une grande majorité des études (77 %) ont rapporté des preuves fondées sur les relations avec des variables externes, tandis que seulement 20 % des études ont présenté des preuves fondées sur le contenu, 9 % des preuves fondées sur le processus de réponse et 5 % des preuves fondées sur la structure interne.

Les preuves fondées sur les conséquences ont été exclues étant donné qu'elles n'ont pas véritablement été prises en considération dans les études recensées.

3.7.4.1 Les preuves fondées sur les relations avec des variables externes

En plus d'être nombreuses, les preuves de validité fondées sur les relations avec des variables externes sont diversifiées. Pour les fins de ces analyses, elles ont été réunies en trois grandes catégories et en sept types de variables externes. Le Tableau XIV présente la proportion d'études qui ont recueilli des preuves fondées sur les variables externes conformément à cette classification.

Tableau XIV. Distribution des études selon trois catégories de preuves et sept types de variables externes.

Catégorie de preuve / Type de variable externe	Méthodes répertoriées	Pourcentage
Variables liées au rendement scolaire		
Variable convergente	Analyse corrélacionnelle	47 %
	Régression linéaire	
	Modèle d'équations structurelles	
	Statistiques descriptives	
	Test du chi-carré	
Variable dite « critère »	Analyse corrélacionnelle	12 %
	Régression linéaire	
Variable divergente	Analyse corrélacionnelle	2 %
Facteur ou condition expérimental		
Différence dans le temps	Test de comparaison de moyennes	18 %
	Analyse de variance	
	Test non paramétriques de comparaison	
	Statistiques descriptives	
Différences entres des groupes de différents niveaux d'habileté	Test de comparaison de moyennes	9 %
	Analyse de variance	
Différences entre des groupes / traitement	Analyse de variance	5 %
	Test de comparaison de médianes	
Variables non liées au rendement scolaire		

Catégorie de preuve / Type de variable externe	Méthodes répertoriées	Pourcentage
Autres	Analyse corrélacionnelle Test de comparaison de moyennes Analyse de variance Régression linéaire	11 %

La première catégorie concerne les preuves qui examinent la corrélation entre les notes obtenues lors d'une tâche de cartographie conceptuelle et les notes obtenues lors d'une autre tâche d'évaluation. Plus particulièrement, cette catégorie inclut les preuves fondées sur des variables convergentes, divergentes et de critère, ces trois types de preuves se distinguant principalement selon la nature des instruments de collecte de données qui sont utilisés pour obtenir les scores représentant la ou les variables externes. Dans le cas des preuves fondées sur des variables convergentes ou divergentes, les scores sont le résultat d'épreuves de rendement scolaire (N = 18, 51 %), d'autres modèles de notation de la tâche (N = 12, 34 %) ou d'autres types de dispositifs de collecte de données pour évaluer l'organisation des connaissances (N = 5, 14 %). Dans le cas des variables dites de critère, les notes des cartes conceptuelles sont comparées aux notes finales du cours concerné (55 %) ou encore aux notes obtenues lors d'épreuves standardisées (44 %).

Peu importe l'instrument utilisé, les résultats des études qui mettent à profit ces trois types de preuves ne s'accordent pas entre eux. Dans un article paru avant même la publication de leur ouvrage, Novak et coll. (1983) font état de valeurs de corrélation d'ordre faible ou modéré ($r = -0,02$ à $0,34$) entre la performance des répondants à des tâches de cartographie conceptuelle et le rendement de ces mêmes répondants à six mesures d'apprentissage : les composantes en lecture et en mathématique du *Scholastic Aptitude Test*, les composantes verbales et quantitatives du *School and College Ability Test*, les notes de l'examen final du cours et les notes moyennes du cours. À l'opposé, Rice et coll. (1998) observent des valeurs de corrélation nettement plus élevées entre les notes attribuées aux cartes des répondants et les notes des examens de rendement scolaire (en moyenne, $r = 0,75$), ainsi qu'entre les notes des tâches de cartographie conceptuelle et les notes de différentes composantes du *Basic Skills Assessment Program* ($r = 0,82$), et du *Scholastic Aptitude Test* (en moyenne, $r = 0,87$).

Pour Novak et coll. (1983) les faibles valeurs de corrélation indiqueraient que la cartographie conceptuelle évalue une dimension de la connaissance qui est différente de celle qui est évaluée par les épreuves traditionnelles. Cette interprétation est acceptée par plusieurs et reprise dans quelques études (Dorsey, Campbell, Foster et Miles, 1999; Evans Iii, Jentsch, Hitt Ii, Bowers et Salas, 2001; Liu, 1994; Liu et Hinchey, 1996; Ruiz-Primo, Schultz, et coll., 2001), mais elle n'est pas la seule explication possible. Les faibles valeurs de corrélation peuvent aussi être le fruit d'obstacles d'ordre méthodologique. Les valeurs observées sont peut-être plus faibles que ce qu'elles devraient être réellement parce que le coefficient de corrélation n'a pas été corrigé pour tenir compte de l'erreur de mesure (Coleman, 1995). Autant les scores de la tâche de cartographie conceptuelle que ceux obtenus par l'instrument d'évaluation externe ont une erreur de mesure correspondante qui, idéalement, doit être prise en considération lors de l'estimation du coefficient de corrélation (Spearman, 1904). Ce problème d'estimation soulève également d'autres questions d'ordre méthodologique, car l'erreur de mesure cache peut-être en elle-même des problèmes liés à l'instrumentation ou à sa mise en application.

Par ailleurs, il suffit de comparer les modalités d'évaluation des études répertoriées pour comprendre qu'elles sont difficilement comparables. Dans l'étude de Novak et coll. (1983), par exemple, les répondants ont construit librement leurs cartes et celles-ci ont été notées à partir du modèle de notation structurelle. Dans l'étude de Rice et coll. (1998) en contrepartie, les cartes ont été produites à partir de listes de concepts et ont été corrigées selon un modèle de notation de la justesse des propositions. Qui plus est, dans cette seconde étude, les tâches de cartographie conceptuelle et les examens de rendement scolaire ont été élaborés par la chercheuse de manière à optimiser la représentativité des connaissances dans les deux types d'instruments.

Non seulement l'erreur de mesure associée aux notes obtenues dans la tâche de cartographie conceptuelle peut affecter l'estimation d'un coefficient de corrélation, mais l'erreur de mesure associée aux scores obtenus par les instruments externes peut également contribuer à diminuer les valeurs estimées. Les résultats des études recensées sont d'ailleurs assez vulnérables à ce phénomène, étant donné que les propriétés psychométriques des instruments ayant été utilisés pour évaluer les variables externes sont peu documentées et méconnues. La plupart des dispositifs recensés sont en effet des outils « maison » qui ont été élaborés uniquement pour les fins des études concernées et ceux-ci n'ont pas toujours été mis à l'essai

de manière rigoureuse. Dans les faits, seulement 10 % des études choisissent des instruments standardisés et seulement 55 % des études présentent au minimum quelques propriétés psychométriques des instruments utilisés, la plupart se limitant à préciser la valeur estimée du coefficient alpha. Quel que soit le type de variable considéré, le bien-fondé des preuves fondées sur les relations avec des variables externes dépend de la qualité des instruments qui sont employés pour construire ces variables.

La deuxième catégorie de preuves fondées sur les relations avec des variables externes se manifeste lorsque le ou les chercheurs manipulent des facteurs expérimentaux afin d'étudier leur effet sur les notes attribuées aux tâches de cartographie conceptuelle. Ce type de preuves s'incarne en particulier dans les devis de recherche à mesures répétées, comme les études composées d'un prétest et d'un post-test, et dont l'objectif est d'évaluer la progression des apprentissages. D'autres études répertoriées s'intéressent aussi aux différences entre des groupes d'individus, que ce soit entre des groupes qui subissent des traitements différents ou des groupes d'individus qui sont naturellement de divers niveaux d'habiletés.

Malgré la diversité des approches méthodologiques, les constats sont à peu près les mêmes entre les études et n'ont rien de surprenant: le score obtenu lors d'une tâche de cartographie conceptuelle augmente de manière significative à la suite d'une séquence d'apprentissage pertinente (Besterfield-Sacre et coll., 2004; Defranco, Jablow, Piovo et Richmond, 2015; Dorsey et coll., 1999; Herl et coll., 1999; Michael, 1995; Schau et Mattern, 1997; Schau et coll., 2001) et les individus dont le niveau de connaissances par rapport à la tâche est le plus élevé (les experts) obtiennent des résultats plus élevés que les individus dont le niveau de connaissances est initialement plus faible (Herl et coll., 1999; Markham, Mintzes et Jones, 1994).

Les résultats de deux articles viennent pourtant nuancer ces observations. Dans une étude réalisée auprès de résidents en médecine, West et coll. (2002) observent une augmentation significative des scores obtenus au moyen du modèle de notation de Novak et Gowin à la suite d'une formation de trois heures, mais ils n'observent aucune différence significative des notes attribuées selon un modèle de notation de la justesse de propositions. Ils en concluent que les modalités de notation n'ont pas le même niveau de sensibilité aux apprentissages réalisés lors de la formation. Comme le modèle de notation de Novak et Gowin est aussi composé d'un

indicateur au sujet de la justesse des propositions, il faut en comprendre que ce sont possiblement les autres indicateurs qui ont le plus contribué à l'augmentation des notes. Autrement dit, la différence entre les cartes produites avant et après la séquence d'apprentissage ne se manifesterait pas nécessairement dans la formulation des propositions, mais dans l'organisation hiérarchique des connaissances, dans la création de relations transversales et dans l'insertion d'exemples.

Dans la deuxième étude, Dowd, Duncan et Reynolds (2015) ont observé que même si, en moyenne, le nombre de concepts et de propositions augmente entre la première passation et la seconde passation, il y a une proportion non négligeable d'individus pour lesquels la carte finale contient moins d'éléments que la carte initiale. Cette réduction de la taille de la carte peut bien entendu signifier un « désapprentissage » ou une perte de motivation chez ces participants, mais elle peut aussi traduire un processus délibéré en vue de rendre la carte plus claire et précise. Si tel est le cas, rien ne garantit que le score reflète effectivement le niveau de maîtrise des connaissances de ces individus.

Enfin, la troisième et dernière catégorie de preuves fondées sur la relation avec des variables externes concerne des variables qui ne sont pas liées directement à l'apprentissage. Parmi les études recensées, une première s'est penchée sur la relation entre les notes attribuées aux cartes et le degré de familiarité des répondants avec la cartographie conceptuelle (McClure et coll., 1999) et trois autres études (DeFranco et coll., 2015; DeFranco, Jablokow, Bilen et Gordon, 2012; Jablokow et coll., 2015) se sont intéressées à la relation entre le style cognitif et le rendement de chaque individu lors de tâches de cartographie conceptuelle. Dans aucun cas, les chercheurs n'ont réussi à établir une relation évidente entre les variables étudiées.

3.7.4.2 Les preuves fondées sur le contenu

Les preuves fondées sur le contenu constituent la deuxième source de preuves de validité en importance dans les études recensées. Cette source de preuves est pourtant unique, car elle s'appuie la plupart du temps sur des informations tacites : ces dernières ne sont pas présentées formellement comme des preuves de validité dans les textes, mais elles peuvent y être associées. Cela rend d'ailleurs l'analyse du contenu des textes plus ardue.

Tout naturellement, plusieurs mécanismes pour assurer la validité du contenu sont mis en œuvre avant la passation de la tâche et sont essentiellement d'ordre procédural. Ces mécanismes peuvent autant concerner l'élaboration et la validation du contenu de la tâche (Anderson et Huang, 1989; Cetin et coll., 2016; Coleman, 1995; Rice et coll., 1998; Ruiz-Primo, Schultz, et coll., 2001; Yin et coll., 2005) que l'élaboration et la validation de la carte de référence (Hatami et coll., 2016; Jimenez Snelson, 2011; Luckie, Harrison et Ebert-May, 2011; Schau et coll., 2001; Taricani et Clariana, 2006). En guise d'exemples, Rice et coll. (1998) expliquent qu'ils ont choisi et apparié les concepts et les propositions critiques des cartes conceptuelles ainsi que les items des épreuves traditionnelles avec les objectifs définis dans un tableau de spécifications. Ruiz-Primo, Schultz, et coll. (2001) ont quant à eux demandé à des experts de contenu de vérifier que les concepts et les mots de liaison sélectionnés pour les diverses variantes de tâches concordent avec le contenu d'un ouvrage de référence. Ces mêmes experts ont aussi participé à l'élaboration de la carte de référence.

Certains mécanismes permettant d'assurer la validité du contenu peuvent également être mis en œuvre après la passation de la tâche. L'analyse des cartes produites par les répondants, et en particulier l'analyse de la fréquence des propositions s'avère une stratégie privilégiée par certains. Yin et coll. (2005), par exemple, ont remarqué que la modalité de passation et les contraintes de la tâche peuvent avoir un effet sur la représentation externe des connaissances. Leurs analyses ont en effet mis en évidence que certaines des propositions les plus fréquentes dans la tâche du type « produire une carte à partir d'une liste de concepts » n'étaient tout simplement pas admises dans la tâche du type « produire une carte à partir d'une liste de concepts et de mots de liaison » et elles ont montré que la liste de mots de liaison éveillait l'attention des élèves à l'égard de propositions auxquelles ils n'avaient pas pensé dans la modalité plus ouverte. Suivant une démarche d'analyse semblable, Ingec (2009) s'est rendu compte que certaines connaissances ciblées dans un examen d'items à choix multiple ne correspondaient à aucune proposition des cartes conceptuelles des élèves, et ce même si les deux instruments de collecte de données avaient pour objectif d'évaluer les mêmes apprentissages.

3.7.4.3 Les preuves fondées sur les processus de réponses

Quatre études parmi les 44 textes répertoriés ont été consacrées en partie ou totalement à l'étude des processus de réponses lors de la passation de tâches de cartographie conceptuelle¹⁴. Ces études se recoupent en deux volets : celles qui ont examiné la relation entre les processus de pensée et les modalités de passation; celles qui ont analysé la relation entre les processus de pensée et le niveau de performance à la tâche.

Les deux études concernant le premier volet ont mis en évidence que les individus ne sollicitent pas nécessairement les mêmes stratégies cognitives selon la modalité de passation. En particulier, l'étude de Ruiz-Primo, Shavelson, et coll. (2001) a montré que la tâche du type « produire une carte à partir d'une liste de concepts » ne cible pas les mêmes dimensions de la connaissance et ne mobilise pas les mêmes activités cognitives que les deux modalités plus restrictives, du type « compléter une carte ». Cette première modalité, plus ouverte, fait mieux ressortir les erreurs conceptuelles des individus et la maîtrise partielle des connaissances, comparativement aux deux autres modalités qui ne permettent que de vérifier si les réponses sont bonnes ou mauvaises. Cette modalité exige aussi un plus grand répertoire de stratégies de planification et de réalisation de la tâche. En contrepartie, les individus ont mis en œuvre plus de stratégies de régulation de la tâche lorsqu'ils devaient compléter les nœuds ou les arcs vides d'une carte.

L'étude de Yin et coll. (2005) a quant à elle comparé deux variantes de la tâche de production d'une carte conceptuelle, soit à partir d'une liste de concepts, soit à partir d'une liste de concepts et de mots de liaison. Comme dans l'étude de Ruiz-Primo, Shavelson, et coll. (2001), ces chercheurs ont constaté que les répondants n'expriment pas autant de connaissances partielles lorsqu'ils sont obligés d'utiliser la liste de mots de liaison. Les individus qui réalisent une carte à partir d'une liste de concepts et de mots de liaison cherchent à agencer les concepts et les relations qui vont bien ensemble, plutôt que d'exprimer toutes les idées qui leur viennent

¹⁴ Comme il sera vu dans le premier article, présenté au cinquième chapitre, d'autres études issues du domaine de l'apprentissage se sont aussi intéressées aux processus mis en œuvre lors de la réalisation d'une tâche de cartographie conceptuelle.

à l'esprit. Le processus de réponses est alors moins spontané, ce qui a pour conséquence de ralentir la construction de la carte.

Trois études, dont l'étude de Ruiz-Primo, Shavelson, et coll. (2001), ont aussi examiné les relations entre les processus de réponses et les performances des répondants. Les résultats à cet égard sont mitigés. De manière générale, il semble très difficile de faire de telles relations : le nombre de stratégies observé est trop grand par rapport au nombre de cas étudiés (Schau et coll., 2001). Cela dit, il y a quelques résultats qui méritent d'être soulignés. Ruiz-Primo, Shavelson, et coll. (2001) ont observé des démarches s'apparentant plus à des approches du type « essais et erreurs » chez les individus novices et des approches plus structurées chez les individus qui maîtrisent mieux le domaine de connaissances. En s'intéressant à l'analyse de séquences des actions des répondants, Chiu et Lin (2012) ont quant à eux mis en évidence une relation entre les performances des répondants et leur habileté à construire des relations hiérarchiques.

Les études au sujet des processus de réponses sont encore rares dans un contexte d'évaluation des apprentissages¹⁵ et elles ne se sont penchées que sur certaines modalités de passation et de notation de tâches de cartographie conceptuelle. Le premier article de cette thèse approfondira la question des processus de réponses lorsque les cartes sont construites à partir de deux supports de passation, papier et informatisé.

3.7.4.4 Les preuves fondées sur la structure interne

La dernière source de preuves discutée dans ce chapitre concerne la relation entre les variables observables et le ou les scores finaux qui représentent ce qui est évalué par l'épreuve. Cette source de preuves de validité a été traitée très peu souvent dans les études recensées et quand elle l'a été, ce n'est pas sans certaines confusions du point de vue méthodologique. En effet, les chercheurs assimilent parfois à tort les indices de cohérence interne des scores (voir la section 3.7.3.2) à des indicateurs de l'unidimensionnalité des scores et donc à preuves de validité

¹⁵ Comme il sera vu lors du premier article de la thèse, d'autres études issues du domaine de la pédagogie ou de l'apprentissage ont aussi étudié les processus de réponses des individus pendant qu'ils construisent leur carte conceptuelle.

fondées sur la structure interne (Sijtsma, 2009). En écartant ces études, il n'en reste qu'une seule, celle de Liu (1994), qui s'est penchée sur cette source de preuves de validité. Rappelons-le, dans cette étude qui a déjà été décrite à la section 3.7.3.6, Liu (1994) a modélisé les scores issus de cartes conceptuelles à l'aide d'un modèle de la théorie de la réponse à l'item, puis il a examiné quelques propriétés psychométriques de la tâche qu'il a proposée. Afin de vérifier la structure interne de scores, en particulier, il a mis en relation les notes partielles de chacun des quatre indicateurs du modèle de notation de Novak et Gowin avec deux scores finaux : les scores estimés par le modèle de la TRI (modèle 1) et les notes totales conventionnelles de ce modèle de notation (modèle 2). Les résultats sont sensiblement les mêmes pour les deux modèles : les valeurs de corrélation entre les scores finaux et les notes partielles sont statistiquement différentes de zéro pour tous les indicateurs, sauf le nombre d'exemples valides, et les scores finaux sont le plus fortement corrélés avec le nombre de relations valides ($r_1 = 0,80$; $r_2 = 0,80$), puis avec le nombre de niveaux hiérarchiques valides ($r_1 = 0,65$, $r_2 = 0,75$). Les analyses corrélationnelles indiquent toutefois des valeurs de corrélation relativement élevées entre les notes partielles, ce qui laisse entendre qu'elles ne sont pas nécessairement indépendantes. En d'autres mots, les valeurs de corrélation observées sont possiblement affectées par l'effet des autres variables considérées dans le modèle de notation. Étant donné que Liu a rendu disponibles toutes les données dans son article de conférence, il a donc été possible de refaire les analyses de corrélation entre les scores de la TRI et les notes partielles en calculant les coefficients de corrélation partielle¹⁶ plutôt que les coefficients absolus, comme l'a fait Liu (1994).

Ces analyses a posteriori fournissent un portrait quelque peu différent des relations entre les variables. En comparaison aux résultats originaux de Liu, seulement deux indicateurs contribuent de manière significative à expliquer la variance des scores totaux et les valeurs de corrélation sont plus faibles: la valeur de corrélation partielle est de 0,61 entre le score estimé et le nombre de relations valides et de 0,45 entre le score estimé et le nombre de niveaux hiérarchiques valides.

¹⁶ Il n'est pas possible d'estimer des coefficients de corrélation partielle avec les notes totales du modèle de notation structurelle parce que les variables ne sont pas linéairement indépendantes. Les analyses ont été effectuées à l'aide de la bibliothèque « ppcor » du logiciel statistique R.

Ces nouveaux résultats mettent en évidence l'importance d'employer des statistiques appropriées afin de tirer des conclusions justes et valides. Ils suscitent également un questionnement fondamental au sujet du modèle de notation structurelle et de modèles apparentés qui combinent plusieurs indicateurs : est-ce que les indicateurs servent véritablement à évaluer des aspects différents de l'apprentissage ou est-ce qu'ils évaluent grosso modo la même chose? En modélisant les scores issus de cartes conceptuelles à partir du modèle de Rasch pour données dichotomiques (Rasch, 1960) et du modèle à crédit partiel (Masters, 1982), mais surtout en examinant la dépendance locale entre les items, le troisième article présente quelques preuves empiriques pour alimenter cette réflexion.

3.8 Synthèse

Ce chapitre a présenté une recension systématique des écrits au sujet des usages de la cartographie conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages. Dans un premier temps, il a retracé les différents usages et modalités de cet outil de représentation des connaissances dans 150 publications à caractère scientifique et dans un deuxième temps, il a fait étalage des différentes preuves de fidélité et de validité recueillies à l'égard de cet outil jusqu'à maintenant. Non seulement cette recension a mis en évidence la diversité des stratégies d'évaluation en lien avec ce dispositif, mais elle a aussi souligné certaines zones d'ombre, c'est-à-dire certains thèmes ou modalités moins bien explorées et certains écarts entre les usages qui sont faits de la cartographie conceptuelle et les résultats probants de recherche. Quatre constats ressortent de cette recension.

Premièrement, il faut bien se rendre compte que les résultats des études antérieures sont difficilement généralisables et transférables. En effet, une grande majorité des études s'appuient sur de petits échantillons de répondants ($N < 100$) et sur un échantillonnage de convenance, ce qui limite forcément la généralisabilité des résultats. De plus, les modalités de passation, de correction et de notation sont tellement diversifiées qu'il est difficile d'établir quelque point de comparaison entre les études analysées. Selon Ruiz-Primo (2014), il n'est pas facile de mettre en œuvre une collecte de données qui fait intervenir la carte conceptuelle : elle nécessite du temps et des ressources, surtout lorsqu'il faut présenter pour la première fois cet outil de représentation de connaissances à des élèves. Les défis concernant la mise en œuvre de telles

démarches de recherche peuvent ainsi être l'un des facteurs concernant les petites tailles d'échantillon des études répertoriées. Cela dit, d'un point de vue scientifique, il y a là lieu à se questionner au sujet des choix méthodologiques appropriés compte tenu des ressources disponibles.

Deuxièmement, cette recension a mis en évidence des zones d'ombre, entre autres des écarts entre les études qui utilisent la carte conceptuelle comme dispositif d'évaluation des connaissances et les études qui se sont penchées sur l'appréciation des propriétés psychométriques de tâches de cartographie conceptuelle. À l'évidence, ce sont les tâches les plus ouvertes, à savoir la production d'une carte conceptuelle à partir de zéro ou à partir d'une liste de concepts, qui sont les plus utilisées dans la pratique. Les propriétés de ces variantes de tâches n'étant pas encore pleinement démystifiées, il y a encore lieu d'effectuer des études psychométriques afin de soutenir la fidélité des scores et la validité des interprétations des scores lorsque les données sont collectées au moyen de ces variantes de tâches.

Troisièmement, les études exploitant les supports de passation électroniques sont encore peu fréquentes, et ce, malgré la disponibilité grandissante des outils informatisée pour construire, stocker et échanger des cartes conceptuelles. Et les études qui se sont intéressées au support de passation comme objet de recherche sont encore plus rares. L'utilisation d'un outil informatisé soulève pourtant plusieurs questions, notamment en ce qui a trait aux démarches et stratégies préconisées par les individus lors de la construction de leur carte conceptuelle, et par le fait même, sur la formation que devraient avoir les individus.

Quatrièmement, les études concernant l'appréciation des propriétés psychométriques se sont principalement concentrées sur certaines méthodes d'estimation de la fidélité des scores et certaines sources de preuves de validité, au détriment d'autres types de preuves, ce qui ne permet pas d'avoir un portrait complet des qualités techniques de tâches de cartographie conceptuelle. Même si elle s'inscrit dans une perspective résolument pragmatique, l'approche de validation fondée sur l'argumentation ne dit pas qu'il faut préférer des types de preuves à d'autres seulement parce qu'il est plus simple de collecter les données empiriques pour appuyer ou réfuter certaines hypothèses. Il est permis de croire qu'un apport plus grand de preuves fondées sur le contenu, sur les processus de réponses et sur la structure interne pourrait apporter un

éclairage nouveau et complémentaire au sujet des tâches de cartographie conceptuelle et des modèles de notation.

Bien entendu, les résultats et les constats formulés ci-dessus sont tributaires des choix méthodologiques et conceptuels faits au cours de la recension des écrits. D'abord, la recherche documentaire et les décisions prises à l'égard de l'inclusion et de l'exclusion des textes ont pu avoir une influence sur les résultats de cette recherche. C'est dans le but de réduire le biais d'échantillonnage que la recherche dans les bases de données a été la plus ouverte possible et que les résumés de chacun des 1984 textes inventoriés ont été lus. Cela a permis d'avoir une meilleure idée du contenu des textes, avant de décider de les inclure ou non dans le corpus final. Cependant, il aurait pu être profitable de circonscrire davantage la recension des écrits afin de cibler des questions bien précises auxquelles répondre. La grande diversité des textes faisait en sorte que leur contenu ne couvrait pas toujours les dimensions étudiées dans cette recension.

L'accès payant à certaines revues a aussi empêché de consulter certaines publications potentiellement pertinentes. En effet, la lecture du résumé n'était pas toujours suffisante pour déterminer la pertinence d'un texte, et dans tel cas, il fallait consulter le texte. Or, ce n'est qu'au moment de la consultation du texte qu'il était possible d'établir si l'étude s'appuyait sur une définition de la cartographie conceptuelle ou sur des modalités d'évaluation conformes à celles consenties dans cette thèse. Cette analyse n'a pas pu être réalisée pour les textes non disponibles et pour ceux dont l'accès était payant¹⁷. Cela dit, rien ne garantit que ces textes aient été pertinents. Après tout, moins de 10 % des textes consultés ont été retenus pour les analyses plus poussées. Dans les circonstances, le nombre de textes qui auraient été pertinents parmi les publications non disponibles n'aurait probablement pas apporté grand-chose de plus aux interprétations et aux conclusions de cette recension.

Tout compte fait, et ce malgré quelques limites d'ordre méthodologique, il y a des avantages évidents d'avoir systématisé la recherche documentaire : il a ainsi été possible

¹⁷ Face au monopole de la publication scientifique par quelques grandes multinationales et aux coûts sans cesse grandissants des droits de consultation de ces publications, les bibliothèques, comme celles l'Université de Montréal, sont contraintes à revoir leurs abonnements aux revues scientifiques (Gagnon, 2017). La privatisation du savoir scientifique est un fait préoccupant auquel est regrettamment confrontée la communauté scientifique.

d'assurer l'échantillonnage des textes et de tracer un portrait plus représentatif des usages de la cartographie conceptuelle comme dispositifs de collecte de données en vue de l'évaluation des apprentissages.

3.9 Un bref rappel au sujet des problématiques ciblées dans les articles de la thèse

Même si cette recension brosse un portrait très large des études au sujet de la cartographie conceptuelle dans un contexte d'évaluation des apprentissages, cette thèse n'a pas la prétention de s'attaquer à chacun des problèmes et défis répertoriés. Au travers de trois articles scientifiques, elle vise à approfondir certaines problématiques jugées prometteuses.

Le premier article découle directement des troisième et quatrième constats discutés ci-dessus. Il y a encore à ce jour peu d'études qui se sont penchées sur la production d'une carte conceptuelle à l'ordinateur, et encore moins d'études au sujet des stratégies cognitives mobilisées dans ce contexte. Cet article vise à combler ce vide en tentant de mieux comprendre le processus de réponses d'élèves du secondaire lorsque ceux-ci produisent leur carte conceptuelle à la main ou à l'ordinateur, une meilleure compréhension de ces processus étant utile notamment pour planifier la formation préalable à la cartographie conceptuelle.

Toujours dans l'esprit d'explorer les possibles intégrations technologiques aux modalités d'évaluation à l'aide de tâches de cartographie conceptuelle, le deuxième article explore une avenue qui n'a été que survolée dans cette recension : la production semi-automatisée d'une carte conceptuelle à partir de textes. Comme il a été vu, la carte de référence est un dispositif très courant lors de la correction et de la notation de cartes conceptuelles. Cela dit, les méthodes pour créer une carte de référence ne sont pas toujours bien documentées et elles reposent largement sur la représentation que se font des experts de contenu d'un domaine de connaissances. Cette approche a ses limites et c'est pourquoi il y a un intérêt à explorer d'autres avenues pour produire une telle carte. En s'appuyant sur les techniques du traitement automatique du langage naturel, cet article met en avant-scène la dimension linguistique de la carte conceptuelle.

Enfin, le troisième article vise à combler les connaissances au sujet de la structure interne des modèles de notation de cartes conceptuelles, en modélisant les scores à l'aide d'un modèle de la famille de modèles de Rasch. Comme il sera vu dans l'article, l'apport de la modélisation est considérable parce qu'elle permet d'explorer des hypothèses (comme la dimensionnalité des données) qui ne pourraient pas être mises à l'épreuve aussi facilement avec des modèles de notation basés sur des scores observés.

En définitive, les trois articles s'inscrivent dans un processus de validation de la carte conceptuelle comme instrument de collecte des données en vue de l'évaluation des apprentissages. Dans le but de tirer des inférences valides, il faut veiller dans un premier temps à ce que le dispositif lui-même collecte des données pertinentes à l'égard de l'objet d'évaluation (représentativité du construit) et que ces données ne soient pas altérées par d'autres variables non contrôlées (effet de contamination). De là l'importance de mieux comprendre les processus de réponse et de prévoir une formation complète et adaptée à la démarche d'évaluation pour réduire ou éliminer l'effet de variables indésirables. Dans un deuxième temps, il est attendu que les notes attribuées aux cartes produites par des individus (des élèves, des étudiants, etc.) soient exemptes d'erreurs de mesure attribuables aux comportements idiosyncrasiques des correcteurs dans une telle démarche. Il y a donc là un intérêt à explorer des mécanismes de production d'une carte de référence qui permettent de représenter les connaissances qui font consensus, mais aussi de neutraliser les connaissances qui sont particulières à chaque correcteur. En troisième lieu, il est fondamental de s'interroger sur le processus de transformation des données en scores et sur la place des modèles de mesure dans ce processus. Dans une démarche exploratoire, la modélisation des scores est une approche avantageuse pour examiner à quel degré un modèle permet de se rapprocher de mesures ayant des qualités psychométriques recherchées.

Chapitre 4 Le cadre méthodologique de la collecte de données

Cette recherche est composée de trois articles scientifiques présentés distinctement dans chacun des trois chapitres qui suivent. Cela dit, le cœur de celle-ci repose sur une collecte de données réalisée en salle de classe avec trois groupes-classes d'élèves de quatrième secondaire en science et technologie. Ce chapitre décrit les grandes lignes de cette collecte de données, alors que chacun des articles donne plus de détails quant aux considérations méthodologiques propres à chacune des expérimentations.

Forcément, certaines informations sont abrégées ou sont laissées de côté dans un article de recherche afin de se conformer au format et à la nature du support de communication. Ce chapitre vise donc à donner un portrait plus complet de la collecte de données, sans toutefois répéter les informations particulières aux expérimentations. Pour cette raison, certains détails qui se trouvent généralement dans un cadre méthodologique, comme les méthodes de traitement des données et les méthodes d'analyses ne s'y trouvent pas. Ce chapitre s'en tient aux considérations méthodologiques générales de la collecte de données.

4.1 Plan et contraintes de la recherche

Comme il a été mentionné dans le chapitre d'introduction, le projet original de cette recherche s'orientait autour des objectifs du premier et du troisième article, alors que le projet du deuxième article s'est greffé seulement après la réalisation des tâches de cartographie conceptuelle. Le plan de la collecte de données avait donc essentiellement pour but de se conformer aux divers besoins et contraintes du premier et troisième article. Pour bien comprendre comment chaque article s'intègre à la recherche, il s'avère nécessaire de présenter d'abord les grandes lignes du déroulement de la collecte de données.

Les activités présentées aux participants au cours de la collecte de données peuvent être divisées en deux temps : la formation préalable à la cartographie conceptuelle et la réalisation de la tâche principale. Étant donné que la carte conceptuelle n'est pas un outil connu et utilisé par tous, la formation s'avère une étape incontournable pour que les élèves se familiarisent avec

cette technique de représentation des connaissances (Novak, 2003; Novak et Canas, 2008) et idéalement, cette formation doit inclure des activités d'entraînement afin que les apprenants maîtrisent mieux cette technique (Anderson et Huang, 1989; Bell, 2018). Non seulement la formation doit s'adapter aux contraintes de la salle de classe, mais elle doit également prendre en considération les diverses modalités de production des cartes conceptuelles.

Afin d'atteindre les objectifs spécifiques de la première étude et de la troisième étude, différentes méthodes de collecte de données ont été mises en œuvre au moment de réaliser la tâche principale de cartographie conceptuelle. Un sous-groupe de participants a effectué la tâche au cours de séances individuelles (un à un avec le chercheur) dans le but de suivre leurs démarches de productions de cartes conceptuelles. Les données collectées au cours de ces séances ont servi dans le cadre des analyses présentées dans le premier article. Le reste des participants a réalisé la tâche en salle de classe. Les cartes produites par les participants ont alors servi lors de l'examen des propriétés psychométriques de la tâche de cartographie conceptuelle (objectif du troisième article). Une description plus détaillée de chacune des méthodes et des instruments qui y sont associés est présentée dans les sections qui suivent.

Plusieurs autres contraintes découlent du plan général de la recherche et des objectifs spécifiques de chaque article. De façon générale, le plan de la collecte de données a été élaboré afin d'intégrer le mieux possible les activités de recherche dans le cours des activités pédagogiques « normales » prévues en salle de classe (plus de détails au sujet du déroulement de la collecte de données sont donnés à la section 4.5). Cette approche a comme bénéfice de perturber au minimum les activités d'enseignement et d'apprentissage, mais elle s'accompagne des contraintes supplémentaires sur le plan de la recherche. Par exemple, les thèmes sur lesquels ont porté les tâches de cartographie conceptuelle ont été choisis par l'enseignant-ressource en collaboration avec le chercheur, afin d'harmoniser les tâches de cartographie conceptuelle avec la planification de l'enseignement. Cela a fait en sorte que les trois volets de cette recherche sont dépendants de ces choix pédagogiques.

Une autre contrainte concerne la production de cartes conceptuelles de manière traditionnelle sur support papier et de manière informatisée. Même si cette contrainte découle directement du premier objectif, qui consiste à documenter les démarches de production de cartes conceptuelles dans ces deux modalités, elle a été prise en considération tout au long de la

collecte de données. Les élèves participant à la recherche ont donc été séparés entre ces deux modalités dès le début de la démarche. Idéalement, les participants auraient été répartis aléatoirement et également dans l'une ou l'autre des modalités afin de minimiser les risques d'une contamination des données par des facteurs non contrôlés (par ex., le niveau de rendement scolaire)¹⁸. Mais concrètement, ce sont les groupes-classes qui ont été partagés entre les deux modalités. Dès lors, il était attendu que tous les élèves d'une même classe suivent la formation préalable à la cartographie conceptuelle et réalisent les tâches de productions de cartes conceptuelles selon la même modalité¹⁹. Comme il est expliqué ci-dessous (à la section 4.2), les élèves de l'un des trois groupes-classes se sont vu attribuer la modalité manuscrite et les deux autres, la modalité informatisée. Donc globalement, la répartition des participants entre les deux modalités s'est retrouvée à être inégale.

De manière à atteindre le premier objectif de recherche, un sous-groupe de participants a spécialement été recruté pour réaliser la tâche principale lors de séances individuelles afin de suivre leurs démarches de productions de cartes conceptuelles. Deux méthodes de collecte de données ont été privilégiées à cette fin : la verbalisation de la pensée et l'enregistrement des actions faites lors de la tâche. L'enregistrement vidéo ou la capture d'écran se veulent des moyens de suivre objectivement les actions faites par les participants lors de la construction de leur carte, la verbalisation de la pensée se veut un moyen de récolter des informations supplémentaires en vue d'inférer les stratégies mises en œuvre par ceux-ci.

Le dernier volet de cette recherche concerne la modélisation des scores issus de cartes conceptuelles à l'aide de modèles de Rasch. Obtenir un échantillon de données de taille suffisante est une condition importante pour assurer la robustesse des estimations et pour mener à bien de telles modélisations et les analyses qui y sont associées. Comme Linacre (1994) l'explique, il est possible d'utiliser le modèle de Rasch pour données dichotomiques avec aussi peu que 30 observations. Or, plus la taille d'échantillon est petite, moins les estimations sont

¹⁸ Une répartition aléatoire des participants entre les deux modalités aurait aussi permis de réaliser certaines analyses supplémentaires lors de la modélisation des scores, comme des analyses de fonctionnement différentiel des items.

¹⁹ Dans les faits, certains élèves qui devaient réaliser les tâches de manière informatisée les ont faites de manière traditionnelle (voir section 4.2).

précises, moins elles sont robustes et moins les analyses statistiques sont puissantes. Il n'est pas rare qu'une taille d'échantillon plus grande soit recommandée, notamment pour assurer la stabilité des estimations issues de la modélisation, mais aussi celle des analyses statistiques subséquentes. Par exemple, Barrett et Kline (1981) recommandent une taille d'échantillon d'au moins 50 personnes en sciences sociales lors de l'exécution d'analyses en composantes principales. Cette technique est souvent utilisée pour apprécier la dimensionnalité des données lors de la modélisation des scores à partir d'un modèle de mesure de la famille de modèle de Rasch. Pour des données à plusieurs catégories de réponses, Linacre (1994) souligne qu'aussi peu de 50 observations peuvent suffire. La taille d'échantillon qui convient pour de tels modèles dépend toutefois de la répartition des observations dans chaque catégorie de scores (Linacre, 2004).

Le plan de la collecte de données de cette recherche a donc été élaboré en tenant compte des nombreuses contraintes énoncées ci-dessus. À la lumière de ces contraintes, les prochaines sections décrivent les considérations méthodologiques relatives aux participants, à l'instrumentation et à la procédure générale de la collecte de données.

4.2 Les participants

Les participants ciblés pour la collecte de données devaient être des enseignants et des élèves au secondaire en science et technologie et compte tenu de la nature exploratoire de la recherche, il n'était pas nécessaire de sélectionner les participants sur la base de caractéristiques démographiques, sociales ou scolaires. Cette recherche s'est donc appuyée sur un échantillon de convenance. Cela dit, la collecte de données devait se dérouler préférentiellement dans un ou des établissements possédant suffisamment d'ordinateurs pour réaliser des tâches de cartographie conceptuelle sous format électronique (afin de répondre à l'objectif du premier article).

La démarche de sollicitation des participants a débuté en contactant un enseignant en science et technologie en quatrième secondaire d'une école secondaire du réseau privé de la région de Montréal. Lors d'un stage en enseignement du chercheur principal, cet enseignant-ressource avait manifesté son intérêt à participer à un ou des projets de recherche. Cette personne enseigne depuis de nombreuses années à des élèves inscrits dans un programme axé sur la

technologie de l'information et de la communication et qui ont en leur possession un ordinateur portable en salle de classe. Cet enseignant-ressource, ayant confirmé qu'il souhaitait participer à la recherche qui lui était proposée, a alors informé la direction de l'établissement de la nature du projet en leur acheminant notamment les formulaires de consentement à l'attention des élèves, des enseignants ainsi que la lettre d'information destinée aux parents. À la suite d'une rencontre avec l'enseignant-ressource et le chercheur principal, la direction a accepté le projet de recherche. L'enseignant-ressource a ensuite sondé ses trois collègues en enseignement des sciences et technologies de quatrième secondaire pour connaître leur intérêt à participer à l'étude. Ces derniers ont décliné l'offre, notamment parce que les activités proposées ne coïncidaient pas avec leur planification de l'enseignement²⁰. L'enseignant-ressource a rempli le formulaire de consentement prévu à l'attention des enseignants (voir Annexe A).

Ainsi, ce sont les élèves des trois groupes-classes de l'enseignant-ressource qui ont été sollicités pour cette recherche. Le déroulement de la recherche a été présenté à chaque groupe-classe, puis les élèves ont reçu un formulaire de consentement²¹ (voir Annexe B). Dans ce formulaire, ils étaient d'abord invités à indiquer s'ils souhaitaient participer ou non à la recherche. Puis, ils devaient mentionner s'ils avaient un intérêt à participer aux séances de verbalisation de la pensée.

Au total, ce sont 100 élèves, soit respectivement 32, 31, 37 élèves des groupes A, B, C qui ont consenti à participer à la recherche. Huit d'entre eux ont également manifesté un intérêt à réaliser les séances de verbalisation de la pensée. Tous ces élèves étaient présents lors de la première séance, dédiée à la formation à la cartographie conceptuelle.

Comme l'indique le Tableau XV, ce ne sont pas tous ces participants qui ont réalisé les tâches prévues dans la collecte de données : 87 élèves ont réalisé la tâche d'entraînement et 85 élèves ont fait la tâche principale. En plus des absences imprévues au cours, plusieurs élèves du

²⁰ Les quatre enseignants ne suivant pas la même planification, il aurait été difficile de trouver des éléments de contenu d'apprentissage qui n'avaient pas été couverts par tous.

²¹ Étant âgés d'au moins quatorze ans, les élèves recrutés pour les fins de la recherche pouvaient donner eux-mêmes leur consentement. Cela dit, une lettre d'information a été acheminée aux parents afin qu'ils puissent prendre connaissance du projet de recherche (voir Annexe C).

groupe B ont été absents lors de la seconde tâche en raison d'une activité extrascolaire. Le Tableau XV décortique également la distribution des participants selon les groupes-classes, les supports de passation et les tâches. Selon le plan original, les élèves des groupes A et B avaient été choisis pour réaliser les tâches à l'ordinateur, et les élèves du groupe C sur support papier, mais comme certains élèves des deux premiers groupes n'avaient pas leur ordinateur portable en salle de classe au moment de la passation de la tâche, ces derniers ont effectué la tâche de façon manuscrite. Cela explique que certains participants des groupes A et B ont effectué la tâche sur support papier.

Tableau XV. Répartition des participants entre la modalité traditionnelle et la modalité informatisée lors des deux tâches de cartographie conceptuelle

Groupe	Modalité		Total
	Traditionnelle	Informatisée	
Tâche d'entraînement			
A	3	23	26
B	0	28	28
C	33	0	33
Total	36	51	87
Tâche principale			
A	4	24	28
B	2	19	21
C	36	0	36
Total	42	43	85

Parmi les cent élèves qui ont consenti à participer à la recherche, huit participants ont manifesté leur intérêt pour réaliser la tâche principale en verbalisant leur pensée à haute voix. Tous ces élèves ont été retenus pour ce volet particulier de la recherche. Or, compte tenu de la répartition inégale de l'ensemble des participants entre les deux modalités de passation, six d'entre eux ont effectué les activités préparatoires (formation et tâche d'entraînement) avec le support informatisé et deux d'entre eux, avec le support papier. Afin d'avoir le même nombre de cas pour chaque modalité, deux participants qui ont produit leur première carte en format informatisé ont été invités à réaliser la tâche principale de manière traditionnelle sur support papier.

4.3 Le matériel informatique

Étant donné que l'école où s'est déroulée la recherche possède un volet technologique, plusieurs salles de classe sont aménagées spécialement afin de faciliter l'utilisation des ordinateurs portables et un service de soutien technique est offert directement à l'établissement. Ce service veille à l'entretien du matériel informatique ainsi qu'à la mise à jour des logiciels qui sont installés dans leurs ordinateurs.

Dans cette école, il est attendu que les élèves inscrits dans ce volet technologique utilisent quotidiennement leur ordinateur portable et conséquemment, qu'ils développent leurs habiletés technologiques au cours des activités pédagogiques proposées par les enseignants. Les élèves qui ne sont pas inscrits dans ce volet technologique suivent des cours et des activités dédiés au développement de ces habiletés. Dans les circonstances, il est attendu que le nouveau d'habileté informatique ne soit pas un frein à la collecte de données. Toutefois, aucune mesure de contrôle des habiletés informatiques n'a été effectuée.

Avant d'entreprendre la recherche, les techniciens en informatique ont installé le logiciel de cartographie conceptuelle CMapTools sur l'ensemble des ordinateurs des élèves. Ce logiciel possède plusieurs avantages pour cette recherche. Son accès est libre et gratuit, en plus d'être disponible pour divers systèmes d'exploitation. Il s'offre aussi en français. Par ailleurs, c'est l'un des rares logiciels qui est mis à jour fréquemment, ce qui réduit considérablement les problèmes de compatibilité entre le logiciel et le système d'exploitation des ordinateurs. D'un point de vue pédagogique, CMapTools permet de créer une variété de tâches et de les gérer relativement efficacement. Ses modalités de correction, quoique limitées, sont pourtant suffisantes pour extraire les listes de propositions produites par les élèves et pour créer un inventaire de propositions.

4.4 L'instrumentation

Deux dispositifs de collecte de données ont été exploités dans cette recherche : les tâches de cartographie conceptuelle et la verbalisation de la pensée. Cette section décrira brièvement ces outils tout en explicitant les liens entre les données qui ont été collectées et chacun des articles qui suivront.

4.4.1 Les conditions d'élaboration des tâches de cartographie conceptuelle

Les deux tâches proposées dans cette recherche, à savoir la tâche d'entraînement et la tâche principale, ont été élaborées de manière collaborative entre le chercheur et l'enseignant-ressource, de sorte à les harmoniser le mieux possible avec le reste des activités pédagogiques prévues par ce dernier. D'un point de vue méthodologique, cette démarche avait aussi pour but d'assurer un certain contrôle sur la validité de contenu des tâches.

Au début de la démarche, l'enseignant a ciblé trois thématiques parmi les quatre grandes problématiques du programme de science et technologie de quatrième secondaire qui pouvaient être traitées lors des tâches de cartographie conceptuelle : l'effet de serre, la circulation océanique et les relations trophiques. Comme dans plusieurs autres études (voir la recension des écrits au troisième chapitre), c'est la modalité de type « produire une carte à partir d'une liste de concepts » qui a été choisie dans le cadre de cette recherche. Cette modalité a l'avantage d'assurer une certaine uniformité dans les productions sans pour autant restreindre la formulation des relations entre les concepts, ce qui est considéré comme un aspect important de l'extériorisation des connaissances. Les concepts des tâches ont été choisis de façon collaborative entre l'enseignant et le chercheur en prenant appui sur le contenu du manuel scolaire, *Observatoire* (Cyr, Verreault et Forget, 2008). Le chercheur a ensuite produit une fiche de consignes aux élèves et cette fiche a été révisée par l'enseignant en guise de vérification finale du contenu de la tâche²².

4.4.2 La tâche d'entraînement

La première tâche proposée aux élèves a porté sur l'effet de serre (voir le formulaire de consignes en Annexe D). L'effet de serre est un phénomène complexe dans lequel interviennent d'autres phénomènes tels que le rayonnement solaire, les échanges thermiques et l'atmosphère terrestre qui sont eux-mêmes réputés difficiles à cerner parce qu'ils font appel à des représentations abstraites de phénomènes physiques et des relations entre ces phénomènes. Dans

²² Les tâches ont été élaborées au fur et à mesure de la recherche et les thèmes ont été choisis afin de s'accommoder à la planification de l'enseignement. Compte tenu de contraintes de nature pédagogique et logistique, les tâches proposées aux élèves n'ont pas été mises à l'essai avant la collecte de données.

cette tâche, il a été demandé aux élèves d'« Expliquer le mécanisme de l'effet de serre », et pour y arriver, onze concepts ont été retenus par l'enseignant et le chercheur : chaleur du Soleil, couche d'ozone, dioxyde de carbone, énergie thermique de la Terre, gaz à effet de serre, lumière du Soleil, processus naturel, rayons infrarouges, rayons ultraviolets, surface de la Terre, température de la Terre.

4.4.3 La tâche principale

La thématique de la seconde tâche a été choisie de façon à se démarquer de la première activité. Alors que cette dernière obligeait les élèves à rendre compte de leur compréhension conceptuelle, la seconde tâche a été conçue de façon que l'élève objectivise ses connaissances déclaratives. La thématique des relations trophiques se conformait bien à cet objectif pédagogique et la tâche a été élaborée en suivant la même approche que pour la tâche d'entraînement. Les élèves ont été invités à construire une carte au sujet des « Écosystèmes et des relations trophiques » en utilisant les dix-sept concepts suivants²³ : animaux, autotrophes, cadavres et déchets d'autres organismes, chaînes alimentaires, champignons et bactéries, consommateurs, consommateurs primaires, consommateurs secondaires, consommateurs tertiaires, décomposeurs, niveaux trophiques, photosynthèse, producteurs, relations trophiques, réseaux trophiques, végétaux (voir Annexe E).

En rétrospective, les deux tâches ont été élaborées de telle sorte qu'elles couvrent différents contenus et types d'apprentissage. Pour reprendre les définitions proposées par Safayeni, Derbentseva et Cañas (2005) et présentées dans le cadre conceptuel (voir le deuxième chapitre), la première tâche concerne des relations de type dynamique, c'est-à-dire des relations de cause à effet ou des relations corrélationnelles, entre des concepts et la seconde, des relations de type statique entre des concepts. Cette différence quant au type de contenu entre les deux tâches est accidentelle. Elle ne découle pas de considérations d'ordre méthodologique ; elle est le résultat du fait d'avoir laissé l'enseignant-ressource choisir les thèmes des tâches selon sa planification de l'enseignement.

²³ La majorité des concepts ont été rédigés au pluriel puisque les notions d'écosystèmes et de relations trophiques font référence aux interactions entre des groupes d'individus, et non aux individus pris séparément.

4.4.4 La verbalisation de la pensée et l'enregistrement des actions

La méthode de verbalisation de la pensée, aussi appelée la méthode de type « penser à haute voix », est le deuxième dispositif de collecte de données mis en œuvre dans cette étude. En recherche, ce dispositif s'avère particulièrement utile pour accéder à ce que pense un individu pendant qu'il effectue une tâche (Ericsson et Simon, 1993) et il a déjà été utilisé avec succès dans quelques recherches concernant la cartographie conceptuelle (Dogusoy-Taylan et Cagiltay, 2014; Gurlitt et Renkl, 2008; Hilbert et Renkl, 2008; Ruiz-Primo, Shavelson, et coll., 2001).

Selon plusieurs chercheurs, il serait en effet assez facile pour des répondants de penser à haute voix lorsque ceux-ci ont eu la chance de s'entraîner à objectiver leur pensée et lorsque le climat est propice à la réalisation de la tâche (Ericsson et Simon, 1993; Van Someren, Barnard et Sandberg, 1994). Quand tout se déroule bien, le rôle du chercheur est d'ailleurs minimal pendant la réalisation de la tâche ; il a seulement la responsabilité de rappeler à l'individu de penser à voix haute s'il constate que le participant oublie de verbaliser ses actions.

Pour s'assurer de garder le plus de traces possible dans la présente recherche, les actions faites par les participants ont aussi été enregistrées au cours des séances de verbalisation. L'enregistrement des actions et des paroles a été fait au moyen d'une caméra vidéo lorsque la tâche a été effectuée sur support papier, alors que les données ont été enregistrées par l'intermédiaire d'un logiciel spécialisé, CamStudio²⁴ lorsque la tâche a été réalisée à l'ordinateur. Les données de ces enregistrements ont ensuite été transcrites sous un même format (document Word) avant d'être traitées et analysées. Le premier article de cette thèse fournit plus de détails au sujet des méthodes exploitées pour le traitement et les analyses des données issues des huit séances de verbalisation de la pensée.

4.5 La procédure de collecte de données

Dans le but de ne pas nuire au bon déroulement des activités d'apprentissage et d'enseignement prévues par l'enseignant, la collecte de données s'est étalée sur quelques séances non consécutives de cours et les tâches ont été élaborées de manière à se conformer aux

²⁴ Voir <http://camstudio.org/>

activités régulières de la salle de classe. Plus précisément, la démarche de collecte de données s'est scindée en trois séances consacrées respectivement à la formation des élèves (environ 25 minutes), à la passation de la tâche d'entraînement (environ 40 minutes) et à la réalisation de la tâche principale (environ 40 minutes). Au début de la deuxième séance et de la troisième séance, le chercheur principal a aussi pris quelques minutes (environ 10 minutes) pour faire un retour sur l'activité précédente et pour présenter les consignes relatives à la tâche concernée.

La passation de la tâche d'entraînement et la passation de la tâche principale ont été précédées d'un enseignement disciplinaire, exécuté par l'enseignant responsable des élèves. Cet enseignement s'est appuyé sur une variété de dispositifs pédagogiques (exposé magistral, travaux individuels, retour en groupe, etc.). Dès lors, il était attendu que les participants maîtrisent suffisamment les notions ciblées dans les tâches de production de cartes conceptuelles. Or, au cours des tâches, certains élèves ont mentionné qu'ils ne maîtrisaient pas encore tout à fait les concepts²⁵.

4.5.1 La formation à la cartographie conceptuelle

Comme il a été question lors de la recension des écrits (au troisième chapitre), les modalités de formation sont toutefois nombreuses et variées et les chercheurs ne s'entendent pas à propos de l'approche à adopter. La démarche de formation suivie dans cette recherche est une adaptation du protocole de formation proposé par Ruiz-Primo et ses collaborateurs (2001). Cette approche a le mérite d'avoir été mise à l'essai dans un contexte similaire à la présente collecte de données et, selon le constat de ces chercheurs, elle est suffisante pour amener les élèves à construire convenablement une carte, c'est-à-dire à produire une carte qui respecte les consignes de la tâche. Cette démarche de formation est divisée en quatre étapes.

La première étape a pour objectif d'introduire quelques notions importantes, comme les usages et bénéfices de la cartographie conceptuelle et les principales composantes d'une carte. Dans le cadre de cette recherche, cette introduction a été faite sous forme d'un échange dynamique, en s'appuyant sur une carte tirée de Morin et Paquette (2014) et d'un support de présentation PowerPoint (voir Annexe F). À cette étape, il a été demandé aux élèves s'ils avaient

²⁵ Un élève a même mentionné que ce n'était qu'à l'approche d'un examen qu'il consolidait ses apprentissages.

déjà travaillé avec des organisateurs graphiques des connaissances et plusieurs ont affirmé avoir construit des schémas semblables dans d'autres cours.

La deuxième étape vise à présenter la démarche de production d'une carte conceptuelle. Le chercheur a expliqué aux élèves comment s'y prendre pour identifier les relations entre des concepts, pour former des propositions, pour reconnaître de bonnes cartes et, pour la modalité papier-crayon, comment retracer une carte du brouillon vers la version finale. Le chercheur a également insisté sur le respect de quatre règles concernant la production d'une carte : 1) le sens des mots de relation doit être le plus précis possible ; 2) les relations doivent absolument comporter un mot lien ; 3) chaque concept doit être utilisé une seule fois ; 4) les éléments de la carte doivent être organisés clairement. À la suite de cette courte présentation, les élèves ont eu la chance de se familiariser avec la cartographie conceptuelle en effectuant une courte tâche de production d'une carte à partir de cinq concepts dans le but d'expliquer le fonctionnement des barrages électriques. Ces concepts avaient déjà été vus en classe et ils étaient considérés comme étant maîtrisés par les élèves. Cette activité a été réalisée sur papier et en équipes de deux ou trois élèves afin que ceux-ci puissent échanger lors de la réalisation de la tâche. À la fin de l'activité, un élève a été invité à tracer sa carte au tableau de façon que ces collègues de classe, l'enseignant et le chercheur puissent lui donner une rétroaction et échanger au sujet du processus de cartographie conceptuelle.

La troisième étape de la formation consiste à mettre en application les consignes et principes vus précédemment lors d'une tâche d'entraînement individuelle. À la différence de Ruiz-Primo et ses collaborateurs (2001) qui ont effectué leur tâche d'entraînement lors de la période de cinquante minutes prévues pour les quatre phases de la formation, une période entière de classe de 50 minutes a été dédiée à la réalisation de la tâche sur l'effet de serre. Au début de la période, l'enseignant et le chercheur ont présenté les consignes de la tâche aux élèves puis leur ont demandé de produire individuellement leur carte. Pendant la tâche, les élèves pouvaient poser des questions à l'enseignant ou au chercheur, mais ces derniers répondaient uniquement aux questions concernant la construction d'un schéma et s'abstenaient de donner des pistes à l'égard des notions scientifiques traitées par la tâche. Enfin, les élèves devaient remettre leur production une fois qu'elle était terminée.

Les consignes et le déroulement de la tâche ont varié quelque peu selon la modalité de production de la carte conceptuelle. Les élèves qui ont accompli la tâche de manière traditionnelle sur support papier ont reçu une feuille décrivant les consignes et la liste de concepts préétablis, des étiquettes amovibles (couramment appelées des Post-its) et une feuille de grand format permettant de construire leur carte. Il a été conseillé aux élèves d'écrire les concepts sur les étiquettes afin de faciliter leur déplacement lors de la phase d'élaboration de la carte, mais il leur a été demandé de retracer directement sur la page le contenu des étiquettes amovibles afin de ne pas perdre les traces de leur travail. Les autres élèves ont travaillé individuellement à l'aide de leur ordinateur portable sur lequel le logiciel Cmap Tools était déjà installé. Lors de la présentation des consignes, plusieurs élèves ont mentionné avoir déjà utilisé Cmap Tools ou d'autres logiciels de conception de cartes conceptuelles. Cela dit, le chercheur a fait une courte démonstration des fonctions de base de ce logiciel, c'est-à-dire les fonctions nécessaires pour produire une carte selon les attentes de cette recherche. Ces fonctions incluent la création, la modification et la suppression d'encadrés représentant des concepts, de liens et de mots de liaison, ainsi que le déplacement d'objets et la sauvegarde du fichier. Selon les observations directes faites pendant la réalisation de la tâche, les élèves n'ont pas éprouvé de difficulté particulière avec le logiciel. L'enseignant leur a transmis la tâche par l'intermédiaire d'un portail informatique interne, et en retour, les élèves devaient déposer leur production finale dans un répertoire prévu à cet effet.

La quatrième étape du protocole de formation de Ruiz-Primo et coll. (2001) consiste en un retour sur la tâche d'entraînement. Dans les jours qui ont suivi la passation de la tâche, les cartes produites par les élèves ont été examinées et annotées afin de vérifier la conformité des productions avec les consignes établies précédemment. De manière générale, il a été observé que les cartes respectaient les consignes données lors de la première séance, sauf que certains participants ont employé à quelques reprises des conjonctions de coordination (mais, et, ou) en guise de mots de relations. Les données d'un participant n'ont aussi pas été traitées dans les analyses de l'article sur la notation des cartes conceptuelles parce qu'il n'avait pas apposé d'expressions de liaison sur les relations entre les concepts. Après coup, il a été constaté que ce participant n'avait pas complété les activités de formation. À la suite de ces observations et avant la passation de la tâche principale, il a alors été expliqué aux élèves qu'il est préférable de

préciser le sens des expressions de liaison et que ces expressions prennent souvent la forme d'un verbe.

4.5.2 La passation de la tâche principale

La réalisation de la seconde carte conceptuelle a été exécutée dans des conditions similaires à celles de la tâche d'entraînement. Les premières minutes de la période en classe ont permis de donner une rétroaction verbale sur les productions de la première tâche. Tous les élèves présents, sauf les participants aux séances de verbalisation de la pensée, ont procédé à l'élaboration de la carte en salle de classe après un court rappel des consignes. Pendant ce temps, les huit élèves exclus ont été invités à effectuer du travail personnel ou ils ont fait la même tâche tout en verbalisant leur pensée à haute voix.

4.5.3 Le protocole de verbalisation de la pensée

Les séances de verbalisation de la pensée se sont déroulées à divers moments, à la convenance des participants, étant donné que le chercheur ne pouvait accueillir qu'un seul élève à la fois. Deux rencontres ont eu lieu à l'heure du dîner, au même moment que la période de récupération de l'enseignant, et les six autres rencontres se sont déroulées pendant le temps de classe, soit pendant la période allouée à la passation de la tâche principale soit lors de la période de classe suivant la passation de la tâche en grand groupe.

Tel que mentionné à la section 4.2, les huit participants ont été séparés également entre les deux modalités de passation. Ainsi, deux élèves qui ont effectué les activités préparatoires de manière informatisée, dont la tâche d'entraînement, ont accepté de réaliser la tâche principale sur support papier. Le chercheur leur a donné les mêmes consignes relatives à la modalité manuscrite que celles qui ont été données aux autres élèves qui ont effectué leurs tâches de cartographie conceptuelle selon cette modalité.

Les séances ont été réalisées dans un local distinct et exempt de sources externes de distraction. Pour les tâches sur support papier, la salle a été aménagée de telle façon qu'une caméra vidéo puisse capturer la démarche de construction de la carte conceptuelle et puisse enregistrer la verbalisation de la pensée. La passation informatisée a quant à elle été réalisée sur

un ordinateur portable fourni par le chercheur sur lequel étaient installés le logiciel CMap Tools et le logiciel de capture d'écran CamStudio.

Les séances se sont déroulées en suivant les recommandations décrites à la section 4.4.4 (voir aussi le premier article pour des informations complémentaires). Le protocole de verbalisation de la pensée a été mis à l'essai auprès d'une personne (un adulte) ne participant pas directement à cette recherche afin de s'assurer du bon déroulement de la collecte de données. La démarche de verbalisation de la pensée a été expliquée aux participants puis ces derniers ont eu la chance de s'exercer lors de la passation d'une courte activité d'entraînement à la verbalisation de la pensée (voir la fiche aide-mémoire du chercheur en Annexe G). Les élèves ont ensuite entrepris la tâche. Conformément aux bonnes pratiques de cette méthode, l'intervention de l'intervieweur a été minimale au cours des séances : il a seulement été question de rappeler aux participants de penser à haute voix lorsqu'ils omettaient de s'exprimer durant une période dépassant 15 secondes. À la fin de la séance, un court retour sur la tâche a été fait avec l'élève pour découvrir des améliorations potentielles à sa carte.

4.5.4 Les procédures éthiques

Cette recherche s'est conformée aux procédures et aux normes éthiques en vigueur à l'Université de Montréal (voir le certificat d'approbation éthique en Annexe K). Comme il est prévu dans ces normes, les élèves ont pu consentir par eux-mêmes à leur participation à la recherche étant donné qu'ils étaient âgés d'au moins 14 ans et que les risques de la recherche étaient faibles. Aucune compensation n'était prévue pour leur participation à l'étude. Le formulaire de consentement a été lu attentivement avec les élèves et ceux-ci pouvaient prendre tout le temps nécessaire pour décider d'y participer. La plupart des élèves ont néanmoins choisi de signer le formulaire sur-le-champ. Un seul élève a refusé de participer aux activités prévues dans la recherche. Comme il en avait été convenu préalablement dans le cadre de la recherche, des activités d'apprentissage comparables ont été planifiées par l'enseignant et lui ont été présentées pendant les activités de la recherche. Une lettre à l'attention des parents a aussi été distribuée afin d'informer les parents du déroulement de la recherche (voir Annexe C).

Les cartes produites par les élèves (version papier ou électronique) et tous les autres documents pertinents ont été ramassés et ils sont conservés dans un endroit sécuritaire. Les

données ont été anonymisées lors du traitement des données dans le but de ne pas pouvoir identifier les participants. Seul le chercheur principal a accès aux informations nominatives et comme convenu, aucune de ces informations n'a et ne sera transmise à d'autres personnes. Aussi, étant donné la nature exploratoire de la recherche, ni l'enseignant ni les élèves n'ont reçu les résultats individuels, c'est-à-dire les notes décernées, aux différentes tâches administrées durant la collecte de données.

4.6 Quelques précisions au sujet de la collecte de données de la deuxième étude

L'étude présentée dans le deuxième article, concernant la production semi-automatisée d'une carte conceptuelle à partir de textes, a vu le jour à la suite de la collecte de données principale, de telle sorte qu'elle n'a pas influencé directement les considérations méthodologiques discutées ci-dessus. Au contraire, ce sont les choix et contraintes de la collecte de données principale qui ont exercé une influence sur les considérations méthodologiques de l'étude décrite dans le second article. Cette section présente de manière très sommaire le fonctionnement du programme informatique conçu dans cette étude, de manière à la situer par rapport à la collecte de données principale. Le cadre méthodologique de la deuxième étude est décrit en détail dans le deuxième article.

La deuxième étude présente le fonctionnement d'un programme informatique qui vise à extraire des propositions d'un corpus de textes en vue de constituer une carte conceptuelle pouvant servir de carte de référence. Le programme exige différentes sources de données : certaines sont obligatoires, d'autres sont facultatives selon les configurations choisies. La liste des énoncés de concepts est la seule source de données qui est obligatoire dans toutes les configurations du programme. Le logiciel se sert de cette liste pour rechercher des relations entre les énoncés de concepts dans le corpus de textes analysé. En option, il est aussi possible de fournir une liste de synonymes pour chacun des concepts, de manière à améliorer le fonctionnement du programme lors de la détection de concepts dans les textes.

Le programme offre deux options pour constituer le corpus de textes : l'utilisateur peut fournir lui-même des textes en format « .txt » ou le programme peut effectuer une recherche de

textes dans l'encyclopédie libre Wikipédia en se servant de la liste initiale de concepts (et de leurs synonymes). L'encyclopédie Wikipédia a été choisie parce qu'elle est l'une des ressources libres qui renferme le plus d'information, et cette information est jugée de bonne qualité dans plusieurs domaines (Okoli, 2009; Reavley et coll., 2012; Temple et Fraser, 2014). Ces deux options ne sont pas mutuellement exclusives, alors le programme peut analyser simultanément des textes provenant de ces deux sources des données. La liste de concepts et de synonymes ainsi que le corpus de textes constituent les données nécessaires à l'exécution du programme. Ce dernier extrait une liste de propositions qui sert ensuite à former une carte conceptuelle.

La démarche de production des cartes conceptuelles, qui est décrite succinctement ci-dessus et de manière approfondie lors du second article, a été mise à l'essai à deux reprises, soit une fois pour chacune des tâches de cartographie conceptuelle proposées lors de la démarche de collecte de données principale de cette recherche. Les listes de concepts de chacune des deux tâches de cartographie conceptuelle, soit la tâche d'entraînement sur l'effet de serre, et la tâche principale sur les relations trophiques ont été reprises intégralement aux fins de cette mise à l'essai. Des synonymes ont été ajoutés à cette liste par le chercheur principal. Le corpus de textes a été constitué à partir de textes issus de trois manuels scolaires en science et technologie (les trois seuls manuels recensés au moment de l'étude), ainsi qu'à partir des textes repérés lors de la recherche d'articles Wikipédia.

Enfin, dans le but d'apprécier le rendement du programme informatique, une carte conceptuelle de référence a été constituée pour chaque thème ciblé lors de ces mises à l'essai. Ces cartes ont été élaborées par le chercheur principal après avoir extrait systématiquement toutes les relations entre les concepts dans les textes issus des manuels scolaires. Ces cartes ont ensuite servi de mesure étalon avec lesquelles les cartes produites par le système ont pu être comparées. La carte portant sur les relations trophiques produite lors de cette analyse de contenu a également servi lors de la correction des cartes produites par les élèves lors de la tâche principale, en vue de réaliser les modélisations et les analyses présentées dans le troisième article.

Même si le programme informatique a été conçu de manière tout à fait indépendante de la collecte de données principale, les mises à l'essai ont porté sur les deux tâches de cartographie conceptuelle proposées lors de cette collecte. Il faut le rappeler, l'idée de concevoir un tel

programme a pris naissance pour répondre à un problème bien concret, soit celui de produire une carte de référence en vue de la correction des cartes produites par les élèves. Beaucoup plus de détails concernant ce volet particulier de la recherche sont donnés dans le deuxième article de cette thèse.

4.7 La synthèse des données

La collecte de données décrite ci-dessus comporte plusieurs volets et plusieurs sources de données en vue de soutenir les trois objectifs de recherche. Le Tableau XVI récapitule les différentes données collectées en fonction des objectifs poursuivis dans chaque article scientifique.

Tableau XVI. Synthèse des données collectées en fonction des objectifs de recherche

<p>Premier article : documenter les démarches de cartographie conceptuelle, et ce dans deux modalités, à savoir lorsque la carte est produite à l'ordinateur ou lorsqu'elle est produite de manière manuscrite.</p> <p>Deuxième article : concevoir et mettre à l'essai une stratégie d'extraction de propositions assistée par ordinateur en vue de la production d'une carte conceptuelle et dans le but éventuel de créer une carte de référence pouvant servir d'outil de correction des cartes produites par des élèves.</p>	<ul style="list-style-type: none"> • Captation de l'écran lors de la production d'une carte conceptuelle à l'ordinateur • Captation vidéo des actions faites lors de la production d'une carte conceptuelle sur support papier • Verbalisation de la pensée • Cartes conceptuelles produites par les élèves lors des séances de cartographie conceptuelle. • Liste de concepts des tâches de cartographie conceptuelle • Liste de synonymes • Textes issus des manuels scolaires • Textes issus des articles Wikipédia
---	--

<i>Objectifs</i>	<i>Sources de données</i>
<p>Troisième article : Illustrer les avantages d'appliquer la modélisation de Rasch, et plus particulièrement le modèle pour données dichotomiques et le modèle à crédit partiel, pour apprécier les propriétés psychométriques de tâches de cartographie conceptuelle.</p>	<ul style="list-style-type: none"> • Carte de référence produite à la suite de l'analyse de contenu des textes issus des ouvrages didactiques consultés lors de la collecte de données particulière du second article. • Cartes conceptuelles produites par les élèves lors de la tâche principale

4.8 La relation entre la collecte de données et les trois articles

Chacun des trois articles présentés dans les chapitres suivants dépend, d'une manière ou d'une autre, de la collecte de données décrite ci-dessus. Comme il a déjà été mentionné, le premier article fait l'analyse des données obtenues lors des séances de verbalisation de la pensée pour documenter les démarches de production des cartes conceptuelles dans les deux supports de production étudiés. Le second article s'appuie sur les tâches proposées aux élèves et sur le contenu de leur manuel de cours pour mettre à l'essai le logiciel de production semi-automatisée d'une carte conceptuelle. Le troisième article présente les résultats de la modélisation des scores issus des cartes conceptuelles produites par les élèves lors de la tâche principale. Bref, les trois articles sont liés de près ou de loin à la démarche de collecte de données décrite dans ce chapitre, c'est pourquoi il a été jugé bon de décrire les grandes lignes de celle-ci avant de présenter les aspects méthodologiques propres à chaque expérimentation.

Les articles de recherche

Chapitre 5 Étude des démarches déployées par des élèves de quatrième secondaire en science et technologie lors de la production d'une carte conceptuelle sur support papier ou à l'ordinateur²⁶

Maxim Morin

Jean-Guy Blais

Université de Montréal

Mots clés : carte conceptuelle, verbalisation de la pensée, processus de pensée, processus de réponse, logiciel de production de cartes conceptuelles.

Depuis déjà plusieurs années, on assiste à l'émergence et à la popularisation de logiciels spécialisés dédiés à la production des cartes conceptuelles. L'utilisation de tels logiciels semble avoir plusieurs avantages en comparaison à la construction de carte de manière traditionnelle sur support papier, mais les différences entre ces deux modalités, notamment les différences entre les démarches déployées par les individus dans une modalité ou dans l'autre, n'ont pas encore été documentées. Cet article se consacre aux différences entre les démarches de huit élèves de quatrième secondaire en science et technologie qui ont produit une carte conceptuelle, soit de façon traditionnelle avec un crayon sur un support papier, soit à l'ordinateur avec un logiciel dédié. En particulier, dans cette recherche exploratoire, les participants ont été appelés à penser à haute voix dans le but de mieux comprendre les différentes stratégies mises en œuvre au courant de la réalisation de la tâche. En plus de mettre en évidence quelques séquences d'actions et stratégies inhérentes à chaque modalité, les analyses de cette étude font ressortir des difficultés rencontrées par les participants lors de la construction de leur carte conceptuelle. Ces observations rappellent notamment l'importance de la formation préalable à cette technique de représentation visuelle des connaissances.

²⁶ Cet article n'a pas encore été soumis aux fins de publications.

5.1 Introduction

La carte conceptuelle est un outil de représentation des connaissances d'une grande utilité en éducation (Novak et Canas, 2008), mais aussi d'une grande polyvalence. Reconnu d'abord comme pouvant favoriser l'apprentissage (Nesbit et Adesope, 2006), cet outil peut aussi servir de dispositif de collecte de données en vue d'évaluer les apprentissages (Ruiz-Primo, Schultz, et coll., 2001; Scallon, 2004; Yin et Shavelson, 2008; Yin et coll., 2005) ou d'outil pour soutenir la planification de l'enseignement (Edmondson, 1995).

Une carte conceptuelle (voir l'exemple de la Figure 6) est un outil graphique composé de nœuds et d'arcs qui symbolisent respectivement des concepts et les relations entre ceux-ci (Jonassen, Reeves, Hong, Harvey et Peters, 1997). Les connaissances y sont exprimées par des énoncés de sens, nommés les propositions, qui sont formés de deux ou plusieurs concepts unis entre eux par des relations significatives.

La production d'une carte conceptuelle n'étant pas connue et maîtrisée par tous, une formation préalable à cette technique de représentation des connaissances est un incontournable pour assurer la qualité des productions des individus (Horton et coll., 1993). Pourtant, les modalités de formation répertoriées dans les écrits scientifiques et pédagogiques sont très variées, mais ne répondent vraisemblablement pas aux mêmes objectifs. Même si une formation courte et ciblée, comme celle décrite dans l'étude de Yin et coll. (2005), peut être suffisante pour amener des individus à respecter les règles de construction d'une carte conceptuelle, une telle formation ne donne pas nécessairement à ces individus l'occasion de développer certaines habiletés spécifiques à cette technique comme dans les modalités de formation plus longues, étalées sur plusieurs séances (Bell, 2018; Novak et coll., 1983; Rice et coll., 1998).

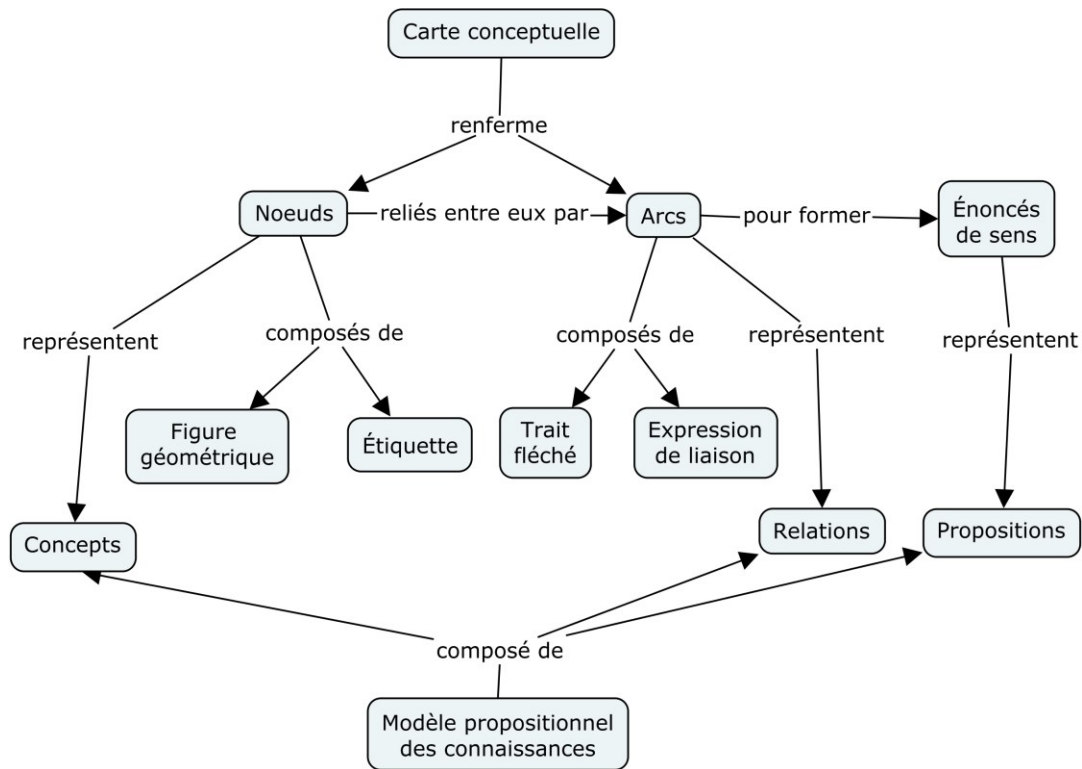


Figure 6. Exemple d'une carte conceptuelle

Lorsque Novak a introduit la cartographie conceptuelle à la fin des années 1970, la production d'une carte ne pouvait se faire que sur support papier. De nos jours cependant, les individus ont accès à une panoplie d'outils, dont de nombreux logiciels conviviaux pour visualiser, construire, stocker et échanger des cartes conceptuelles. Quelle que soit l'utilisation de la carte conceptuelle, il existe probablement des différences entre les supports de production et il importe de bien comprendre ces différences afin d'adapter la tâche aux ressources disponibles. Or, à ce jour, aucune étude ne s'est penchée sur les différences concernant la démarche de production des cartes lorsque divers supports sont utilisés. C'est donc ce dont il sera question dans la présente étude.

L'article est scindé en quatre sections. La première section présente le cadre théorique et se sous-divise en trois parties. Elle débute avec une recension de cinq articles scientifiques concernant des recherches sur les démarches de production d'une carte conceptuelle. Ensuite, elle expose le cadre conceptuel qui est inspiré du modèle du processus rédactionnel de Hayes et

Flower (1980). Elle se termine avec la présentation de l'objectif de recherche. La deuxième section décrit le cadre méthodologique et la troisième section comporte les analyses des résultats. La quatrième et dernière section présente la discussion des résultats et la conclusion.

5.2 Cadre théorique

5.2.1 Les études sur les démarches de production d'une carte conceptuelle

Loin d'être une démarche linéaire, la production d'une carte conceptuelle sollicite une multitude de connaissances, d'habiletés et de stratégies cognitives et métacognitives (Novak et Gowin, 1984). Même si certaines études se sont penchées sur des éléments bien précis de cette technique de représentation de connaissances, comme sa fonction métacognitive (Novak, 1990b ; Rye et Rubba, 1998), rares sont celles qui se sont intéressées aux démarches déployées lors de la conception d'une carte conceptuelle. Les cinq études décrites ci-après fournissent un panorama des travaux de recherche sur le sujet.

Dans la première étude, Ruiz-Primo, Shavelson, et coll. (2001) ont examiné la correspondance entre les contraintes de tâches de cartographie conceptuelle, les activités cognitives déployées lors de la réalisation de ces tâches et le niveau de performance à la tâche. Huit participants de niveaux d'habileté variés, soit deux enseignants de chimie et six élèves du secondaire, ont été appelés à réaliser une tâche de cartographie conceptuelle sur le thème des « noms et formules chimiques » à trois reprises (une fois pour chacune des trois modalités examinées), et ce tout en exprimant leur pensée à haute voix. Les transcriptions des séances de verbalisation de la pensée ont été analysées afin de faire ressortir les propositions et idées exprimées par les participants (appelé le codage du micro-niveau par les chercheurs), ainsi que les stratégies mises en œuvre au cours de la tâche (le codage du macro-niveau). Ces stratégies incluent des stratégies de planification (c'est-à-dire des manifestations de l'élaboration d'un plan qui ont lieu avant de commencer la tâche), de démarrage (par ex., choisir le concept le plus général), de poursuite (par ex., choisir un concept en lien avec le concept précédemment utilisé), de monitoring (par ex., réviser les propositions de la carte en vue de trouver des améliorations) et d'arrêt de la tâche (par ex., réviser la liste initiale de concepts pour s'assurer qu'ils sont tous utilisés). Globalement, les résultats de cette étude montrent que la modalité la plus ouverte, qui

consiste à produire une carte à partir d'une liste de concepts, ne cible pas les mêmes dimensions de la connaissance et ne mobilise pas les mêmes activités cognitives que les deux modalités les plus restrictives, du type « compléter une carte ». En effet, les participants ont fourni plus d'explications et expriment plus d'erreurs d'ordre conceptuel lorsqu'ils ont produit eux-mêmes leur carte conceptuelle, alors qu'ils ont fait appel à des stratégies d'élimination d'options de réponses ou de choix au hasard d'une réponse lorsqu'ils ont été appelés à compléter une carte à partir d'une liste de concepts ou de mots de liaison.

Dans une deuxième étude, Hilbert et Renkl (2008) ont, eux aussi, fait appel à la verbalisation de la pensée pour étudier les processus cognitifs déployés lors de la production d'une carte conceptuelle. Une analyse par grappes (en anglais, *cluster analysis*) des données collectées²⁷ auprès de 38 étudiants issus de différents programmes de niveau universitaire leur a permis de caractériser les démarches des répondants lors d'une tâche de cartographie conceptuelle sur le thème des cellules souches. En particulier, ils ont découvert que, parmi les répondants qui ont démontré un faible niveau de maîtrise des connaissances disciplinaires, plusieurs ont omis d'écrire les expressions de liaison sur les liens qu'ils ont tracés et plusieurs ont mobilisé très peu de processus de planification et de contrôle. D'après Hilbert et Renkl (2008), le fait d'avoir observé ces lacunes chez près de la moitié des répondants rappelle l'importance de former convenablement les personnes à la cartographie conceptuelle afin de tirer profit de cette technique de représentation des connaissances.

Dans une troisième étude répertoriée, Gurlitt et Renkl (2010) présentent deux expérimentations qui s'intéressent à la cartographie conceptuelle comme technique d'activation des connaissances. Dans la première expérimentation, ils ont analysé et comparé les protocoles de verbalisation de la pensée de 20 étudiants de différents programmes universitaires qui ont été soumis à deux modalités de production d'une carte conceptuelle²⁸. La première variante consiste

²⁷ Pour la liste complète, le lecteur peut consulter l'article original. Plusieurs données collectées ne sont pas considérées dans les interprétations finales, puisqu'elles ne sont pas discriminantes ou jugées secondaires par les auteurs.

²⁸ Les tâches ont porté sur deux thèmes : l'économie et la météorologie. Les modalités et les thèmes des tâches ont été présentés aux participants dans différents ordres afin de contrôler les effets de ces deux facteurs sur les résultats de cette étude.

à compléter les relations lorsque les liens sont déjà fournis dans la carte et l'autre variante consiste à tracer et à étiqueter les relations lorsque seulement les concepts sont fournis à l'avance. En conformité avec les recherches antérieures sur l'activation des connaissances, ils observent que les répondants expriment plus de liens entre les connaissances antérieures et les nouvelles connaissances dans la première modalité, dite à cohérence élevée. Autrement dit, le fait de donner à l'avance les liens (non étiquetés) entre les concepts activerait davantage les connaissances antérieures des individus. Dans la seconde expérimentation, Gurlitt et Renkl ont évalué l'effet de la modalité d'une tâche de cartographie conceptuelle sur l'apprentissage et sur le sentiment d'auto-efficacité. Pour ce faire, 42 étudiants ont réalisé une seule tâche dans l'une ou l'autre des modalités proposées lors de la première expérimentation, suivie d'une activité d'apprentissage de vingt minutes dans un environnement hypertexte²⁹. Les résultats de cette seconde expérimentation coïncident grandement avec ceux de la première. Dans l'ensemble, les répondants qui ont effectué la tâche dans la modalité à cohérence élevée ont obtenu un rendement d'apprentissage plus élevé et ont attesté d'un sentiment d'auto-efficacité plus élevé que ceux soumis à l'autre tâche. Les auteurs en concluent que la modalité la plus restrictive favorise plus l'apprentissage que la modalité moins restrictive lorsqu'elle est utilisée pour activer les connaissances antérieures.

Dans la quatrième étude, Jeong (2014) a étudié les démarches de construction d'une carte causale, un outil apparenté³⁰ à la carte conceptuelle, auprès de 17 étudiants universitaires. À l'aide d'un logiciel spécialement conçu à cet effet, les actions des répondants ont été captées et

²⁹ Un hypertexte (hypertext) est un ensemble de documents électroniques liés entre eux par des hyperliens. Dans un contexte d'apprentissage, cela permet à un individu de parcourir plusieurs documents et de chercher de l'information de façon non linéaire.

³⁰ Il existe plus d'une version de la carte causale. La définition de la carte causale d'Eden (1988), aussi appelée la carte cognitive, est l'une d'entre elles. Une carte causale selon cette définition comprend des nœuds, qui renferment des idées exprimées souvent sous forme de phrases ou de paragraphes, qui sont reliés entre eux par des liens non étiquetés. Ces liens renvoient à la relation implicite « conduit à ». La définition utilisée par Jeong est quelque peu différente. Les nœuds désignent des concepts ou des principes et les liens peuvent être composés de divers attributs comme le type de relation (corrélationnelle ou causale), la direction de la relation et la probabilité « causale ».

enregistrées automatiquement par l'ordinateur pour ensuite, être analysées en suivant la méthode de l'analyse séquentielle de données³¹. Lors de ses analyses, Jeong a constaté que les étudiants qui ont démontré une meilleure maîtrise des connaissances ont aussi adopté globalement des démarches plus structurées, notamment en alternant entre des opérations de déplacements de concepts et de créations de liens, tandis que les participants dont le niveau de maîtrise des connaissances a été plus faible se sont retrouvés à déplacer des objets de manière beaucoup plus fréquente et désordonnée.

Enfin, dans la cinquième et dernière étude recensée, Dogusoy-Taylan et Cagiltay (2014) ont mis à profit une méthode de collecte de données mixte, combinant l'oculométrie (en anglais, *eye tracking*), l'analyse des actions et la verbalisation de la pensée, pour étudier les processus cognitifs mobilisés lors de la production d'une carte conceptuelle sur le thème de la cellule ou de la matière. Leurs analyses semblent indiquer que les démarches de six experts de contenu diffèrent de celles des 29 étudiants universitaires en enseignement (les novices) en plusieurs points. En ce qui a trait au suivi du regard, la durée des fixations est en moyenne plus longue chez les répondants experts, mais le nombre de visites du regard sur chaque objet de la carte est plus faible. En ce qui concerne l'analyse des actions et de la verbalisation, les répondants experts ont fait plus de déplacements d'objets, principalement vers la fin de la tâche ; ils ont déployé plus de stratégies de contrôle et de vérification du contenu de la carte tout au long de la tâche ; et ils ont exprimé moins de conceptions erronées. Finalement, les experts ont manifesté une plus grande propension à verbaliser un plan initial. Les analyses faites par ces chercheurs ne s'appuient toutefois pas sur des inférences statistiques, alors il faut considérer avec prudence leurs interprétations des résultats.

Malgré certaines différences sur le fond et sur la forme, les cinq études recensées ont plusieurs éléments en commun, entre autres sur le plan méthodologique. À l'exception de l'étude de Jeong (2014), elles tentent toutes d'inférer des processus cognitifs à partir

³¹ La méthode de l'analyse séquentielle des données consiste à calculer les probabilités de transitions entre des actions successives et de déterminer, à l'aide d'un test z (au seuil de signification de $p < 0,01$), les séquences d'actions qui caractérisent la démarche d'un répondant. L'ouvrage de Bakeman et Quera (2011) offre une bonne introduction à cette méthode d'analyse.

d'observations directes ou d'enregistrements de la verbalisation de la pensée. Cette approche a toutefois certaines limites.

Tout d'abord, la notion de processus cognitifs à laquelle font référence directement ou indirectement les articles recensés est généralement considérée comme vague (Groome, 1999). En psychologie, cette notion renvoie à l'ensemble des processus mentaux qui permettent à un individu de traiter l'information, comme la perception, l'attention, le langage, la résolution de problèmes et la prise de décision, pour en nommer que quelques-uns. La définition de processus cognitifs est un problème très large et par conséquent difficile à opérationnaliser, de sorte que la présente étude n'a pas la prétention de s'attaquer à un problème de cette taille.

Malgré cela, il y a toujours des bénéfices potentiels à sonder la pensée des individus comme l'ont fait les auteurs de quatre des études répertoriées, si ce n'est que pour mieux comprendre les décisions qui guident leurs actions. À cet égard, le protocole de verbalisation de la pensée de type concomitant (en anglais, *concurrent think aloud protocol*) demeure une méthode de collecte et d'analyse de données privilégiée pour suivre le cheminement de la pensée (Ericsson et Simon, 1993). Cela dit, cette approche a aussi quelques limites (Cabello et O Hora, 2002 ; Charters, 2003). Les paroles exprimées par les individus ne traduisent pas toujours pleinement leur pensée et leurs actions, c'est pourquoi il est souvent recommandé de procéder à une « triangulation » des observations à l'aide d'autres méthodes de collecte de données (Charters, 2003). Ruiz-Primo, Shavelson, et coll. (2001), par exemple, ont complété leurs informations à l'aide d'entretiens rétrospectifs alors que Dogusoy-Taylan et Cagiltay (2014) se sont fiés à l'oculométrie et à l'analyse des actions effectuées par les répondants.

Malgré les avantages reconnus de la verbalisation de la pensée pour étudier les processus de réponses, il faut s'assurer d'analyser et d'interpréter les données au moyen de stratégies adaptées. Parmi les études recensées, les chercheurs ont entre autres procédé à des analyses de corrélation (Gurlitt et Renkl, 2010 ; Hilbert et Renkl, 2008 ; Ruiz-Primo, Shavelson, et coll., 2001) afin de mettre en relation les données issues des séances de verbalisation de la pensée avec d'autres données, comme des indicateurs de performance à la tâche ou à une tâche concurrente. La robustesse des conclusions découlant de ces techniques d'analyse dépend toutefois de la taille d'échantillon et des qualités psychométriques des mesures qui sont utilisées. Or, la taille d'échantillon de ces études est assez faible, soit aussi peu que huit dans l'étude de

Ruiz-Primo, Shavelson, et coll. (2001), et les propriétés psychométriques des instruments *ad hoc* ne sont pas toujours présentées (Gurlitt et Renkl, 2010 ; Hilbert et Renkl, 2008). Pour des raisons logistiques évidentes, il est en fait très difficile d'effectuer des séances de verbalisation de la pensée avec un nombre assez grand d'individus pour assurer la robustesse des résultats tels que ceux provenant d'analyses corrélationnelles.

Pour l'ensemble des raisons évoquées ci-dessus, la présente étude prend une certaine distance par rapport aux recherches antérieures. Plutôt que d'examiner la relation entre le niveau d'expertise (ou de performance) d'un individu et les processus cognitifs mobilisés lors de la tâche, elle cherche à mieux documenter le déroulement de la tâche en triangulant notamment les données.

5.2.2 Les actions et les stratégies mobilisées lors de la production d'une carte

Cette étude s'intéresse en particulier aux actions et aux stratégies mises à profit au cours de la production d'une carte conceptuelle. Une action est tout simplement l'observation directe d'une opération faite lors de la réalisation de la tâche, comme le déplacement d'un objet. Ces actions peuvent être regroupées en différentes catégories représentant les opérations les plus communes d'une démarche de construction d'une carte conceptuelle. Dans son étude, Jeong (2014), propose six catégories d'actions : la création d'un lien, la redirection d'un lien, la modification d'un attribut, la suppression d'un lien, le déplacement d'un nœud, ou la formulation d'un commentaire. Seule la dernière catégorie ne s'applique pas au contexte de la présente étude, car la formulation d'un commentaire est une caractéristique spécifique du logiciel mis au point par Jeong.

La notion de stratégie, quant à elle, peut prendre plusieurs sens en éducation (par exemple, stratégies d'apprentissage, stratégies cognitives, stratégies métacognitives, etc.), mais dans le contexte de cette étude, elle renvoie à sa définition dans le langage courant : « un ensemble organisé d'actions dans le but d'accomplir un but précis » (Antidote 9, 2017). Conséquemment, le codage des stratégies se situe à un niveau d'inférence supérieur par rapport au codage des actions parce qu'il exige de porter une attention particulière aux buts sous-jacents aux actions observées.

Même si quelques études (Dogusoy-Taylan et Cagiltay, 2014 ; Hilbert et Renkl, 2008) ont fait allusion à certaines classes de stratégies, celle de Ruiz-Primo, Shavelson, et coll. (2001) est la seule à s'être penché concrètement sur les stratégies mobilisées par les répondants. Dans leur étude, ils divisent la démarche de production d'une carte conceptuelle en quatre phases (la planification, le démarrage, la progression et l'arrêt), à laquelle ils ajoutent une catégorie supplémentaire, la régulation. D'après ce modèle, il est donc attendu que l'individu exprime d'abord son plan, puis il met en œuvre des mécanismes de démarrage de la tâche, ensuite il commence à construire sa carte, etc. Ce compartimentage des stratégies peut toutefois sembler un peu artificiel, car le déroulement d'une tâche n'est pas nécessairement scindé en une progression d'étapes clairement définies. À cet égard, il est intéressant de faire un parallèle entre le modèle des processus de production d'une carte conceptuelle de Ruiz-Primo, Shavelson, et coll. (2001) et certains modèles du processus rédactionnel, dont l'un des modèles les plus influents de ce domaine (Piolat, 2004), est celui de Hayes et Flower (1980).

5.2.3 Le modèle du processus rédactionnel de Hayes et Flower (1980)

S'inscrivant dans une approche cognitiviste, le modèle de Hayes et Flower reconnaît que le processus d'écriture est en constante interaction avec deux autres composantes, soit l'environnement de la tâche (comme les caractéristiques de la tâche et les outils disponibles) et les informations emmagasinées dans la mémoire à long terme du rédacteur. Le processus d'écriture quant à lui se divise en quatre sous-processus: la planification, la mise en texte, la révision et le contrôle. D'abord, le premier sous-processus correspond à la récupération et à l'organisation des idées contenues dans la mémoire à long terme, de même qu'à l'établissement des buts de la tâche. Puis la mise en texte concerne la génération de mots, syntagmes, phrases et paragraphes cohérents et respectant les règles linguistiques. Ensuite, la révision englobe les mécanismes de relecture et de modification du texte. Enfin, le contrôle correspond aux ressources cognitives nécessaires à la régulation de la tâche et des trois autres sous-processus.

Un avantage du modèle de Hayes et Flower (1980) est qu'il peut être facilement adapté à d'autres types de tâches, comme la production d'une carte conceptuelle. En effet, les composantes et les processus de ce modèle sont assez génériques et peuvent être redéfinis de sorte à prendre en considération les particularités et les contraintes de la tâche. Pour ce qui est

de la cartographie conceptuelle, les processus de planification, de révision et de contrôle peuvent quasiment être pris tels quels ; la « mise en texte » est le seul processus qui ne s'applique pas directement au contexte de la cartographie conceptuelle. La mise en relation des concepts, c'est-à-dire le processus qui consiste à tracer un lien entre deux concepts et à y apposer une expression de liaison, est certainement le processus de cette technique de représentation de connaissances qui s'en approche le plus. C'est dans l'optique d'avoir un modèle assez générique que le modèle de Hayes et Flower (1980) a été choisi dans le cadre de cette recherche.

L'adoption du modèle de Hayes et Flower (1980) a aussi quelques autres avantages. D'une part, ce modèle reconnaît clairement que les caractéristiques de l'environnement de la tâche ont un rôle à jouer dans les démarches des individus. Cette recherche s'intéresse après tout aux différences entre les démarches d'individus qui produisent une carte conceptuelle de manière traditionnelle ou de manière informatisée. D'autre part, ce modèle s'oppose à une conception hiérarchique des quatre sous-processus déployés au cours de la tâche. Ainsi, d'après ce modèle, la planification ne se produit pas seulement au début d'une activité, comme le propose Ruiz-Primo, Shavelson, et coll. (2001), mais elle peut survenir à tout moment, se répéter et se réactualiser, au besoin.

5.3 Objectif de la recherche

Comme il a été mentionné précédemment, les études au sujet des processus de production de cartes conceptuelles sont encore peu nombreuses, et, à ce jour, aucune d'entre elles ne s'est intéressée au rôle du support de production dans la démarche de cartographie conceptuelle. Pourtant, une meilleure compréhension de ces processus peut s'avérer utile afin d'adapter adéquatement la formation préalable des répondants, ce qui est d'ailleurs capital dans un contexte d'évaluation des apprentissages pour assurer la justesse et la pertinence des données collectées au moyen de tâches de cartographie conceptuelle, et ultimement, ce qui permettrait de prendre une décision « valide ». Cette recherche vise donc à documenter les démarches de cartographie conceptuelle, et ce dans deux modalités, à savoir lorsque la carte est produite à l'ordinateur ou lorsqu'elle est produite de manière manuscrite. Pour ce faire, une méthode de collecte de données mixte est préconisée, conciliant la verbalisation de la pensée et l'observation

des traces, afin d'obtenir des informations au sujet des actions et des stratégies déployées lors de la construction d'une carte. La prochaine section décrit le cadre méthodologique de l'étude.

5.4 Méthodologie

5.4.1 Participants

La collecte de données de cette étude a été réalisée auprès des groupes d'élèves d'un enseignant en science et technologie œuvrant dans une école secondaire de la grande région de Montréal, dans la province de Québec au Canada. Les participants ont été recrutés sur une base volontaire et sans aucune compensation et, dans le but de faciliter le recrutement, l'enseignant a accepté que la production des cartes et leur verbalisation soient faites pendant les heures de classe. Au total, huit élèves de quatrième secondaire en sciences et technologie ont participé aux séances de verbalisation de la pensée, quatre pour chacune des modalités de passation : de manière traditionnelle à l'écrit et de manière informatisée. Ce nombre de participants peut sembler faible, mais dans une étude exploratoire, comme celle-ci, le traitement en profondeur des données collectées se veut un moyen de compenser la petite taille de l'échantillon de sujets.

5.4.2 Formation des participants

Tous les élèves de la classe dont les huit participants ont pris part à trois activités de formation pilotées par le chercheur principal en collaboration avec l'enseignant. Dans le but de les familiariser progressivement à la cartographie conceptuelle, (1) une activité d'introduction à la carte conceptuelle (environ 10 minutes), (2) une activité de construction d'une carte en équipe (environ 25 minutes) au sujet des barrages hydroélectriques et (3) un exercice individuel de cartographie conceptuelle concernant le phénomène de l'effet de serre suivi de rétroactions en groupe (50 minutes) leur ont été proposés. L'activité de construction d'une carte en équipe a été réalisée seulement sur support papier tandis que l'exercice individuel a été effectué sur

support papier par deux des huit participants³². Ces trois activités ont été pilotées par le chercheur principal, en collaboration avec l'enseignant responsable des élèves.

À la suite de ces activités préparatoires, les huit participants ont été conviés à tour de rôle à réaliser la tâche principale. Afin d'avoir quatre participants dans chacune des modalités lors de la verbalisation de la pensée, il a été demandé à deux élèves (identifiés comme étant les participants 5 et 8 dans ce qui suit) de réaliser la tâche principale sur support papier. Les six autres participants ont utilisé les mêmes outils que ceux qui leur ont été fournis lors des activités précédentes. Le Tableau XVII indique le support de production des cartes conceptuelles des huit participants lors des deux tâches.

Tableau XVII. Support de production des cartes conceptuelles des huit participants lors de la tâche d'entraînement et de la tâche principale

<i>Participants</i>	<i>Support de production</i>	
	Tâche d'entraînement	Tâche principale
1	Informatisé	Informatisé
2	Informatisé	Informatisé
3	Informatisé	Informatisé
4	Informatisé	Informatisé
5	Informatisé	Papier
6	Papier	Papier
7	Papier	Papier
8	Informatisé	Papier

5.4.3 Description de la tâche principale

Les activités proposées aux élèves dans le cadre de cette étude ont été développées conjointement avec l'enseignant de manière à ce qu'elles s'intègrent dans la planification des apprentissages du cours. Pour la tâche principale, l'enseignant a établi une liste de 17 concepts

³² Comme certains élèves ont accepté de participer aux séances de verbalisation de la pensée après avoir entamé la formation, il n'a pas été possible de former dès le départ quatre participants dans chaque modalité.

relatifs au thème des relations trophiques³³ en s'inspirant du manuel scolaire utilisé pour le cours.

5.4.4 Matériel et données

En modalité papier-crayon, les élèves ont reçu une feuille comprenant les consignes et la description de la tâche, une feuille de grande dimension (27,9 cm X 43,2 cm) ainsi qu'un paquet d'étiquettes amovibles. Leurs actions et leur verbalisation ont été enregistrées au moyen d'une caméra vidéo. En modalité informatisée, les élèves ont réalisé leur carte à l'aide du logiciel en version gratuite CMap Tools (Cañas et coll., 2004), avec lequel ils s'étaient déjà familiarisés. Les actions ont été captées au moyen d'un logiciel de captures d'écran gratuit nommé CamStudio³⁴ et la verbalisation, au moyen d'un enregistreur sonore. Les données obtenues lors de cette collecte comprennent donc les enregistrements vidéo, les enregistrements sonores et les cartes tracées sur support papier lorsque la tâche a été effectuée de manière manuscrite et les enregistrements de la capture d'écran, les enregistrements audio et le fichier électronique de la carte produite par le participant lorsque la tâche a été réalisée à l'ordinateur.

5.4.5 Déroulement des séances de verbalisation

D'après Ericsson et Simon (1993), il est généralement assez facile d'amener des individus à verbaliser leur pensée lorsque la tâche se déroule dans des conditions calmes et propices. Une période d'environ 50 minutes a été prévue et une salle particulière de l'école a été spécialement aménagée pour réaliser chacune des rencontres. Ceux-ci ont duré entre 15 et 38 minutes. Avant d'entamer la tâche, chaque participant a été entraîné à verbaliser sa pensée en construisant des liens entre les concepts déjà vus lors de la tâche d'entraînement faite en classe. Lors de la tâche, et conformément aux bonnes pratiques de la verbalisation de la pensée (Trickett et Trafton, 2009), le chercheur est intervenu minimalement lors de la réalisation de la tâche, sauf pour rappeler au participant de s'exprimer à voix haute après chaque pause silencieuse d'environ 15 secondes.

³³ Les relations trophiques concernent les relations alimentaires entre les êtres vivants.

³⁴ Voir <http://camstudio.org/>

5.4.6 La transcription et le codage

La complémentarité entre les actions et la verbalisation est une dimension importante de cette recherche et par conséquent, la transcription des enregistrements vise à illustrer la coïncidence des paroles et des actions réalisées par les participants. Pour y arriver, les règles suivantes ont été établies lors de la transcription des enregistrements :

- Les concepts ont été écrits en MAJUSCULE.
- Les actions qui précèdent et qui coïncident avec la verbalisation sont décrites entre des crochets, avant et après la verbalisation.
- Les informations qui ont été inférées à partir des observations directes (par exemple, un concept) se retrouvent entre des parenthèses.

La Figure 7 illustre d'ailleurs un exemple d'unité d'analyse, telle que définie selon ces règles.

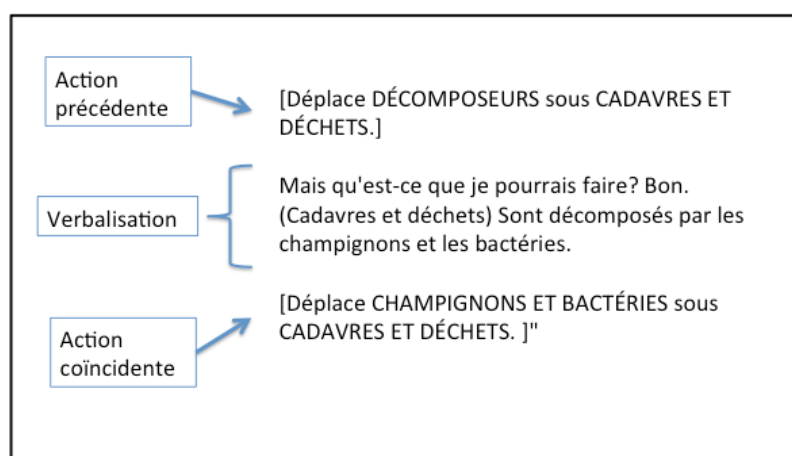


Figure 7. Exemple d'une unité de l'analyse des stratégies

Initialement, les transcriptions des démarches des participants ont été découpées en unités d'analyses, puis codées et analysées par le chercheur principal à partir d'une première adaptation des grilles de codage de Ruiz-Primo, Shavelson, et coll. (2001), soit une grille pour

la verbalisation et une grille pour les stratégies³⁵. Un exercice de contre-codage a également été effectué par deux autres chercheurs : un pour chacune des grilles. Pour le codage de la verbalisation, l'accord interjuges, estimé au moyen de coefficient kappa de Cohen a été de 0,73, ce qui est jugé satisfaisant selon les critères établis par Landis et Koch (1977). Le codage des stratégies a quant à lui conduit à un très bon niveau d'accord interjuges (kappa de Cohen de 0,88).

Au cours de l'analyse de contenu préliminaire, il est pourtant apparu évident que ces grilles de codage ne répondaient pas entièrement aux objectifs de cette recherche. Entre autres, elles ne permettaient pas de mettre en lumière la progression temporelle du processus de réponses des participants³⁶. Ce n'est qu'à la suite de cette analyse préliminaire que les grilles ont été remaniées de manière à séparer clairement le lexique des codes associés aux actions observables et le lexique des stratégies³⁷.

Le découpage des unités d'analyse a aussi été retravaillé. Chaque verbatim a été scindé en deux séries d'unités, une pour chacun de ces lexiques. Les démarches des participants ont en outre été chronométrées lors de ce nouveau découpage des enregistrements des séances de verbalisation de manière à reconstituer fidèlement le déroulement de la tâche. Dans les transcriptions finales, chaque unité d'analyse est donc accompagnée d'un marqueur de temps désignant le début et la fin de l'évènement.

Au cours de cette même analyse préliminaire, un troisième lexique de codes a également émergé pour représenter les difficultés et les contraintes rencontrées par les participants. Ce codage, qui s'harmonise d'ailleurs avec la première composante du modèle de Hayes et Flower (1980), à savoir l'environnement de la tâche, a été réalisé à partir des mêmes unités verbales que celles des stratégies.

³⁵ Voir les grilles de codage initiales de la verbalisation et des stratégies en Annexe H et Annexe I.

³⁶ Diverses méthodes d'analyse séquentielle, notamment celles basées sur des modèles d'analyse log-linéaire ont été explorées à cette étape. Ces méthodes d'analyse n'ont pas été retenues parce qu'elles alourdissaient la présentation des résultats sans nécessairement apporter d'information supplémentaire quant à leur interprétation.

³⁷ Voir la grille de codage finale des stratégies en Annexe J.

Quel que soit le lexique concerné, et comme c'est souvent le cas avec des données issues de la verbalisation de la pensée (Charters, 2003), le codage a nécessité quelques va-et-vient entre le perfectionnement du lexique, le découpage des transcriptions et le codage. Les lexiques de codes présentés ci-dessous sont l'aboutissement de ce processus. Pour des raisons logistiques, il n'a pas été possible de refaire un contre-codage à partir des nouvelles grilles de codage³⁸. Les prochaines sous-sections décrivent plus en détail le contenu de chaque lexique.

5.4.6.1 Les indicateurs des actions

Les actions faites par les participants sont caractérisées par un changement apparent et visible de la carte, comme le déplacement d'un objet, la modification d'un texte, etc. Elles sont divisées en onze indicateurs distincts (voir le Tableau XVIII), dont cinq qui sont inspirées de l'étude de Jeong (2014). Les autres ont été définies au cours de la transcription des verbatim et lors des analyses préliminaires.

Tableau XVIII. Lexique des actions

Id	Indicateurs
1	Écrire un concept sur une étiquette amovible
2	Retranscrire un concept sur la feuille
3	Déplacer un concept
4	Déplacer un mot-lien
5	Lier
6	Formuler
7	Ajouter un concept
8	Modifier un concept
9	Modifier une relation
10	Supprimer un concept
11	Supprimer un lien ou une relation

³⁸ Il s'est écoulé près d'un an entre les premières analyses et les analyses finales. Malheureusement, les analystes qui ont fait le codage lors des analyses préliminaires n'étaient plus disponibles pour réaliser le codage des transcriptions.

5.4.6.2 Les indicateurs des stratégies

Les stratégies sont catégorisées selon le modèle du processus d'écriture de Hayes et Flower (1980) et les indicateurs présentés au Tableau XIX sont inspirés à la fois de ce modèle, des recherches qui se sont basées sur ce modèle (Diarra, 2013) et des lexiques de codes de deux études recensées concernant le processus de production d'une carte conceptuelle (Hilbert et Renkl, 2008; Ruiz-Primo, Shavelson, et coll., 2001). Comme c'est le cas dans quelques études similaires (Diarra, 2013), le lexique ne tient pas compte du dernier sous-processus du modèle de Hayes et Flower, à savoir le processus de contrôle. Même si ce dernier est essentiel à la réalisation de la tâche, leurs manifestations ne sont pas nécessairement dissociables des autres processus. Les trois autres catégories de stratégies se séparent en plusieurs indicateurs.

D'abord, les stratégies de planification sont divisées en deux sous-catégories, soit la récupération en mémoire et l'organisation. Le modèle de Hayes et Flower (1980) comporte une troisième sous-catégorie, l'établissement de buts, mais comme aucune manifestation de cette stratégie n'a été observée, elle n'est pas répertoriée dans le lexique de codage.

Deuxièmement, les stratégies de la « mise en relation des concepts » sont divisées en trois indicateurs. Le premier indicateur concerne la formulation d'une relation entre deux concepts et le second, la formulation de plusieurs relations verbalisées de manière contiguë, comme les relations hiérarchiques. Le troisième indicateur a été inséré dans la grille pour refléter une stratégie particulière qui consiste à tracer une accolade pour représenter simultanément plusieurs relations entre des concepts.

Enfin, les stratégies de révision couvrent les deux processus énumérés dans le modèle de Hayes et Flower (1980), soit la relecture et l'édition, de même que les stratégies relatives à l'évaluation/diagnostic de la production. Ces stratégies, comme celles des deux autres catégories, sont expliquées plus en détail lors des analyses.

Tableau XIX. Lexique des stratégies

Stratégies et sous-stratégies	Id	Indicateurs
Planification		
Récupération d'idées	P.1	Exprimer verbalement des relations entre des concepts
	P.2	Définir un concept
Organisation	P.3	Identifier le thème principal
	P.4	Identifier les concepts principaux
	P.5	Réunir deux concepts en vue de former ultérieurement une proposition
	P.6	Regrouper plusieurs concepts en vue de former ultérieurement des relations
	P.7	Déplacer des concepts et des relations en vue d'ajouter d'autres propositions
Mise en relation		
	F.1	Formuler une relation entre des concepts
	F.2	Formuler plusieurs relations contiguës
	F.3	Tracer une accolade
Révision		
Relecture	R.1	Relire des propositions
	R.2	Corriger une ou plusieurs relations
Modification/Édition	R.3	Corriger l'étiquette d'un concept
	R.4	Remplacer une ou plusieurs propositions par d'autres propositions
	R.5	Supprimer une ou des relations
Évaluation/Diagnostic	R.6	Évaluer le contenu de la carte
	R.7	Chercher des améliorations potentielles
	R.8	Vérifier le nombre de concepts

5.4.6.3 Les codes des difficultés et contraintes rencontrées

Le dernier lexique de codage concerne les difficultés rencontrées par les participants et les contraintes exprimées par ceux-ci au cours de la réalisation de la tâche (voir Tableau XX). Ces codes sont issus d'un codage ouvert; ils ont été déterminés au fur et à mesure de l'analyse

du contenu et proviennent tantôt d'observations directes (D.1 à D.3)³⁹, tantôt de commentaires verbalisés par les participants (D.4 à D.8). Le dernier indicateur (D.9) a trait à une situation particulière qui se produit lorsque le contenu de la carte ne reflète pas l'idée qui a tout juste été exprimée.

Tableau XX. Lexique des difficultés et contraintes

Id	Indicateur
D.1	Difficulté menant à la création d'une bulle vide
D.2	Difficulté liée à la méconnaissance du logiciel
D.3	Difficulté liée à la superposition de concepts
D.4	Commentaire lié à l'utilisation d'un verbe comme expressions de liaison
D.5	Commentaire lié à la mise en relation d'une expression de liaison avec une autre expression de liaison
D.6	Commentaire lié au croisement des liens
D.7	Commentaire lié à l'utilisation d'un ouvrage de référence
D.8	Commentaire lié à l'utilisation de l'espace de la page
D.9	Représentation non fidèle d'une idée

5.4.7 Stratégie d'analyse

Dans cette étude, l'analyse de résultats commence par la description sommaire des cartes des participants et est suivie par la présentation de graphiques reconstituant la chronologie des événements. Décrire l'enchaînement des actions et des stratégies constitue un défi considérable lors de l'analyse de protocoles verbaux et les graphiques se sont avérés des outils particulièrement utiles pour présenter les démarches de participants. Une série d'analyses subséquentes basées sur des tableaux de fréquences viennent préciser les caractéristiques de chaque démarche.

³⁹ Des illustrations de ces difficultés sont présentées en Annexe L.

Le découpage et le codage préliminaire ont été réalisés à l'aide du logiciel de traitement qualitatif NVivo (Version 10, 2012) et plusieurs analyses supplémentaires ont été effectuées à l'aide du tableur Excel et du logiciel de traitement statistique R.

5.5 Analyse des résultats

5.5.1 Productions finales des participants

Le Tableau XXI trace le portrait général des cartes produites par les participants. Force est de constater que certains participants ne s'en sont pas tenus aux 17 concepts de la liste initiale : le participant 3 a ajouté deux concepts (« la lumière du soleil [sic] » et « d'angendrer le processus chimique [sic] ») et le participant 8 n'a pas utilisé le concept « consommateur ». Ce dernier a d'ailleurs fait quelques opérations non conformes aux principes de cartographie conceptuelle énoncés en salle de classe avec les élèves participant à l'étude. Il a écrit le concept « décomposeurs » à deux endroits dans sa carte et il a tracé une accolade pour représenter que la « chaîne alimentaire » englobe l'ensemble des relations entre les autres concepts.

Le Tableau XXI comprend également le nombre de relations formulées par les participants ainsi que le nombre de relations valides⁴⁰ contenues dans la carte. L'écart entre les participants est d'ailleurs assez grand pour ce qui est de ce dernier indicateur. Le participant 5 a formulé seulement neuf relations valides parmi les seize relations inscrites dans sa carte (9 sur 16, 56 %), alors que le participant 4 en a écrit 19 (79 %).

⁴⁰ Les relations ont été évaluées dans le cadre d'un autre volet de cette recherche, qui se penche sur la notation d'une carte conceptuelle produite par un élève.

Tableau XXI. Caractéristiques des cartes produites par les participants

Modalité	N° du participant	N ^{bre} de concepts distincts	N ^{bre} de relations	N ^{bre} de relations valides
Informatisée	1	17	20	12
	2	17	18	15
	3	19	19	11
	4	17	24	19
Papier-crayon	5	17	16	9
	6	17	19	14
	7	17	19	14
	8	16	25	17

5.5.2 Le déroulement de la tâche

Dans ce qui suit, les démarches des participants sont illustrées au moyen de graphiques qui reconstituent la chronologie des actions et des stratégies déployées. L'axe vertical de chaque graphique représente l'une des onze actions ou des six sous-catégories de stratégies, alors que l'axe horizontal constitue l'échelle du temps. Chaque action ou stratégie est illustrée par une boîte rectangulaire, dont les côtés marquent le début et la fin de l'évènement. Une analyse comparative des graphiques des huit participants a fait ressortir trois types de démarche de cartographie conceptuelle. Dans le but de ne pas alourdir le texte, un seul exemple de chacun des trois types de démarche est présenté dans le corps du texte (Figure 8 à Figure 10) ; les autres graphiques se trouvent en annexe (voir Annexe M).

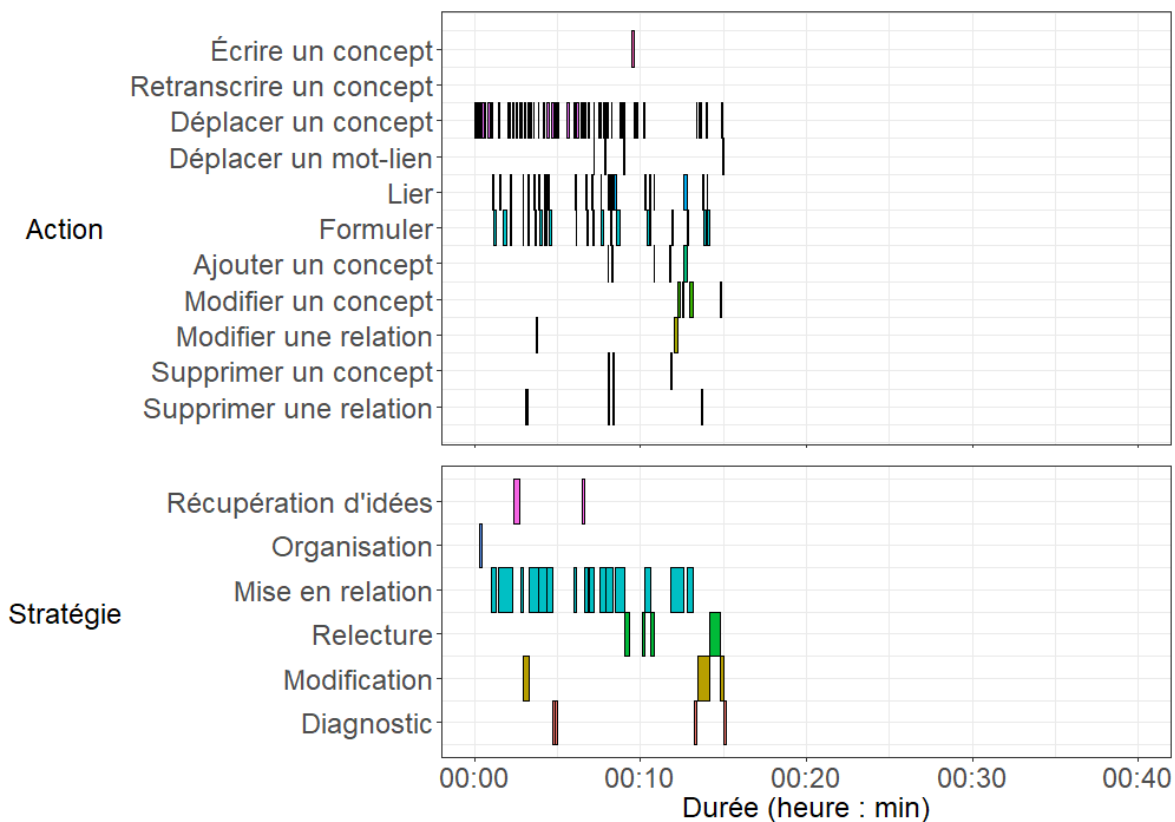


Figure 8. Démarche du participant 3

Dans le premier type de démarche, le participant met en relation les concepts au fur et à mesure que les relations lui viennent à l'esprit. Cette démarche a été observée chez les participants 1, 3 et 8, et est illustrée à la Figure 8. Plus précisément, dans cette démarche le participant construit sa carte en alternant majoritairement entre trois opérations, soit le déplacement de concepts, la formation de liens et l'écriture des expressions de liaison. Qui plus est, il déploie très peu de stratégies de planification, c'est-à-dire de processus de récupération et d'organisation d'idées ; il commence à former des relations très tôt dans la tâche ; et il manifeste très peu de stratégies de révision.

Au contraire de la démarche dépeinte à la Figure 8, le deuxième type de démarche (voir la Figure 9) est caractérisé par une séparation claire entre une phase de planification et une phase d'écriture des relations. Dans cette seconde démarche, observée pour trois participants de la modalité papier-crayon (5, 6 et 7), l'individu passe une assez longue période de temps à

organiser ses idées et à positionner les concepts sur la feuille, pour ensuite écrire les expressions de liaison entre ceux-ci. Un fait intéressant est que la majorité des relations qui sont rédigées lors de la mise en relations des concepts ont préalablement été exprimées à haute voix lors de la phase de planification des idées. Le participant 7, par exemple, dont la démarche est représentée à la Figure 9, a écrit tous les concepts sur les étiquettes amovibles, les a positionnés dans la page, puis les a retranscrits sur la feuille, avant de commencer à écrire les expressions de liaison.

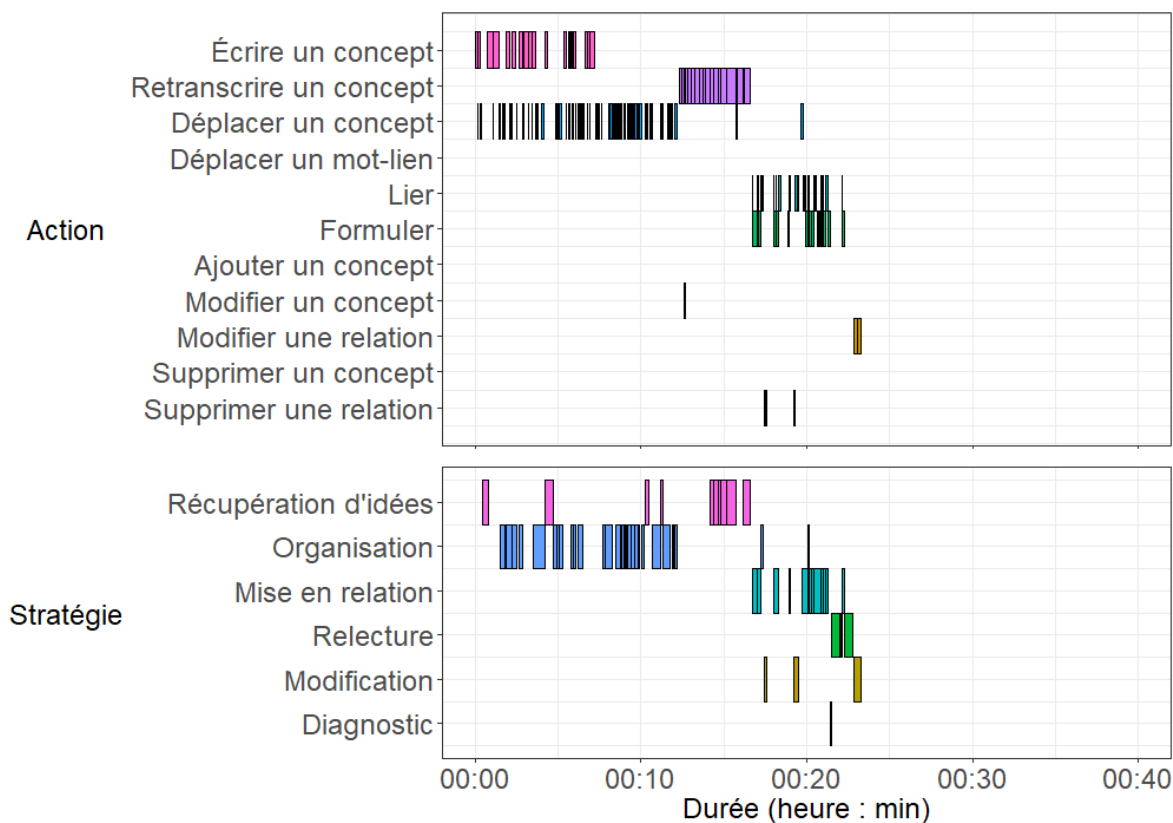


Figure 9. Démarche du participant 7

Le dernier type de démarche se situe à mi-chemin entre les deux démarches précédentes et concerne les participants 2 et 4 (voir la Figure 10). Il débute par une phase d'organisation qui consiste essentiellement à regrouper les concepts en catégories de concepts, puis est suivi d'une phase de mise en relations des concepts qui s'apparente à celle du premier type de démarche. Même si la phase d'organisation des concepts en début de tâche est moins soutenue que celle

observée pour les trois participants de la modalité papier-crayon, il n'en demeure pas moins qu'elle a une incidence sur le reste du déroulement de la tâche et en particulier, sur la séquence de mise en relation des concepts. Les deux participants concernés ont en effet commencé par former les propositions entre les concepts d'un même groupe, puis ils sont passés à celles entre les concepts des différents groupes. Comme il sera vu plus loin dans le texte, cette approche leur a occasionné quelques embûches, notamment parce qu'ils ont eu à remanier simultanément plusieurs objets de la carte.

Les prochaines sous-sections approfondissent le traitement des données relatives aux actions, aux stratégies et aux difficultés rencontrées.

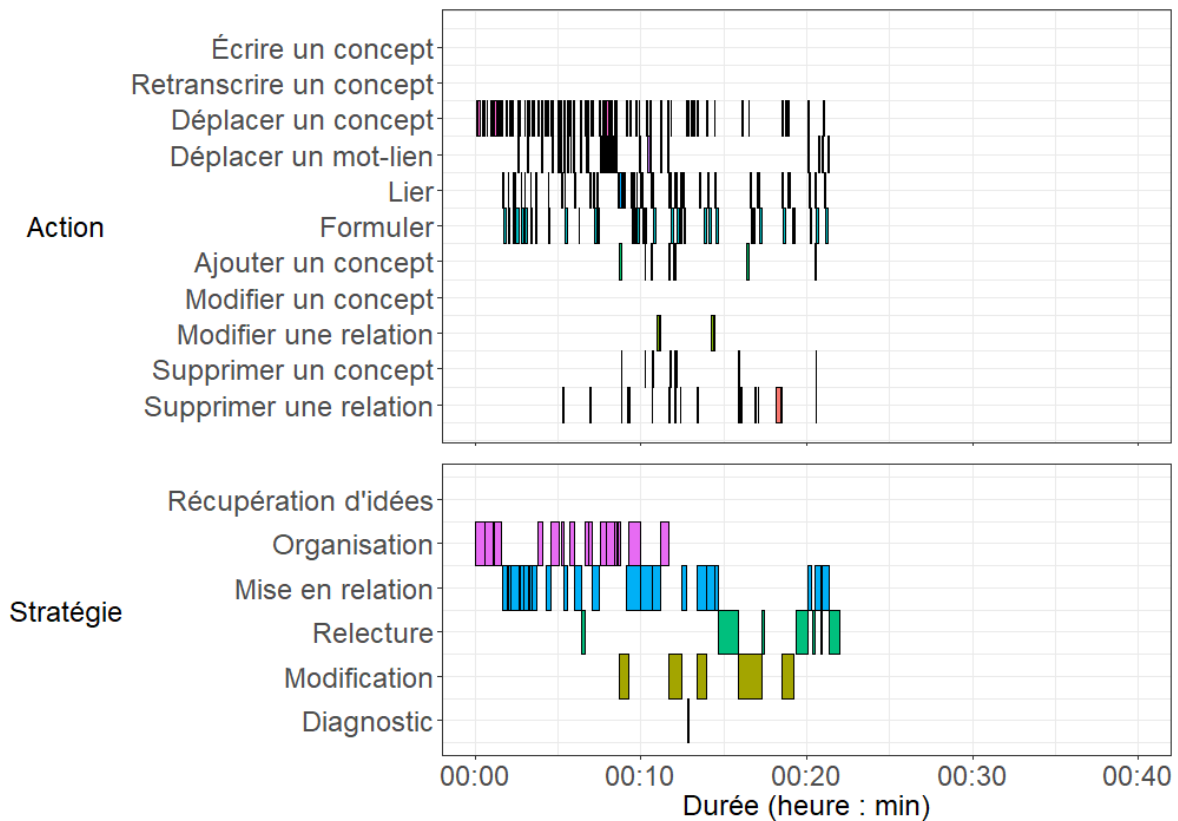


Figure 10. Démarche du participant 4

5.5.3 Les actions

Cette section analyse la distribution de fréquences des actions de chaque participant et fait ressortir les recoupements entre les démarches de ceux-ci (voir le Tableau XXII). Certaines observations sont communes à tous les participants, alors que d'autres sont particulières à la modalité de passation de la tâche ou encore au type de démarche préconisé.

Dans l'ensemble, le déplacement de concepts est l'action répertoriée la plus fréquente, suivie de la création de liens, puis de l'écriture des expressions de liaisons. Même si, habituellement, l'écriture d'une expression de liaison vient immédiatement après la création d'un lien, ces deux opérations sont parfois séparées. Cela se produit notamment lorsqu'un individu trace un lien entre deux concepts et le supprime aussitôt ou encore lorsqu'il remet à plus tard l'écriture de l'expression de liaison. En ce qui concerne les participants de la modalité informatisée, cela se produit aussi lorsqu'ils créent simultanément un concept vide (c'est-à-dire une bulle vide) et une relation vide (c'est-à-dire un lien sans étiquette) par inadvertance.

Les résultats du Tableau XXII font également ressortir quelques différences entre les modalités de production de la carte. Évidemment, seuls les participants de la modalité papier-crayon ont eu à écrire et à retranscrire les concepts de la liste⁴¹. Les participants de l'autre modalité ont quant à eux fait plus d'actions concernant l'ajout, la modification ou la suppression de contenu. Ces actions ne sont pas nécessairement liées à des activités de révision du contenu, plusieurs d'entre elles sont le résultat des difficultés techniques rencontrées pendant la réalisation de la tâche, comme il sera vu plus loin.

⁴¹ Il faut noter que le participant 3 a eu à réécrire un concept après l'avoir effacé, par erreur.

Tableau XXII. Fréquence des actions

Id	Code	Participant							
		Modalité informatisée				Modalité papier-crayon			
		1	2	3	4	5	6	7	8
1	Écrire un concept	0	0	1	0	17	17	18	16
2	Retranscrire un concept	0	0	0	0	17	17	17	16
3	Déplacer un concept	91	134	92	159	98	54	93	62
4	Déplacer un mot-lien	19	68	5	62	0	0	0	0
5	Lier	27	50	25	49	16	20	21	31
6	Écrire un mot de liaison	21	39	23	37	16	19	18	30
7	Ajouter un concept	1	11	5	8	0	0	0	1
8	Modifier un concept	0	2	4	0	0	0	1	0
9	Modifier une relation	7	10	2	4	0	3	2	7
10	Supprimer un concept	1	11	3	8	0	0	0	1
11	Supprimer un lien ou une relation	6	31	6	25	0	1	3	6
Total		173	356	166	352	163	131	173	170

En outre, les démarches des participants 2 et 4 se démarquent à plusieurs égards de celles des autres participants. En général, ces deux participants ont effectué beaucoup plus d'actions que les autres et en particulier, ils ont déplacé plus d'objets, ils ont créé plus de liens entre des concepts, ils ont écrit plus d'expression de liaison et ils ont modifié plus souvent le contenu de leur carte. Ces résultats s'expliquent en grande partie par les épisodes de remaniement de groupes de concepts et de relations, lors desquels le déplacement de chaque concept ou mot-lien était comptabilisé séparément, mais ils s'expliquent aussi par les opérations supplémentaires engendrées par les difficultés techniques, comme la création et la suppression de bulles vides.

5.5.4 La fréquence des stratégies

Le second volet de cette analyse examine les stratégies de planification, de mise en relation des concepts et de révision déployées par les participants.

5.5.4.1 Les stratégies de planification

Le Tableau XXIII présente plus particulièrement la fréquence des indicateurs associés aux deux catégories de stratégies de la planification, à savoir la récupération et l'organisation d'idées.

Tableau XXIII. Distribution des stratégies de planification mobilisées par les huit participants

Id	Indicateurs des stratégies	Participant							
		Modalité informatisée				Modalité papier-crayon			
		1	2	3	4	5	6	7	8
Récupération d'idées									
P.1	Exprimer verbalement des relations entre des concepts	0	0	0	0	2	1	7	2
P.2	Définir un concept	0	2	2	0	0	0	4	2
	Sous-total	0	2	2	0	0	1	11	4
Organisation d'idées									
P.3	Identifier le thème principal	0	0	0	0	1	0	0	0
P.4	Identifier les concepts principaux	0	0	0	0	0	1	0	0
P.5	Réunir deux concepts en vue de former une proposition	5	6	1	4	10	15	17	3
P.6	Regrouper plusieurs concepts	1	5	0	5	4	18	12	1
P.7	Déplacer des concepts et des relations en vue d'ajouter d'autres propositions	1	5	0	6	2	1	0	0
	Sous-total	7	16	1	15	19	35	29	4

Dans l'ensemble, les manifestations de l'organisation des idées sont beaucoup plus fréquentes que celles de la récupération d'idées ; même que ces dernières sont absentes des démarches des quatre participants de la modalité informatisée. Cela ne veut pas dire pour autant que les participants ne font pas de récupération d'idées. Ce que ces résultats ne montrent pas, mais qui est beaucoup plus apparent lors du visionnement des enregistrements vidéo, c'est que la récupération d'idées se fait souvent de manière simultanée avec d'autres opérations, comme l'organisation d'idées ou la mise en relation des concepts.

Le visionnement des enregistrements vidéo rappelle aussi que certains indicateurs prennent un sens différent lorsqu'ils sont remis dans leur contexte. La verbalisation de relations

(P.1), par exemple, ne signifie pas exactement la même chose pour chacun des participants. Chez les participants 5 et 8, cette stratégie sert à anticiper de nouvelles relations, avant même de positionner les concepts dans la feuille, tandis que chez les participants 6 et 7, cette stratégie est un moyen de vérifier l'organisation des concepts, déjà positionnés dans la page, avant de commencer à écrire les expressions de liaison entre ceux-ci. La mise en œuvre de cette stratégie est d'ailleurs relativement longue pour ces deux participants, respectivement de 1 minute 42 secondes et de 1 minute 34 secondes.

Toutes choses étant égales par ailleurs, il demeure que l'organisation des concepts est la stratégie dominante parmi les stratégies de planification ; elle a été adoptée par tous les participants. De même, la fréquence totale des manifestations de stratégies de planification reflète bien les trois types de démarches répertoriées. Les participants dont la démarche est centrée sur l'anticipation des relations (5 à 7) ont mobilisé le plus souvent des stratégies de planification, suivis des participants dont la démarche est centrée sur la catégorisation des concepts (2 et 4). Ces derniers sont aussi ceux qui ont réorganisé simultanément les concepts et les liens (P.7) de leur carte. Enfin, les participants qui ont construit leur carte sur-le-champ (1, 3 et 8) ont généralement mobilisé moins de stratégies de planification que les autres participants.

5.5.4.2 Les stratégies relatives à la mise en relation des concepts

Le Tableau XXIV expose la distribution de fréquences des stratégies de mise en relation des concepts. Même si les participants ont formulé la majorité des propositions une à la fois (F.1), chacun d'entre eux a aussi écrit quelques propositions de manière contiguë (F.2). Ce second processus est très fréquent lorsqu'un individu énonce une relation de type hiérarchique, comme la relation entre le concept général de « consommateurs » et les concepts subordonnés « consommateurs primaires », « consommateurs secondaires », « consommateurs tertiaires ». Dans ce cas, les propositions sont peut-être tracées une à une, mais elles sont issues d'une seule et même idée.

Tableau XXIV. Distribution des stratégies de mise en relation mobilisées par les huit participants.

Id.	Indicateurs des stratégies	Participant							
		Modalité informatisée				Modalité papier-crayon			
		1	2	3	4	5	6	7	8
F.1	Formuler une proposition	13	12	12	18	6	10	9	19
F.2	Formuler plusieurs relations contiguës	3	2	3	2	4	4	3	3
F.3	Créer une accolade	0	0	0	0	0	0	0	1
	Sous-total	16	14	15	20	10	14	12	24

Le dernier indicateur de mise en relations de concepts (F.3) appartient seulement au participant 8. À un certain moment, ce dernier a dérogé des principes de construction d'une carte conceptuelle en utilisant une accolade pour représenter la relation entre le concept de « chaînes alimentaires » et les autres concepts de sa carte.

5.5.4.3 Les stratégies de révision

Le Tableau XXV donne la fréquence des manifestations de chacune des trois sous-stratégies de révision. De manière générale, les participants qui ont manifesté le plus grand nombre de stratégies d'organisation (5, 6 et 7) ont fait moins de processus de révision, et en particulier moins de processus de relecture et de correction.

La relecture est la seule stratégie de révision qui a été utilisée par l'ensemble des participants et elle domine largement les stratégies de révision, en comparaison aux autres stratégies (R.6 à R.8) qui sont plutôt rares. Sa mise en œuvre peut être d'une très courte durée, à peine deux secondes (participant 1), ou elle peut être d'une plus longue durée, allant parfois jusqu'à deux minutes ou plus (participants 6 et 8). Dans le premier cas, la relecture sert essentiellement à préparer la prochaine action tandis que, dans le second cas, il s'agit d'un mécanisme de révision en profondeur du contenu de la carte.

Tableau XXV. Distribution des stratégies de révision mobilisée par les huit participants

Id	Indicateurs des stratégies	Participant							
		Modalité informatisée				Modalité papier-crayon			
		1	2	3	4	5	6	7	8
	Relecture								
R.1	Relire des propositions	11	5	4	7	1	1	3	7
	Modification								
R.2	Corriger une ou plusieurs relations	5	5	0	1	0	1	2	4
R.3	Corriger l'étiquette d'un concept	0	0	1	0	0	1	0	0
R.4	Remplacer une ou plusieurs propositions par d'autres propositions	1	4	2	4	0	0	0	1
R.5	Supprimer une ou des relations	2	0	0	0	0	0	1	0
	Sous-total	8	9	3	5	0	2	3	5
	Diagnostic								
R.6	Évaluer le contenu de la carte	2	0	1	0	0	0	1	2
R.7	Chercher des améliorations potentielles	0	2	0	0	0	1	0	0
R.8	Vérifier le nombre de concepts	0	0	0	0	1	0	0	0
	Sous-total	2	4	4	1	1	1	1	2

5.5.5 Les difficultés techniques et les contraintes

Le dernier volet de ces analyses porte sur les difficultés et les contraintes rencontrées au cours de la tâche (voir le Tableau XXVI). Les trois catégories de difficultés techniques (D.1 à D.3) ont seulement été répertoriées chez les participants de la modalité informatisée. La première, qui est la plus fréquente, concerne les situations lors desquelles le participant tente d'effectuer une opération, par exemple de relier deux concepts, mais fait un clic au mauvais endroit et un concept vide apparaît. Dans une telle situation, le participant doit généralement effacer la bulle vide et recommencer son opération. La seconde difficulté, soit les problèmes liés à la méconnaissance du logiciel, a seulement été observée chez le participant 2. Au cours de la tâche, celui-ci a passé environ 30 secondes à chercher un moyen d'inverser une relation, notamment en explorant le menu et les options graphiques du logiciel CMapTools, mais, incapable de trouver une solution satisfaisante, il s'est résigné à effacer la relation et à la réécrire en traçant le lien dans le sens inverse. Enfin, le participant 4 s'est retrouvé à opter pour une

solution similaire lorsque plusieurs éléments de sa carte se sont mélangés entre eux, après qu'il ait déplacé un groupe de concepts et de relations au-dessus d'un autre groupe. N'arrivant pas à séparer les concepts, les liens et les expressions de liaison de chaque groupe, le participant n'a pas vu d'autres solutions que d'effacer un bon nombre de relations et de reconstituer une partie du contenu de sa carte.

Tableau XXVI. Distribution des difficultés et des contraintes rencontrées par les huit participants

Id	Indicateurs	Participant							
		Modalité informatisée				Modalité papier-crayon			
		1	2	3	4	5	6	7	8
D.1	Difficulté menant à la création d'une bulle vide	1	6	2	3	0	0	0	0
D.2	Difficulté liée à la méconnaissance du logiciel	0	1	0	0	0	0	0	0
D.3	Difficulté liée à la superposition de concepts	0	0	0	2	0	0	0	0
D.4	Commentaire lié à l'utilisation d'un verbe comme expressions de liaison	6	1	1	0	0	0	0	0
D.5	Commentaire lié à la mise en relation d'une expression de liaison avec une autre expression de liaison	0	3	0	0	0	1	0	0
D.6	Commentaire lié au croisement des liens	0	1	0	0	2	0	0	0
D.7	Commentaire lié à l'utilisation d'un ouvrage de référence	1	0	1	0	0	0	0	0
D.8	Commentaire lié à l'utilisation de l'espace de la page	0	0	0	0	2	1	2	2
D.9	Représentation non fidèle d'une idée	0	1	0	1	0	0	0	0

En plus des difficultés qui ont pu être observées directement, le Tableau XXVI répertorie divers commentaires des participants liés aux consignes et contraintes de la tâche de cartographie conceptuelle. Les trois premiers commentaires (D.4 à D.6), en particulier, expriment une réflexion ou une question au sujet des consignes de la tâche et reflètent peut-être, à un certain degré, un manque d'expérience avec la cartographie conceptuelle. Le commentaire D.4, par exemple, met en évidence la difficulté de certains individus, en particulier

le participant 1, à formuler des expressions de liaison comprenant un verbe. Le commentaire D.7 indique que les participants n'étaient pas tout à fait prêts pour la tâche et qu'ils maîtrisaient plus ou moins certains concepts : deux participants ont formulé le souhait de consulter leur manuel scolaire lors de la tâche. Enfin, le commentaire D.8 est unique à la modalité papier-crayon. Les quatre participants ont exprimé, à un moment ou à un autre, qu'il était difficile d'exploiter convenablement l'espace de la feuille dans laquelle il traçait la carte conceptuelle. Après avoir passé 20 minutes à construire sa carte, le participant 8 s'est même demandé, s'il n'a pas intérêt à la recommencer sur une nouvelle feuille.

Enfin, le dernier indicateur inventorié (D.9) relève d'une situation particulière observée lors des rencontres avec les participants 2 et 4. Tous les deux ont à un certain moment fait des actions qui ne représentent pas fidèlement l'idée qu'ils ont exprimée oralement. L'extrait du participant 4, rapporté à la Figure 11, donne un exemple d'un tel écart entre la pensée et le contenu de la carte. Le participant veut exprimer l'idée que l'ensemble des relations entre les consommateurs primaires, secondaires et tertiaires constitue un exemple d'une chaîne alimentaire, mais ce n'est pas ce qui apparaît dans sa carte conceptuelle. Il formule plutôt la « *chaîne alimentaire Ex consommateurs primaires* ». Le contenu se trouve donc en décalage avec son intention de départ.

Ce dernier exemple, en particulier, rappelle que le contenu d'une carte conceptuelle ne représente pas complètement ni toujours fidèlement les connaissances d'un individu. Le participant 8, en réfléchissant aux avantages de la tâche, mentionne d'ailleurs que : « Je trouve que ça m'a beaucoup aidé de le faire et de le dire. Parce que beaucoup de liens que je n'ai pas faits sur papier, mais j'ai fait oralement ». Pour tirer profit pleinement de la carte conceptuelle, il importe donc de bien réfléchir à ses usages de même qu'aux conditions dans lesquelles celles-ci sont mises à contribution.

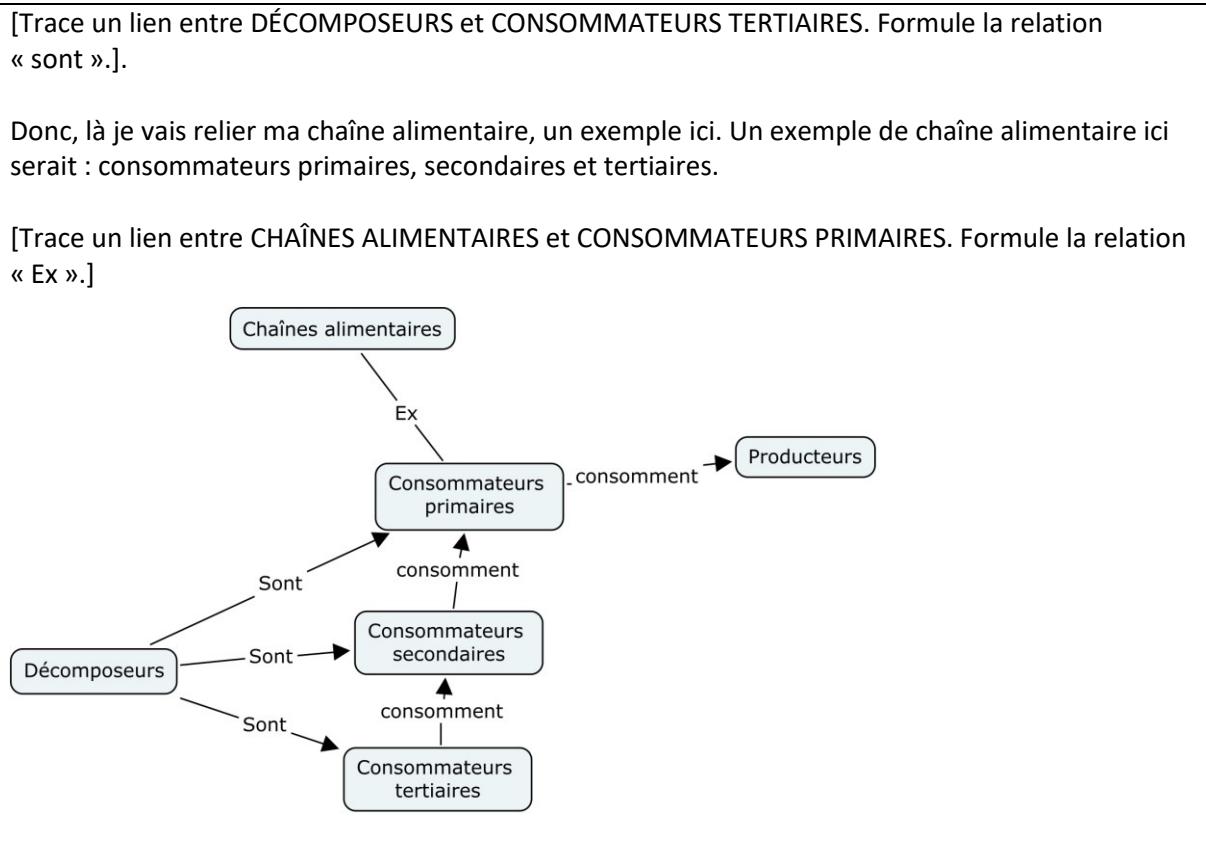


Figure 11. Extrait du participant 4 illustrant un écart entre la pensée et le contenu de la carte

5.6 Discussion

L'étude présentée dans cet article avait pour but de documenter les démarches de production d'une carte conceptuelle en modalité manuscrite et en modalité informatisée. Une meilleure compréhension de ces processus s'avère une condition essentielle pour adapter adéquatement la formation préalable à la cartographie conceptuelle aux particularités inhérentes

au support de production. En combinant les données issues de la verbalisation de la pensée et celles de l'observation directe des actions des participants, il a été possible de faire ressortir les caractéristiques de trois types de démarches : une démarche qui consiste à élaborer les propositions sur-le-champ, une démarche centrée sur l'anticipation des relations et une démarche centrée sur la catégorisation des concepts.

Comme il a été décrit lors de l'analyse des résultats, les participants ont tendance à rédiger les propositions sur-le-champ lorsqu'ils effectuent la tâche à l'ordinateur, alors que les individus vont repousser à plus tard la mise en relation des concepts lorsqu'ils font la même tâche de manière manuscrite. Cela fait en sorte que l'individu, à l'écrit, a le temps de se former une image mentale plus complète de sa carte conceptuelle avant de commencer à tracer les liens entre les concepts, et ce qui explique peut-être, qu'il manifeste moins d'opérations de révision. À l'inverse, dans la modalité informatisée, cette représentation mentale se forme au moment même de la construction de la carte, et cela expliquerait pourquoi les participants ayant utilisé cette modalité révisent et corrigent plus souvent le contenu de leur carte au cours de la tâche. La facilité à modifier le contenu d'une carte à l'ordinateur est d'ailleurs un autre facteur contribuant probablement aux différences observées relativement à la fréquence des stratégies de modification du contenu (Watson et coll., 2016). Ces observations concordent d'ailleurs avec les propos d'un enseignant rapportés dans l'article de Royer et Royer (2004), indiquant que les élèves ont manifesté plus d'intérêt à réviser leur carte à l'ordinateur qu'à l'écrit.

À plusieurs égards, les résultats de cette étude rejoignent ceux de travaux antérieurs, dont ceux dans le domaine de la recherche sur les processus d'écriture. En effet, il existe plusieurs parallèles à faire entre la production d'une carte conceptuelle et la production d'un texte écrit. Dans un cas comme dans l'autre, l'utilisation d'un outil informatisée fait en sorte qu'un individu met en œuvre moins de stratégies de planification au début de la tâche ; il se lance plus rapidement dans la mise en texte (ou la mise en forme de la carte) ; et procède à des stratégies de planification plus ciblées (Haas, 1989). La mise en texte (ou la mise en relation des concepts) est quant à elle plus fragmentaire et est plus souvent interrompue par d'autres processus (Diarra, 2013 ; Van Waes et Schellens, 2003). Enfin, en ce qui concerne la révision, même s'il semble que l'outil informatique favorise les opérations d'ajout, de suppression ou de remplacement dans la production (Diarra, 2013), les individus portent plus d'attention à la révision des éléments en

surface du texte ou de la carte (Bridwell-Bowles, Johnson et Brehe, 1987 ; Hult, 1988 ; Lutz, 1987 ; Van Waes et Schellens, 2003) et moins à la révision en profondeur du contenu de la production.

Bien entendu, de tels résultats méritent d'être nuancés. D'abord, la présente étude, qui prend appui sur un échantillon de convenance composé seulement de huit cas, ne prétend pas avoir atteint le seuil de saturation des données. Un nombre plus grand de participants et un échantillonnage plus hétérogène (les participants avaient entre 14 et 16 ans et étaient issus de la même école et du même niveau de scolarité) auraient pu aboutir à une plus grande variété de démarches de cartographie conceptuelle. Il est toutefois intéressant de noter que, même dans les conditions limitées de cette recherche, quelques démarches différentes ont été observées. Ensuite, il faut rappeler que la transcription et le codage final n'ont été effectués que par un seul chercheur pour des raisons logistiques. Si le temps et les ressources l'avaient permis, il aurait été utile d'effectuer un exercice de contre-codage pour vérifier le degré de fidélité du codage, mais aussi pour clarifier le lexique de codes. Une autre stratégie pour s'assurer de la cohérence du codage aurait été de vérifier la fidélité intracodeur. Puis, il faut s'apercevoir que certains facteurs et variables n'ont pas été contrôlés dans cette étude. Dans leur étude, Ruiz-Primo, Shavelson, et coll. (2001) ont par exemple observé des démarches différentes selon le niveau de maîtrise préalable des connaissances disciplinaires. Lim, Lee et Grabowski (2009) ont quant à eux montré que les individus bénéficient davantage de la cartographie conceptuelle comme stratégie d'apprentissage lorsqu'ils démontrent un niveau plus élevé d'habiletés d'autorégulation des apprentissages. Contrôler le niveau de connaissances antérieures et les habiletés d'autorégulation des apprentissages serait en toute logique un moyen d'approfondir l'étude des démarches des participants. Les habiletés en informatique des individus, ainsi que leur attitude et leur niveau d'anxiété face à l'utilisation de l'ordinateur sont également des variables qui méritent d'être contrôlées lors de prochaines études sur les différences entre l'évaluation traditionnelle sur support papier et l'évaluation assistée par ordinateur (Erdogan, 2009 ; McDonald, 2002 ; Shudong, Hong, Young, Brooks et Olson, 2007). Enfin, il serait pertinent d'examiner scrupuleusement les caractéristiques spécifiques du support de passation. Les comportements des participants auraient pu être différents si un autre logiciel de

cartographie conceptuelle avait été utilisé (Weinerth et coll., 2014) ou encore si les participants n'avaient pas reçu d'étiquettes amovibles.

À cet égard, Weinerth et coll. (2014) ont bien montré que les écrits au sujet de la cartographie conceptuelle se sont peu intéressés à la facilité d'utilisation (*usability*) des programmes informatiques servant à produire des cartes conceptuelles et que, de manière générale, les études qui y ont accordé de l'attention l'ont fait de façon superficielle. La facilité d'utilisation fait pourtant partie des principes directeurs en matière de testing et d'évaluation assistée par ordinateur (*computer-based testing*) depuis de nombreuses années (The International Test Commission, 2006). Une réflexion similaire peut avoir lieu quant aux diverses façons de produire une carte conceptuelle sur support papier. Il apparaît donc important de tenir compte du contexte (par ex., le contenu des tâches) et des conditions dans lesquels ces stratégies sont mises en œuvre.

La maîtrise de la technique de cartographie conceptuelle est d'ailleurs un élément contextuel à prendre en considération. Le cas particulier du participant 8 soulève quelques questionnements à cet égard, notamment au sujet de la transférabilité des habiletés spécifiques d'une modalité à une autre. La démarche qu'il a suivie lors de la tâche principale, faite sur support papier, ainsi que les difficultés qu'il a rencontrées laissent penser que celui-ci a peut-être essayé de transférer telles quelles les stratégies développées lors des activités préparatoires faites à l'ordinateur.

La formation est sans contredit un enjeu clé afin de s'assurer que les individus maîtrisent la technique de la cartographie conceptuelle. Non seulement la formation doit permettre à l'individu de développer certaines stratégies propres à cette technique de représentation des connaissances, mais elle doit aussi leur permettre de maîtriser certaines stratégies spécifiques à la modalité de production de la carte. Dans le cadre de cette étude, même si la formation offerte aux élèves était plus longue que celles proposées dans plusieurs autres travaux de recherche (Dogusoy-Taylan et Cagiltay, 2014 ; Gurlitt et Renkl, 2010 ; Hilbert et Renkl, 2008 ; Ruiz-Primo, Schultz, et coll., 2001 ; Ruiz-Primo, Shavelson, et coll., 2001 ; Yin et coll., 2005), plusieurs observations indiquent que la formation n'était pourtant pas complète. Ces observations rejoignent celles de Hilbert et Renkl (2008), tel que discuté dans la section présentant le cadre théorique. Les programmes de formation plus longs, étalés sur plusieurs

séances et composés d'activités variées, semblent être plus appropriés pour s'assurer que les individus maîtrisent suffisamment la technique de la cartographie conceptuelle (Bartels, 1996 ; Bell, 2018 ; Rice et coll., 1998).

La formation à la cartographie conceptuelle est d'ailleurs importante, peu importe le contexte. En situation d'apprentissage, comme l'ont souligné, le respect des consignes, comme l'étiquetage des relations, ou encore la mise à contribution de certaines stratégies de planification et de contrôle du contenu sont essentielles pour profiter pleinement des avantages pédagogiques de cette technique de représentation des connaissances (Hilbert et Renkl, 2008). Un enseignant a donc la responsabilité de mettre en place une séquence d'apprentissage appropriée pour ses étudiants ou élèves de manière à les aider à maîtriser ces consignes et ces stratégies.

En contexte d'évaluation des apprentissages, la maîtrise de la technique de la cartographie conceptuelle est quant à elle en enjeu critique lié à la validité des interprétations des résultats. Idéalement, les conditions de passation d'une tâche d'évaluation ne doivent pas interférer avec ce qui est évalué. Ni l'expérience antérieure par rapport à la cartographie conceptuelle ni le choix des modalités ne doivent faire obstacle à l'extériorisation des connaissances de l'individu, sans quoi il s'agit de menaces potentielles à la validité (Sireci et Zenisky, 2015). La formation préalable est un mécanisme qui sert autant que possible à minimiser la contamination des données par des facteurs indésirables.

5.7 Conclusion

En somme, les principales contributions de cet article sont d'avoir mis en évidence les caractéristiques de trois types de démarches de production d'une carte conceptuelle, mais aussi de rappeler, à l'aide d'exemples concrets, l'importance d'offrir une formation appropriée aux individus qui tracent une carte conceptuelle. Former des individus à la construction de cartes conceptuelles demeure un défi important, mais nécessaire afin de limiter la contamination des données collectées au moyen de cartes conceptuelles (Shemwell, Fu, Figueroa, Davis et Shavelson, 2010). Plusieurs chercheurs ont soulevé l'utilité de la cartographie conceptuelle (voir notamment la méta-analyse de Nesbit et Adesope, 2006), soit comme outil d'apprentissage, soit comme outil de collecte de données en vue d'évaluer les apprentissages, mais ces prétendues

qualités ne peuvent se concrétiser sans s'attarder aux modalités de production d'une carte conceptuelle, incluant celles associées à la formation.

Chapitre 6 Production semi-automatisée d'une carte conceptuelle en science et technologie⁴²

Maxim Morin

Jean-Guy Blais

Université de Montréal

Mots clés: génération automatique de cartes conceptuelles, carte conceptuelle, analyse de texte, exploration de données, traitement automatique du langage naturel

Cet article décrit la conception et la mise à l'essai d'un programme informatique de production semi-automatisée de cartes conceptuelles. Cette technique dérivée de l'extraction d'information a pour but de produire une représentation simple et signifiante du contenu d'un ou de plusieurs textes sous forme d'un schéma de connaissances. Le programme a été mis à l'essai à partir de deux corpus de textes choisis pour couvrir deux thèmes du Programme de formation de l'école québécoise de quatrième secondaire en science et technologie. Les résultats de l'étude sont très encourageants et montrent le potentiel d'une telle approche pour faciliter la construction d'une carte conceptuelle, une tâche qui est généralement réalisée par des humains. Les résultats suscitent également quelques réflexions quant aux modalités courantes d'élaboration et d'évaluation de cartes conceptuelles.

6.1 Introduction

La cartographie conceptuelle est une technique d'organisation et de représentation des connaissances largement répandue en éducation qui a fait ses preuves à la fois comme stratégie d'enseignement, stratégie d'apprentissage et stratégie d'évaluation des apprentissages (Cañas et

⁴² Publié dans *Mesure et évaluation en éducation* : Morin, M. et Blais, J.-G. (2017). Production semi-automatisée d'une carte conceptuelle en science et technologie. *Mesure et évaluation en éducation*, 40(1), 61-99. doi: <https://doi.org/10.7202/1041004ar>.

Ce chapitre est une version modifiée du manuscrit final. Les titres, les figures et les tableaux ont été renumérotés afin de s'harmoniser avec le format et les contraintes d'une thèse par articles.

al., 2003 ; Marchand et d'Ivernois, 2004). Jusqu'à tout récemment, l'élaboration d'une carte conceptuelle était une tâche réalisée par un ou des humains, par exemple un expert ou un panel d'experts. Les progrès technologiques en matière d'intelligence artificielle, de traitement automatique du langage naturel (TALN) et d'exploration des données (*data mining*) ont depuis peu ouvert la porte à des techniques de production automatisée ou semi-automatisée de cartes à partir de documents textuels. Les applications potentielles de telles techniques sont nombreuses, que ce soit pour résumer un texte (Chen, Kinshuk, Wei & Chen, 2008), pour corriger et noter automatiquement des questions à réponse élaborée (Clariana & Koul, 2004 ; Koul, Clariana & Salehi, 2005), pour créer une base de connaissances pour des systèmes tutoriels intelligents (Nielsen, Ward, Martin & Palmer, 2008) ou encore pour servir d'outil de soutien de correction lors de l'évaluation de cartes conceptuelles élaborées par des élèves.

Le présent article se penche sur la conception et la mise à l'essai d'un programme informatique permettant la production semi-automatisée d'une carte conceptuelle. Comme il sera expliqué plus loin, cette démarche s'inscrit dans une recherche plus vaste dont les objectifs sont, de façon générale, d'étudier l'usage de la cartographie conceptuelle comme instrument d'évaluation des apprentissages d'élèves et, de manière plus précise, d'étudier les propriétés psychométriques des données collectées au moyen des cartes conceptuelles.

La suite de cet article se divise en sept sections : le cadre théorique, le contexte de la recherche, le fonctionnement du programme, le cadre méthodologique, les résultats de la mise à l'essai, la discussion et la conclusion.

6.2 Cadre théorique

Les appellations « carte conceptuelle » ou « cartographie conceptuelle » sont employées de plusieurs façons dans les écrits scientifiques ou pédagogiques. La première sous-section de ce cadre théorique définit ce qui est entendu par cartographie conceptuelle dans cette recherche. La seconde sous-section décrit plus spécifiquement les divers procédés et techniques qui peuvent être mis à profit lors de la production automatisée, tandis que la troisième sous-section fait état de quelques projets de recherches dans ce domaine. Une synthèse terminera cette section sur le cadre théorique.

6.2.1 Définition de la cartographie conceptuelle

D'abord utilisée comme technique de synthèse d'entretiens de recherche par Joseph D. Novak (1977), la cartographie conceptuelle s'est rapidement révélée comme étant une stratégie d'enseignement et d'apprentissage efficace en sciences (Novak, Gowin & Johansen, 1983) et dans plusieurs autres domaines (Nesbit & Adesope, 2006). Pour Novak, une carte conceptuelle est un réseau composé de nœuds et d'arcs organisés de manière semi-hiérarchique, qui a comme objectif « de représenter les relations significantes entre les concepts sous forme de propositions » (Novak & Gowin, 1984, p. 13). D'un point de vue graphique, le nœud est formé d'une figure géométrique et l'arc est un trait fléché. Du point de vue sémantique, le nœud exprime un concept, c'est-à-dire « une régularité perçue (ou un patron) d'évènements ou d'objets, ou de répertoires d'évènements ou d'objets, désignée par une étiquette » (Novak & Cañas, 2008, p. 10). Quant à l'arc, il est accompagné d'une expression de liaison définissant la relation entre deux concepts. La mise en relation de deux ou plusieurs concepts forme des propositions, c'est-à-dire des énoncés de sens significants.

La démarche de construction d'une carte conceptuelle est importante aux yeux de Novak. Ce spécialiste des sciences de l'éducation encourage l'apprenant à participer activement à la construction de ses apprentissages, notamment en l'invitant à choisir lui-même ses libellés de concepts et à énoncer les relations entre ceux-ci. Qui plus est, Novak prône une organisation semi-hiérarchique des concepts et la formation de relations dites transversales, c'est-à-dire des relations entre des concepts éloignés dans la carte conceptuelle.

La définition de la cartographie conceptuelle promue par Novak est probablement la plus répandue, mais ce n'est pas la seule. Sa définition ou d'autres sont souvent adaptées, implicitement ou explicitement, au contexte pédagogique ou de recherche. À la fin des années 1990 et au début des années 2000, une équipe de chercheurs du *Center for Research on Evaluation, Standards and Student Testing* (CRESST) a étudié diverses modalités de la cartographie conceptuelle pour évaluer les apprentissages d'élèves, principalement dans le domaine de l'éducation scientifique (Herl, O'Neil, Chung & Schacter, 1999; Klein, Chung, Osmundson & Herl, 2002; Ruiz-Primo, Schultz, Li & Shavelson, 2001; Ruiz-Primo & Shavelson, 1996; Ruiz-Primo, Shavelson, Li & Schultz, 2001; Yin, Vanidez, Ruiz-Primo, Ayala & Shavelson, 2005). Leur définition de la cartographie conceptuelle est essentiellement la même

que celle de Novak, sauf qu'ils ne requièrent pas que la carte soit organisée de manière hiérarchique. Leurs travaux de recherche se démarquent aussi par l'éventail de modalités d'élaboration des cartes conceptuelles. Les chercheurs du CRESST s'intéressent aux tâches du type « remplir les trous d'une carte » ainsi qu'aux tâches qui consistent à élaborer une carte à partir de concepts ou de mots de liaison préétablis (Ruiz-Primo, Schultz et al., 2001 ; Ruiz-Primo, Shavelson et al., 2001 ; Yin & Shavelson, 2008 ; Yin et al., 2005). Puisqu'elles sont destinées à l'évaluation des apprentissages, ces modalités sont en général plus contraignantes que celles de Novak. Dans ces conditions, les cartes produites par les élèves sont plus uniformes, ce qui permet d'appliquer ultérieurement des modèles de notation quantitative.

Bref, la carte conceptuelle est un outil d'explicitation et d'organisation des connaissances qui sert essentiellement à l'humain (Valerio & Leake, 2006). Le contenu d'une carte est très souvent exprimé en langage naturel, ce qui en fait une représentation « informelle » des connaissances, au contraire d'autres représentations externes des connaissances comme le graphe conceptuel (Sowa, 2000) qui, lui, est de type « formel » et spécialement conçu pour être traité par un logiciel. La composition graphique et textuelle de la carte conceptuelle en fait un outil tout à fait adapté à la représentation de connaissances conceptuelles sous la forme de relations binaires, c'est-à-dire de triplets « concept, relation, concept ». Elle s'adapte moins bien à d'autres types de relations. Les relations unaires ne sont tout simplement pas permises, tandis que les relations ternaires sont inévitablement divisées en deux relations binaires (*Jean donne un livre, livre à Johanne*) ou remplacées par une seule relation binaire (p. ex., *Jean donne un livre à Johanne*). La Figure 12 illustre des relations de chaque ordre.

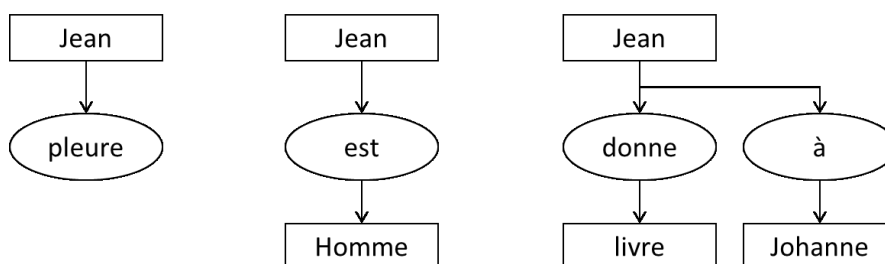


Figure 12. Exemple de relations unaires, binaires et ternaires

6.2.2 Processus d'extraction d'information

La production automatisée d'une carte conceptuelle à partir de documents est essentiellement un problème d'extraction d'information au cours duquel un système extrait et organise des données textuelles en une représentation traitable par l'ordinateur et pertinente pour une application particulière. Typiquement, ces représentations de concepts et relations sont des structures relativement formelles (p. ex., des ontologies informatiques) et servent de bases de données pour d'autres programmes informatiques. En revanche, la production automatisée d'une carte a très souvent une visée éducationnelle et didactique. Selon Villalon et Calvo (2008), la carte générée doit refléter les caractéristiques fondamentales de la cartographie conceptuelle, et doit être une représentation de connaissances simple et compréhensible pour ses utilisateurs.

Une stratégie de production automatisée ou semi-automatisée d'une carte conceptuelle est un assemblage cohérent de techniques et de procédés statistiques ou linguistiques. Cette stratégie peut être divisée en quatre étapes : 1) le prétraitement des textes, 2) la reconnaissance des concepts, 3) la reconnaissance des relations, et 4) la synthèse de la carte.

Étant donné que l'ordinateur n'a pas la capacité de lire un texte comme un humain, il est nécessaire de réaliser une série d'opérations linguistiques permettant de construire une représentation compréhensible des textes pour la machine. Il s'agit de l'étape de prétraitement des textes. Les procédés courants sont les suivants :

- la *tokenisation* : processus de segmentation des textes en phrases et en mots ;
- la *normalisation des mots* : processus de représentation des mots par leur forme canonique (lemmatisation) ou leur racine (racinisation ou, en anglais, *stemming*) ;
- la *désambiguïsation des mots* : processus d'identification du sens d'un mot dans une phrase quand le mot est polysémique. Dans le contexte de la production d'une carte conceptuelle, ce processus se traduit entre autres par l'abstraction d'un concept à partir de plusieurs mots ou groupes de mots qui signifient la même chose, notamment des synonymes ;
- la *résolution d'anaphores* : processus par lequel les pronoms personnels sont remplacés par leur antécédent dans le texte ;

- l'*analyse morphosyntaxique* : processus d'analyse de la phrase qui reconnaît la forme morphologique de chaque mot (catégorie de mot, genre, nombre, personnes, etc.) et la structure de la phrase selon un formalisme donné. Ce formalisme est nommé une grammaire. On distingue généralement les grammaires en constituants et les grammaires de dépendance. La Figure 13 donne un exemple d'un arbre syntaxique selon une grammaire de dépendance. Le verbe gouverneur «représentées» est la racine de cet arbre, le mot «relations» est le sujet de la phrase et la préposition «par» introduit le complément d'objet indirect. Chaque mot de la phrase est étiqueté ainsi.

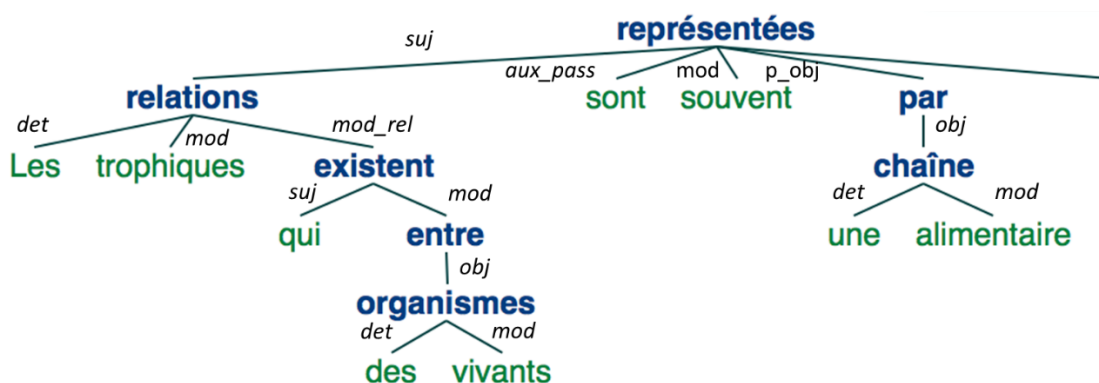


Figure 13. Arbre de dépendance syntaxique de la phrase

La seconde étape d'une démarche de construction automatisée d'une carte conceptuelle consiste à identifier les termes composés d'un ou de plusieurs mots qui désignent les concepts ou de potentiels concepts. Selon les besoins de l'utilisateur, le système peut repérer des termes prédéterminés, traiter les variations morphologiques des mots, proposer de nouveaux concepts ou même tenter de regrouper des termes similaires sous un seul concept. Le niveau de sophistication du logiciel croît rapidement selon les mécanismes choisis.

La troisième étape consiste à reconnaître les relations entre les concepts. Alors que des méthodes statistiques (p. ex., compter le nombre de mots séparant deux concepts) permettent de détecter la présence d'une relation, ce sont des méthodes linguistiques, fondées sur l'analyse

morphosyntaxique de la phrase, qui reconnaissent les expressions de liaison formant les propositions.

La dernière étape du processus d'extraction a pour but de synthétiser l'information retenue lors des étapes précédentes. Selon Novak et Cañas (2008), une carte conceptuelle doit comporter entre 15 et 25 nœuds, mais il n'est pas rare que l'inventaire fait par le système dépasse ces balises. Lorsque cela se produit, les techniques pour synthétiser l'information sont nombreuses. Elles peuvent consister à sélectionner les concepts et les relations les plus signifiants ou encore à optimiser certaines propriétés structurelles du réseau (p. ex., éliminer le morcèlement de la carte en réseaux distincts). Une fois les propositions choisies, elles peuvent être assemblées en une carte conceptuelle.

6.2.3 Recension des écrits

Un logiciel de production semi-automatisée de cartes conceptuelles est donc une combinaison de différentes techniques en vue d'effectuer une tâche précise. Quatre stratégies issues du domaine de l'éducation ou des sciences informatiques sont présentées dans cette section.

Clariana et Koul (2004) ont mis sur pied une stratégie d'extraction d'une variante d'un réseau de concepts, nommée un PFNet (abréviation de *pathfinder network*), en vue de la notation des productions écrites d'élèves. Dans leur approche, chaque production écrite est transformée en un PFNet, pour ensuite être comparée au réseau PFNet construit à partir du texte d'un expert de contenu. Plus spécifiquement, la stratégie mise en œuvre est la suivante. Un expert de contenu détermine une liste de concepts et de synonymes. Puis, le logiciel ALA-Reader relève ces termes dans les textes et détermine toutes les paires de concepts qui apparaissent dans les mêmes phrases. Enfin, les cooccurrences des paires de concepts sont analysées au moyen d'un logiciel d'analyse de grappes (*cluster analysis*) nommé PCKNOT (Schvaneveldt, 1990) de sorte à produire une représentation simplifiée du texte, le PFNet. Ces réseaux servent enfin à la notation des productions écrites des élèves. Dans leur court article, les auteurs portent leur attention sur l'analyse des propriétés psychométriques des productions écrites, et non sur les performances de leur stratégie. Ainsi, ils donnent très peu d'information sur le rendement du logiciel.

Valerio et ses collaborateurs (Valerio & Leake, 2006 ; Valerio, Leake & Cañas, 2012) ont développé une application pour assister la synthèse, la compréhension et le référencement de documents. D'un point de vue technique, leur système est plus sophistiqué que celui de Clariana et Koul (2004). Il procède à la reconnaissance de termes composés de plusieurs mots, à la résolution d'anaphores et à l'identification de relations de synonymie. La détection de structures syntaxiques du type « nom, verbe, nom » sert ensuite à reconnaître les expressions de liaison. Enfin, même s'ils explorent de façon informelle certaines méthodes de synthèse des concepts et relations (p. ex., choisir les n concepts les plus fréquents), Valerio et ses collaborateurs n'intègrent aucune méthode dans leur solution finale. Malgré l'absence d'information au sujet de la taille des textes, tout laisse croire que ceux-ci sont suffisamment courts et que les cartes générées ne comptent pas au-delà de 30 concepts, c'est-à-dire le nombre maximal acceptable de concepts considéré par ces chercheurs.

Dans le domaine des sciences informatiques, Zouaq et Nkambou (2008) ont développé un logiciel de production de cartes conceptuelles, appelée TEXCOMON, dans le but de constituer une ontologie informatique. L'analyse morphosyntaxique en grammaire de dépendance est au cœur de leur approche. Ils s'en servent pour identifier des termes composés de plusieurs mots et pour reconnaître les relations de type « prédicat, argument ». Pour ce faire, l'analyseur syntaxique repère les mots qui modifient les noms (p. ex., l'adjectif « sémantique » modifie le nom « web » dans le terme « web sémantique ») et les triplets du type « sujet, verbe, complément d'objet », couramment appelés des patrons SVO. Le logiciel TEXCOMON identifie également divers patrons lexico-syntaxiques (Hearst, 1992) qui témoignent de relations d'hyponymie et d'hyperonymie entre les concepts. Enfin, ce logiciel ne requiert pas de procédé de synthèse de l'information étant donné qu'il a pour but de construire une ontologie informatique, et non des cartes conceptuelles. Une ontologie informatique contient généralement beaucoup plus de termes et de relations qu'une carte conceptuelle, et les corpus analysés par TEXCOMON sont beaucoup plus volumineux. Sept corpus composés de 10 à 36 textes et de 728 à 1 578 phrases ont été étudiés lors de leur mise à l'essai.

Qasim et ses collaborateurs (2013) se sont inspirés des procédés de TEXCOMON pour la création d'une carte conceptuelle portant sur le thème des sciences informatiques. Ils emploient les mêmes stratégies pour reconnaître les concepts et les relations. Par contre, ils

échafaudent une solution complexe combinant plusieurs méthodes statistiques pour faire la synthèse du contenu des textes. En comparaison aux trois études précédentes, leur tâche de synthèse est colossale : 65 textes sont analysés simultanément par le logiciel. Pour y arriver, ils recourent à une statistique de pondération de l'information pour déterminer les expressions de liaison les plus significatives et à une technique d'analyse de grappes nommée la propagation d'affinité pour réduire l'inventaire de concepts et de relations. Il n'est pas possible d'évaluer le rendement effectif de ces procédés puisque les auteurs ne mentionnent pas le nombre de concepts et de relations qui ont été repérés dans la phase initiale ni combien ont été retenus dans la carte finale.

6.2.4 Synthèse

Même si elles partagent quelques points en commun, les quatre démarches de production de cartes conceptuelles ont des caractéristiques qui les distinguent. D'abord, chaque démarche adapte quelque peu la définition de la cartographie conceptuelle à son contexte d'utilisation. Dans les études décrites ci-dessus, les réseaux PFNet de l'étude de Clariana et Koul ne comportent pas de mots de relation, tandis que les réseaux générés lors des mises à l'essai de Zouaq et Nkambou (2008) ou encore de Qasim et ses collaborateurs (2013) comptent plus de 25 concepts. Aucune étude n'organise les concepts de manière hiérarchique. Les principes guidant la construction d'une carte conceptuelle dépendent de l'usage qui en est fait.

Deuxièmement, le degré d'automatisation varie d'un système à l'autre. Chacune des solutions discutées comporte des étapes manuelles. Dans Clariana et Koul (2004), c'est l'utilisateur qui fournit la liste de concepts et, dans trois des quatre systèmes, c'est l'humain qui a la responsabilité d'assembler la carte finale. Seulement Valerio et Leake (2006) ont eu recours aux fonctionnalités du logiciel CmapTools pour former une représentation graphique initiale de la carte conceptuelle. Vu que l'utilisateur intervient à une étape ou l'autre de la démarche, il apparaît plus juste de parler de production semi-automatisée d'une carte conceptuelle ou encore de production d'une carte conceptuelle assistée par ordinateur.

Troisièmement, les indicateurs de performance et les niveaux de tolérance ne sont pas les mêmes d'un logiciel à l'autre puisqu'ils ne répondent pas aux mêmes buts. Les programmes informatiques incorporant des techniques de traitement automatique du langage naturel sont

inévitablement faillibles. Cela dit, il y a bien des situations où un traitement incomplet ou légèrement erroné d'un texte aura un effet négligeable sur le rendement final du logiciel. Par exemple, pour le référencement de textes, une représentation même partielle du texte est suffisante pour relier un texte à un ensemble de mots clés et pour distinguer les textes entre eux. À l'inverse, il y a d'autres situations plus exigeantes d'un point de vue technique, comme la notation de productions écrites, où il est attendu que le logiciel produise une représentation riche et juste de chaque texte écrit.

Dans le même ordre d'idées, le rendement du logiciel dépend aussi du nombre, de la qualité et de la pertinence des textes analysés. Toutes choses étant égales par ailleurs, le système restitue plus de concepts et de relations lorsque le corpus de textes est volumineux et lorsque les techniques de traitement du langage sont adaptées au style, au format et à la qualité de l'écriture. Les propositions inventoriées sont quant à elles plus significatives lorsque les textes touchent spécifiquement le thème ciblé. En somme, autant la définition de la cartographie conceptuelle que la stratégie d'extraction doivent s'aligner avec les objectifs poursuivis et le contexte d'utilisation du logiciel. La prochaine section s'attarde au contexte dans lequel le présent logiciel a été développé.

6.3 Contexte

La production automatisée d'une carte conceptuelle à partir de textes est rarement une finalité en soi ; elle sert d'autres buts. Dans le cadre de ce projet, elle s'inscrit dans une recherche plus large qui étudie les propriétés psychométriques des données collectées au moyen de cartes conceptuelles construites par des élèves en vue de l'évaluation de leurs apprentissages. Plus précisément, deux tâches du type « élaborer une carte conceptuelle à partir d'une liste préétablie de concepts » et conformes avec la définition de la cartographie conceptuelle du CRESST ont été soumises à des élèves de quatrième secondaire en science et technologie.

L'usage d'une carte de référence est une pratique courante dans plusieurs modalités de correction et de notation de cartes conceptuelles. Elle peut servir d'outil d'appoint ou encore de clé de correction et permet d'obtenir d'ailleurs une plus grande fidélité interjuges (McClure, Sonak & Suen, 1999).

L'approche la plus courante pour élaborer une carte de référence est de demander à un ou des experts de construire une carte conceptuelle. Cette approche a pourtant ses lacunes : les représentations externes des connaissances varient d'un expert à l'autre (Acton, Johnson & Goldsmith, 1994) et ne se réconcilient pas nécessairement à celles de novices (Chi, Feltovich & Glaser, 1981 ; Chi, Glaser & Farr, 1988 ; Glaser, 1991), surtout si cette représentation n'est pas construite dans un but « didactique ». Il est possible de recourir à un panel d'experts pour réduire les singularités, mais cette stratégie est irréaliste dans plusieurs contextes. La production d'une carte conceptuelle à partir de textes tente d'offrir une solution de rechange aux approches fondées sur la construction directe d'une carte de référence par des experts de contenu. Comme il sera décrit ci-dessous, la carte conceptuelle sera construite principalement à partir de textes provenant de manuels scolaires. Ces textes sont composés et révisés par des experts du domaine de connaissances, mais, surtout, ils sont rédigés dans un but « didactique » ; leur contenu découle d'un consensus entre des experts et est plus près de ce qui est attendu de l'élève.

6.4 Objectif de la recherche

L'objectif du présent article est donc de mettre en place une stratégie d'extraction de propositions assistée par ordinateur en vue de la production d'une carte conceptuelle et dans le but éventuel de créer une carte de référence pouvant servir d'outil de correction des cartes produites par des élèves. Pour y arriver, il est attendu que les propositions extraites soient représentatives du domaine de connaissances, valides et adaptées au niveau de scolarité des élèves. La démarche de production d'une carte conceptuelle, telle que décrite à la section suivante, a donc été développée en conformité avec cet objectif et le type de tâches de cartographie conceptuelle soumises aux élèves.

6.5 Démarche de production d'une carte conceptuelle

Cette section présente la démarche de production d'une carte conceptuelle à partir de textes (voir Figure 14). Une application informatique a été développée sous le langage de programmation Python afin d'automatiser les processus de sélection des textes, de prétraitement des textes, de reconnaissances des concepts et de reconnaissances des relations. Cette application est intitulée EPTAO et désigne l'extraction (E) de propositions (P) à partir de textes

(T) et assistée (A) par ordinateur (O). Les prochaines sous-sections décrivent les composantes techniques de la démarche et d'EPTAO.

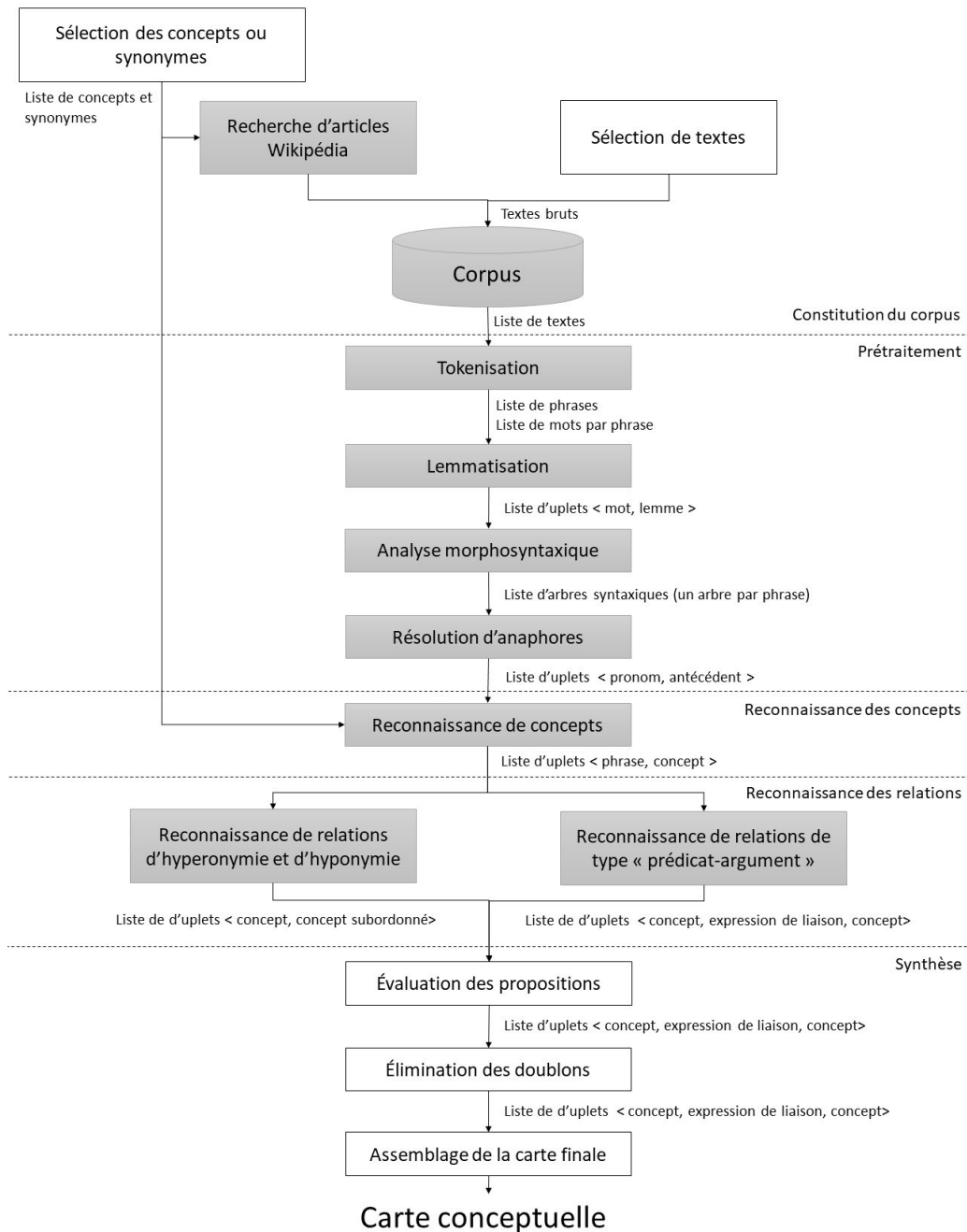


Figure 14. Démarche de production de la carte, fonctionnement d'EPTAO et modèle de données

6.5.1 Constitution du corpus de textes

EPTAO offre deux mécanismes pour former un corpus : l'injection de textes choisis par une personne ou la recherche automatisée d'articles Wikipédia associés aux concepts fournis par l'utilisateur. Tous les documents sont formatés en fichiers de texte brut.

6.5.2 Prétraitement des textes

Plusieurs techniques de traitement automatique du langage naturel (TALN) contribuent à la préparation des documents. La tokenisation des textes est faite directement par l'analyseur morphosyntaxique. La lemmatisation, qui sert principalement à définir toutes les flexions morphologiques des concepts, se fait au moyen du *Lexique des formes fléchies du français* (Sagot, 2010). Ce lexique a en général une bonne couverture du français (plus de 530 000 entrées), mais certains mots spécialisés peuvent ne pas s'y trouver (p. ex., les mots « autotrophes » et « hétérotrophes »). Lorsque nécessaire, certains mots ont été ajoutés manuellement au lexique.

Chaque phrase est analysée deux fois par EPTAO : une première fois selon une grammaire en constituants, à l'aide du modèle pour le français de l'analyseur Stanford Parser (Green, De Marneffe, Bauer & Manning, 2011), et une seconde fois selon une grammaire de dépendance, à l'aide d'une adaptation francisée de l'analyseur MSTParser (Candito, Nivre, Denis & Anguiano, 2010). Les annotations de la grammaire en constituants sont utilisées dans le processus de résolution d'anaphores, et celles de la grammaire de dépendance font partie du processus de reconnaissance des relations.

EPTAO comporte également un module de résolution des pronoms de la troisième personne (il, elle, ils, elles, leur, leurs). Ce module est basé sur l'algorithme de Hobbs (1978), et plus spécifiquement, il exploite les relations syntaxiques de la grammaire en constituants et les caractéristiques morphologiques (genre et nombre) des mots pour déterminer l'antécédent qui correspond au pronom à l'examen. Cet algorithme purement linguistique est l'un des plus anciens, mais il offre des performances comparables aux techniques plus récentes et sophistiquées (Tetreault, 1999).

6.5.3 Reconnaissance des concepts

En conformité avec les tâches de cartographie conceptuelle soumises aux élèves, les concepts et leurs synonymes sont fournis à l'avance par un humain dans cette étude, ce qui simplifie grandement le processus d'identification des concepts. Le logiciel n'a qu'à reconnaître les mots dans les textes ou leurs flexions morphologiques ; il n'a pas à construire lui-même une banque de termes candidats.

Au cours de la mise au point du logiciel, il a été constaté que certains concepts sont précédés de locutions telles que « la majorité de » ou « une partie de ». Dans une expression comme « la majorité des rayons solaires », c'est donc le nom collectif (majorité) et non le concept (rayons solaires) qui est à la tête du groupe nominal. Pour éviter des embûches lors de la reconnaissance des relations, le logiciel a été modifié afin qu'il reconnaisse autant l'expression « la majorité des rayons solaires » que le concept seul « rayons solaires ». Cet ajustement est en fait dérivé de la méthode du plus court parcours, décrite dans la prochaine sous-section.

6.5.4 Reconnaissance des relations

Dans une tâche du type « élaborer une carte à partir d'une liste de concepts préétablie », les élèves ont le libre choix de formuler les expressions de liaison. Le logiciel extrait donc les expressions de liaison le plus naturellement possible, à savoir comment elles sont exprimées dans les textes. Pour y arriver, EPTAO intègre les deux mécanismes proposés par Zouaq et Nkambou (2008), à savoir les patrons lexico-syntaxiques de Hearst pour la reconnaissance des relations d'hyponymie et d'hyperonymie ainsi que les patrons SVO pour la reconnaissance des relations de type « prédicat, argument ».

Dans un premier temps, les sept patrons de Hearst ont été adaptés au français et deux patrons supplémentaires ont été ajoutés à la suite d'une analyse préliminaire du corpus. Le Tableau XXVII dresse la liste de ces patrons, donne leur formulation en expression rationnelle (en anglais, *regular expression*) et les exemplifie. Il est à noter que les patrons n^{os} 1 et 2 sont rassemblés puisqu'ils sont équivalents en français.

Tableau XXVII. Adaptation des patrons lexico-syntaxiques de Hearst (1992)

N°	Expression rationnelle	Exemples
1-2	NP ₀ tel que [NP ₁ , NP ₂ ,...] [et/ou] NP _n	Un félin tel que le chat ou le lynx
3	NP [, NP]* [,] ou [autre/autres] NP ₀	Le chat, le lynx ou autres félins
4	NP [, NP]* [,] et [autre/autres] NP ₀	Le chat, le lynx et autres félins
5	NP ₀ [,] [comprenant/y compris] [NP,]* [et/ou] NP	Les félins, y compris le chat, ou le lynx
6	NP ₀ [,] en particulier [NP,]* [et/ou] NP	Les félins, en particulier le chat et le lynx
7	NP ₁ [est une sorte de/est un type de] NP ₀	Le chat est une sorte de félin
8	NP ₀ suivant[e][s] : [NP,]* [et/ou] NP	Les félins suivants : le chat, le lynx
9	NP ₀ ([NP,]* [et/ou] NP)	Les félins (le chat, le lynx)

Dans un second temps, la méthode SVO a été révisée pour satisfaire aux conditions de la présente étude. Les triplets « sujet, verbe, complément d'objet » ne désignent pas toutes les formes de patrons de dépendance (Stevenson & Greenwood, 2009). Dans la phrase « Un orchestre est un ensemble de musiciens instrumentistes », le sujet « orchestre » est associé au nom collectif et complément d'objet « ensemble ». La méthode SVO ne constitue pas de relation entre les concepts d'« orchestre » et de « musiciens ». La construction de chaînes de dépendance à l'aide de la méthode du plus court parcours entre deux nœuds (Bunescu & Mooney, 2005) est une technique permettant d'accroître la reconnaissance de relations. Dans l'exemple précédent, une chaîne de dépendance peut être construite entre « ensemble » et « musiciens » puisque ces termes font partie du même groupe nominal dans l'arbre syntaxique de la phrase. En combinant les résultats du patron SVO et de la chaîne de dépendance, il est donc possible de former la proposition suivante : « un orchestre *est un ensemble de* musiciens ». Des cas similaires se produisent lorsqu'un concept est dans le sous-arbre syntaxique du complément d'objet. Selon cette approche modifiée, l'expression de liaison est définie comme étant le plus court parcours entre deux concepts et qui traverse un verbe dans l'arbre syntaxique de la phrase.

6.5.5 Synthèse de la carte

La dernière étape de la démarche comporte la synthèse des propositions extraites par le logiciel et la création de la carte conceptuelle finale. Cette étape n'est pas traitée par EPTAO et est faite manuellement. La synthèse des propositions consiste essentiellement à regrouper les mots de relation qui portent le même sens et à les représenter par une seule proposition. Le Tableau XXVIII donne un exemple de cette démarche. Dans le corpus sur l'effet de serre, les concepts « couche d'ozone » et « rayons ultraviolets » sont unis entre eux par six relations, regroupées en trois catégories. Les deux premières catégories dénotent respectivement le rôle protecteur de la couche d'ozone et sa capacité à absorber les rayons ultraviolets. Elles désignent des propositions scientifiquement valides. La proposition de la troisième catégorie partage le même champ sémantique que celle de la première catégorie, mais elle est non valide d'un point de vue scientifique, du moins lorsqu'elle est isolée du reste de la phrase d'où elle est extraite¹.

Tableau XXVIII. Exemple de catégorisation de mots de liaison

Mots de liaison
Sens 1
– protège la biosphère d'une partie du
– joue un rôle essentiel de protection (filtre) contre les
Sens 2
– a pour effet d'absorber la plus grande partie du
– est un filtre chimique qui absorbe les
– désigne la partie de l'atmosphère qui contient une
concentration
élevée de molécules d'ozone et qui absorbe une partie des
Sens 3
– offrait si peu de protection contre les

Une fois que la liste de propositions distinctes est générée, celles-ci peuvent être représentées graphiquement. Cette opération a été réalisée par le premier auteur, et ce, sans privilégier un mode d'organisation spécifique.

6.6 Méthodologie

La mise à l'essai d'un logiciel a pour but de tester ses performances et celles de ses composantes dans diverses conditions. Dans ce qui suit, les performances d'EPTAO sont testées auprès de deux corpus de textes. Ces corpus correspondent aux thèmes des deux tâches de cartographie conceptuelle élaborées par un enseignant participant à l'étude plus vaste dans laquelle s'inscrit cette recherche.

6.6.1 Description des tâches

La première tâche porte sur les mécanismes de l'effet de serre et de phénomènes associés. En s'inspirant du manuel scolaire de son cours, l'enseignant a retenu 11 concepts pour la réalisation de cette tâche, chacun étant accompagné de 1 à 4 synonymes. La liste complète est exposée dans le Tableau XXXIX en annexe. La seconde tâche concerne les relations trophiques (c.-à-d. les relations alimentaires) entre les êtres vivants. L'enseignant a formé une liste comprenant 17 concepts pour cette activité (voir Tableau XL en annexe).

6.6.2 Création des corpus

Un corpus de textes a été formé pour chacune des deux tâches. Les textes de base sont tirés de manuels scolaires en science et technologie de quatrième secondaire et sont conformes au plus récent curriculum. Seulement trois ouvrages ont été recensés : *Observatoire* (Cyr, Verreault & Forget, 2008), *Science-Tech au secondaire* (Barbeau et al., 2008) et *Synergie* (Chartré et al., 2008). Et seule l'information pertinente aux thèmes des tâches a été extraite de ces manuels. Le Programme de formation de l'école québécoise (MELS, 2007) de science et technologie en quatrième secondaire adopte une approche intégrative des apprentissages en les organisant autour de quatre grands thèmes (p. ex., les changements climatiques). Selon les activités pédagogiques proposées, les mêmes concepts et apprentissages peuvent être réinvestis à travers plusieurs thèmes. Cela fait que certaines informations sont éparpillées et parfois répétées à plusieurs endroits dans le manuel scolaire. Pour les fins de cette étude, toutes les données pertinentes ont été rapatriées dans un même document.

La recherche d'articles Wikipédia se veut quant à elle un moyen d'enrichir le corpus et, ultimement, les résultats de la production d'une carte conceptuelle. Cette recherche a ajouté

17 textes à la première tâche et 15 textes à la seconde tâche. Le Tableau XXIX décrit quelques caractéristiques des deux corpus soumis au logiciel. Le nombre de documents et le nombre de mots sont beaucoup plus élevés pour les articles Wikipédia que pour les textes tirés des manuels scolaires.

Tableau XXIX. Description des corpus des deux mises à l'essai

Corpus	Nbre de documents	Nbre de mots		
		Min	Max	Total
Effet de serre				
Manuels scolaires	3	760	2 251	4 474
Articles Wikipédia	17	170	5 227	30 814
Relations trophiques				
Manuels scolaires	3	869	1 682	3 695
Articles Wikipédia	15	140	5 687	36 112

6.6.3 Stratégies d'évaluation du rendement du logiciel

L'évaluation du rendement du logiciel est une étape importante dans le cycle de développement d'une stratégie de production d'une carte conceptuelle (Villalon & Calvo, 2008). Les stratégies d'évaluation varient d'une étude à l'autre. Qasim et ses collaborateurs (2013) évaluent les performances des sous-composantes (p. ex., la résolution d'anaphores, l'identification des concepts) de leur système, alors que Villalon et Calvo (2008) suggèrent de comparer quantitativement la carte extraite de textes avec une carte produite par un panel d'experts de contenu. Clariana et Koul (2004) évaluent quant à eux les performances prédictives de leur système en établissant dans quelle mesure les scores dérivés des cartes des élèves (en les comparant à la carte de l'expert) sont équivalents aux scores décernés par des correcteurs.

Un système doit idéalement être scruté à divers degrés de granularité. Une analyse très fine des performances sert à établir les pistes d'amélioration potentielles, tandis qu'une analyse à l'échelle macroscopique donne un aperçu plus juste du potentiel du programme dans son contexte d'utilisation. Les différentes stratégies d'évaluation mises en œuvre dans cette étude sont détaillées dans les sous-sections suivantes.

6.6.3.1 Stratégie d'évaluation du processus de reconnaissance des relations

Dans le but d'effectuer une analyse fine des performances du logiciel, tous les textes ont été annotés de façon manuelle par le premier auteur de cet article, également formé en enseignement au secondaire en science et technologie. La codification des textes a été réalisée en deux étapes. Dans un premier temps, toutes les anaphores, pronominales ou non, ont été signalées, puis remplacées par leur antécédent lorsqu'il exprimait un concept de la liste initiale. Ensuite, tous les concepts et les relations valides ont été codifiés et isolés à l'intérieur des documents. À certaines occasions, il était crucial de tenir compte du contexte de la phrase pour juger de l'exactitude d'une proposition potentielle. Par exemple, il n'aurait pas été juste d'isoler la proposition « *Le dioxyde de carbone est le plus abondant des gaz à effet de serre* » de la phrase « *Le dioxyde de carbone est le plus abondant des gaz à effet de serre produits par les activités humaines* ». Une proposition valide est donc une proposition qui, prise isolément, est valide d'un point de vue scientifique.

Les annotations manuelles ont ensuite servi de critère pour l'évaluation du processus de reconnaissance des concepts et relations. Dans le cadre des mises à l'essai, deux statistiques, soit le taux de restitution et le taux de précision, sont calculées pour évaluer les performances du programme. Ces statistiques sont définies comme suit :

$$\text{Restitution} = \frac{\text{Nombre de propositions valides reconnues par le programme}}{\text{Nombre de propositions reconnues par l'expert}} \quad (6.1)$$

$$\text{Précision} = \frac{\text{Nombre de propositions valides reconnues par le programme}}{\text{Nombre de propositions reconnues par le programme}} \quad (6.2)$$

Les valeurs de ces deux statistiques se situent entre 0 et 1. Un taux de restitution de 1 indique une adéquation parfaite entre les propositions restituées par le système et celles reconnues par l'expert. Un taux de précision de 1 indique que toutes les entités reconnues par le système sont valides.

6.6.3.2 Stratégie d'évaluation du processus de remplacement des anaphores pronominales

EPTAO comporte un module de substitution des anaphores pronominales de la troisième personne basé sur l'algorithme de Hobbs (voir la sous-section concernant les mécanismes de prétraitement des textes). Dans le but d'évaluer le rendement réel et attendu de cet algorithme, trois variantes du taux de restitution sont rapportées dans les résultats. Les deux premières variantes servent à comparer les taux de restitution des propositions avant et après la résolution d'anaphores. Pour la troisième variante, une correction est appliquée au calcul du taux de restitution comme si tous les pronoms personnels de la troisième personne avaient été remplacés correctement par le système. Cette dernière variante permet ainsi de mesurer les gains potentiels dans le cas où le module aurait fonctionné parfaitement.

6.6.3.3 Stratégie d'évaluation des sources d'erreurs

À certaines occasions, le logiciel commet des erreurs et suggère des relations qui ne font pas partie du texte. Une analyse préliminaire des faux positifs a permis de les catégoriser en quatre sources d'erreurs : 1) le remplacement d'un pronom par un mauvais antécédent, 2) l'équivocité des concepts et de leurs synonymes, 3) l'assemblage ambigu d'un patron SVO et d'une chaîne en dépendance, ou 4) une simplification abusive de la phrase.

6.6.3.4 Stratégie d'évaluation du processus de la synthèse des propositions

La sélection de textes est une étape importante de la démarche proposée. Un examen approfondi de l'inventaire des propositions extraites par EPTAO offre alors la possibilité d'évaluer la pertinence des textes choisis au début de la démarche. En particulier, deux indicateurs sont étudiés : la répétition et la pertinence des propositions distinctes. Le premier peut être considéré comme un indicateur, certes imparfait, d'un consensus au sujet des connaissances importantes dans un domaine spécifique. Le second indique à quel degré le texte contient des informations qui sont proches ou éloignées du thème principal de la carte conceptuelle.

6.6.3.5 Stratégie d'analyse de la carte finale

Lors de la dernière étape de la démarche, les informations consignées dans l'inventaire des propositions distinctes sont assemblées sous forme d'une carte conceptuelle. Cette carte issue des propositions extraites par EPTAO peut alors être comparée à une carte de référence. Villalon et Calvo (2008) proposent de former une carte de référence à l'aide d'un panel d'experts. En conformité avec la démarche proposée dans cette étude, la carte de référence est tirée des manuels scolaires puisque ces documents sont jugés plus conformes aux attentes quant aux apprentissages réalisés par les élèves. Cette comparaison est faite sur la base d'indicateurs quantitatifs, comme le taux de restitution des propositions distinctes, et d'indicateurs qualitatifs, comme le respect des principes de construction d'une carte conceptuelle.

En résumé, ce cadre méthodologique propose d'étudier sous plusieurs angles le rendement du logiciel. La prochaine section présente les résultats de cette mise à l'essai.

6.7 Résultats

Les résultats de chaque stratégie d'analyse sont présentés séparément dans les sous-sections suivantes.

6.7.1 Évaluation du processus de reconnaissance des propositions

Le Tableau XXX donne les taux de restitution et de précision pour les deux mises à l'essai. Globalement, les taux réels de restitution (39 %) et de précision (66 %) sont assez modestes pour la tâche sur l'effet de serre, tandis que les résultats sont plus encourageants pour la seconde tâche, avec des taux réels respectivement de 60 % et 76 %. Comme il sera expliqué plus loin, le thème de l'effet de serre est peut-être en lui-même une source de difficultés pour le processus d'extraction.

Les résultats varient selon la source d'information. Les manuels scolaires renferment un plus grand nombre de propositions, et ce, même s'ils comptent beaucoup moins de mots que les articles provenant de Wikipédia (voir Tableau XXIX). Les taux de restitution des deux tâches sont relativement stables entre les textes des manuels scolaires et les articles Wikipédia, mais le taux de précision des textes provenant des manuels scolaires est toujours plus élevé que celui

des articles Wikipédia. Le contenu, le style d'écriture et l'organisation des articles ne sont pas nécessairement uniformes, ce qui occasionne certaines difficultés lors de leur traitement.

Tableau XXX. Taux de restitution et taux de précision de l'extraction des propositions des deux mises à l'essai

Source	Nbre de propositions	Taux de restitution			Taux de précision
		Avant	Après (Réel)	Après (Corrigé)	
Effet de serre					
Manuels scolaires	38	0,39	0,42	0,47	0,84
Articles Wikipédia	32	0,34	0,34	0,44	0,50
Total	70	0,37	0,39	0,46	0,66
Relations trophiques					
Manuels scolaires	52	0,58	0,62	0,65	0,89
Articles Wikipédia	39	0,56	0,59	0,69	0,64
Total	91	0,57	0,60	0,67	0,76

6.7.2 Évaluation du processus de résolution d'anaphores

Toujours selon les résultats du Tableau XXX, l'algorithme de Hobbs contribue très peu à l'extraction de propositions. D'une part, le système n'a pas réussi à remplacer tous les pronoms de la troisième personne par leurs antécédents. Un seul pronom a été remplacé lors de la première tâche, tandis que seulement trois pronoms l'ont été dans le corpus de la seconde tâche. Il reste un grand nombre de pronoms qui n'a pas été traité par le système. Comme l'indique la différence entre le taux corrigé et le taux après résolution, des gains additionnels de 7 % auraient pu être enregistrés si tous ces pronoms avaient été traités. Les valeurs modérées des taux corrigés (entre 0,46 et 0,67) indiquent également que la prise en charge d'autres formes d'anaphores (p. ex., les pronoms démonstratifs) pourrait accroître grandement le taux de restitution.

6.7.3 Évaluation des sources d'erreurs

Comme nous l'avions anticipé, le logiciel fournit des propositions qui sont fausses. Le Tableau XXXI donne la fréquence d'apparition des quatre sources d'erreurs pour les

propositions faussement identifiées. Les deux corpus ont été affectés par différentes sources d'erreurs et ce sont les articles Wikipédia qui ont mené au plus grand nombre d'erreurs. Le corpus sur l'effet de serre a mené à 9 erreurs de « simplification abusive de la phrase ». Ces erreurs se produisent lorsqu'une relation binaire est extraite d'une phrase qui comporte en fait une relation ternaire. Le corpus sur les relations trophiques est plutôt marqué d'erreurs d'équivocité des concepts et synonymes. À 10 occasions, le concept identifié dans une proposition ne correspond pas exactement au concept de la liste de départ. Cela se produit entre autres lorsque les termes « champignon » ou « bactérie » pris isolément ne désignent pas « champignon et bactérie ».

Alors que les erreurs de simplification abusive de la phrase émanent de la structure des phrases, les erreurs d'équivocité sont issues d'une liste de concepts et de synonymes ambigus. C'est pourquoi les secondes sont beaucoup plus faciles à corriger que les premières. Il suffit de réviser la liste de concepts et synonymes. Quant aux erreurs de simplification abusive, il faut probablement revoir le programme informatique afin qu'il puisse analyser efficacement les relations ternaires.

Tableau XXXI. Fréquence des erreurs d'extraction dans les deux corpus

Corpus/Documents	Remplacement d'un pronom par un mauvais antécédent	Équivocité des concepts et synonymes	Assemblage ambigu d'un patron SVO et d'une chaîne en dépendance	Simplification abusive de la phrase	Total
Effet de serre					
Manuels scolaires	0	0	0	3	3
Articles	0	2	3	6	11
Wikipédia					
Total	0	2	3	9	14
Relations trophiques					
Manuels scolaires	0	0	3	1	4
Articles	1	10	2	0	13
Wikipédia					
Total	1	10	5	1	17

6.7.4 Évaluation de la redondance et de la pertinence des idées

La dernière étape de la stratégie d'extraction consiste à faire une synthèse des propositions identifiées. Au cours de cette étape, un expert de contenu, dans ce cas-ci le premier auteur, regroupe les propositions extraites par le logiciel en idées communes, puis il détermine si ces idées sont pertinentes d'un point de vue didactique, c'est-à-dire à l'égard de la tâche et du niveau de scolarité des élèves. Deux propositions représentent une même idée si elles partagent le même sens d'un point de vue scientifique (voir Tableau XXVIII pour un exemple). Dans le but d'évaluer le rendement du logiciel, ces deux opérations ont été effectuées avec les propositions extraites par le logiciel ainsi qu'avec les propositions extraites manuellement des textes. Le Tableau XXXII présente la fréquence d'apparition des propositions non pertinentes ou pertinentes dans le corpus et dans la carte de référence. Il est important de noter que le nombre d'idées provenant du corpus entier ne correspond pas au nombre d'idées identifiées dans les manuels scolaires et dans les articles Wikipédia. Ces deux ensembles de textes ne sont pas mutuellement exclusifs. Une même idée est comptabilisée une seule fois dans le corpus entier, même si elle est répétée dans les deux ensembles pris séparément. Par exemple, l'idée « les gaz à effet de serre emprisonnent une partie des rayons infrarouges » est répétée quatre fois dans les manuels scolaires et deux fois dans les articles Wikipédia.

Comme l'illustre le Tableau XXXII, seul le corpus sur les relations trophiques a généré des propositions qui sont scientifiquement valides, mais non pertinentes d'un point de vue didactique. Un exemple d'une telle proposition est « les animaux *sont l'un des types d'eucaryotes à s'être développés sur un mode multicellulaire comme les végétaux* ». Il n'y a rien de faux dans cette proposition, mais celle-ci est trop précise pour des élèves de quatrième secondaire en science et technologie. Le contenu des ouvrages didactiques est en effet plus homogène et plus ciblé que celui des articles Wikipédia. Ces articles contiennent parfois des propositions trop pointues pour le groupe d'élèves ciblé par la tâche. Dans un contexte scolaire, l'enseignant devra vraisemblablement modifier ou éliminer les propositions qui ne correspondent pas aux attentes du programme de formation.

Tableau XXXII. Pertinence et fréquence d'apparition des propositions distinctes dans le corpus et dans les inventaires de propositions extraites par EPTAO

Sources	N ^{bre} de propositions distinctes	Non pertinente		Pertinente	
		1 fois	2 fois ou plus	1 fois	2 fois ou plus
Effet de serre					
Corpus entier	43	0	0	31	12
Manuels scolaires	21	0	0	10	11
Articles Wikipédia	27	0	0	21	6
Inventaire de propositions	17	0	0	11	6
Relations trophiques					
Corpus entier	49	4	0	24	21
Manuels scolaires	31	0	0	12	19
Articles Wikipédia	27	2	0	12	13
Inventaire de propositions	33	5	0	15	15

6.7.5 Analyse de la carte finale

Une fois que la révision de l'inventaire des idées communes est terminée, les propositions sont assemblées en une carte conceptuelle. Par souci d'économie d'espace, seules les cartes portant sur les relations trophiques sont présentées et décrites dans ce qui suit. Pour plusieurs raisons, certaines ayant déjà été évoquées, cette tâche a mené à de meilleurs résultats. La Figure 15 et la Figure 16 illustrent respectivement la carte de référence, obtenue à l'aide des annotations manuelles des textes, et la carte avec les propositions extraites par EPTAO. Les concepts et relations ont été organisés de manière à faire ressortir les ressemblances et les différences entre ces deux cartes. En prenant la carte de référence comme critère de comparaison, le logiciel a extrait 24 des 31 idées pertinentes (77 %) contenues dans les textes didactiques sur les relations trophiques⁴³. Un nœud (Animaux) de la liste initiale de concepts est

⁴³ En guise de comparaison, 11 des 21 idées (52 %) contenues dans les textes traitant de l'effet de serre ont été identifiées correctement.

totallement absent des deux cartes. Même si ce concept ou ses synonymes étaient présents dans les documents, aucune relation valide n'a été découverte.

Un code de couleurs a été utilisé dans la Figure 15 et la Figure 16 pour discerner les différences entre les deux cartes. Plus précisément, les expressions de liaison encadrées en gris foncé désignent celles qui se trouvent dans une seule des deux cartes conceptuelles. Dans la carte de la Figure 15, ces relations uniques découlent toutes de situations où un pronom n'est pas remplacé par son antécédent. Dans la carte de la Figure 16, les relations distinctes sont toutes issues des articles Wikipédia.

Selon Villalon et Calvo (2008), les cartes produites automatiquement devraient ressembler autant que possible aux cartes élaborées à la main. Les cartes de la Figure 15 et la Figure 16 s'éloignent un peu des principes généralement admis de construction d'une carte conceptuelle. Premièrement, les deux cartes renferment plus d'une proposition pour certaines paires de concepts. Deuxièmement, elles renferment des relations redondantes. Par exemple, les propositions « les consommateurs *peuvent se nourrir de* décomposeurs » et « les champignons et bactéries *peuvent être mangés par des* consommateurs » signifient essentiellement la même chose. Troisièmement, la construction de propositions en combinant les chaînes de dépendance et les patrons SVO peut générer des expressions de liaison plus longues qu'à l'accoutumée (voir les exemples dans Novak & Cañas, 2008). Les répercussions de ces observations sont discutées plus longuement dans la discussion, présentée dans la prochaine section.

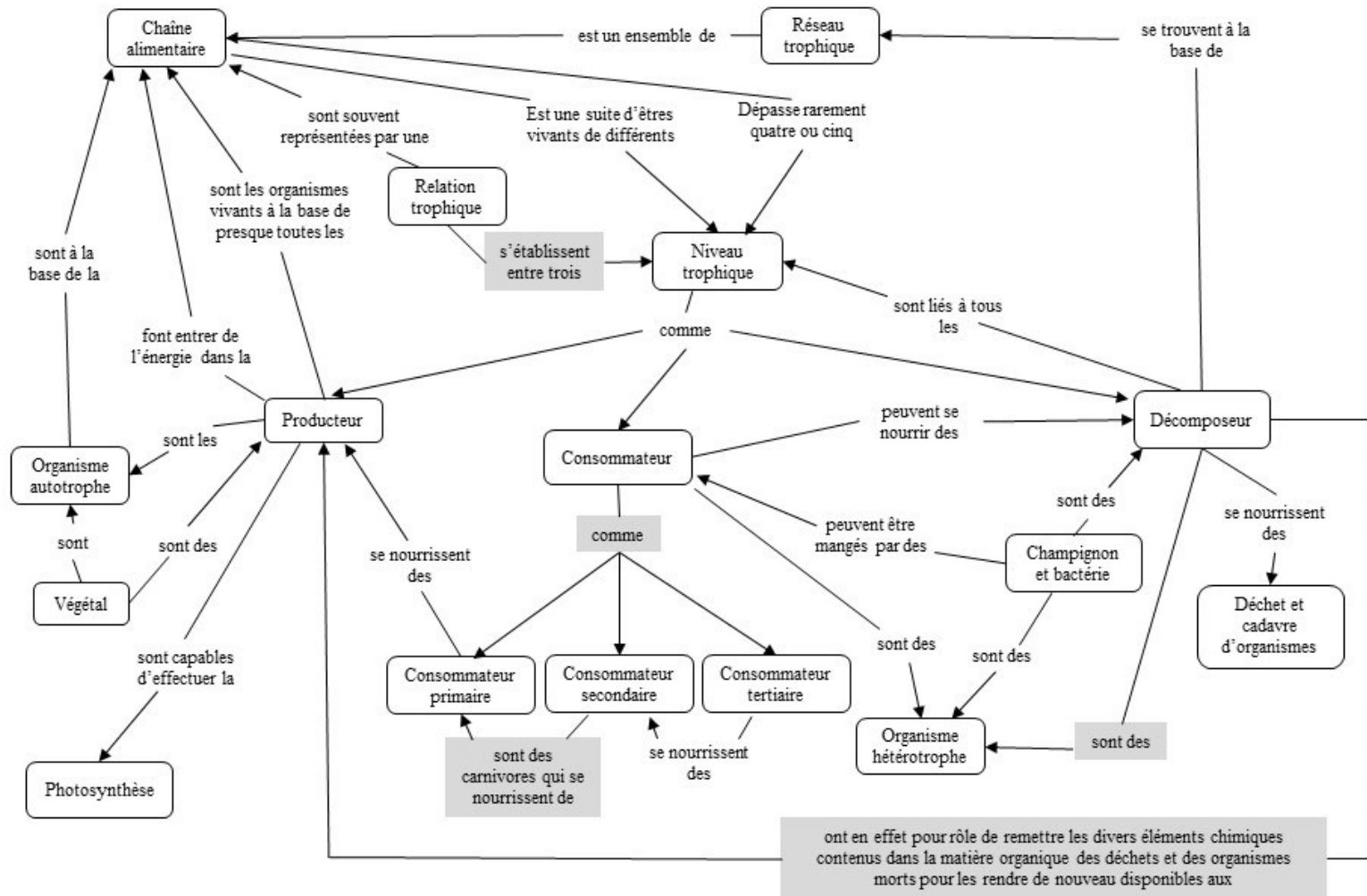


Figure 15. Carte conceptuelle de référence sur le thème « relations trophiques »

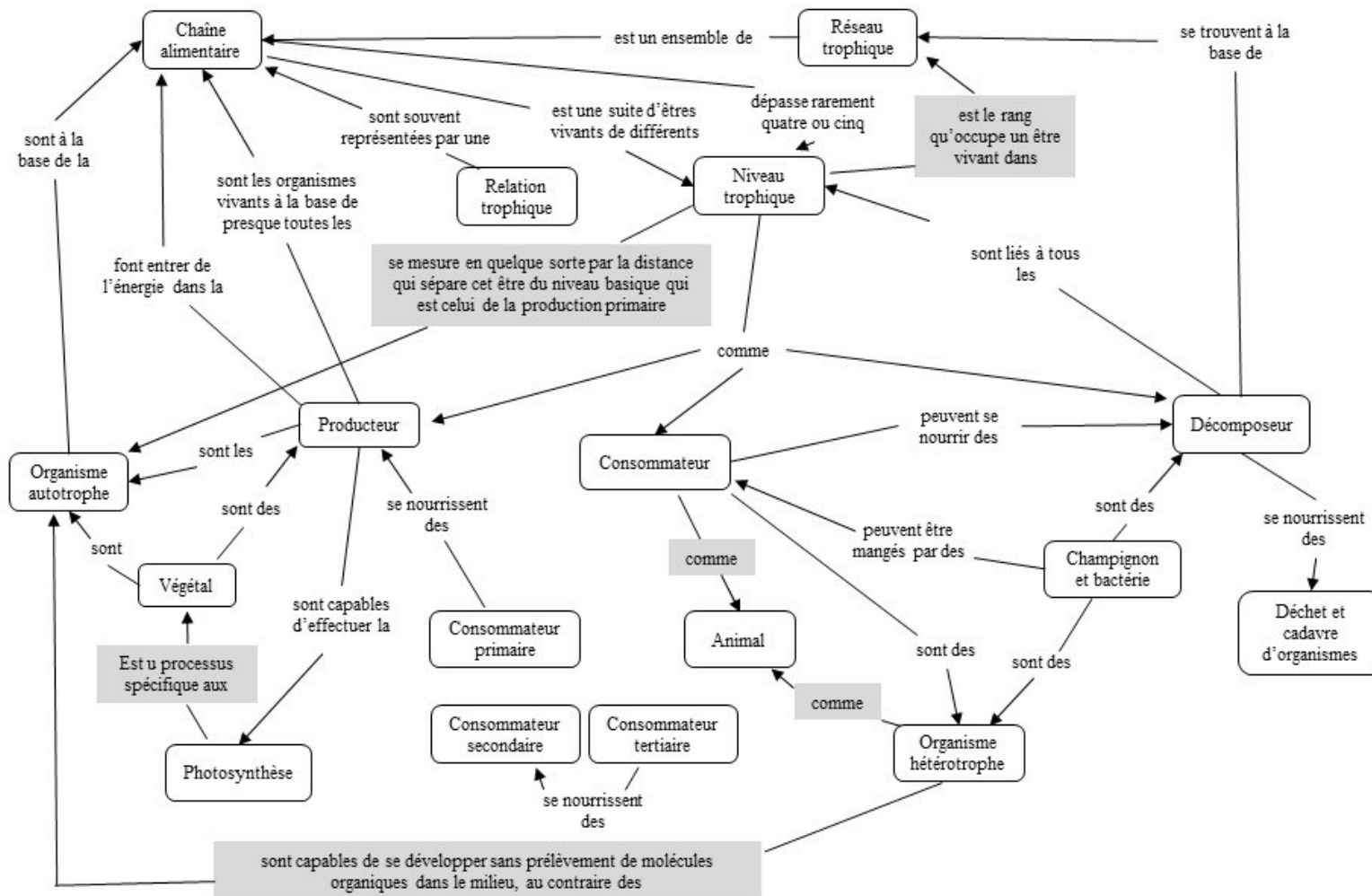


Figure 16. Carte produite sur la base des propositions extraites par EPTAO

6.8 Discussion

La section précédente a décrit les performances globales et spécifiques d'EPTAO et de ses composantes. Globalement, les taux réels de restitution (entre 39 et 60 %) et de précision (entre 66 et 76 %) sont relativement modestes. Cela dit, il est difficile de situer les performances d'EPTAO par rapport à celles d'autres recherches similaires parce qu'aucune d'entre elles n'a, autant que nous sachions, évalué de manière systématique le taux de restitution et de précision des relations extraites⁴⁴. En comparaison au taux de restitution des propositions, le taux de restitution des propositions distinctes (idées) est plus encourageant. Les mêmes idées se répètent d'un texte à l'autre de telle sorte qu'au moment de la synthèse des propositions, une grande partie des idées contenues dans les textes se retrouvent dans la carte conceptuelle finale.

Dans une démarche de recherche et développement, l'évaluation de la performance des composantes spécifiques du logiciel est une source d'information capitale pour établir les améliorations potentielles de la démarche. Du point de vue technique, il y a lieu de se questionner au sujet du faible rendement de l'algorithme de Hobbs (2 à 3 %) pour remplacer les pronoms de la troisième personne. En comparaison, Qasim et ses collaborateurs (2013) ont obtenu un rendement de 9 % avec l'algorithme de Lappin et Leass (1994). De plus, une analyse plus détaillée des textes laisse aussi croire qu'il faudrait se pencher sur la résolution d'autres types d'anaphores, dont celles composées de pronoms démonstratifs (celui, celle, etc.) ou les anaphores nominales (p. ex., « [...] des rayons solaires. Ces rayons [...] »). La résolution d'anaphores est une composante essentielle de la reconnaissance des concepts, même s'ils sont fournis par l'utilisateur.

Certains choix méthodologiques peuvent aussi être revus à la lumière des résultats de la recherche. Entre autres, la contribution des articles Wikipédia est plutôt mitigée. Certes, ces articles ont généré quelques propositions pertinentes qui ne se trouvent pas dans les manuels scolaires. En revanche, ils sont à l'origine d'un plus grand nombre de propositions non

⁴⁴ Qasim, Jeong, Heu et Lee (2013), par exemple, évaluent le taux de restitution et de précision des concepts, mais pas des relations.

pertinentes à l'égard du niveau de scolarité ciblé et ils occasionnent plus d'erreurs lors de la reconnaissance des relations. Les articles Wikipédia sont des textes de nature encyclopédique et ne sont pas rédigés dans une visée didactique comme les textes tirés des manuels scolaires. Le style d'écriture, la longueur des phrases et le contenu HTML sont tous des facteurs susceptibles d'influencer le processus d'extraction des propositions contenues dans ces articles.

La présente étude soulève aussi quelques pistes de réflexion d'ordre pédagogique. La première piste concerne un principe de construction d'une carte conceptuelle généralement admis et peu contesté, soit celui de formuler une et une seule relation à partir de deux concepts. L'analyse systématique du contenu des textes montre pourtant qu'il est possible de formuler plus d'une idée indépendante à partir d'une même paire de concepts. D'un point de vue pédagogique, il ne semble donc pas y avoir de raisons d'empêcher un élève d'exprimer deux relations ou plus entre les deux mêmes concepts.

La seconde piste de réflexion concerne la représentation d'une même idée à partir de paires de concepts différentes. Dans certaines modalités d'évaluation de cartes conceptuelles (p.ex., voir Ruiz-Primo, Schultz et al., 2001), seules les propositions qui se trouvent conjointement dans la carte de l'élève et dans la carte de référence sont jugées valides. Qu'arrive-t-il lorsque l'élève choisit une paire de concepts différente pour exprimer une proposition qui se trouve dans la carte de référence ?

La troisième piste de réflexion porte sur la longueur et le contenu des mots de relation. Normalement, une expression de liaison comporte quelques mots et est composée d'un verbe. Parmi les expressions extraites par le logiciel, certaines sont très longues (plus de 30 mots) et sont composées de compléments d'objet dont certains comportent des groupes nominaux. Ces groupes du nom auraient, eux-mêmes, pu faire partie de la liste de concepts. Puisque cette liste est établie à l'avance et figée, EPTAO cherche tous les parcours raisonnables (dans les arbres syntaxiques) entre deux concepts contenus dans la même phrase. Dans la pratique, de telles relations, si jugées trop longues, peuvent tout simplement être ignorées ou reformulées par un élève ou un enseignant. Cela dit, cette approche offre une autre option pour représenter les relations ternaires (p. ex., les producteurs *font entrer de l'énergie dans* la chaîne alimentaire). Elle évite de les reformuler comme une suite de deux relations binaires (p. ex., les producteurs

font entrer de l'énergie dans la chaîne alimentaire) et réduit le risque d'interpréter ces deux relations binaires de manière indépendante.

Une dernière piste de réflexion, plus large, concerne la pertinence de la cartographie conceptuelle pour représenter certains domaines de connaissances. La cartographie conceptuelle est assujettie à une « syntaxe » particulière qui ne s'adapte peut-être pas à toutes les situations. La représentation de connaissances sous forme de relations binaires est un premier exemple des limites de cet outil. L'ajout d'un quantificateur venant modifier le concept est un second exemple. La carte conceptuelle, du moins dans sa forme courante, n'est pas adaptée à la représentation de relations dynamiques entre les concepts (Safayeni, Derbentseva & Cañas, 2005), ce qui explique pourquoi certains ajustements ont dû être apportés à EPTAO pour extraire des propositions dans les textes concernant l'effet de serre.

Enfin, le rendement d'un logiciel de production automatisée de cartes conceptuelles n'est pas seulement tributaire du logiciel, mais aussi des domaines de connaissances représentés dans les textes qui sont analysés par le système. Le but de l'activité, le thème et les contenus d'apprentissage sont ainsi tous des facteurs à considérer lors du choix d'un outil de représentation des connaissances.

6.9 Conclusion

Cet article a présenté la conception et la mise à l'essai d'une démarche de production semi-automatisée d'une carte conceptuelle à partir de documents textuels. Le développement d'une telle stratégie a pour but d'offrir une solution de rechange à la méthode de constitution d'une carte conceptuelle par un expert ou un panel d'experts de contenu.

Les résultats de cette étude sont prometteurs. Le logiciel EPTAO, conçu et testé dans cette recherche, a démontré qu'il est capable d'extraire les connaissances contenues dans les textes et ainsi d'assister un humain dans la création d'une carte conceptuelle. Pour l'instant, les propositions extraites par EPTAO n'aboutissent peut-être pas à une représentation suffisante du domaine de connaissances pour des usages contraignants, comme la comparaison de cartes produites par des élèves et d'une carte de référence en vue de l'évaluation des apprentissages. Cependant, la carte finale construite à partir des propositions extraites par le logiciel offre un

bon point de départ pour divers autres usages (la rétroaction, la synthèse des apprentissages, etc.).

Cela dit, il y a encore plusieurs avenues à explorer pour améliorer le rendement d'EPTAO. Du point de vue technique, quelques composantes du logiciel peuvent être perfectionnées ou encore d'autres composantes de la démarche peuvent être automatisées. D'un point de vue méthodologique, il y a certainement un intérêt à mettre en place de nouvelles stratégies pour augmenter le corpus, car EPTAO n'a quasiment pas de limites quant au nombre de textes qu'il peut traiter. Ces améliorations pourront à leur tour faciliter le travail des utilisateurs d'EPTAO et accroître potentiellement les usages d'un tel logiciel.

Chapitre 7 Modélisation des scores issus de cartes conceptuelles à l'aide de modèles de Rasch⁴⁵

Maxim Morin

Jean-Guy Blais

Université de Montréal

Mots clés : carte conceptuelle, modélisation des scores, modèle de Rasch, modèle à crédit partiel, évaluation des apprentissages en sciences.

Une carte conceptuelle est un outil de représentation externe des connaissances qui est composé de nœuds représentant des concepts et de liens étiquetés représentant les relations entre ces concepts. Quoiqu'il soit principalement utilisé comme outil d'apprentissage, cet outil peut aussi servir comme dispositif de collecte de données en vue de l'évaluation des apprentissages. Dans les écrits scientifiques, les modèles de notation de cartes conceptuelles sont d'ailleurs nombreux et variés. Pourtant, ces modèles reposent presque tous sur des méthodes de notation basées sur les scores observés et font rarement appel à des modèles de mesure pour modéliser les scores et pour apprécier les propriétés psychométriques de tâches de cartographie conceptuelle. La présente étude cherche à combler ce manque en modélisant les scores issus de cartes conceptuelles de 84 élèves de quatrième secondaire en science et technologie à l'aide du modèle de Rasch pour données dichotomiques (Rasch, 1960) et du modèle à crédit partiel (Masters, 1982). Au cours de quatre modélisations successives, l'examen attentif de la condition d'indépendance locale et de la condition d'unidimensionnalité qui caractérisent les modèles de mesure de la famille des modèles de Rasch fait ressortir certaines particularités de la structure interne des scores qui n'ont pas été signalés auparavant. Ces résultats remettent en question certaines des approches conventionnelles de notation de cartes conceptuelles.

⁴⁵ Cet article n'a pas encore été soumis pour publication.

7.1 Introduction

La cartographie conceptuelle est une technique de représentation des connaissances très polyvalente et aux usages multiples en éducation (Cañas et coll., 2003; Cañas et Novak, 2008; Marchand et d'Ivernois, 2004). Dans le contexte de l'évaluation des apprentissages, plusieurs ont reconnu son potentiel pour collecter des données concernant des aspects de l'apprentissage qui ne sont pas captés à l'aide d'instruments traditionnels (Edmondson, 2005; Novak et Canas, 2008; Ruiz-Primo, Schultz, et coll., 2001). Cela dit, pour que la cartographie conceptuelle soit acceptée comme méthode de collecte de données, il faut s'assurer que le modèle de notation qui y est associé permette d'obtenir des scores fidèles et de tirer des interprétations valides, deux propriétés essentielles pour un instrument d'évaluation des apprentissages (Yin et Shavelson, 2008).

Un grand nombre de modèles de notation de carte conceptuelle répertoriés sont basés sur un calcul de la somme pondérée des scores observés (voir les recensions de Anohina et Grundspenkis, 2009; Ruiz-Primo et Shavelson, 1996). Cette approche est particulièrement flexible et accommode bien les tâches ouvertes comme la cartographie conceptuelle, sauf qu'elle restreint la liste de méthodes disponibles pour apprécier leurs propriétés psychométriques, ainsi que la portée des interprétations des résultats. En effet, les approches de notation basées sur des scores observés ne reposent pas véritablement sur une théorie de la mesure et par conséquent, ne sont donc pas assujetties aux postulats de modèles de mesure. Il serait bien sûr possible de faire un rapprochement entre ces approches et le modèle du score vrai de la théorie classique des tests (pour une description détaillée, voir Bertrand et Blais, 2004; Laveault et Grégoire, 2002), mais les auteurs des études recensées adoptent rarement ce modèle. En l'absence de conditions d'application, les données collectées au moyen de cartes conceptuelles sont acceptées telles quelles, et ne sont soumises à aucun test empirique pour assurer qu'elles sont conformes avec les inférences qui sont faites lors de l'opération d'évaluation des apprentissages.

Les modèles de mesure de la famille de modèles de Rasch (Rasch, 1960), comme ceux de la théorie de la réponse à l'item (Lord et Novick, 1968), s'appuient sur des conditions d'application dites fortes (Hambleton et Jones, 1993), ce qui offre un cadre théorique et méthodologique supérieur en vue d'examiner la qualité des données. La modélisation de Rasch,

en particulier, est une approche utilisée dans de nombreuses applications de construction d'échelles de mesure en sciences sociales et en sciences de la santé (voir par ex., Smith, 2004; Tennant, McKenna et Hagell, 2004) et peut s'avérer particulièrement intéressante lors du développement ou du perfectionnement d'un instrument de collecte de données en vue d'évaluer les apprentissages. La carte conceptuelle, un dispositif de collecte de données dont les propriétés psychométriques sont encore méconnues, est un bon exemple où l'utilité de cette approche est potentiellement supérieure à la simple somme des scores pondérés.

Cet article vise à illustrer les avantages d'appliquer la modélisation de Rasch aux scores issus d'une carte conceptuelle. Il est divisé en quatre sections. La première section constitue le cadre théorique et présente les trois principaux concepts de la recherche : la carte conceptuelle, les modalités de notation de cartes conceptuelles et la modélisation de Rasch. La section suivante décrit le cadre méthodologique. La troisième section présente les résultats des quatre modélisations effectuées dans l'étude et la dernière section consiste en la discussion et la conclusion de l'article.

7.2 Cadre théorique

7.2.1 La carte conceptuelle et ses composantes

La cartographie conceptuelle tire ses origines des modèles de représentation de type propositionnel des connaissances en psychologie cognitive (par ex., Ausubel et coll., 1978 ; Collins et Quillian, 1969). Elle repose sur deux unités fondamentales de la cognition : le concept et la proposition. Un concept est une représentation mentale de classes ou de catégories d'idées ou d'éléments (Gerrig et Zimbardo, 2002) et il est généralement désigné par une étiquette. La seconde unité cognitive, la proposition, correspond au plus court énoncé de sens auquel on peut attribuer une valeur de vérité et prend généralement la forme de relations entre des concepts ou de relations entre un concept et une propriété (Bernstein et coll., 2003).

La carte conceptuelle est un outil de représentation externe des connaissances. Elle se compose de concepts, de liens et de mots de liaison. Les concepts sont compris dans des figures géométriques et sont désignés par des mots ou des groupes de mots. Les concepts et les relations s'assemblent en énoncés de sens nommés des propositions. Une proposition comprend

généralement deux concepts et une relation, mais elle comporte parfois plus d'éléments. La Figure 17 illustre une carte conceptuelle réduite au sujet de la notion de « concept ».

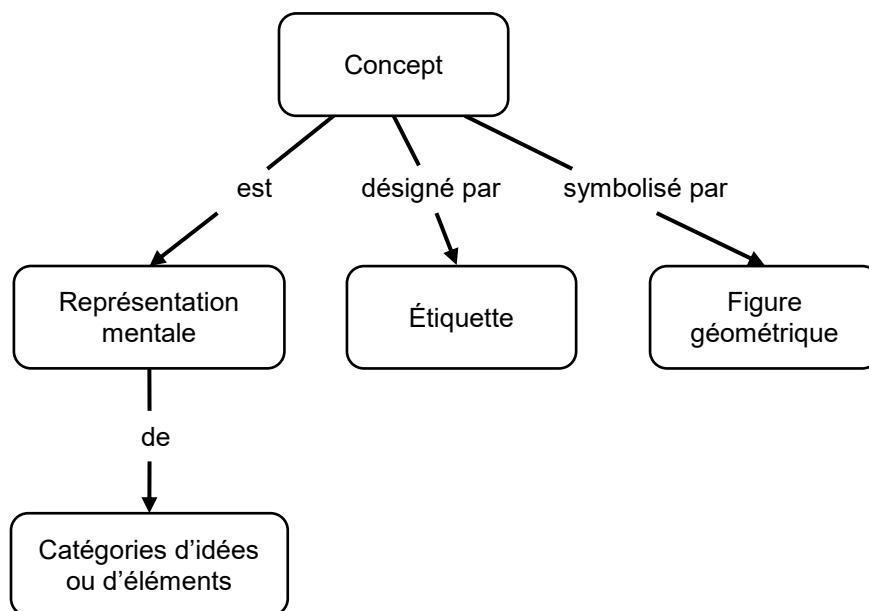


Figure 17. Exemple simple d'une carte conceptuelle

Pour Novak, à qui l'on attribue la paternité de la cartographie conceptuelle, cette technique de représentation des connaissances repose sur deux caractéristiques fondamentales supplémentaires : l'organisation semi-hiérarchique des concepts et la formation de relations entre des concepts éloignés, dites des relations transversales. Ces deux caractéristiques ne font toutefois pas l'unanimité dans les écrits scientifiques et plusieurs chercheurs ne les ont donc pas adoptées dans leurs travaux, surtout dans le contexte de l'évaluation des apprentissages (voir, Rice et coll., 1998 ; Rye et Rubba, 2002 ; Yin et Shavelson, 2008).

Malgré le fait qu'elle ait démontré son utilité dans plusieurs contextes, la cartographie conceptuelle ne se prête pas à toutes les situations d'apprentissages ou d'évaluation parce qu'elle ne convient pas à représenter tous les types de connaissances. De façon générale, elle cadre mieux avec les relations de type statique (Safayeni et coll., 2005) comme les relations d'inclusion de classes (*félins* se composent de *chats*, *tigres*...), d'appartenance commune (*chat*

et *tigre* font partie des *félins*) ou d'intersection de sens (le *triangle* a un côté de moins que le *carré*) et moins bien avec les relations de type dynamique comme les relations de cause à effet (l'augmentation de la *concentration des gaz à effet de serre* accroît la *température de la Terre*) ou les relations de nature probabiliste/corrélacionnelle.

7.2.2 La notation d'une carte conceptuelle

Les écrits scientifiques font état d'une très grande diversité de modèles pour noter des cartes conceptuelles (voir les recensions de Anohina et Grundspenkis, 2009 ; Ruiz-Primo et Shavelson, 1996). Ces modèles se distinguent par rapport aux indicateurs observables retenus, à leur pondération et à la manière de combiner les scores partiels en un score total.

Novak et Gowin (1984) ont proposé ce qui est probablement le premier modèle de notation et celui qui est considéré, pour certains, comme l'un des plus détaillés sur le plan théorique (Ruiz-Primo et Shavelson, 1996). Leur modèle, parfois nommé la notation structurelle, est composé de quatre indicateurs qui reflètent différents mécanismes de la théorie de l'assimilation de connaissances d'Ausubel (Ausubel et coll., 1978). Selon ce modèle, le correcteur accorde 1 point à chaque relation valide, 10 points à chaque niveau hiérarchique valide, 5 points à chaque relation transversale valide et 1 point à chaque exemple valide. Le score total est tout simplement la somme des scores partiels décernés à chaque indicateur.

Novak et Gowin (1984) l'admettent eux-mêmes, la pondération des indicateurs de leur modèle est fondée sur leur propre expérience et est tout à fait arbitraire ; c'est pourquoi ils conseillent de l'adapter au besoin. Leur pondération est néanmoins souvent adoptée telle quelle et, à notre connaissance, aucune étude psychométrique n'a cherché à déterminer la pondération optimale de ces indicateurs.

La notation de la qualité des propositions est une seconde famille de modèles de notation des cartes conceptuelles qui a retenu l'attention de plusieurs chercheurs (McClure et coll., 1999 ; Ruiz-Primo, Schultz, et coll., 2001 ; Srinivasan et coll., 2008). Cette approche consiste à juger et à noter la qualité de chaque proposition, tantôt sur une échelle dichotomique (0 = proposition incorrecte ; 1 = proposition correcte), tantôt sur une échelle à plusieurs catégories de scores (par ex., 0 = proposition incorrecte ou scientifiquement non pertinente, 1 = proposition partiellement incorrecte, 2 = proposition correcte, mais scientifiquement « mince », 3 = proposition

scientifiquement correcte et formulée en termes scientifiques, Yin et coll., 2005), puis à combiner les scores individuels en un score total. Ce modèle a l'avantage d'être basé sur un seul indicateur et il n'est donc pas nécessaire de pondérer plusieurs indicateurs comme pour la notation structurelle.

Cela dit, les modèles de notation de la qualité des propositions comportent parfois une pondération des propositions, de manière à les départager selon leur degré de pertinence à l'égard de la tâche. Rye et Rubba (2002), par exemple, attribuent 4 points à une relation valide et principale et 2 points à une relation valide, mais secondaire. Yin et Shavelson (2008) adoptent quant à eux une approche beaucoup plus stricte, en corrigeant et en notant exclusivement les propositions dérivées d'une liste de paires de concepts jugées critiques⁴⁶. Cette approche, qui est équivalente à attribuer une pondération de zéro à toutes les propositions formées à partir de la liste de paires de concepts critiques, a un avantage considérable en comparaison aux autres modèles de notation de la qualité des propositions : les paires de concepts critiques sont connues à l'avance et peuvent ainsi être traitées comme des items d'un test, ce qui ouvre la porte à l'application de plusieurs modèles de mesure ou de certaines techniques d'appréciation des propriétés psychométriques qui ne sont pas accessibles autrement.

C'est ainsi que Yin et Shavelson (2008) arrivent à examiner les propriétés psychométriques d'une tâche de cartographie conceptuelle à l'aide de la théorie de la généralisabilité. Plus précisément, dans leur étude, ils essaient d'établir les conditions minimales pour atteindre un coefficient de généralisabilité d'au moins 0,80, le seuil qu'ils considèrent comme la valeur minimale acceptable pour la fidélité des scores d'une tâche d'évaluation à grande échelle. Au cours de leur expérimentation, chaque participant a effectué la même tâche à deux occasions et chaque fois, la carte est corrigée et notée en tenant compte de seulement seize paires de concepts critiques. Les résultats de leur recherche montrent toutefois que ces conditions ne sont pas suffisantes pour atteindre le niveau de fidélité souhaité. Les chercheurs obtiennent une valeur du coefficient de généralisabilité de 0,70 et ils estiment que, pour atteindre le seuil de 0,80, il serait nécessaire d'évaluer les apprentissages en une occasion au moyen d'une

⁴⁶ Yin et Shavelson définissent une proposition comme une triade « concept, relation, concept ». Leur approche peut néanmoins s'étendre à des propositions plus longues, formées de trois concepts ou plus.

tâche composée d'au moins 30 paires de concepts critiques ou encore en deux occasions à l'aide d'une tâche composée d'au moins 18 paires de concepts critiques.

Les modèles de la théorie de la généralisabilité, qui sont ni plus ni moins des extensions de modèles basés sur des scores observés, sont utiles pour estimer et séparer les sources d'erreur de la mesure, mais ils offrent très peu d'information au sujet des propriétés internes du modèle de notation. La modélisation ne permet pas, par exemple, de déterminer si certaines propositions devraient être ajoutées ou éliminées de la notation. Nonobstant les limites de la théorie de la généralisabilité, le modèle de notation proposé par Yin et Shavelson (2008) demeure très polyvalent et peut être adapté à d'autres modèles de mesure, notamment aux modèles à variables latentes, qui proposent plus d'options pour apprécier les propriétés psychométriques d'une tâche d'évaluation des apprentissages.

Un modèle à variable latente est un modèle mathématique qui formalise explicitement la relation entre une ou des variables latentes, c'est-à-dire une variable qui n'est pas observée directement (par ex., l'intelligence générale) et les données collectées. Cette variable hypothétique est donc inférée à partir d'observations. La modélisation basée sur des variables latentes repose par conséquent sur des conditions fortes, et est accompagnée de tests empiriques pour vérifier l'adéquation du modèle ou de ses conditions d'applications (Borsboom, 2005), ces tests étant utiles pour examiner les propriétés de la modélisation ou pour comparer un ou plusieurs modèles concurrents.

La plupart des modèles de mesure à variable latente couramment utilisés et appliqués en éducation (par ex., les modèles unidimensionnels de la théorie de la réponse à l'item ou les modèles de la famille de Rasch) s'appuient d'ailleurs sur les deux mêmes conditions d'application : l'unidimensionnalité et l'indépendance locale des items. La première condition stipule qu'une seule variable (latente) explique les variations au niveau des variables observables tandis que la seconde condition stipule qu'aucun facteur, autre que la variable latente évaluée, n'explique la covariance entre les scores aux items. Ces conditions représentent un idéal. En pratique, les données violent toujours, à un certain degré, ces hypothèses, et il est dès lors important de les vérifier empiriquement. Un écart trop important par rapport aux conditions d'application peut avoir un impact sur la validité de l'interprétation des scores. Cet écart peut biaiser l'estimation des paramètres d'items et l'estimation des niveaux d'habiletés, en

plus d'influencer toutes les analyses statistiques réalisées a posteriori (Baghaei, 2008). Il est d'ailleurs bien connu que la violation de la condition d'indépendance locale des items a un effet sur les propriétés d'un test, comme sur la fidélité des scores (Sireci, Thissen et Wainer, 1991).

Les applications de modèles de mesure basés sur des variables latentes à la notation de cartes conceptuelles sont rares. En fait, nous en avons recensé une seule dans laquelle les scores de la notation structurelle ont été modélisés à partir d'un modèle de la théorie de la réponse à l'item (Liu, 1994). Dans cette étude, les cartes d'élèves sont corrigées selon la méthode de Novak et Gowin, puis les scores partiels de chacun des quatre indicateurs sont transformés en catégories de scores. Chaque indicateur forme ainsi un item dont les scores partiels sont répartis entre 8 et 10 catégories de scores selon des règles empiriques définies a posteriori. Les scores aux items sont ensuite modélisés à partir du modèle gradué pour réponses ordonnées (Samejima, 1969) dans le logiciel MULTILOG (Thissen, 1991). Dans une série d'analyses a posteriori visant à récolter des preuves de validité et de fidélité de la cartographie conceptuelle comme tâche d'évaluation alternative en sciences, l'auteur de la recherche observe des valeurs de corrélation statistiquement significatives entre les scores obtenus lors de la modélisation, les scores bruts de la notation structurelle et les scores bruts d'un examen de rendement scolaire, de même qu'il estime la fidélité des scores à 0,85⁴⁷. Ces résultats amènent l'auteur à conclure que les scores produits avec l'approche proposée sont valides et fidèles.

La solidité de ces conclusions dépend toutefois de certaines conditions qui n'ont pas été contrôlées dans ces analyses, comme la robustesse des estimations, l'adéquation entre les données et le modèle ou les conditions d'application du modèle de mesure. La modélisation proposée par Liu comporte en effet un grand nombre de catégories de scores et dans les circonstances, le nombre de répondants (N = 92) n'est probablement pas suffisant pour obtenir des valeurs stables des paramètres estimés. La modélisation proposée fait aussi l'hypothèse que les items formés des quatre indicateurs de la notation structurelle sont indépendants entre eux. Cette hypothèse reste aussi à être vérifiée.

L'existence de tests empiriques permettant de vérifier les conditions d'application d'un modèle de mesure est un avantage important de la modélisation basée sur des variables latentes

⁴⁷ Cette valeur est obtenue à partir de la moyenne des erreurs de modélisation et de la variance des scores.

par rapport à la modélisation basée sur des scores observés, notamment lorsque vient le temps d'apprécier les propriétés psychométriques d'une tâche d'évaluation des apprentissages. Dans notre recherche, nous devons, bien entendu, choisir un modèle de mesure qui est cohérent avec le modèle de notation des cartes conceptuelles, mais ce modèle doit aussi être accompagné des outils méthodologiques nécessaires pour vérifier ses conditions d'application. La famille de modèles de Rasch répond bien à ces besoins.

7.2.3 La modélisation selon la famille des modèles de Rasch

La famille des modèles de Rasch est un ensemble de modèles de mesure qui fait l'hypothèse d'une relation non linéaire, décrite par une fonction logistique, entre les scores observés et le niveau d'habileté de l'individu. Parmi ces modèles, une variante bien connue, le modèle de Rasch pour données dichotomiques (Rasch, 1960), modélise les réponses aux items qui sont catégorisées selon un choix binaire, 0 ou 1. Une autre variante, le modèle à crédit partiel (Masters, 1982), est tout simplement la généralisation du modèle pour données dichotomiques lorsque les réponses sont divisées en plus de deux catégories. Puisque le modèle à crédit partiel est équivalent au modèle pour données dichotomiques lorsque les items comportent seulement deux catégories de scores, nous présentons seulement ce modèle plus général dans ce qui suit.

Prenons une variable aléatoire en valeur entière X_{ni} dont la valeur observée est un nombre entier se situant entre 0 et un score maximal m_i . La probabilité d'observer la réalisation $X_{ni} = x \in \{0, 1, \dots, m_i\}$ correspond à :

$$P_{ni}(X_{ni} = x | \theta_n, b_{ki}) = \frac{\exp \sum_{k=1}^x (\theta_n - b_{ki})}{1 + \sum_{j=1}^{m_i} \exp \sum_{k=1}^j (\theta_n - b_{ki})} \quad (7.1)$$

Où θ_n symbolise la variable latente visée par l'instrument et b_{ki} désigne la difficulté relative de se situer dans la catégorie de score k par rapport à la catégorie de score $k-1$. Les valeurs prises par les paramètres θ_n et b_{ki} indiquent respectivement la position de l'individu et la position de la transition entre les catégories de scores k et $k-1$ sur le continuum de la variable latente. Ces valeurs sont exprimées naturellement selon l'échelle du logarithme du rapport des chances, nommée le logit.

La modélisation de Rasch repose sur les deux conditions d'application mentionnées précédemment : l'unidimensionnalité et l'indépendance locale des items. Le respect des conditions d'application n'est pas le seul aspect à considérer lors de la modélisation des scores. Il faut aussi porter une attention particulière à l'adéquation entre les données et le modèle (Smith, 1999), ainsi qu'à la robustesse des estimations des paramètres du modèle (De Ayala, 2009).

7.3 Objectif de la recherche

Comme il a été mentionné ci-dessus, la présente étude vise à illustrer les avantages d'appliquer la modélisation de Rasch, et plus particulièrement le modèle pour données dichotomiques et le modèle à crédit partiel, pour apprécier les propriétés psychométriques de tâches de cartographie conceptuelle. En l'absence de travaux antérieurs solides, il nous apparaît prudent de commencer avec un modèle de notation simple des cartes conceptuelles comme celui de la qualité des propositions critiques et de vérifier dans quelle mesure les données obtenues au moyen de ce modèle de notation se conforment aux conditions de la modélisation de Rasch.

7.4 Méthodologie

7.4.1 Participants

La collecte de données a été réalisée dans une école secondaire de la grande région de Montréal au Canada. Au total, 85 élèves de quatrième secondaire en sciences et technologie et leur enseignant ont participé à l'étude.

7.4.2 Formation des participants

Les activités ont été réparties en trois séances afin que les élèves se familiarisent progressivement à l'outil de représentation des connaissances. Les élèves ont d'abord participé à une activité d'introduction à la cartographie conceptuelle. Ils ont ensuite construit une carte conceptuelle lors d'une tâche pratique et une autre carte lors de la tâche principale de l'étude. Ces activités ont été pilotées par le chercheur principal et en collaboration avec l'enseignant

responsable des élèves. À la fin des deux premières activités, les élèves ont reçu des rétroactions pour les guider dans la construction de leur carte conceptuelle.

7.4.3 Description de la tâche principale

Chacune des trois activités a été développée conjointement avec l'enseignant de manière à ce qu'elle s'intègre dans la planification des apprentissages du cours. L'enseignant a choisi le thème des relations trophiques⁴⁸ comme tâche principale. Une liste de 17 concepts a été fournie aux élèves pour réaliser cette tâche. Les concepts, inspirés du contenu du manuel scolaire des élèves, ont été choisis par l'enseignant en collaboration avec le chercheur principal. Une période d'environ 40 minutes a été accordée aux élèves pour construire leur carte conceptuelle. Près des deux tiers d'entre eux ont produit leur carte à l'aide du logiciel CMap Tools (Cañas et coll., 2004) et le reste des élèves ont conçu leur carte de manière manuscrite, en recourant à des étiquettes amovibles (*Post-it*). Même si les élèves étaient invités à représenter leurs connaissances sans recourir à un quelconque document, certains ont consulté leur manuel scolaire ou encore le web au cours de la réalisation de la tâche.

7.4.4 Élaboration de la carte de référence

Une carte de référence a été élaborée à partir de la liste des concepts choisis par l'enseignant et dans le but d'établir les propositions critiques. Il existe plusieurs méthodes pour construire une carte de référence dans les écrits, mais l'approche la plus courante consiste à demander à un ou des experts de contenu de participer à l'élaboration de cette carte (Ruiz-Primo, Schultz, et coll., 2001 ; Yin et coll., 2005). Cette méthode a pourtant ses limites. Les experts de contenu possèdent des représentations idiosyncrasiques des connaissances (Acton et coll., 1994) qui ne se réconcilient pas nécessairement à celles de novices (Chi, Feltovich et Glaser, 1981 ; Chi, Glaser et Farr, 1988 ; Glaser, 1991).

Une autre approche qui nous semble prometteuse est de produire une carte de référence à partir des propositions contenues dans un corpus de textes didactiques. Cette méthode vise à extraire les relations formulées entre deux concepts dans une même phrase. Chaque proposition

⁴⁸ Les relations trophiques concernent les relations alimentaires entre les êtres vivants.

extraite est ainsi représentée sous forme de paires de concepts reliés par une expression de liaison. Trois manuels scolaires en science et technologie et conformes au plus récent curriculum de formation générale au Québec ont été analysés selon cette méthode. Ces ouvrages constituent des matériaux de premier choix : ils sont composés et révisés par des experts du domaine de connaissances, mais surtout, ils sont rédigés dans un but didactique. Leur contenu est ainsi plus près de ce qui est attendu de l'élève. La codification des textes a été réalisée par le premier auteur de cet article, également formé en enseignement au secondaire en sciences et technologie.

La carte extraite des textes (voir la Figure 18) compte 16 concepts, 28 paires de concepts ayant au moins une relation et 31 propositions distinctes. Le concept « Animal », faisant partie de la liste de concepts de départ, n'apparaît pas dans la carte de référence puisqu'il ne forme aucune proposition avec les autres concepts de la liste dans les textes analysés.

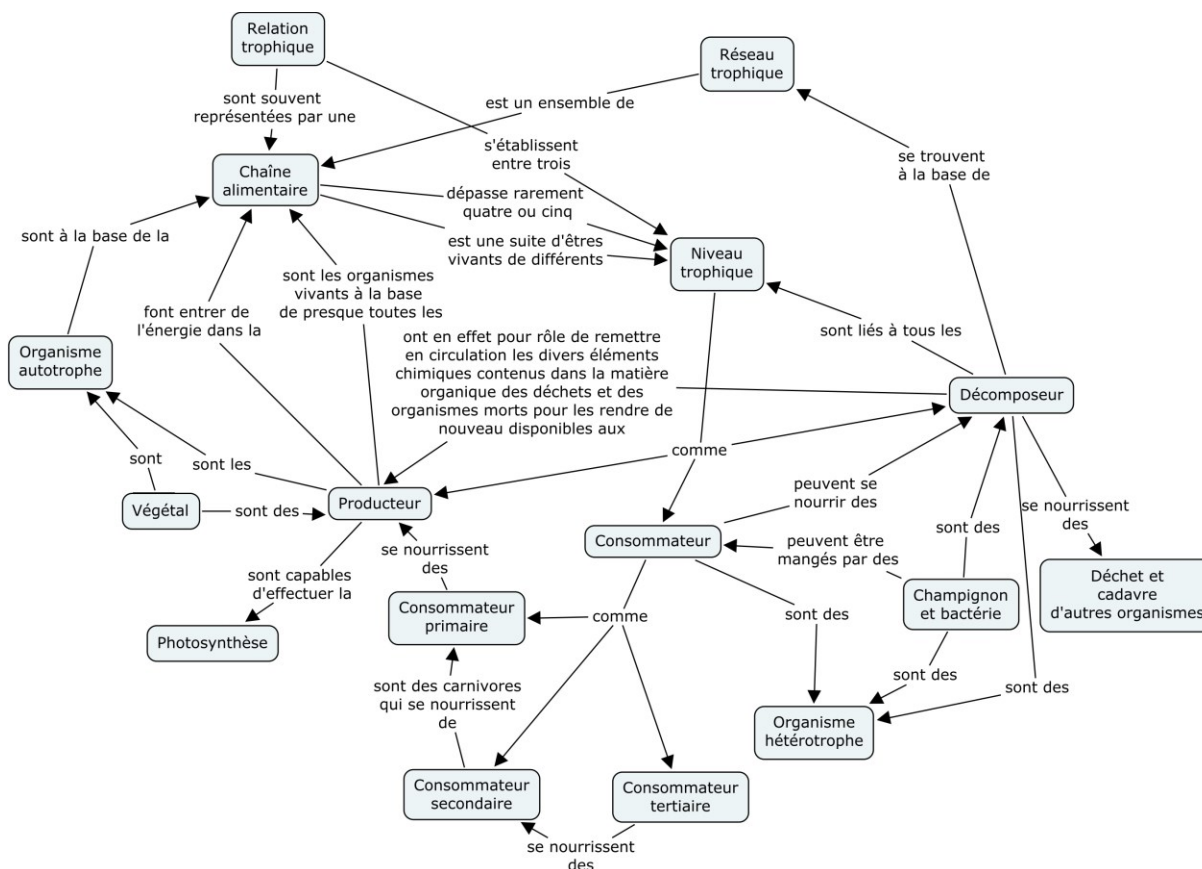


Figure 18. Carte conceptuelle sur le thème « Relations trophiques » tiré des trois manuels scolaires⁴⁹

Une analyse préliminaire des cartes produites par les élèves nous a toutefois poussés à réviser la carte de référence, car ceux-ci ont réussi à trouver d'autres manières d'organiser et de représenter leurs connaissances. Comme il a été fait dans quelques études antérieures (Ruiz-

⁴⁹ Les relations ne sont pas nécessairement en accord grammatical avec les concepts qu'elles unissent. Les concepts sont inscrits systématiquement au singulier, mais les expressions de relation respectent la formulation du texte duquel elles sont extraites. Ainsi, la phrase « Les décomposeurs se trouvent à la base des réseaux trophiques » est représentée par les énoncés de concepts « Décomposeur » et « Réseau trophique » au singulier, et par la relation « se trouvent à la base des » telle qu'extraite dans le texte original.

Primo, Schultz, et coll., 2001 ; Yin et coll., 2005), nous avons eu à réajuster notre carte de référence pour inclure certaines des représentations alternatives des participants.

Les paires de concepts du Tableau XXXIII ont été ajoutées aux paires de concepts issues de la carte présentée à la Figure 18. Nous avons d'abord ajouté cinq paires de concepts formées à partir du concept « Animal ». Même si ce concept est absent de la carte de référence initiale, il fait partie de nombreuses propositions formulées par les élèves⁵⁰. Les paires de concepts 1 à 5 du Tableau XXXIII ont été ajoutées à la liste de paires de concepts critiques. De façon similaire, nous avons découvert que certaines paires de concepts absentes de la carte (n° 6 à 11) servent à énoncer des propositions équivalentes à celles d'autres paires de concepts présentes dans la carte. Il sera vu plus loin dans le texte qu'il est important de considérer toutes les manières de formuler une même idée.

Tableau XXXIII. Paires de concepts ajoutées à la carte de référence initiale

<i>No.</i>	<i>Paire de concepts</i>	<i>Nbre de propositions dans les cartes des participants</i>	<i>Nbre de propositions valides</i>
1	Animal, Consommateur	49	45
2	Animal, Consommateur primaire	27	24
3	Animal, Consommateur secondaire	33	20
4	Animal, Consommateur tertiaire	25	16
5	Animal, Hétérotrophe	35	31
6	Photosynthèse, Végétal	73	65
7	Cadavre et déchet d'organismes, Champignon et bactérie	50	44
8	Niveau trophique, Réseau trophique	26	17
9	Relation trophique, Réseau trophique	43	5
10	Consommateur primaire, Végétal	17	17
11	Autotrophe, Photosynthèse	9	3

⁵⁰ Le concept « Animal » ne forme aucune relation avec les autres concepts dans les trois textes analysés lors de la production de la carte de référence.

La nouvelle carte de référence, qui comporte maintenant 39 paires de concepts distinctes (28 paires tirées de la carte initiale et 11 paires ajoutées par la suite) a servi à l'analyse des cartes produites par les élèves.

7.4.5 Correction et notation des cartes des élèves

Dans le but de simplifier la correction, nous avons suivi la procédure de notation de l'inventaire de propositions de Ruiz-Primo, Schultz, et coll. (2001) qui consiste à rassembler les propositions identiques, puis à noter une seule fois chaque proposition distincte. Les propositions ont été notées selon une échelle de score dichotomique : soit la relation est valide (1), soit la relation est absente ou invalide (0). Ce sont ces données dichotomiques qui ont été analysées lors de la première modélisation.

7.4.6 L'appréciation de la modélisation de Rasch

Cette section détaille les différentes techniques et statistiques utilisées pour apprécier les conditions d'application du modèle de Rasch, l'adéquation entre les données et le modèle, ainsi que la fidélité des scores. La modélisation de Rasch a été réalisée à l'aide du logiciel Winsteps (Linacre, 2017) alors que certaines analyses secondaires telles que la vérification de l'indépendance locale des items et l'examen de la dimensionnalité ont été effectuées dans le programme d'analyse statistique R (R Development Core Team, 2017).

7.4.6.1 Critères d'appréciation de la condition d'indépendance locale des items

Dans le contexte de la modélisation de Rasch, la condition d'indépendance locale des items indique que la probabilité de répondre correctement à un item, conditionnelle à la position sur le continuum de la variable latente, ne dépend pas de la probabilité conditionnelle de répondre correctement à un autre item. Cela revient à dire qu'une fois que l'effet du trait latent est soustrait, la corrélation entre les résidus⁵¹ de la modélisation doit être égale ou proche de zéro pour chaque paire d'items. Conformément à la technique proposée par Yen (1984), nous utilisons donc la corrélation entre les résidus des paires d'items pour détecter les situations de

⁵¹ Un résidu est la différence entre une valeur observée et une valeur prédite par le modèle.

dépendance locale entre les items. Plus spécifiquement, nous identifions les items pour lesquels les valeurs r de la corrélation sont statistiquement différentes de zéro ($H_0 : r = 0,0$, $H_1 : r \neq 0,0$), après avoir corrigé le seuil de signification pour les comparaisons multiples selon la méthode de Holm (1979).

7.4.6.2 Critères d'appréciation de la condition d'unidimensionnalité

Une méthode courante pour détecter la violation potentielle de la condition d'unidimensionnalité consiste à réaliser une analyse en composantes principales⁵² (ACP) des résidus standardisés (Linacre, 1998). Comme Linacre le souligne, le modèle de Rasch impose une structure unidimensionnelle aux réponses, peu importe la dimensionnalité des données ; les corrélations entre les résidus doivent alors être proches de zéro si la condition d'unidimensionnalité est respectée. L'ACP aide à établir s'il y a un ou plusieurs facteurs secondaires qui expliquent les valeurs observées de corrélations entre les résidus des items. Dans cette recherche, nous déterminons le nombre de composantes à extraire de l'ACP à l'aide de la méthode de l'analyse parallèle de Horn (1965)⁵³ et nous considérons qu'une variable observable contribue à une composante lorsque la valeur de son coefficient de saturation dépasse 0,30 en valeur absolue.

7.4.6.3 Critères d'appréciation de l'adéquation entre les données et le modèle

Dans une démarche de construction d'une échelle de scores, il est assez commun d'examiner les propriétés des items afin de vérifier l'adéquation entre les données et le modèle. Conformément à cette pratique, nous calculons dans un premier temps la difficulté des items et la discrimination des items. La difficulté de l'item correspond tout simplement à la moyenne

⁵² Voir Tabachnick et Fidell (2013) pour une description exhaustive de cette technique statistique multivariée.

⁵³ L'analyse parallèle de Horn est une méthode éprouvée pour établir le nombre de composantes à extraire (Zwick et Velicer, 1986). Cette procédure consiste à comparer les valeurs propres (qui représentent la variance expliquée) estimées à partir des données originales à la moyenne de valeurs propres obtenues lors d'ACP de matrices de données générées au hasard et contenant le même nombre d'observations et de variables que les données originales. Les composantes retenues correspondent alors aux valeurs propres des données originales qui dépassent celles des données générées aléatoirement.

des scores à l’item. Un nombre élevé d’items trop faciles ou trop difficiles indique que la tâche n’est peut-être pas adaptée aux répondants. La discrimination est quant à elle estimée au moyen du coefficient de corrélation item-total corrigée⁵⁴. Une valeur faible ou négative signale généralement un problème d’adéquation de l’item à l’échelle des scores (Laveault et Grégoire, 2002). Dans cette étude, nous signalons les items dont l’indice de corrélation item-total corrigée est inférieur à 0,10.

Dans un deuxième temps, nous faisons appel aux statistiques *Infit* et *Outfit* pour estimer le degré d’ajustement des items au modèle de Rasch⁵⁵. Plus précisément, nous avons recours à la version du carré moyen (en anglais, *mean square*) de ces statistiques dont la valeur attendue est de 1 et l’étendue est de 0 à $\infty+$. Une valeur empirique qui s’écarte trop de la valeur attendue indique une mauvaise adéquation. Dans nos analyses, nous accordons une attention particulière à la version non pondérée, c’est-à-dire à la statistique *Outfit* pour laquelle Wu et Adams (2013) ont dérivé les seuils critiques à l’aide de ses propriétés asymptotiques. Ses seuils sont établis selon l’équation suivante :

$$1 \pm 2\sqrt{2/N} \quad (7.2)$$

; où N est le nombre d’observations. Les valeurs à l’extérieur de ces seuils signalent un mauvais ajustement des données au modèle.

7.4.6.4 Critère d’appréciation de la fidélité des scores

La fidélité est un indicateur de la précision des scores d’un test, ou plus spécifiquement de la cohérence et de la reproductibilité des scores (Laveault et Grégoire, 2002). L’indice de fidélité de séparation des personnes⁵⁶, aussi nommé l’indice de fidélité des personnes, est une statistique dérivée de la modélisation de Rasch et analogue aux indices traditionnels de fidélité d’un test. L’étendue théorique de ses valeurs est de 0 à 1, et plus sa valeur est élevée, plus il est probable que les répondants dont le niveau d’habileté est élevé soient véritablement différents

⁵⁴ Le score à l’item est soustrait du score total pour éviter la colinéarité.

⁵⁵ Voir Smith (1999) pour une discussion complète des méthodes d’analyse de l’ajustement des données au modèle.

⁵⁶ Smith Jr (2001) fournit une description plus détaillée de cet indice et d’autres indices de fidélité dans le cadre de la modélisation de Rasch.

(sur l'échelle de mesure) des répondants dont le niveau d'habileté est faible. Un indice de fidélité de 0,70 est la valeur minimalement requise pour diviser les personnes en au moins deux niveaux d'habiletés distincts (appelés des strates), c'est-à-dire deux niveaux d'habiletés séparés par au moins trois erreurs-types de mesure, alors qu'une valeur de 0,80 est nécessaire pour constituer trois strates de personnes (Smith Jr, 2001).

7.5 Résultats

La présente étude étant essentiellement exploratoire, nous avons effectué une série de modélisations et d'analyses afin de mieux comprendre les mécanismes sous-jacents à la construction d'une échelle de scores avec des données issues de cartes conceptuelles. Les résultats de quatre modélisations sont présentés dans les sous-sections suivantes. Sauf indication contraire, les analyses sont basées sur les réponses de 84 répondants. Un répondant n'a pas respecté les consignes de production d'une carte conceptuelle : les liens entre les concepts n'étaient pas accompagnés de mots de liaison. Ce n'est qu'après la collecte de données qu'il a été réalisé que ce participant était absent lors des deux premières séances d'activités préparatoires. Les réponses de ce participant ont donc été exclues des analyses suivantes.

7.5.1 Première modélisation : Paire de concepts comme des items

L'inventaire des 39 propositions distinctes a servi à modéliser les scores issus de cartes conceptuelles à l'aide du modèle de Rasch pour données dichotomiques. Les statistiques d'items de cette modélisation sont rapportées dans le Tableau XXXIV. Quatre paires de concepts n'ont pas été modélisées parce qu'aucun élève n'a formulé de propositions justes à partir de celles-ci⁵⁷.

⁵⁷ Dans la modélisation de la TRI ou dans la modélisation de Rasch, le paramètre de difficulté d'un item est inestimable lorsque tous les individus y ont répondu correctement ou incorrectement, autrement dit lorsque la proportion de bonnes réponses est de 1,0 ou de 0,0.

Tableau XXXIV. Statistiques d'items de la première modélisation

No.	Paire de concepts	Score moyen	Corrélation item – total corrigée	Paramètre de difficulté	<i>Infit</i>	<i>Outfit</i>
1	Animal, Consommateur	0,53	0,18	-1,29	1,04	1,01
2	Animal, Consommateur primaire	0,28	-0,01	-0,11	1,12	2,03
3	Animal, Consommateur secondaire	0,24	0,16	0,16	1,01	0,95
4	Animal, Consommateur tertiaire	0,19	0,23	0,46	0,96	0,86
5	Animal, Hétérotrophe	0,36	0,04	-0,53	1,12	1,17
6	Autotrophe, Photosynthèse	0,04	-0,02	2,38	1,04	1,16
7	Autotrophe, Producteur	0,48	0,12	-1,08	1,09	1,10
8	Autotrophe, Végétal	0,49	-0,06	-1,13	1,22	1,33
9	Cadavre et déchet d'organismes, Champignon et bactérie	0,52	0,04	-1,24	1,15	1,14
10	Cadavre et déchet d'organismes, Décomposeur	0,20	0,18	0,38	1,00	0,90
11	Chaîne alimentaire, Niveaux trophique	0,32	0,19	-0,30	1,01	0,95
12	Chaîne alimentaire, Producteur	0,04	0,01	2,38	1,03	1,01
13	Chaîne alimentaire, Relation trophique	0,09	0,08	1,31	1,01	1,40
14	Chaîne alimentaire, Réseau trophique	0,16	0,23	0,63	0,95	0,86
15	Champignon et bactérie, Décomposeur	0,86	0,34	-3,29	0,94	0,84
16	Champignon et bactérie, Hétérotrophe	0,04	0,01	2,38	1,03	1,04
17	Consommateur, Consommateur primaire	0,55	0,48	-1,40	0,83	0,80
18	Consommateur, Consommateur secondaire	0,53	0,46	-1,29	0,85	0,82
19	Consommateur, Consommateur tertiaire	0,54	0,46	-1,35	0,85	0,81
20	Consommateur, Décomposeur	0,01	0,05	3,51	1,00	0,76
21	Consommateur, Hétérotrophe	0,41	0,18	-0,75	1,03	0,98
22	Consommateur, Niveau trophique	0,45	0,28	-0,91	0,96	0,94

No.	Paire de concepts	Score moyen	Corrélation item – total corrigée	Paramètre de difficulté	<i>Infit</i>	<i>Outfit</i>
23	Consommateur primaire, Consommateur secondaire	0,28	0,30	-0,11	0,93	0,84
24	Consommateur primaire, Producteur	0,11	0,19	1,17	0,99	0,75
25	Consommateur primaire, Végétal	0,20	0,11	0,38	1,02	1,27
26	Consommateur secondaire, Consommateur tertiaire	0,28	0,29	-0,11	0,93	0,85
27	Décomposeur, Hétérotrophe	0,11	0,10	1,17	0,99	1,20
28	Décomposeur, Niveau trophique	0,48	0,40	-1,08	0,87	0,83
29	Niveau trophique, Producteur	0,47	0,44	-1,02	0,85	0,80
30	Niveau trophique, Relation trophique	0,06	0,26	1,83	0,92	0,59
31	Niveau trophique, Réseau trophique	0,20	0,15	0,38	1,02	0,95
32	Photosynthèse, Producteur	0,07	-0,19	1,63	1,11	2,50
33	Photosynthèse, Végétal	0,76	0,30	-2,53	0,99	0,93
34	Producteur, Végétal	0,75	0,18	-2,46	1,06	1,10
35	Relation trophique, Réseau trophique	0,06	-0,04	1,83	1,05	1,97
36*	Décomposeur, Producteur	-	-	-	-	-
37*	Décomposeur, Réseau trophique	-	-	-	-	-
38*	Autotrophe, Chaîne alimentaire	-	-	-	-	-
39*	Champignon et bactérie, Consommateur	-	-	-	-	-
	Moyenne	0,32	0,17	0,00	1,00	1,07
	Écart-type	0,23	0,16	1,56	0,09	0,39
	Minimum	0,01	-0,19	-3,29	0,83	0,59
	Maximum	0,86	0,48	3,51	1,22	2,50

* Item dont le paramètre de difficulté ne peut pas être estimé

Plus précisément, le Tableau XXXIV indique le score moyen aux items, le paramètre de difficulté, la corrélation item-total corrigée ainsi que les statistiques d'ajustements *Infit* et *Outfit*. De façon générale, les items sont assez difficiles pour les participants. Le score moyen est assez faible ($M = 0,32$; É.-T., 0,23) et les paramètres de difficulté des items ont en moyenne des valeurs plus élevées ($M = 0,00$; É.-T., 1,56) que les estimations des habiletés ($M = -1,16$; É.-T., 0,82). La Figure 19, couramment nommée une carte de Wright, illustre conjointement la distribution des estimations d'habiletés (à gauche) et la distribution des paramètres de difficulté des items (à droite). Idéalement, les deux distributions sont alignées et la distribution des paramètres d'items couvre toute l'étendue des estimations de niveau d'habileté des répondants. Nous observons plutôt qu'un bon nombre d'items sont très difficiles pour le groupe de participants et qu'il manque des items faciles pour les répondants les plus faibles. D'ailleurs, la fidélité de séparation des personnes de cette modélisation est assez faible, soit de 0,71, ce qui est tout juste au-dessus du seuil minimalement requis pour former deux niveaux d'habiletés distincts, tel que suggéré par Smith Jr (2001).

Carte de Wright

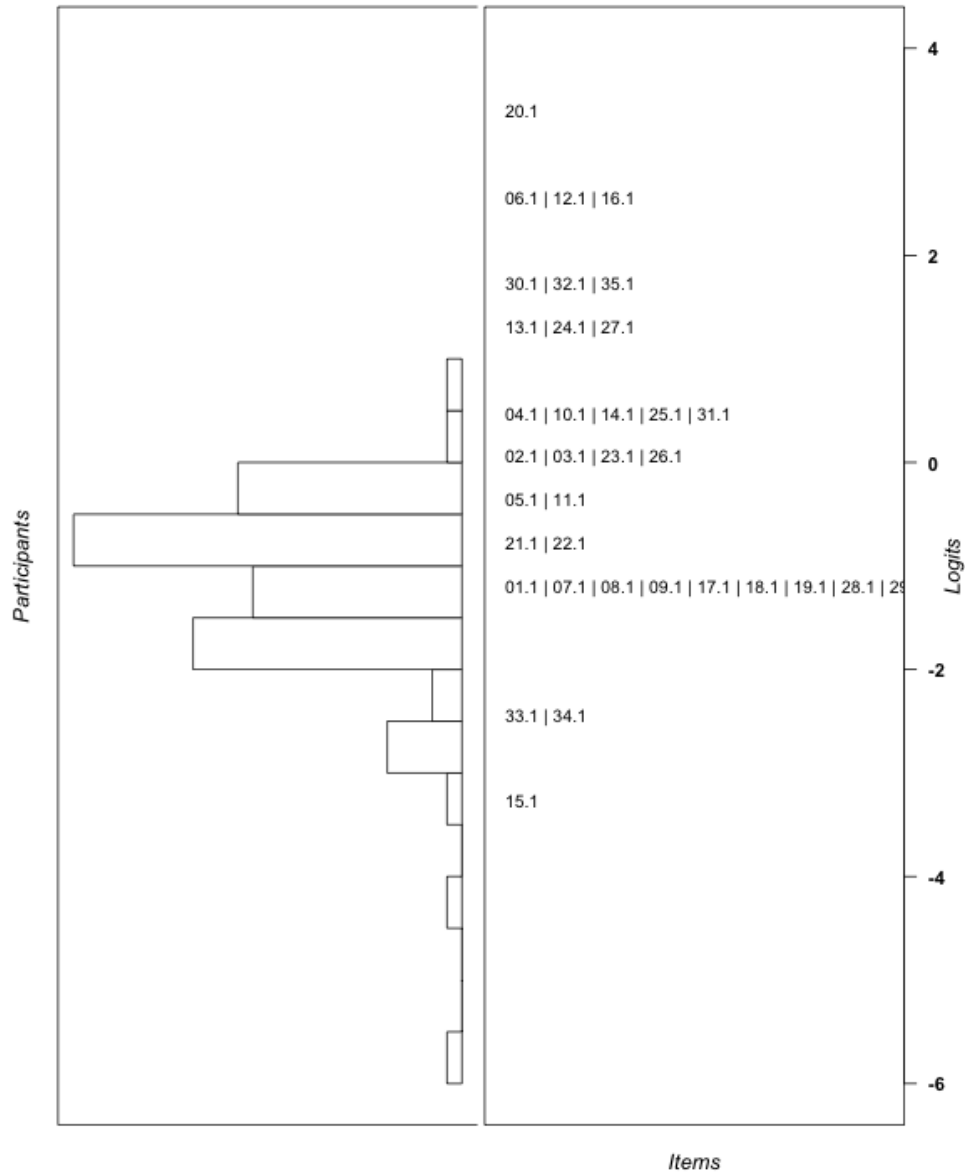


Figure 19. Distribution des estimations de niveaux d'habileté et des paramètres de difficulté des items pour la première modélisation

Dans le Tableau XXXIV, il y a aussi plusieurs indices montrant que les données ne s'ajustent pas pleinement au modèle. Parmi les 35 items modélisés, on compte 11 items dont

l'indice de discrimination est inférieur au seuil de 0,10. Plusieurs d'entre eux sont formés de paires de concepts qui, lorsqu'elles sont formulées en propositions, ont une signification similaire à une proposition formée d'une autre paire de concepts. Par exemple, les propositions « les *producteurs* font de la *photosynthèse* » et « les *végétaux* font de la *photosynthèse* » sont essentiellement équivalentes. À l'inverse, les items pour lesquels les indices de corrélation item-total corrigée sont les plus élevés (17, 18, 19, 22, 28, 29) désignent des relations de type hiérarchique. Ces résultats suggèrent que l'échelle de scores est largement influencée par certains regroupements d'items. Ce constat sera corroboré dans les prochaines analyses.

Par ailleurs, six items ont des valeurs de la statistique *Outfit* qui ne se situent pas entre les seuils critiques de 0,69 et 1,31 (dérivés selon l'équation 7.2 avec la valeur $N = 84$) : un seul item (30) se situe sous la barre de 0,69 et cinq items (2, 8, 13, 32, 35) ont une statistique supérieure à 1,31. En comparaison aux valeurs de la statistique *Outfit*, toutes les valeurs des statistiques *Infit* sont plus près de la valeur attendue de 1, ce qui suggère la présence de valeurs aberrantes dans le jeu de données, c'est-à-dire de réponses qui ne sont pas conformes au modèle.

Dans les circonstances, nous avons effectué quelques analyses supplémentaires afin de vérifier que les interprétations ne sont pas influencées par ces valeurs aberrantes. Pour ce faire, nous avons remplacé les réponses par des valeurs manquantes et nous avons répété la modélisation des scores. Cette démarche a d'ailleurs été réalisée séparément pour les personnes et les items signalés comme étant potentiellement problématiques lors des analyses de l'ajustement des données au modèle. Puisque les résultats de la modélisation et les interprétations des résultats étaient virtuellement les mêmes avant ou après avoir remplacé des réponses par des valeurs manquantes, nous avons retenu toutes les observations dans la modélisation.

Le Tableau XXXV présente les estimations des corrélations entre les résidus standardisés. Dans le but de simplifier la présentation des résultats, seules les paires d'items pour lesquelles les corrélations sont statistiquement différentes de zéro y sont affichées. Celles-ci sont regroupées en trois sources de violation de la dépendance locale selon une analyse du contenu des propositions. La source la plus importante reflète les relations de type hiérarchique (source I), pour lesquelles un concept général (par ex., consommateur) est associé à un ou plusieurs concepts subordonnés (par ex., consommateurs primaires). La deuxième source

(source II), très semblable à la première, illustre la propagation d'une même idée à travers une chaîne de concepts. Plus spécifiquement, elle consiste au déploiement de l'énoncé « le consommateur d'ordre supérieur se nourrit d'un consommateur d'ordre inférieur » dans une chaîne de deux propositions : « les *consommateurs tertiaires* se nourrissent des *consommateurs secondaires* » et « les *consommateurs secondaires* se nourrissent des *consommateurs primaires* ». La troisième source (III) reflète le fait que la même idée peut être représentée plus d'une façon dans la carte. Plusieurs autres paires de concepts se trouvent dans la même situation, même si elles ne sont pas identifiées dans le Tableau XXXV. Enfin, l'avant-dernière paire d'items présentée dans le Tableau XXXV n'a été catégorisée dans aucune de ces trois sources de violation : elle est formée d'items très difficiles et dont les estimations des paramètres de difficulté sont peut-être instables.

Tableau XXXV. Indices de corrélations statistiquement significatifs entre les résidus standardisés

Paire 1	Paire 2	Corrélation entre les résidus standardisés	Source de la violation
Consommateur, Consommateur secondaire	Consommateur, Consommateur tertiaire	0,97*	I
Consommateur, Consommateur primaire	Consommateur, Consommateur secondaire	0,94*	I
Consommateur primaire, Consommateur secondaire	Consommateur secondaire, Consommateur tertiaire	0,92*	II
Consommateur, Consommateur primaire	Consommateur, Consommateur tertiaire	0,91*	I
Animal, Consommateur secondaire	Animal, Consommateur tertiaire	0,74*	I
Décomposeur, Niveau trophique	Niveau trophique, Producteur	0,72*	I
Consommateur, Niveau trophique	Niveau trophique, Producteur	0,63*	I
Chaîne alimentaire, Relation trophique	Relation trophique, Réseau trophique	0,59*	III
Consommateur, Niveau trophique	Décomposeur, Niveau trophique	0,55*	I
Photosynthèse, Producteur	Relation trophique, Réseau trophique	0,47*	-
Animal, Hétérotrophe	Consommateur, Hétérotrophe	-0,51*	III

*p < 0,001 (ajustée pour les comparaisons multiples à l'aide de la correction de Holm).

Source I : relations de type hiérarchique; Source II : relations qui répètent la même idée; Source III : relations composées de paires d'items distincts, mais qui signifient la même chose

Les résultats de l'ACP corroborent les résultats de l'analyse des corrélations entre les résidus standardisés. En suivant la procédure de l'analyse parallèle (voir la Figure 20), six valeurs propres observées (valeurs réelles) se situent au-delà des intervalles de confiance calculés autour des valeurs propres générées à partir de matrices de données au hasard, c'est-à-dire au-delà de la limite supérieure de 95 % de l'analyse parallèle.

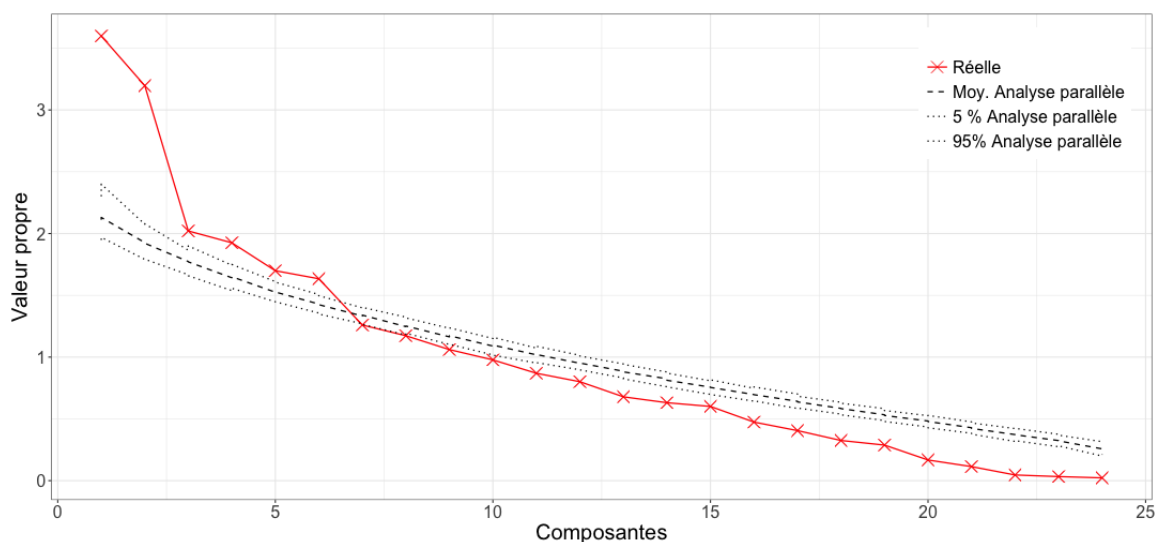


Figure 20. Diagramme des éboulis de l'analyse en composantes principales des résidus.

Note : comparaison des valeurs réelles (identifiées par un X) par rapport aux intervalles de confiance autour des valeurs propres générées au hasard.

Le Tableau XXXVI donne les valeurs des coefficients de saturation pour les six premières composantes, et ce, uniquement pour les items dont au moins un coefficient est inférieur à -0,30 ou supérieur à 0,30. Les résultats sont organisés de sorte à faire ressortir les regroupements naturels des items selon les composantes de l'ACP. En reprenant la typologie élaborée lors de l'analyse des corrélations des résidus standardisés (voir Tableau XXXV), les composantes 1, 3 et 4 représentent des regroupements d'items associés à la première source de violation, et la composante 2, à la seconde source. Les composantes 5, 6 et 7 sont associées à des items très difficiles et isolées. Aucun item n'est associé à la troisième source. Les trois premières composantes comportent respectivement 3, 2 et 3 items, ce qui est insuffisant pour

les traiter comme des habiletés distinctes; la formation de ces composantes corrobore les problèmes de dépendance locale des items.

Tableau XXXVI. Coefficients de saturation des six premières composantes de l'ACP des résidus standardisés de la première modélisation

Paire	Composante					
	1	2	3	4	5	6
Consommateur, Consommateur secondaire	-0,96	-0,09	-0,07	0,01	0,02	-0,11
Consommateur, Consommateur tertiaire	-0,96	-0,04	-0,09	0,02	0,02	-0,12
Consommateur, Consommateur primaire	-0,94	-0,08	-0,03	0,03	0,00	0,15
Consommateur primaire, Consommateur secondaire	0,08	0,92	-0,20	0,11	0,07	0,06
Consommateur secondaire, Consommateur tertiaire	0,14	0,92	-0,15	0,09	0,12	0,05
Décomposeur, Niveau trophique	0,15	-0,17	0,89	0,02	-0,06	0,03
Consommateur, Niveau trophique	0,18	-0,26	0,47	0,03	0,09	-0,12
Niveau trophique, Producteur	0,06	-0,22	0,84	-0,02	0,06	0,08
Animal, Consommateur tertiaire	0,03	-0,09	0,02	-0,93	0,00	-0,04
Animal, Consommateur secondaire	0,05	-0,19	-0,05	-0,77	0,08	-0,03
Animal, Hétérotrophe	0,06	-0,17	-0,01	0,05	-0,91	-0,05
Consommateur, Décomposeur	0,07	0,09	0,06	0,04	0,04	0,97

Pour récapituler, les résultats des analyses pour cette première modélisation montrent que les données ne satisfont pas pleinement les conditions de la modélisation des Rasch. Nous avons identifié trois sources de violation de l'indépendance locale, ce qui nous indique que ce ne sont pas toutes les paires de concepts qui peuvent être définies comme des items indépendants. Les prochaines modélisations tentent de remédier au problème de dépendance locale en introduisant deux procédés de formation d'items. Le premier consiste à réunir les idées redondantes en un seul item (source III) et le second à regrouper les relations de type hiérarchique⁵⁸ (sources I et II).

⁵⁸ Il est à noter que les sources I et II de violation de l'indépendance locale sont traitées de la même manière.

Afin de mieux évaluer les deux nouveaux mécanismes de formation d'items, nous avons réalisé trois modélisations : une pour chaque mécanisme et une combinant les deux mécanismes simultanément. Dans ce qui suit, nous présentons des résultats sommaires pour les deux premières modélisations et des résultats plus détaillés pour la troisième.

7.5.2 Deuxième modélisation : regroupement de propositions redondantes

Un individu qui construit une carte conceptuelle va rarement répéter deux fois la même idée. Prenons l'exemple de deux propositions similaires : « les *producteurs* sont des *autotrophes* » et « les *végétaux* sont des *autotrophes* ». La première a été formulée correctement par 48 % des répondants, la seconde par 49 % des répondants, mais seulement 14 % d'entre eux ont inscrit les deux propositions dans leur carte. Si l'on part du principe que ces propositions désignent une seule et même idée, nous proposons de les réarranger de sorte à les combiner en un seul item dichotomique; un score de 1 est attribué lorsqu'au moins une des deux propositions est énoncée correctement. Ainsi, pour le même exemple, la proportion de bonnes réponses de l'item créé suivant cette règle est de 0,84⁵⁹. Cet item est beaucoup plus facile ($b_{12} = -2,51$ logits) comparativement aux deux items de la modélisation initiale ($b_1 = -1,08$ et $b_2 = -1,13$).

En fin de compte, huit nouveaux items ont été créés à partir de ce mécanisme, et ce qui a grandement modifié les résultats de la modélisation. Tous les indices de discrimination des items sont maintenant positifs et l'indice de fidélité est de 0,75, et ce même si cette seconde modélisation comporte 12 items de moins que la précédente. Ces changements ont aussi eu pour effet de réduire l'écart entre les valeurs des distributions des paramètres d'items ($M = 0,00$; É.-T. = 1,92) et les estimations des niveaux d'habiletés ($M = -0,49$; É.-T. = 1,27).

7.5.3 Troisième modélisation : regroupement de propositions de type hiérarchique

La troisième modélisation a pour but d'éliminer la dépendance locale pour les items partageant une même relation de type hiérarchique. Une stratégie courante pour contrôler l'effet de la dépendance locale consiste à regrouper les items sous la forme d'un superitem à plusieurs

⁵⁹ Après avoir arrondi la valeur estimée.

catégories de scores (Sireci et coll., 1991), puis de les modéliser avec un modèle de mesure approprié comme le modèle à crédit partiel (Masters, 1982). Nous avons effectué ces transformations pour tous les regroupements d'items de type I et II. À la suite d'analyses préliminaires, deux des trois superitems assemblés (voir I.1 et II.2 dans le Tableau XXXVII) ont été convertis en item à échelle de notation dichotomique puisqu'il n'y avait pas suffisamment d'observations dans les catégories de scores intermédiaires.

L'effet de ces changements est important : la valeur de l'indice de fidélité de séparation chute à 0,63, ce qui est en deçà du seuil minimal pour former deux niveaux d'habiletés distincts. Comme il sera expliqué lors de la prochaine modélisation, le regroupement des items de type hiérarchique a pour effet d'éliminer de grandes sources de dépendance locale dans les données.

7.5.4 Quatrième modélisation : Modèle final

Cette quatrième et dernière modélisation réunit les procédés des deux modélisations précédentes. Après avoir effectué les divers regroupements des paires de concepts, la solution finale comporte 18 items. Un répondant a aussi été exclu du calibrage étant donné que ses données s'ajustaient très mal lors de cette dernière modélisation (*Outfit* de 5,6), et ce au détriment des estimations des paramètres de deux items.

L'ajustement des données pour les 18 items est bien meilleur que ce qui a été observé lors de la première modélisation. Comparativement à la première modélisation, la distribution des paramètres de difficulté des items couvre mieux l'étendue des estimations des niveaux d'habiletés des répondants (voir la Figure 21).

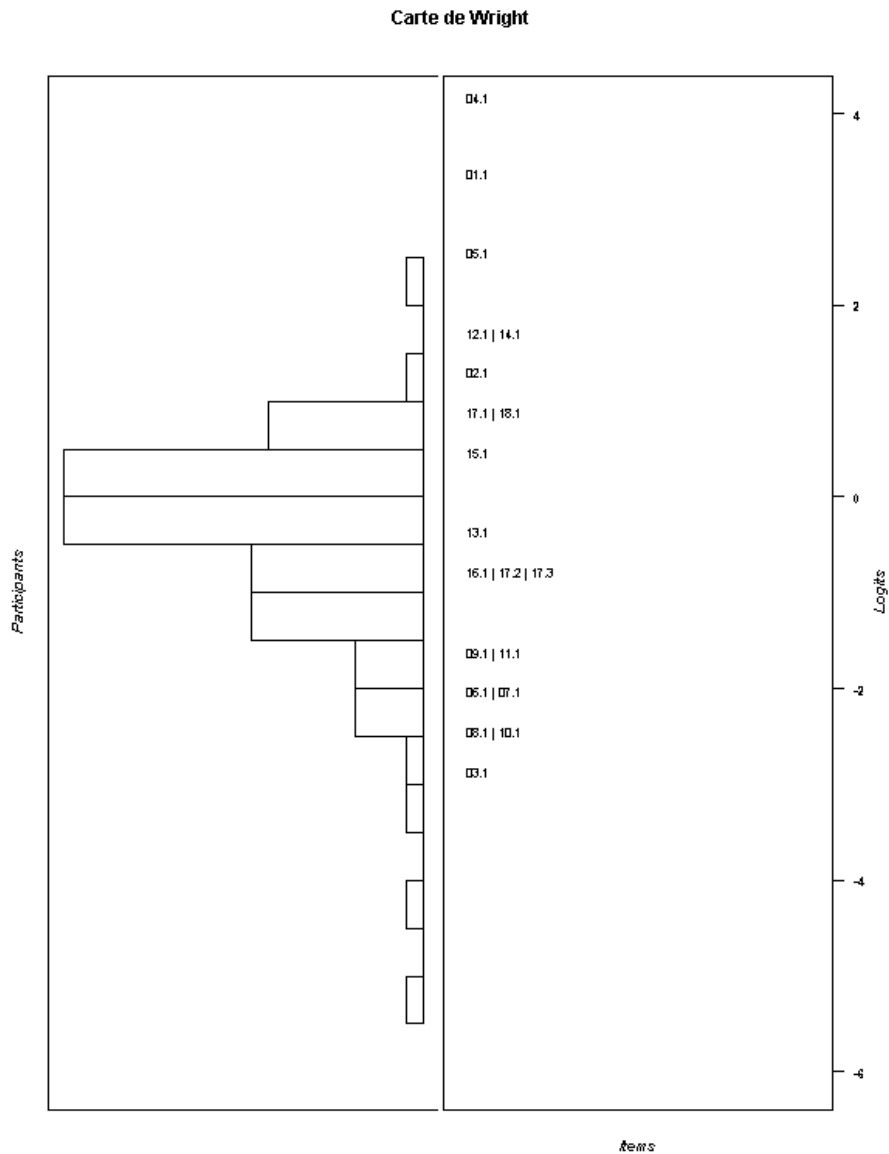


Figure 21. Distribution des estimations de niveaux d'habileté et des paramètres de difficulté des items pour la quatrième modélisation

Le Tableau XXXVII donne les diverses statistiques des nouveaux items. : la corrélation item-total corrigée est positive pour tous les items et les statistiques *Outfit* de tous les items, sauf deux, se situent entre les seuils critiques. Pour ces deux items (4 et 5), qui sont d'ailleurs très difficiles, les valeurs de l'indice *Outfit* sont inférieures à 1, ce qui signale un surajustement des

données au modèle. Puisqu'un surajustement des données ne dégrade pas pour autant l'échelle de scores (Wright, Linacre, Gustafson et Martin-Lof, 1994), ces items ont été conservés dans la modélisation.

Enfin, les analyses de corrélations entre les résidus standardisés et l'analyse en composantes principales des résidus ne signalent rien d'alarmant. Seulement deux valeurs de corrélations sont statistiquement différentes de zéro : la valeur de la corrélation est de -0,44 ($p < 0,001$) entre les résidus standardisés des items 5 et 7, et de -0,44 entre les items 17 et 18. Même si ces valeurs sont statistiquement significatives, elles sont beaucoup moins élevées que les valeurs de corrélations observées lors de la première modélisation (voir Tableau XXXV). Similairement, l'analyse des coefficients de saturation de l'APC ne révèle pas de regroupement notable parmi les items (voir Tableau XXXVIII). Pour chacune des quatre composantes extraites lors de l'analyse parallèle, il y a un seul item, facile ou difficile, qui possède un coefficient de saturation dépassant le seuil de 0,30. Enfin, la fidélité de séparation des personnes est estimée à 0,66.

Tableau XXXVII. Statistiques d'items de la quatrième modélisation

Item	Paire de concepts	Score moyen	Corrélation item – total corrigé	Paramètre de difficulté	<i>Infit</i>	<i>Outfit</i>
1	Chaîne alimentaire, Producteur	0,04	0,04	3,19	1,06	0,99
2	Chaîne alimentaire, Réseau trophiques	0,17	0,19	1,40	0,97	0,86
3	Champignon et bactérie, Décomposeur	0,87	0,38	-2,79	0,97	0,72
4	Consommateur, Décomposeur	0,01	0,07	4,34	1,01	0,65
5	Niveaux trophique, Relations trophique	0,06	0,23	2,63	0,89	0,55
6	Producteur, Végétal	0,76	0,15	-1,86	1,13	1,19
19*	Décomposeur, Producteur	-	-	-	-	-
20*	Décomposeur, Réseaux trophique	-	-	-	-	-
21*	Autotrophe, Chaînes alimentaire	-	-	-	-	-
22*	Champignon et bactérie, Consommateur	-	-	-	-	-
16 ^I	Consommateur, Consommateurs primaire Consommateur, Consommateurs secondaire Consommateur, Consommateurs tertiaire	0,54	0,21	-0,62	1,05	1,13
17 ^I	Niveau trophique, Consommateur Niveau trophique, Producteur Niveau trophique, Décomposeur	1,42	0,21	-0,3	1,07	0,98
18 ^{II}	Consommateur primaire, Consommateur secondaire Consommateur secondaire, Consommateur tertiaire	0,27	0,19	0,69	1,00	0,98
7 ^{III}	Animal, Consommateur Animal, Consommateur primaire Animal, Consommateur secondaire	0,81	0,39	-2,22	0,92	0,86

Item	Paire de concepts	Score moyen	Corrélation item – total corrigé	Paramètre de difficulté	<i>Infit</i>	<i>Outfit</i>
8 ^{III}	Animal, Consommateur tertiaire Autotrophe, Producteur Autotrophe, Végétal	0,83	0,19	-2,43	1,08	1,28
9 ^{III}	Cadavre et déchet d'organismes, Champignon et bactérie	0,69	0,30	-1,42	1,00	0,93
10 ^{III}	Cadavre et déchet d'organismes, Décomposeur Photosynthèse, Autotrophe Photosynthèse, Producteur Photosynthèse, Végétal	0,83	0,30	-2,43	1,00	0,91
11 ^{III}	Hétérotrophe, Animal Hétérotrophe, Consommateurs	0,75	0,42	-1,78	0,85	0,91
12 ^{III}	Hétérotrophe, Champignon et bactérie Hétérotrophe, Décomposeur	0,14	0,09	1,60	1,02	1,18
13 ^{III}	Niveaux trophiques, Chaîne alimentaire Niveaux trophiques, Réseau trophique	0,48	0,28	-0,33	0,97	1,04
14 ^{III}	Relations trophiques, Chaîne alimentaire Relations trophiques, Réseaux trophiques	0,12	0,12	1,83	1,05	0,90
15 ^{III}	Consommateur primaire, Producteur Consommateur primaire, Végétal	0,31	0,24	0,49	0,96	1,07
	Moyenne	0,51	0,22	0,00	1,00	0,95
	Écart-type	0,39	0,11	2,15	0,07	0,19
	Minimum	0,01	0,04	-2,79	0,85	0,55
	Maximum	1,42	0,42	4,34	1,13	1,28

Note. ^I Item composé de concepts partageant une relation de type hiérarchique. ^{II} Item composé d'une chaîne de relations. ^{III} Item composé de propositions redondantes. * Item dont le paramètre de difficulté ne peut pas être estimé.

Tableau XXXVIII. Coefficients de saturation de l'ACP des résidus standardisés de la quatrième modélisation

Item	Composante			
	1	2	3	4
18	0,97	0,01	-0,02	0,09
9	0,02	0,96	-0,11	0,04
16	-0,01	-0,1	0,96	-0,07
7	0,1	0,05	-0,06	0,97

Les résultats de cette dernière modélisation contrastent grandement avec ceux de la première modélisation avec le modèle pour données dichotomiques. En plus des différences observées lors des analyses précédentes, l'introduction des deux mécanismes de formation d'items altère les estimations des niveaux d'habileté des répondants. La valeur de la corrélation entre les niveaux d'habileté estimés du premier modèle et ceux du dernier modèle est de 0,86.

7.6 Discussion et Conclusion

L'élaboration d'un modèle de notation pour une tâche complexe, comme pour la cartographie conceptuelle, n'est pas une opération simple. La modélisation de Rasch nous a permis d'examiner de manière plus détaillée les propriétés internes de la notation de la qualité des propositions qui avaient, jusqu'à ce jour, été analysées uniquement à l'aide de modèles de mesure basés sur des scores observés.

Les résultats de la première modélisation de Rasch montrent que les paires de concepts critiques ne correspondent pas nécessairement à des items indépendants entre eux, comme le suggèrent Yin et Shavelson (2008). Ces chercheurs, comme bien d'autres, définissent les manifestations observables de leur modèle de notation à partir des composantes structurelles de la carte conceptuelle alors que notre étude montre qu'il faut également s'attarder à la signification des concepts et des propositions.

Nous avons d'abord mis en évidence que les propositions issues de relations de type hiérarchique (source I) ou de relations d'association (source II) ne se comportent pas comme

des unités de connaissances indépendantes, mais plutôt comme des regroupements de connaissances. En psychologie cognitive, il est bien connu que certaines informations ne sont pas encodées dans la mémoire sous la forme de propositions isolées, mais elles sont plutôt stockées sous forme de « paquets » de propositions. Ces « paquets » portent le nom de schémas (Rumelhart, 1980) et sont particulièrement utiles pour expliquer certains mécanismes d'apprentissage ou de récupération d'information de la mémoire à long terme que ne peuvent expliquer à eux seuls les modèles de représentation propositionnelle des connaissances (Anderson, 1980). Dans le cadre de notre recherche, nous avons traité certains paquets de propositions comme des superitems afin de tenir compte des problèmes de dépendance locale entre les propositions d'un même groupe.

À cet égard, nous adoptons une approche qui est diamétralement opposée à celle de Novak et Gowin en ce qui a trait à la notation des relations hiérarchiques. Leur modèle donne beaucoup d'importance à ce type de relations en accordant un point à chaque proposition valide, puis 10 points à chaque niveau hiérarchique valide. Ainsi, un ensemble de propositions comme celui composé des relations entre le concept « consommateur » et les concepts subordonnés « consommateur primaire », « consommateur secondaire » et « consommateur tertiaire » peut valoir jusqu'à 23 points ($= 3 \times 1 \text{ point} + 2 \times 10 \text{ points}$) selon le modèle de la notation structurelle. Dans notre approche, ces mêmes propositions sont notées soit sur une échelle à quatre catégories de score (0 à 3), soit sur une échelle dichotomique lorsqu'elles sont assemblées en superitem. Étant donné que les expressions de liaison entre ces concepts désignent déjà une relation hiérarchique, il n'est pas nécessaire de leur attribuer des points à deux reprises : à la fois pour les aspects sémantiques, à la fois pour les aspects structuraux. Cela expliquerait peut-être le fait que Liu (Liu, 1994; Liu et Hinchey, 1996) ait observé des valeurs de corrélation élevées entre les indicateurs qui ont trait à la justesse des propositions et ceux qui concernent le nombre de relations hiérarchiques valides. Des résultats similaires ont été trouvés plus récemment dans von der Heide (2015). Il y a de fortes chances d'ailleurs que Liu (1994) ait observé des problèmes de dépendance locale s'il avait vérifié les conditions d'application lors de sa modélisation des scores issus de cartes conceptuelles.

Notre étude a fait ressortir un autre problème concernant la notation des cartes conceptuelles et qui, curieusement, n'a jamais été soulevé auparavant : comment tenir compte

de propositions qui veulent dire sensiblement la même chose, mais qui sont composées de paires de concepts différentes ? Dans notre recherche, nous proposons de mettre en commun toutes les propositions équivalentes sous forme d'un item dichotomique. Nous pensons que cette démarche reflète bien les fondements de la cartographie conceptuelle, sans pénaliser ou avantager les participants. Une carte conceptuelle n'est ni une représentation suffisante ni une représentation complète des connaissances. Il existe vraisemblablement plus d'une solution possible (Tardif, 2006). Chaque personne fait des choix délibérés dans la manière d'exprimer ses idées et ne devrait pas être pénalisée si elle formule une proposition quelque peu différemment de celle qui est exprimée dans la carte de référence. Inversement, il y a peu d'intérêt à récompenser plusieurs fois une personne qui inscrit plusieurs propositions équivalentes dans sa carte conceptuelle. C'est pourquoi les propositions qui partagent le même sens sont regroupées en un superitem et que le score maximal de ce superitem est de 1.

Lorsqu'il est comparé à d'autres modèles de notation, le modèle proposé dans cette étude demeure relativement rigide. Certes, il admet que certaines paires de concepts partagent le même sens et peuvent être combinées afin d'éviter la redondance, mais fondamentalement, il oblige l'évaluateur de cibler une liste de paires de concepts au départ et de noter seulement les propositions formées à partir de ces paires de concepts. Pour quelques chercheurs, l'interprétation des cartes conceptuelles devrait être encore plus souple et devrait tenir compte de l'unicité de chaque représentation (Besterfield-Sacre et coll., 2004; Edmondson, 2005; Tardif, 2006). Il convient probablement, dans certains contextes, d'adopter des modalités plus souples qui s'accommodent à l'ensemble des représentations possibles, par exemple lors d'un exercice d'activation des connaissances antérieures en début de séquence d'enseignement et d'apprentissage, mais il peut aussi s'avérer profitable d'établir à l'avance les connaissances que devront manifester les individus lors de situations de vérification des acquis. Les modèles d'interprétation plus souples peuvent particulièrement être utiles lorsque le thème de la carte est très large et ouvert (Shallcross, 2016). Le modèle de notation doit logiquement être adapté au contexte et à l'objectif de l'évaluation.

Et comme le suggèrent les résultats de cette étude, il n'y a pas à coup sûr un seul modèle de notation pour des tâches de cartographies conceptuelles. Le modèle doit être adapté à la tâche et au domaine de connaissances ciblé par celle-ci. L'approche proposée dans cet article est certes

plus exigeante que des méthodes omnibus, comme la notation structurelle de Novak et Gowin (1984), parce qu'elle requiert d'analyser le contenu de la carte de référence a priori et de regrouper certaines propositions en items, mais nous pensons qu'elle permet d'obtenir une appréciation plus exacte et juste des propriétés psychométriques d'une tâche de cartographie conceptuelle.

À cet égard, les procédés de formation d'items élaborés dans cette recherche ne se limitent pas à la famille des modèles de Rasch. Ils sont directement applicables dans les modèles de notation basés sur les scores observés, dans la mesure où l'évaluateur veut tenir compte des différentes manières d'exprimer les mêmes connaissances et éliminer la redondance dans la notation. La modélisation des scores basée sur les modèles de la famille de Rasch a néanmoins été d'une très grande utilité dans cette recherche, car elle nous a permis de pousser plus loin les analyses de données collectées au moyen de tâches de cartographie conceptuelle et d'explorer certaines propriétés, comme l'unidimensionnalité et l'indépendance locale, qui n'avaient jamais été explorées auparavant. À leur tour, ces analyses nourrissent une réflexion au sujet de modalités de correction et notation qui dépasse, à notre avis, l'approche proposée dans cette étude.

Bien entendu, cette recherche est exploratoire et il conviendrait de reproduire la méthodologie dans d'autres contextes (thématiques de la carte, populations d'étudiants, etc.). Notre principale contribution est d'avoir mis en évidence que la notation doit être adaptée au contexte et au contenu de la tâche de cartographie conceptuelle et du domaine de connaissances.

Chapitre 8 Discussion et Conclusion

Malgré que la carte conceptuelle ait une histoire étalée sur quelques décennies en éducation, les savoirs scientifiques à son sujet sont encore épars et incomplets en ce qui concerne le domaine de l'évaluation des apprentissages (voir la recension des écrits). À travers trois articles, cette thèse s'est intéressée à l'utilisation de la carte conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages. Ce dernier chapitre fait un retour sur chacun des articles en faisant ressortir les principales conclusions de la recherche, ses limites, ses retombées et quelques pistes de recherches potentielles. Il propose également deux réflexions à l'égard de l'objet de recherche : une première réflexion quant au contexte d'utilisation de la carte conceptuelle en évaluation des apprentissages; une seconde réflexion quant au processus de validation. Ce chapitre se termine avec une conclusion générale.

8.1 Retour sur le premier article

Le premier volet de cette recherche a concerné les supports de production d'une carte conceptuelle. Traditionnellement, une carte conceptuelle est construite sur une feuille de grand format avec un crayon, mais, de nos jours, il est possible d'exploiter un logiciel dédié pour réaliser la même tâche. Cette deuxième approche aurait comme avantage de faciliter la modification et la réorganisation du contenu (Watson et coll., 2016), deux activités qui sont réputées laborieuses dans la méthode traditionnelle. Compte tenu des différences fondamentales entre le support papier et le support électronique, il est tout à fait raisonnable de supposer que ces deux méthodes ne sont pas interchangeables. Or, jusqu'à maintenant, les études au sujet de la cartographie conceptuelle n'ont pas vraiment accordé d'attention aux différences entre ces supports. Quelques résultats de recherche ont déjà montré que des élèves obtiennent des notes plus élevées lorsqu'ils réalisent une tâche de cartographie conceptuelle à l'ordinateur qu'à l'ordinateur (Brandstädter et coll., 2012 ; Erdogan, 2009), mais aucune étude ne s'est penchée sur les différences entre les démarches mobilisées par des individus pendant qu'ils réalisent la carte.

Dans cet esprit, la première expérimentation avait pour objectif de documenter les démarches lorsque des élèves produisent leur carte de manière traditionnelle sur support papier

ou à l'ordinateur avec un logiciel dédié. Des séances de verbalisation de la pensée ont été effectuées avec huit élèves de quatrième secondaire en science et technologie afin de suivre leur processus de réponses au cours de la réalisation d'une tâche de cartographie conceptuelle. Les analyses approfondies des données ont mis en évidence une diversité de stratégies pour planifier, construire et réviser leur carte conceptuelle et elles révèlent que certaines sont plus fréquentes pour un support de production que pour l'autre.

Dans la modalité traditionnelle, les participants ont passé plusieurs minutes à prévoir l'emplacement des concepts, en déplaçant les étiquettes amovibles sur lesquelles étaient inscrits les énoncés de concepts, avant de commencer à tracer les liens et à écrire les relations entre ceux-ci. Par ailleurs, ces participants ont manifesté moins souvent des processus de révision. À l'ordinateur, les participants ont commencé à construire beaucoup plus tôt les relations entre les concepts et ont adopté une démarche de construction progressive de la carte. Au contraire de la modalité traditionnelle, cette approche semble provoquer plus de stratégies de révision, car les participants doivent ajuster constamment leur carte afin d'accommoder l'ajout de nouveaux concepts et de nouvelles relations. En plus des différences observées entre les démarches des participants, cette étude a permis d'observer quelques difficultés inhérentes à chacune des modalités, comme le défi d'exploiter convenablement tout l'espace de la feuille à l'écrit ou certaines difficultés techniques à l'ordinateur. L'ensemble de ces observations appelle évidemment à une réflexion accrue au sujet de la formation préalable à la cartographie conceptuelle, et notamment à la formation au support de production.

8.2 Retour sur le deuxième article

La seconde expérimentation de ce projet de recherche a été mise au point pour répondre à un besoin particulier en vue des analyses du troisième article : la production d'une carte de référence pour corriger les cartes produites par des élèves. Comme il a été vu dans la recension des écrits, la carte de référence est un outil fort pratique et répandu, mais la méthode traditionnelle pour produire une telle carte, qui consiste à demander à un ou plusieurs experts de contenu de produire ladite carte, n'est pas toujours praticable et ne s'adapte pas à tous les contextes.

Cette deuxième expérimentation, du type recherche et développement, a exploré une autre approche pour produire une carte de référence fondée sur diverses techniques de traitement automatique du langage naturel. Une application informatique, nommée EPTAO, a été développée pour assister l'humain dans la production d'une carte conceptuelle à partir d'un corpus de textes et le logiciel a été soumis à deux mises à l'essai, l'une pour la thématique de la tâche d'entraînement (l'effet de serre) et l'autre pour la thématique de la tâche principale (les relations trophiques).

Comme il a été vu dans l'article, même si les niveaux de rendement de certaines composantes du logiciel ont été modestes, le niveau de performance global du logiciel a été plutôt encourageant. La démarche mise en œuvre a en effet permis de restituer une grande partie des connaissances repérées lors de l'extraction manuelle du contenu des cartes (jusqu'à 77 % des propositions pour la tâche sur les relations trophiques). Faute de balises dans les études similaires recensées, il n'est pourtant pas clair si ce niveau de rendement est suffisant pour l'utilisation ciblée au départ, c'est-à-dire la production d'une carte de référence en vue de corriger des cartes produites par des élèves. Considérant les performances observées, il est raisonnable de penser que les cartes obtenues à la fin de la démarche peuvent être utilisées en contexte d'évaluation formative, mais il semble encore être trop tôt pour s'en servir dans un contexte plus contraignant, comme en évaluation sommative.

Le logiciel EPTAO est pourtant appelé à évoluer : il existe encore plusieurs avenues à explorer pour l'améliorer, pour le rendre multifonctionnel, pour en faciliter l'exécution et pour en améliorer les performances. N'empêche, le deuxième article a présenté un outil et une démarche prometteurs et a fait la démonstration qu'il est possible de produire de manière semi-automatisée des cartes conceptuelles à partir de petits corpus de textes, et de surcroît, à partir de documents rédigés en français.

8.3 Retour sur le troisième article

Le dernier volet de la thèse a porté sur la notation des cartes conceptuelles. Comme il a été vu lors de la recension des écrits et dans le troisième article, la majorité des modèles de notation recensés font appel à des approches basées sur les scores observés et certaines propriétés psychométriques des tâches de cartographie conceptuelle sont encore méconnues. La

troisième expérimentation avait pour objectif d'apprécier les propriétés de telles tâches sous un autre angle, en explorant la modélisation des scores à l'aide deux modèles de la famille de modèles de Rasch : le modèle de Rasch pour données dichotomiques (Rasch, 1960) et le modèle à crédit partiel (Masters, 1982). Plus particulièrement, cette expérimentation s'est penchée sur l'examen de la structure interne des notes, lorsqu'un point est accordé à chaque proposition juste d'une carte conceptuelle.

Les résultats de la modélisation et des analyses a posteriori ont mis en évidence que les « items » d'une carte, c'est-à-dire les propositions dérivées des paires de concepts critiques, ne s'organisent pas toujours conformément aux modèles de notation courants recensés dans les écrits scientifiques. Les chercheurs traitent généralement chaque proposition comme une unité de connaissances indépendante et corrigent chaque unité de manière séparée. Les résultats de la modélisation suggèrent plutôt que certains concepts et relations d'une carte se regroupent naturellement pour former des parcelles de connaissances et qu'il est mieux de traiter ces parcelles comme un tout lors de la correction et la notation des cartes conceptuelles.

Les relations hiérarchiques en sont de bons exemples. Au point de vue de la modélisation de Rasch, il n'y a pas lieu de séparer en trois items distincts la relation entre « consommateurs » et « consommateurs primaires », la relation entre « consommateurs » et « consommateurs secondaires » et la relation entre « consommateurs » et « consommateurs tertiaires ». Ces relations sont mieux représentées par une seule et même unité cognitive. Il y a là d'ailleurs un parallèle intéressant à faire entre les résultats de ces modélisations et les recherches en psychologie cognitive : il est d'ores et déjà connu dans ce domaine que la proposition n'est pas le seul type d'unité cognitive (Anderson, 1980). Certaines unités, comme les schémas, sont composées d'un ensemble de propositions (voir le cadre conceptuel). Les résultats des modélisations suggèrent qu'il est peut-être mieux de noter les unités cognitives plutôt que les propositions des cartes conceptuelles.

En rétrospective, la troisième expérimentation a elle aussi atteint son objectif : en modélisant les scores issus de cartes conceptuelles et en explorant les propriétés de la structure interne des notes, elle a permis de mieux comprendre les avantages et les limites de modèles de notation traditionnels de cartes conceptuelles.

8.4 Les limites de la recherche

Même s'ils ont été présentés de manière séparée, les articles réunis dans cette thèse ont plusieurs points en commun, en commençant par le fait qu'ils sont liés à la même collecte de données. Ces articles se situent donc dans une seule et même démarche d'évaluation et contribuent, sous différents angles, au processus de validation de la carte conceptuelle comme instrument de collecte de données en vue de l'évaluation des apprentissages. Cette section et la prochaine font un retour sur les limites et les retombées de la recherche.

Cette recherche, comme tout projet scientifique, n'est pas exempte de limites. Trois limites découlant de la démarche générale de recherche sont discutées ci-dessous ; les deux premières sont d'ordre méthodologique, la troisième est plutôt d'ordre conceptuel. L'échantillonnage des participants est une première limite de cette recherche. Au départ, il avait été prévu de recruter environ quatre enseignants et douze groupes-classes (environ 400 élèves) pour la collecte de données principale. Or, pour des raisons logistiques, seuls un enseignant et 100 élèves ont accepté de participer à la recherche et 85 élèves ont accompli la tâche principale. Avec cet échantillon de convenance, il a été possible de réaliser les analyses et les modélisations présentées lors du troisième article, mais un échantillon plus grand aurait permis d'explorer d'autres modélisations, comme des modèles de mesure multidimensionnels, ou encore de pousser plus loin les analyses, en effectuant des analyses de fonctionnement différentiel des items⁶⁰, par exemple. Adopter la carte conceptuelle comme outil de collecte de données présente en effet quelques défis et contraintes d'ordre logistique, notamment pour ce qui est de la formation des participants (Ruiz-Primo, 2014). Il n'est donc pas trop surprenant de constater que la grande majorité des études antérieures (voir la recension) reposent aussi sur des échantillons de convenance et de petite taille ($N < 100$).

Une deuxième limite générale découle de l'ajout tardif de la deuxième étude dans le projet de recherche. Le logiciel conçu et mis à l'essai dans cette étude a obtenu de moins bons résultats pour la tâche sur l'effet de serre et des résultats plus encourageants pour la tâche sur

⁶⁰ En fait, certaines de ces modélisations ont été explorées, mais des analyses subséquentes par méthodes de ré-échantillonnage (en anglais, *bootstrap*) éveillaient un doute quant à la robustesse des estimations. C'est pourquoi les résultats n'ont pas été présentés dans cette recherche.

les relations trophiques, et comme il a été signalé à quelques occasions, une différence fondamentale entre ces deux tâches est que la première sert à représenter des connaissances de type dynamique et la seconde, des connaissances de type statique (Safayeni et coll., 2005). Ces résultats laissent croire que certains types de connaissances peuvent être extraits plus facilement que d'autres, et conséquemment que certains thèmes et contenus d'apprentissage s'accommodent mieux à la production semi-automatisée d'une carte conceptuelle. Le type de connaissances représenté est vraisemblablement un facteur qui devra être envisagé lors de nouvelles recherches.

La troisième et dernière limite discutée ici est plutôt d'ordre conceptuel. Le modèle de notation étudié lors de la troisième expérimentation a seulement tenu compte de l'appréciation du contenu de la carte conceptuelle et a fait fi de critères et indicateurs relatifs à l'appréciation de la structure du contenu. Rappelons-le, Cañas et coll. (2015) défendent l'idée qu'une bonne carte conceptuelle dépend de trois critères : la structure de la carte, la qualité du contenu et la qualité globale de la carte. Alors que les deux premiers critères peuvent être clairement établis, du moins être mis en relation avec des indicateurs observables, le troisième critère est naturellement flou, selon les dires de ces auteurs. Il s'agit en quelque sorte du « je ne sais quoi », c'est-à-dire de l'effet « wow » d'une carte conceptuelle. Cela explique probablement pourquoi ce critère n'apparaît pas de manière générale dans les écrits scientifiques. Les deux autres catégories de critères sont beaucoup plus fréquentes.

Il n'y a pourtant pas de consensus à savoir s'il faut obligatoirement inclure ces deux critères pour évaluer les apprentissages. Les résultats du troisième article semblent indiquer que certains indicateurs courants de la structure de la carte, comme ceux concernant les niveaux hiérarchiques, n'apportent pas nécessairement d'information supplémentaire par rapport à l'évaluation de la qualité des propositions. Autrement dit, il serait redondant de prendre en considération le nombre de niveaux hiérarchiques si le modèle de notation renferme déjà un critère relativement à la qualité des propositions. Cela ne veut pas dire pour autant qu'il n'existe pas d'autres indicateurs de la structure de la carte pouvant servir à évaluer les apprentissages. S'il avait été possible de collecter un plus grand nombre de cartes conceptuelles, il aurait été intéressant d'explorer des modèles de mesure pouvant intégrer ces indicateurs, par exemple des modèles de mesure multidimensionnels ou des modèles comportant des co-variables, pour

vérifier dans quelle mesure ils apportent une information complémentaire (ou non) par rapport à la notation du contenu de la carte.

8.5 Les retombées et les recommandations pour des recherches potentielles

Les contributions de cette thèse à l'avancement des connaissances sont multiples. Cette section dégage quatre retombées de cette recherche et, lorsqu'il est opportun, propose des pistes de recherche potentielles. D'abord, et avant même de passer aux résultats des expérimentations, il s'avère judicieux de faire un retour sur le troisième chapitre de cette thèse, qui a proposé une recension exhaustive des écrits au sujet des usages de la cartographie conceptuelle comme dispositif d'évaluation des apprentissages. À ce jour, aucune recension dédiée à la cartographie conceptuelle dans un contexte d'évaluation des apprentissages (Anohina et Grundspenkis, 2009 ; Hartmeyer et coll., 2017 ; Ruiz-Primo et Shavelson, 1996) n'a analysé autant de textes et aucune n'a étudié à fond chaque volet de la démarche d'évaluation des apprentissages, de la formation préalable à la notation des cartes conceptuelles. Non seulement la recension présentée au troisième chapitre visait à bien situer les études antérieures, mais elle se voulait un moyen de déterminer le travail qu'il reste à accomplir pour mieux comprendre les usages de la cartographie conceptuelle en évaluation des apprentissages.

Une deuxième retombée concerne la formation préalable à la cartographie conceptuelle. Comme il a été vu lors de la recension des écrits (au troisième chapitre), la formation est probablement le parent pauvre de la cartographie conceptuelle : les explications relatives aux modalités sont souvent survolées et parfois même omises dans les publications consultées, de même que les modalités sont assez variables d'une étude à l'autre. Et les chercheurs remettent en question très rarement les résultats de leur étude par rapport à la formation préalable (Hilbert et Renkl, 2008).

Les modalités de formation sont d'ailleurs souvent centrées autour du respect des principes et des consignes de la cartographie conceptuelle. Suivant les lignes directrices qui se retrouvent dans les écrits scientifiques (Ruiz-Primo, Schultz, et coll., 2001 ; Yin et coll., 2005), le protocole de formation mis en œuvre dans cette recherche a suivi une approche assez

conventionnelle : la formation a commencé avec un bref exposé au sujet des principes de construction d'une carte conceptuelle, puis elle a continué avec deux activités pratiques, une en équipe et l'autre individuelle. Et comme il est recommandé dans certaines études similaires, les cartes produites lors de l'activité individuelle ont ensuite été analysées pour s'assurer du respect des consignes. S'il n'avait été que de cette analyse, la conclusion aurait été que la formation été exécutée avec succès. En effet, toutes les cartes, sauf une, ont répondu aux critères de qualité établis.

Pourtant, les analyses des données collectées lors des séances de verbalisation de la pensée ont remis en question ces premières conclusions. Même si les huit participants ont de manière générale respecté les consignes de la tâche, ils ont tous rencontré des obstacles à une occasion ou à une autre lors la réalisation de la tâche et ils ont mobilisé très rarement certaines stratégies qui sont considérées comme utiles pour construire des cartes conceptuelles (Hilbert et Renkl, 2008 ; Ruiz-Primo, Shavelson, et coll., 2001). Le respect des consignes de la tâche tout comme la maîtrise de certaines stratégies de planification et de révision sont manifestation des conditions nécessaires pour tirer profit des avantages de la cartographie conceptuelle comme outil favorisant l'apprentissage (Hilbert et Renkl, 2008).

Il en va de même dans le contexte de l'évaluation des apprentissages. Quelles habiletés doivent être maîtrisées (ou suffisamment bien maîtrisées) afin de se servir de la carte conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages ? Dans quelle mesure faut-il faire un enseignement explicite de stratégies de planification, de progression et de révision d'une carte conceptuelle ? Quelles stratégies sont mieux adaptées aux supports de production et quelles stratégies sont transférables ? Lesquelles ne le sont pas ? Dans quelle mesure l'enseignement et le développement de ces habiletés sont-ils importants dans un contexte d'évaluation des apprentissages ? Dans ces questionnements, il y a bien entendu un terrain fertile de recherches potentielles et indispensables pour assurer la validité des interprétations des résultats d'évaluation.

Une troisième retombée de cette recherche a trait au potentiel d'outils comme EPTAO pour soutenir la pratique des enseignants. En effet, il y a plusieurs raisons de penser qu'un tel logiciel puisse remplir bien d'autres fonctions que celles ciblées au départ et ainsi rendre de précieux services à des enseignants et à des évaluateurs. La logique sous-jacente à EPTAO, qui

repose sur la notion des connaissances « partagées », pourrait effectivement s'étendre à d'autres sources de données et servir à d'autres fins. Lors de la mise à l'essai, le logiciel a analysé des textes provenant d'ouvrages pédagogiques, mais, dans une prochaine mise à l'essai, il pourrait aussi traiter des cartes produites par des apprenants (Mühling, 2016, 2017), des données issues d'examens de rendement scolaire (Aajli et Afdel, 2015 ; Al-Sarem, Bellafkih et Ramdani, 2011), des notes de cours (Atapattu, Falkner et Falkner, 2015) ou des textes produits directement par les apprenants (Clariana et Koul, 2004). Certaines de ces sources de données sont possiblement plus riches et abondantes que le corpus analysé lors de la mise à l'essai, composé de textes issus d'ouvrages scolaires.

D'un point de vue pédagogique, il y a aussi un intérêt à diversifier les utilisations d'un tel logiciel. Récemment, Mühling (2016, 2017) a développé et a étudié des algorithmes pour réunir le contenu de plusieurs cartes en vue d'établir la représentation des connaissances de groupes d'apprenants. Une telle représentation pourrait devenir une source d'information riche pour un enseignant, parce qu'elle lui permettrait d'établir un diagnostic des connaissances partagées par ses élèves. Les algorithmes comme ceux de Mühling constitueraient d'ailleurs une valeur ajoutée à EPTAO, car ils faciliteraient le traitement de plus grands ensembles de données.

En dernier lieu, en plus des retombées d'ordre méthodologique, les résultats de la modélisation des scores présentés lors du troisième article ouvrent la porte à une réflexion théorique au sujet de la carte conceptuelle comme outil de représentation propositionnelle des connaissances. Alors que la proposition est décrite comme l'unité de sens fondamentale d'une carte conceptuelle, les résultats de la modélisation semblent indiquer qu'une carte conceptuelle peut également renfermer des unités cognitives de niveau supérieur, composées de plusieurs propositions.

À certains égards, il est même permis de croire que, dans certains contextes, ces unités d'ordre supérieur sont plus significatives que les simples propositions. Un enseignant, par exemple, a probablement plus d'intérêt à vérifier qu'un élève reconnaisse collectivement que les trois niveaux trophiques principaux sont les producteurs, les consommateurs et les décomposeurs que de vérifier que son élève puisse formuler indépendamment les trois propositions suivantes : les « niveaux trophiques sont composés des producteurs », les « niveaux

trophiques sont composés des consommateurs », les niveaux trophiques sont composés des décomposeurs ».

Cette reconnaissance d'unités cognitives d'ordre supérieur rappelle qu'il n'y a pas nécessairement de correspondance parfaite entre les composantes graphiques de la carte conceptuelle et les connaissances qui y sont représentées. Dans cette perspective, les unités de connaissances peuvent au contraire prendre différentes formes graphiques : elles peuvent être composées de deux concepts et d'une relation, mais elles peuvent aussi comporter plusieurs concepts et plusieurs relations.

En ce qui concerne l'évaluation des apprentissages, il est attendu que cette approche centrée sur des unités cognitives plutôt que sur des propositions assure une plus grande cohérence entre le modèle de notation et l'objet d'évaluation. Pour l'évaluateur, cela veut pourtant dire qu'il n'y a pas un seul modèle de notation qui convient à toutes les tâches de cartographie conceptuelle. Le modèle de notation doit tenir compte des particularités du domaine de connaissances et pour y arriver, l'évaluateur doit préalablement procéder à une analyse du contenu de la tâche et déterminer les unités de connaissances propres à celle-ci. Cette manière de faire est encore peu répandue et peu documentée dans les écrits scientifiques (des démarches apparentées ont été trouvées dans Marchand, D'Ivernois, Assal, Slama et Hivon, 2002 ; Nottis, 1996) ; il reste alors encore beaucoup de travail à accomplir pour mieux comprendre les mécanismes sous-jacents et les retombées et les limites de cette approche.

8.6 Réflexion quant au contexte d'évaluation des apprentissages

Tel que souligné au tout long de cette thèse, les modalités de passation, de correction et de notation de tâches de cartographie conceptuelle sont nombreuses et variées. Cela dit, ces modalités ne s'accordent pas avec tous les contextes et avec tous les objectifs d'évaluation. Cette section fait un retour sur les modalités étudiées dans cette recherche et s'interroge sur les contextes dans lesquels elles peuvent être utilisées.

La carte conceptuelle est avant tout un outil utile pour l'enseignant en salle de classe. Un avantage de celle-ci est sa polyvalence : elle peut servir à la fois d'outil d'apprentissage, d'outil d'évaluation des apprentissages et d'outil de planification de l'enseignement. À certaines

occasions, les cartes produites par des étudiants peuvent même servir à plusieurs fins. Par exemple, une carte relative aux connaissances antérieures construite en début de séquence d'apprentissage et d'enseignement peut être jointe à un portfolio afin de documenter la progression des apprentissages. La tâche de cartographie conceptuelle a ainsi une double fonction, soit d'activer les connaissances antérieures et de collecter des données en vue de l'évaluation des apprentissages. Pour le lecteur intéressé, Tardif (2006) propose plusieurs autres usages et modalités de ce dispositif pour l'évaluation des compétences.

Même si la carte conceptuelle sert principalement l'évaluation en salle de classe, rien n'exclut de l'utiliser lors de la passation d'une épreuve à grande échelle. C'est d'ailleurs ce qu'a tenté le programme américain d'évaluation du *National Assessment of Educational Progress* au début des années 2000, lors d'une mise à l'essai des tâches informatisées de cartographie conceptuelle auprès d'élèves du primaire et du secondaire (Fu, Raizen et Shavelson, 2009 ; Ruiz-Primo, Iverson et Yin, 2009). Il va sans dire que les conditions de passation des tâches ainsi que les modalités de correction et de notation ne sont pas les mêmes dans un contexte comme dans l'autre.

Par la force des choses, plusieurs contraintes relatives à la formation préalable, comme le temps alloué à la formation, les ressources disponibles, la possibilité de réaliser des activités préparatoires et la possibilité d'obtenir des rétroactions, se manifestent différemment selon le contexte dans lequel se déroule l'évaluation des apprentissages. Ces contraintes soulèvent d'ailleurs des questions fondamentales quant à la validité de l'interprétation des résultats, surtout lorsque les enjeux de l'évaluation sont élevés. En contexte d'évaluation à grande échelle, par exemple, il est important de se demander si la démarche d'évaluation est juste et équitable entre les individus qui ont eu la chance de produire des cartes conceptuelles en salle de classe avant la passation de l'épreuve et ceux qui subissent une telle tâche pour la première fois. De telles considérations n'appartiennent d'ailleurs pas seulement aux tâches de cartographie conceptuelle, mais elles concernent tous les formats de tâches ou d'items moins couramment utilisés (Parshall et coll., 2010; Scalise et Gifford, 2006; Sireci et Zenisky, 2015; Thomas, 2016). Les observations faites dans le premier article jettent néanmoins des doutes quant aux modalités de formation plus courtes, comme celles qui seraient vraisemblablement mises en œuvre lors de la passation d'une épreuve à grande échelle.

Les modalités de passation soulèvent aussi des enjeux différents selon le contexte d'évaluation. En salle de classe, un enjeu important est de s'assurer que les modalités d'évaluation répondent aux besoins pédagogiques. Ainsi, dans plusieurs situations, l'enseignant peut bien s'accommoder de différentes variantes de tâches et peut laisser les élèves choisir leur support de production. L'évaluation à grande échelle soulève quant à elle des enjeux d'uniformité des conditions de passation et dans la situation où la même épreuve peut être effectuée selon divers supports de passation (par ex., sur support papier et sur support informatisé), elle soulève des enjeux de comparabilité (Raikes et Harding, 2003). Même si la première étude de cette thèse ne s'inscrit pas directement dans une logique de comparaison de supports de passation, elle témoigne de différences entre les processus de pensée mobilisés par des élèves du secondaire lorsqu'ils construisent une carte conceptuelle sur papier ou à l'ordinateur. Et ces différences peuvent à leur tour traduire une menace potentielle à la validité des interprétations des résultats, dans la mesure où elles rendent compte de facteurs qui peuvent perturber l'opération d'évaluation des apprentissages. Plus de travaux de recherche seront nécessaires pour comprendre les relations entre ces processus de pensée, le construit et les notes décernées aux cartes produites par les répondants.

La correction et la notation imposent également des contraintes logistiques particulières lors de l'évaluation à grande échelle. Dans ce contexte, il y a un avantage certain à adopter des modalités qui permettent d'automatiser la correction. Or, à l'heure actuelle, seulement les variantes de tâches du type « compléter les nœuds » (Schau et coll., 2001) et du type « produire une carte à partir d'une liste de concepts et d'une liste d'expressions de liaison » (Buschang, Chung, Delacruz et Baker, 2012 ; Chiu et Lin, 2012 ; Dennis, 1999 ; Dorsey et coll., 1999) conviennent à la correction automatisée. Les variantes de tâches les plus populaires, comme celles qui nécessitent de construire une carte à partir de zéro ou à partir d'une liste de concepts, doivent encore être corrigées par des correcteurs humains. Conlon (2006) est vraisemblablement le seul à avoir proposé un programme informatique pour analyser automatiquement les énoncés de concepts et les expressions de liaisons qui sont rédigés par un participant lors de la production d'une carte. Cependant, Conlon l'admet lui-même, son logiciel est faillible et ne peut pas être utilisé pour évaluer les apprentissages lorsque les enjeux sont élevés. Les applications de techniques du traitement automatique des langues (McNamara, Allen, Crossley, Dascalu et

Perret, 2017) et de méthodes d'apprentissage automatique (Colchester, Hagra, Alghazzawi et Aldabbagh, 2017) gagnent en popularité en éducation et certaines de ces techniques et méthodes sont susceptibles d'apporter de nouvelles solutions à des problèmes concrets comme la correction et la notation de cartes conceptuelles. Après tout, le logiciel E-rater® est utilisé depuis la fin des années 1990 pour corriger des productions écrites du *Graduate Management Admissions Test*® (Burstein, 2003). Les avancées technologiques ouvrent plusieurs portes dans le domaine de l'évaluation des apprentissages, notamment en ce qui a trait à l'évaluation de tâches complexes (Williamson, Bejar et Mislevy, 2006).

Les usages et modalités de la cartographie conceptuelle peuvent également être catégorisés selon qu'ils cadrent dans une visée d'évaluation formative ou d'évaluation sommative. Dans le premier cas, le dispositif sert essentiellement à collecter des données en vue d'améliorer l'apprentissage, alors que dans le deuxième cas, il est utilisé pour reconnaître les acquis. La nature des données requises et la manière des interpréter ne sont pas forcément les mêmes selon les visées de l'évaluation. Par exemple, en contexte d'évaluation formative, l'enseignant peut décider de porter une attention particulière aux éléments absents d'une carte ou aux conceptions erronées pour détecter les difficultés potentielles de ses élèves. La prise en compte des erreurs des apprenants et des obstacles cognitifs s'avère en effet une démarche décisive pour favoriser l'apprentissage, notamment en didactique des sciences (Astolfi et coll., 2008). La signification des éléments absents et des conceptions erronées est différente en contexte d'évaluation sommative. Ces éléments sont d'ordinaire considérés comme de mauvaises réponses.

Les usages et modalités de tâches de cartographie conceptuelle peuvent également être analysés conformément au type de traitement des données. Comme l'a remarqué Tardif (2006), le traitement des données issues de cartes conceptuelles est passablement quantitatif dans les écrits scientifiques. La majorité des modèles d'interprétation ont pour objectif de résumer la performance d'un individu à partir d'une valeur numérique. Ces modèles ne négligent pas pour autant la dimension qualitative du contenu des cartes conceptuelles : plusieurs d'entre eux exigent qu'un correcteur porte un jugement sur la justesse ou la pertinence du contenu.

Les approches essentiellement qualitatives demeurent quant à elles plutôt rares. Quelques exemples de grilles à échelles descriptives ont été recensés dans divers écrits à

caractère scientifique ou pédagogique (par ex., Besterfield-Sacre et coll., 2004 ; Morin et Paquette, 2014), mais il est difficile d'estimer à quel point de tels outils sont utilisés en pratique. La vaste majorité des écrits scientifiques (voir le chapitre 3) font état de modalités de correction et de notation plutôt quantitatives.

Un quatrième et dernier axe d'analyse des usages et modalités de tâches de cartographie conceptuelle concerne l'objet de l'évaluation. Au troisième chapitre de cette thèse, il a été vu que la carte conceptuelle peut rendre compte de l'organisation d'une variété de types de connaissances. En s'intéressant à ce qu'ils nomment des cartes cycliques, Safayeni et coll. (2005) ont notamment attiré l'attention en particulier sur la différence entre les relations de type statique, qui se prêtent à la représentation de l'état des choses ou des relations entre des concepts (comme les relations trophiques entre les organismes vivants), et les relations de type dynamique, qui sont utiles pour expliquer des phénomènes (comme le fonctionnement de l'effet de serre). Or, ces deux types de représentation des connaissances renvoient probablement à des objets d'évaluation différents. Il se peut également que le contenu de la carte conceptuelle ne soit pas l'objet principal de l'évaluation. Cela peut se produire, par exemple, en situation d'évaluation de performances complexes ou d'évaluation des compétences (Tardif, 2006; Vitale et Romance, 2005). Les compétences ciblées par la tâche complexe constituent alors l'objet de l'évaluation; dans ce contexte, la carte conceptuelle est plutôt un dispositif de collecte de données servant à garder des traces (dans un dossier de progression ou dans un portfolio) des ressources mobilisées ou mobilisables à l'égard de la tâche.

En définitive, il faut bien se rendre compte que les modalités de passation, de correction et de notation préconisées dans le cadre de cette recherche ne constituent qu'une partie des modalités d'évaluation possibles. Les tâches de cartographie conceptuelle élaborées pour la collecte de données ont servi essentiellement à reconnaître les acquis des élèves et les modalités de correction et de notation adhérent aux approches passablement quantitatives de traitement des données. La modélisation des scores, telle qu'opérée au troisième article, est d'ailleurs une méthode qui est beaucoup plus courante dans le contexte de l'évaluation à grande échelle.

Cela étant dit, ce contexte est quelque peu artificiel, car il reflète davantage les besoins et contraintes du projet de recherche, et non pas les utilisations potentielles de la carte conceptuelle en évaluation des apprentissages. D'ailleurs, plusieurs résultats des trois études

présentées dans cette thèse dépassent le contexte de cette recherche. D'abord, la réflexion entamée lors du premier article quant à la formation préalable à la cartographie conceptuelle touche naturellement tous les contextes d'évaluation. Ensuite, les apports de techniques et méthodes modernes pour synthétiser les informations recueillies dans un corpus de textes, comme celles proposées dans le deuxième article, sont susceptibles d'ouvrir de nouvelles perspectives en éducation. Enfin, la réflexion entreprise quant à ce qui devrait former une unité de connaissances dans une carte conceptuelle et les mécanismes de formation d'items de test qui en découlent se transportent dans tous les contextes d'évaluation des apprentissages. À plusieurs égards, le champ d'action des résultats de cette recherche va bien au-delà de son contexte de réalisation.

8.7 Réflexion au sujet du processus de validation

Cette thèse s'inscrit naturellement dans une démarche de validation d'un instrument de collecte de données en vue de l'évaluation des apprentissages, à la fois dans le sujet traité et dans la démarche suivie. Dans le domaine de l'évaluation des apprentissages, il n'est pas rare que la carte conceptuelle soit définie à titre de format d'item, un peu comme l'est la question à choix multiple. Par la force des choses, la démarche de validation d'un dispositif de collecte de données ou d'un format d'item ne suit pas les mêmes impératifs que la démarche de validation d'une épreuve. Dans le seizième chapitre du *Handbook of Test Development*, Sireci et Zenisky (2015) soutiennent que les considérations de validité qui découlent des formats d'items innovants sont principalement de deux ordres. Les premières concernent la représentativité du construit. À cet égard, l'utilisation de la carte conceptuelle présuppose qu'il est souhaitable d'évaluer l'organisation des connaissances et que les données collectées permettent en effet de soutenir les inférences par rapport aux connaissances, habiletés ou compétences ciblées par la tâche ou l'épreuve. Les secondes considérations ont trait à la contamination des données. Cette recherche s'est principalement penchée sur les considérations de cet ordre.

En faisant ressortir les différences entre les démarches de production d'une carte conceptuelle sur support papier et sur support informatisé, de même qu'en détectant certaines difficultés rencontrées par des individus lorsqu'ils emploient l'un ou l'autre de ces supports, la première étude a mis en évidence des sources potentielles de contamination des données lors de

la passation de tâches de cartographie conceptuelle. Elle rappelle du même coup l'importance de la formation préalable afin de minimiser la contamination des données lors de la démarche d'évaluation des apprentissages.

Dans le même ordre d'idées, la recherche-développement présentée lors du deuxième article se veut un moyen de diminuer la dépendance des processus de création d'une carte de référence par rapport aux représentations idiosyncrasiques des connaissances d'experts de contenu. L'exploration de solutions de rechange telle que la production de cartes conceptuelles à partir de textes semble donc tout appropriée pour tenter de réduire les risques de contamination des données lors de l'élaboration d'une carte de référence.

Enfin, la modélisation des scores et les méthodes d'analyse des données employées lors du troisième article se veulent des moyens de vérifier à quel degré un modèle de notation des cartes conceptuelles permet de se rapprocher de mesures ayant des qualités psychométriques recherchées, et ainsi d'étudier certaines sources de contamination des erreurs. L'examen de la condition d'indépendance locale et de la condition d'unidimensionnalité a par ailleurs soulevé des questions légitimes quant à la relation entre les unités d'analyse d'une carte conceptuelle et les interprétations qui peuvent être tirées lors de la correction et la notation de celle-ci.

8.8 Conclusion générale

Cette thèse avait pour objectif d'explorer et de mieux comprendre certaines modalités de production, de correction et de notation de tâches de cartographie conceptuelle. Au moyen de trois expérimentations, il a été possible de mieux comprendre les différences entre les démarches mises en œuvre lors de la construction d'une carte conceptuelle de manière traditionnelle sur support papier ou de manière informatisée à l'aide d'un logiciel dédié, d'explorer de nouvelles techniques pour produire une carte de référence et d'étudier la structure interne des notes en modélisant les scores issus de cartes conceptuelles. Chacun de ces trois articles contribue ainsi au processus de validation de la carte conceptuelle comme instrument de collecte de données en vue de l'évaluation des apprentissages, que ce soit en documentant les particularités des démarches des participants qui peuvent signaler des menaces à la validité de l'interprétation de scores, en explorant des mécanismes de production de carte de référence en vue de neutraliser les erreurs de mesure attribuables aux comportements idiosyncrasiques des correcteurs, ou en

explorant des modèles de notation qui permettent de tirer des inférences valides à partir des scores obtenus.

Cette thèse a ainsi contribué à faire avancer et à actualiser les connaissances au sujet de la carte conceptuelle, un dispositif de collecte de données qui s'harmonise bien avec les conceptions contemporaines de l'évaluation et de l'apprentissage. Malgré les avancés faites dans cette thèse, il reste encore beaucoup de travail à accomplir pour mieux comprendre les usages de la carte conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages.

Bibliographie

- Aajli, A. et Afdel, K. (2015). *A new hybrid approach for constructing the concept map based on fuzzy prerequisite relationships*. Communication présentée à Colloquium in Information Science and Technology, CIST. doi: 10.1109/CIST.2014.7016604 Repéré à <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84938061801&doi=10.1109%2fCIST.2014.7016604&partnerID=40&md5=ad42c19a6ba64252a2362191c95370ff>
- Abd-El-Khalick, F. et BouJaoude, S. (1997). An exploratory study of the knowledge base for science teaching. *Journal of Research in Science Teaching*, 34(7), 673-699. doi: 10.1002/(SICI)1098-2736(199709)34:7<673::AID-TEA2>3.0.CO;2-J
- Acton, W. H., Johnson, P. J. et Goldsmith, T. E. (1994). Structural knowledge assessment: comparison of referent structures. *Journal of Educational Psychology*, 86(2), 303-311. doi: <http://dx.doi.org/10.1037/0022-0663.86.2.303>
- Al-Sarem, M., Bellafkih, M. et Ramdani, M. (2011). Leveraging historical assessment records for constructing concept maps. *International Review on Computers and Software*, 6(6), 994-1000.
- Alberts, B. (2009). Redefining science education. *Science*, 323(5913), 437. doi: 10.1126/science.1170933
- Alhashem, F. et Al-Jafar, A. (2015). Assessing teacher's integration of technology and literacy in elementary science classrooms in Kuwait. *Asian Social Science*, 11(18), 71-81. doi: 10.5539/ass.v11n18p71
- American Educational Research Association, American Psychology Association et National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Anderson, J. R. (1980). *Concepts, propositions, and schemata: What are the cognitive units?* Communication présentée à Nebraska Symposium on Motivation. Repéré à <http://www.dtic.mil/dtic/tr/fulltext/u2/a085997.pdf>
- Anderson, T. H. et Huang, S. C. C. (1989). *On using concept maps to assess the comprehension effects of reading*. (Rapport n° 483). Champaign, Illinois: Center for the Study of Reading. Repéré à https://www.ideals.illinois.edu/bitstream/handle/2142/17794/ctrstreadtechrepv01989i00483_opt.pdf?sequence=1
- Anohina-Naumeca, A. et Grundspenkis, J. (2010). Evaluating students' concept maps in the concept map based intelligent knowledge Dans J. Grundspenkis, M. Kirikova, Y. Manolopoulos et L. Novickis (dir.), *Assessment System : Advances in Databases and Information Systems* (p. 8-15). Berlin, Heidelberg: Springer
- Anohina, A., Graudina, V. et Grundspenkis, J. (2008). Using concept maps in adaptive knowledge assessment. Dans G. Magyar, G. Knapp, W. Wojtkowski, W. G. Wojtkowski et J. Zupančič (dir.), *Advances in Information Systems Development* (p. 469-479). New York, NY: Springer.
- Anohina, A. et Grundspenkis, J. (2007). *A concept map based intelligent system for adaptive knowledge assessment*. Communication présentée à Proceedings of the 2007 conference on Databases and Information Systems IV: Selected Papers from the Seventh

- International Baltic Conference. Repéré à
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.456.2656&rep=rep1&type=pdf#page=274>
- Anohina, A. et Grundspenkis, J. (2009). *Scoring concept maps: an overview*. Communication présentée à Proceedings of the International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing, Ruse, Bulgaria. doi: 10.1145/1731740.1731824 Repéré à
http://www.cs.swan.ac.uk/~csmarkd/Literature%20Review%20and%20Papers/ft_gateway-2.cfm.pdf
- Antidote 9. (dir.) (2017). *Dictionnaire des définitions*. Montréal, Canada: Druide informatique.
- Astolfi, J.-P., Darot, É., Ginsburger-Vogel, Y. et Toussaint, J. (2008). *Mots-clés de la didactique des sciences. Repère, définitions, bibliographies*. Louvain-la-Neuve: De Boeck Supérieur.
- Atapattu, T., Falkner, K. et Falkner, N. (2015). *An evaluation methodology for concept maps mined from lecture notes: An educational perspective*. Communication présentée à International Conference on Computer Supported Education, Cham. doi: https://doi.org/10.1007/978-3-319-25768-6_5
- Austin, L. B. et Shore, B. M. (1995). Using concept mapping for assessment in physics. *Physics Education*, 30(1), 41-45. doi: <https://doi.org/10.1088/0031-9120/30/1/009>
- Ausubel, D. P. (1968). *Educational psychology; a cognitive view*. New York: Holt Rinehart and Winston Toronto.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge: a cognitive view*. : Kluwer Academic Publishers.
- Ausubel, D. P., Novak, J. D. et Hanesian, H. (1978). *Educational psychology : a cognitive view*. New York: Holt, Rinehart and Winston.
- Baghaei, P. (2008). Local dependency and Rasch measures. *Rasch Measurement Transactions*, 21(3), 1105-1106.
- Baharom, S. (2012). Concept mapping and the teaching of science. Dans F. Ornek et I. M. Saleh (dir.), *Contemporary science teaching approaches: Promoting conceptual understanding in science*. (p. 115-135). Charlotte, NC, US: IAP Information Age Publishing.
- Bahr, G. S. et Dansereau, D. F. (2001). Bilingual knowledge maps (BiK-Maps) in second-language vocabulary learning. *The Journal of Experimental Education*, 70(1), 5-24. doi: <https://doi.org/10.1080/00220970109599496>
- Bahr, G. S. et Dansereau, D. F. (2005). Bilingual knowledge maps (BiK Maps) as a presentation format: delayed recall and training effects. *The Journal of Experimental Education*, 73(2), 101-118. doi: <https://doi.org/10.3200/JEXE.73.2.101-118>
- Baird, J.-A., Andrich, D., Hopfenbeck, T. N. et Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice*, 24(3), 317-350. doi: 10.1080/0969594X.2017.1319337
- Bakeman, R. et Quera, V. (2011). *Sequential analysis and observational methods for the behavioral sciences*. : Cambridge University Press.
- Barrett, P. T. et Kline, P. (1981). The observation to variable ratio in factor analysis. *Personality Study & Group Behaviour*, 1(1), 23-33.

- Bartels, B. H. (1995). Promoting mathematics connections with concept mapping. *Mathematics Teaching in the Middle School*, 1(7), 542-549. Repéré à <https://search.proquest.com/docview/62657329?accountid=12543>
http://opurl.bib.umontreal.ca:9003/sfx_local?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Aeric&atitle=Promoting+Mathematics+Connections+with+Concept+Mapping.&title=Mathematics+Teaching+in+the+Middle+School&issn=10720839&date=1995-01-01&volume=1&issue=7&spage=542&au=Bartels%2C+Bobbye+Hoffman&isbn=&jtitle=Mathematics+Teaching+in+the+Middle+School&btile=&rft_id=info:eric/EJ516909&rft_id=info:doi/
- Bartels, B. J. (1996). *Examining and promoting mathematical connections with concept mapping*. (Thèse de doctorat inédite, University of Illinois at Urbana-Champaign, US). Repéré à <http://hdl.handle.net/2142/19078>
- Bartram, D. (2006). Testing on the Internet: issues, challenges and opportunities in the field of occupational assessment. Dans D. Bartram et R. K. Hambleton (dir.), *Computer-based testing and the Internet* (p. 13-38). Hoboken, NJ: John Wileys & Sons.
- Baxter, G. P., Elder, A. D. et Glaser, R. (1996). Knowledge-based cognition and performance assessment in the science classroom. *Educational Psychologist*, 31(2), 133-140. doi: 10.1207/s15326985ep3102_5
- Bell, K. A. (2018). *Concept mapping in the middle school mathematics classroom*. (Thèse de doctorat, New Mexico State University, US). Accessible par Proquest. (10633402)
- Bennett, R. E. (2014). Preparing for the future: what educational assessment must do. *Teachers College Record*, 116(11). Repéré à <https://www.learntechlib.org/p/155766/>
- Berlanga, A. J., van Rosmalen, P., Boshuizen, H. P. A. et Sloep, P. B. (2012). Exploring formative feedback on textual assignments with the help of automatically created visual representations. *Journal of Computer Assisted Learning*, 28(2), 146-160. doi: 10.1111/j.1365-2729.2011.00425.x
- Bernstein, A., Penner, L., Clarke-Stewart, A. et Roy, E. (2003). *Psychology*. (6^e éd.). Houghton Mifflin: Boston.
- Bertrand, R. et Blais, J.-G. (2004). *Modèles de mesure: l'apport de la théorie des réponses aux items*. Sainte-Foy: Presses de l'Université du Québec.
- Besterfield-Sacre, M., Gerchak, J., Lyons, M., Shuman, L. J. et Wolfe, H. (2004). Scoring concept maps: An integrated rubric for assessing engineering education. *Journal of Engineering Education*, 93(2), 105-115. doi: 10.1002/j.2168-9830.2004.tb00795.x
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347-364. doi: <https://doi.org/10.1007/BF00138871>
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical theories of mental test scores*.
- Blais, J.-G. (2011). L'évaluation des apprentissages intégrée à l'enseignement avec les technologies de l'information et de la communication : le défi du passage à l'acte. Dans J.-G. Blais et J.-L. Gilles (dir.), *Évaluation des apprentissages et technologies de l'information et de la communication : le futur est à notre porte* (p. 1-10). Québec: Les Presses de l'Université Laval.
- Blais, J.-G. (dir.). (2009). *Évaluation des apprentissages et technologies de l'information et de la communication : enjeux, applications et modèles de mesure*. Québec: Les Presses de l'Université Laval.

- Bond, T. et Fox, C. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: Routledge.
- Borrego, M., Newswander, C. B., McNair, L. D., McGinnis, S. et Parette, M. C. (2009). Using concept maps to assess interdisciplinary integration of green engineering knowledge. *Advances in Engineering Education*, 1(3), 26. Repéré à <https://files.eric.ed.gov/fulltext/EJ1076049.pdf>
- Borsboom, D. (2005). *Measuring the mind: conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.
- Brandstädter, K., Harms, U. et Großschedl, J. (2012). Assessing system thinking through different concept-mapping practices. *International Journal of Science Education*, 34(14), 2147-2170. doi: <https://doi.org/10.1080/09500693.2012.716549>
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer.
- Bridwell-Bowles, L., Johnson, P. et Brehe, S. (1987). Composing and computers: Case studies of experienced writers. Dans A. Matsushashi (dir.), *Writing research: Multidisciplinary inquiries into the nature of writing. Writing in real time: Modelling production processes*. Westport, CT, US: Ablex Publishing.
- Burrows, N. L. et Mooring, S. R. (2015). Using concept mapping to uncover students' knowledge structures of chemical bonding concepts. *Chemistry Education Research and Practice*, 16(1), 53-66.
- Burstein, J. (2003). The E-rater® scoring engine: Automated essay scoring with natural language processing. Dans M. D. Shermis et J. C. Burstein (dir.), *Automated essay scoring: A cross-disciplinary perspective*. New York: Routledge.
- Buschang, R. E., Chung, G. K. W. K., Delacruz, G. C. et Baker, E. L. (2012). Validating measures of algebra teacher subject matter knowledge and pedagogical content knowledge. *Educational Assessment*, 17(1), 1-21. doi: 10.1080/10627197.2012.697847
- Buzan, T. et Buzan, B. (1995). *Dessine-moi l'intelligence*. Paris: Éditions d'Organisation.
- Cabello, F. et O Hora, D. (2002). Addressing the limitations of protocol analysis in the study of complex human behavior. *International Journal of Psychology & Psychological Therapy*, 2(2), 115-130.
- Cañas, A. J., Coffey, J. W., Carnot, M. J., Feltovich, P. J., Hoffman, R. R., Feltovich, J. et Novak, J. D. (2003). A summary of literature pertaining to the use of concept mapping techniques and technologies for education performance support. *The Institute for Human and Machine Cognition*. Repéré à http://cmap.bradercomm.net:8001/rid=1186721000045_1491728413_5664/IHMC%20Literature%20Review%20on%20Concept%20Mapping.pdf
- Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., . . . Carvajal, R. (2004). *CmapTools: A knowledge modeling and sharing environment*. Communication présentée à Concept maps: Theory, methodology, technology. Proceedings of the first international conference on concept mapping. Repéré à <http://eprint.ihmc.us/id/eprint/89>
- Cañas, A. J. et Novak, J. D. (2008). Concept mapping using Cmap Tools to enhance meaningful learning. Dans A. Okada, S. B. Shum et T. Sherborne (dir.), *Knowledge Cartography* (p. 25-46). doi: 10.1007/978-1-84800-149-7_2. Repéré à <http://www.springerlink.com/content/978-1-84800-148-0/fulltext/>
- Cañas, A. J., Novak, J. D. et Reiska, P. (2015). How good is my concept map? Am I a good Cmapper? *Knowledge Management and E-Learning*, 7(1), 6-19. Repéré à

<https://www.scopus.com/inward/record.uri?eid=2-s2.0-84927137420&partnerID=40&md5=3733be1d67f854d0b2063c759b8c9af9>

- Cardinet, J., Johnson, S. et Pini, G. (2010). *Applying generalizability theory using EduG*. New York: Routledge.
- Cathcart, L. A. (2016). *The Salient Map Analysis for Research and Teaching (SMART) method: Powerful potential as a formative assessment in the biomedical sciences*. (ProQuest Information & Learning, US).
- Cetin, B., Guler, N. et Sarica, R. (2016). Using generalizability theory to examine different concept map scoring methods. *Eurasian Journal of Educational Research*, (66), 211-228. Repéré à <https://files.eric.ed.gov/fulltext/EJ1149045.pdf>
- Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal*, 12(2). doi: [HTTPS://DOI.ORG/10.26522/BROCKED.V12I2.38](https://doi.org/10.26522/BROCKED.V12I2.38)
- Chi, M. T. H., Feltovich, P. J. et Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121-152. doi: http://dx.doi.org/10.1207/s15516709cog0502_2
- Chi, M. T. H., Glaser, R. et Farr, M. J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.
- Chiu, C.-H. et Lin, C.-L. (2012). Sequential pattern analysis: Method and application in exploring how students develop concept maps. *Turkish Online Journal of Educational Technology - TOJET*, 11(1), 145-153. Repéré à <http://www.tojet.net/articles/v11i1/11113.pdf>
- Chmielewski, T. C. et Dansereau, D. F. (1998). Enhancing the recall of text: Knowledge mapping training promotes implicit transfer. *Journal of Educational Psychology*, 90(3), 407-413. doi: 10.1037/0022-0663.90.3.407
- Chung, G. K. W. K., Baker, E. L. et Cheak, A. M. (2002). *Knowledge mapper authoring system prototype*. (Rapport n° 575). Los Angeles: National Centre for research on evaluation, standards, and student testing (CRESST), University of California. Repéré à <http://www.cse.ucla.edu/products/reports.php>
- Chung, G. K. W. K., O'Neil, H. F. et Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior*, 15(3-4), 463-493. doi: [https://doi.org/10.1016/S0747-5632\(99\)00032-1](https://doi.org/10.1016/S0747-5632(99)00032-1)
- Clariana, R. B. et Koul, R. (2004, Septembre 14-17). *A computer-based approach for translating text into concept map-like representations*. Communication présentée à First international conference on concept mapping, Navara, Espagne. Repéré à <http://cmc.ihmc.us/papers/cmc2004-045.Pdf>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1), 37-46. doi: <https://doi.org/10.1177/001316446002000104>
- Colchester, K., Hagraas, H., Alghazzawi, D. et Aldabbagh, G. (2017). A survey of artificial intelligence techniques employed for adaptive educational systems within e-learning platforms. 7(1), 47. doi: <https://doi.org/10.1515/jaiscr-2017-0004>
- Coleman, S. L. (1995). *Estimating the reliability and validity of concept mapping as a tool to assess prior knowledge*. (Thèse de doctorat, Virginia Polytechnic Institute and State University, US). Repéré à https://vtechworks.lib.vt.edu/bitstream/handle/10919/38314/LD5655.V856_1994.C656.pdf?sequence=1

- Collins, A. M. et Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428. doi: 10.1037/0033-295X.82.6.407
- Collins, A. M. et Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 240-247. doi: 10.1016/s0022-5371(69)80069-1
- Conlon, T. (2006). Formative assessment of classroom concept maps: the Reasonable Fallible Analyser. *Journal of Interactive Learning Research*, 17(1), 15-36. Repéré à <https://www.learntechlib.org/primary/p/6117/>
- Correia, P., Cabral, G. et Aguiar, J. (2016). Cmaps with errors: Why not? Comparing two cmap-based assessment tasks to evaluate conceptual understanding. Dans A. J. Cañas, P. Reiska et J. Novak (dir.), *Innovating with Concept Mapping. CMC 2016. Communications in Computer and Information Science* (Vol. 635, p. 1-15). Cham: Springer.
- Crocker, L. et Algina, J. (1986). *Introduction to classical and modern test theory*. Mason, Ohio: Cengage Learning.
- Cronbach, L. J. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles*. : John Wiley & Sons.
- Cyr, M.-D., Verreault, J.-S. et Forget, D. (2008). *Observatoire : l'environnement. Manuel de l'élève*. Saint-Laurent, Québec: ERPI.
- Daley, B. J. et Torre, D. M. (2010). Concept maps in medical education : an analytical literature review. *Medical Education*, 44(5), 440-448. doi: 10.1111/j.1365-2923.2010.03628.x
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York, NY: Guilford Press.
- De Ketele, J.-M. (1989). L'évaluation de la productivité des institutions d'éducation. *Cahiers de la Fondation Universitaire: Université et société, le rendement de l'enseignement universitaire*, 3, 73-83.
- De Ketele, J.-M. et Roegiers, X. (2015). *Méthodologie du recueil d'informations: fondements des méthodes d'observation, de questionnaire, d'interview et d'étude de documents*. (5^e éd.). Louvain-La-Neuve: De Boeck Supérieur.
- Defranco, J. F., Jablokow, K., Piovoso, M. J. et Richmond, S. S. (2015). Objectively assessing concept maps for knowledge integration. *International Journal of Engineering Education*, 31(5), 1288-1298. Repéré à https://www.researchgate.net/profile/Kathryn_Jablokow/publication/280575278_Objectively_Assessing_Concept_Maps_for_Knowledge_Integration/links/55e6e36708ae55b89e407706.pdf
- DeFranco, J. F., Jablokow, K. W., Bilen, S. G. et Gordon, A. (2012). *The impact of cognitive style on concept mapping: Visualizing variations in the structure of ideas*. Communication présentée à ASEE Annual Conference and Exposition, Conference Proceedings. Repéré à <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85029104252&partnerID=40&md5=a93d448af3639733d46bed9c81106897>
- Dennis, R. A. (1999). *The effects of feedback cues, relevant information, and scoring procedures on concept mapping performance*. (Thèse de doctorat, University of California, Los Angeles, US). Accessible par Proquest. (9906802)
- Diarra, L. (2013). *Comparabilité entre modalités d'évaluation TIC et papier-crayon: cas de productions écrites en français en cinquième secondaire au Québec*. (Thèse de doctorat, Université de Montréal, Montréal). Repéré à

https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/10109/Diarra_Luc_2012_these.pdf?sequence=2&isAllowed=y

- Dionne, E. (2005). Enjeux en évaluation des apprentissages dans le cadre de programmes d'études formulés par compétences: l'exemple de science et technologie. *Mesure et évaluation en éducation*, 28(2), 49-66.
- Dogusoy-Taylan, B. et Cagiltay, K. (2014). Cognitive analysis of experts' and novices' concept mapping processes: An eye tracking study. *Computers in Human Behavior*, 36, 82-93. doi: <https://doi.org/10.1016/j.chb.2014.03.036>
- Dorsey, D. W., Campbell, G. E., Foster, L. L. et Miles, D. E. (1999). Assessing knowledge structures: Relations with experience and posttraining performance. *Human Performance*, 12(1), 31-57. doi: 10.1207/s15327043hup1201_2
- Dowd, J. E., Duncan, T. et Reynolds, J. A. (2015). Concept maps for improved science reasoning and writing: Complexity isn't everything. *CBE Life Sciences Education*, 14(4). doi: 10.1187/cbe.15-06-0138
- Driscoll, M. P. (2005). *Psychology of learning for instruction*. (3^e éd.). Boston: Pearson.
- Durand, M.-J. et Chouinard, R. (dir.). (2012). *L'évaluation des apprentissages: De la planification à la communication des résultats*. Montréal: Marcel Didier.
- Eden, C. (1988). Cognitive mapping. *European Journal of Operational Research*, 36(1), 1-13. doi: [https://doi.org/10.1016/0377-2217\(88\)90002-1](https://doi.org/10.1016/0377-2217(88)90002-1)
- Eden, C. (1992). On the nature of cognitive maps. *Journal of Management Studies*, 29(3), 261-265. doi: <https://doi.org/10.1111/j.1467-6486.1992.tb00664.x>
- Edmondson, K. M. (1995). Concept mapping for the development of medical curricula. *Journal of Research in Science Teaching*, 32(7), 777-793. doi: 10.1002/tea.3660320709
- Edmondson, K. M. (2005). Assessing science understanding through concept maps. Dans J. J. Mintzes, J. H. Wandersee et J. D. Novak (dir.), *Assessing science understanding: a human constructivist view* (p. 15-40). Burlington: Elsevier Academic Press.
- Erdogan, Y. (2009). Paper-based and computer-based concept mappings: The effects on computer achievement, computer anxiety and computer attitude. *British Journal of Educational Technology*, 40(5), 821-836. doi: 10.1111/j.1467-8535.2008.00856.x
- Ericsson, K. A. et Simon, H. A. (1993). *Protocol analysis : verbal reports as data*. (2^e éd.). Cambridge, MA: MIT Press.
- Evans Iii, A. W., Jentsch, F., Hitt Ii, J. M., Bowers, C. et Salas, E. (2001). *Mental model assessments: Is there convergence among different methods?* Communication présentée à Proceedings of the Human Factors and Ergonomics Society. Repéré à <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0442326540&partnerID=40&md5=630f47fb52d8963ee44a0df07089013a>
- Fisher, K. M. (1990). Semantic networking : the new kid on the block. *Journal of Research in Science Teaching*, 27(10), 1001-1018. doi: <https://doi.org/10.1002/tea.3660271008>
- Francisco, J. S., Nakhleh, M. B., Nurrenbern, S. C. et Miller, M. L. (2002). Assessing student understanding of general chemistry with concept mapping. *Journal of Chemical Education*, 79(2), 248-257. doi: 10.1021/ed079p248
- Fu, A. C., Raizen, S. A. et Shavelson, R. J. (2009). The nation's report card: A vision of large-scale science assessment. *Science*, 326(5960), 1637-1638.
- Gagnon, S. (2017). Journal publishers' Big Deals: Are they worth it? *Against the Grain*, 29(2). Repéré à

https://papyrus.bib.umontreal.ca/xmlui/bitstream/handle/1866/18507/Gagnon_Stephanie_2017_article.pdf

- Gerrig, R. J. et Zimbardo, P. G. (2002). *Psychology and Life*. Boston: Allyn and Bacon.
- Glaser, R. (1991). Expertise and assessment. Dans M. C. Wittrock et E. L. Baker (dir.), *Testing and cognition*. Eglewood Cliffs, NJ: Prentice-Hall.
- Greene, B. A., Lubin, I. A., Slater, J. L. et Walden, S. E. (2013). Mapping changes in science teachers' content knowledge: concept maps and authentic professional development. *Journal of Science Education and Technology*, 22(3), 287-299. doi: <https://doi.org/10.1007/s10956-012-9393-9>
- Gregoriades, A., Pampaka, M. et Michail, H. (2009). Assessing students' learning in MIS using concept mapping. *Journal of Information Systems Education*, 20(4), 419-430. Repéré à <http://jise.org/Volume20/n4/JISEv20n4p419.pdf>
- Groome, D. (1999). *An introduction to cognitive psychology: Processes and disorders*. Hove, England: Taylor & Francis.
- Grundspenkis, J. (2011). Concept map based intelligent knowledge assessment system: experience of development and practical use. Dans D. Ifenthaler, J. M. Spector, Kinshuk, P. Isaias et D. Sampson (dir.), *Multiple Perspectives on Problem Solving and Learning in the Digital Age* (p. 179-197). New York: Springer
- Gurlitt, J. et Renkl, A. (2008). Are high-coherent concept maps better for prior knowledge activation? Differential effects of concept mapping tasks on high school vs. university students. *Journal of Computer Assisted Learning*, 24(5), 407-419. doi: <https://doi.org/10.1111/j.1365-2729.2008.00277.x>
- Gurlitt, J. et Renkl, A. (2010). Prior knowledge activation: how different concept mapping tasks lead to substantial differences in cognitive processes, learning outcomes, and perceived self-efficacy. *Instructional Science*, 38(4), 417-433. doi: <https://doi.org/10.1007/s11251-008-9090-5>
- Haas, C. (1989). How the writing medium shapes the writing process: Effects of word processing on planning. *Research in the Teaching of English*, 23(2), 181-207. Repéré à https://www.jstor.org/stable/40171409?seq=1#page_scan_tab_contents
- Haertel, E. H. (2006). Reliability. Dans R. L. Brennan (dir.), *Educational Measurement* (4^e éd.). Westport, CT: Praeger Publishers.
- Hambleton, R. K. (2006). Psychometric models, test designs, and item types for the next generation of educational and psychological tests. Dans D. Bartram et R. K. Hambleton (dir.), *Computer-based testing and the Internet* (p. 77-90). Hoboken, NJ: John Wileys & Sons.
- Hambleton, R. K. et Jones, R. W. (1993). An NCME Instructional Module on Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47. doi: 10.1111/j.1745-3992.1993.tb00543.x
- Harris, D. (2008). *A comparative study of the effect of collaborative problem-solving in a Massively Multiplayer Online Game (MMOG) on individual achievement*. (Thèse de doctorat, University of San Francisco, US). Accessible par Proquest. (3317694)
- Hartmeyer, R., Stevenson, M. P. et Bentsen, P. (2017). A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assessment in Education: Principles, Policy and Practice*, 1-22. doi: 10.1080/0969594X.2017.1377685

- Hatami, J., Farrokhnia, M. et Hassanzadeh, M. (2016). Select-and-fill-in concept maps as an evaluation tool in science classrooms. Dans A. J. Cañas, P. Reiska et J. D. Novak (dir.), *Innovating with Concept Mapping. CMC 2016. Communications in Computer and Information Science* (Vol. 635, p. 169-180). Cham: Springer.
- Hay, D., Kinchin, I. et Lygo-Baker, S. (2008). Making learning visible: The role of concept mapping in higher education. *Studies in Higher Education*, 33(3), 295-311. doi: <https://doi.org/10.1080/03075070802049251>
- Hay, D. B. (2007). Using concept maps to measure deep, surface and non-learning outcomes. *Studies in Higher Education*, 32(1), 39-57. doi: <https://doi.org/10.1080/03075070601099432>
- Hay, D. B., Kehoe, C., Miquel, M. E., Hatzipanagos, S., Kinchin, I. M., Keevil, S. F. et Lygo-Baker, S. (2008). Measuring the quality of e-learning. *British Journal of Educational Technology*, 39(6), 1037-1056. doi: <https://doi.org/10.1111/j.1467-8535.2007.00777.x>
- Hay, D. B. et Kinchin, I. M. (2006). Using concept maps to reveal conceptual typologies. *Education and Training*, 48(2), 127-142. doi: <https://doi.org/10.1108/00400910610651764>
- Hayes, J. R. et Flower, L. (1980). Identifying the organization of writing processes. Dans W. Gregg et E. R. Steinberg (dir.), *Cognitive processes in writing* (p. 3-30). Hillsdale, NJ: Lawrence Erlbaum.
- Herl, H. E., O'Neil, H. F., Chung, G. K. W. K. et Schacter, J. (1999). Reliability and validity of a computer-based knowledge mapping system to measure content understanding. *Computers in Human Behavior*, 15, 315-333. doi: [http://dx.doi.org/10.1016/S0747-5632\(99\)00026-6](http://dx.doi.org/10.1016/S0747-5632(99)00026-6)
- Hilbert, T. S. et Renkl, A. (2008). Concept mapping as a follow-up strategy to learning from texts: what characterizes good and poor mappers? *Instructional Science*, 36(1), 53-73. doi: 10.1007/s11251-007-9022-9
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70. Repéré à <http://www.jstor.org/stable/4615733>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. doi: 10.1007/bf02289447
- Horton, P. B., McConney, A. A., Gallo, M., Woods, A. L., Senn, G. J. et Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education*, 77(1), 95-111. doi: 10.1002/sce.3730770107
- Hult, C. A. (1988). The computer and the inexperienced writer. *Computers and Composition*, 5(2), 29-38. doi: [http://dx.doi.org/10.1016/8755-4615\(88\)80003-3](http://dx.doi.org/10.1016/8755-4615(88)80003-3)
- Ifenthaler, D. (2010a). Bridging the gap between expert -novice differences: The model-based feedback approach. *Journal of Research on Technology in Education*, 43(2), 103-117. doi: 10.1080/15391523.2010.10782564
- Ifenthaler, D. (2010b). Relational, structural, and semantic analysis of graphical representations and concept maps. *Educational Technology Research and Development*, 58(1), 81-97. doi: 10.1007/s11423-008-9087-4
- Ingec, S. K. (2009). Analysing concept maps as an assessment tool in teaching physics and comparison with the achievement tests. *International Journal of Science Education*, 31(14), 1897-1915. doi: <http://dx.doi.org/10.1080/09500690802275820>

- Jablokow, K. W., Defranco, J. F., Richmond, S. S., Piovoso, M. J. et Bilén, S. G. (2015). Cognitive style and concept mapping performance. *Journal of Engineering Education*, 104(3), 303-325. doi: 10.1002/jee.20076
- Jeong, A. (2014). Sequentially analyzing and modeling causal mapping processes that support causal understanding and systems thinking. Dans D. Ifenthaler et R. Hanewald (dir.), *Digital Knowledge Maps in Education* (p. 239-251). doi: <https://doi.org/10.1007/978-1-4614-3178-7>.
- Jimenez Snelson, L. (2011). *Estimating the reliability of concept map ratings using a scoring rubric based on three attributes of propositions*. (Thèse de doctorat, Brigham Young University, US). Accessible par Proquest. (3427193)
- Jonassen, D. H. (2000). *Computers as mindtools for schools: Engaging critical thinking*. (2^e éd.). Upper Saddle River, NJ: Merrill.
- Jonassen, D. H. (2006). *Modeling with technology: Mindtools for conceptual change*. (3^e éd.). Upper Saddle River, NJ: Pearson.
- Jonassen, D. H., Reeves, T. C., Hong, N., Harvey, D. et Peters, K. (1997). Concept mapping as cognitive learning and assessment tools. *Journal of Interactive Learning Research*, 8(3-4), 289-308.
- Joseph, C., Conradsson, D., Nilsson Wikmar, L. et Rowe, M. (2017). Structured feedback on students' concept maps: The proverbial path to learning? *BMC Medical Education*, 17(1). doi: 10.1186/s12909-017-0930-3
- Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319-342. doi: 10.1111/j.1745-3984.2001.tb01130.x
- Kane, M. T. (2006). Validation. Dans R. L. Brennan (dir.), *Educational measurement* (4^e éd.). Westport, CT: Praeger Publishers.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. doi: 10.1111/jedm.12000
- Kassab, S. E., Fida, M., Radwan, A., Hassan, A. B., Abu-Hijleh, M. et O'Connor, B. P. (2016). Generalisability theory analyses of concept mapping assessment scores in a problem-based medical curriculum. *Medical Education*, 50(7), 730-737. doi: 10.1111/medu.13054
- Keppens, J. et Hay, D. (2008). Concept map assessment for teaching computer programming. *Computer Science Education*, 18(1), 31-42. doi: <https://doi.org/10.1080/08993400701864880>
- Kessler, R. (1995). *Cognitive styles and concept mapping dimensions of hypermedia computer users*. (Thèse de doctorat, Florida Atlantic University, US). Accessible par Proquest. (9528553)
- Kinchin, I. M. (2001). If concept mapping is so helpful to learning biology, why aren't we all doing it? *International Journal of Science Education*, 23(12), 1257-1269. doi: <https://doi.org/10.1080/09500690010025058>
- Kinchin, I. M., De-Leij, F. A. A. M. et Hay, D. B. (2005). The evolution of a collaborative concept mapping activity for undergraduate microbiology students. *Journal of Further and Higher Education*, 29(1), 1-14. doi: <https://doi.org/10.1080/03098770500037655>
- Kinchin, I. M., Hay, D. B. et Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual development. *Educational Research*, 42(1), 43-57. doi: <https://doi.org/10.1080/001318800363908>

- Krabbe, H. (2014). Digital concept mapping for formative assessment (*Digital Knowledge Maps in Education: Technology-Enhanced Support for Teachers and Learners* (p. 275-297).
- Landis, J. et Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. Repéré à <https://www.jstor.org/stable/2529310>
- Larousse en ligne. (s.d.). Dans Proposition. Repéré <http://www.larousse.fr/dictionnaires/francais/proposition/64405>
- Laurier, M. D., Tousignant, R. et Morissette, D. (2005). *Les principes de la mesure et de l'évaluation des apprentissages*. (3^e éd.). Montréal, Québec: Gaëtan Morin Éditeur.
- Laveault, D. et Grégoire, J. (2002). *Introduction aux théories des tests: en psychologie et en sciences de l'éducation*. : De Boeck Supérieur.
- Lim, K. Y., Lee, H. W. et Grabowski, B. (2009). Does concept-mapping strategy work for everyone? The levels of generativity and learners' self-regulated learning skills. *British Journal of Educational Technology*, 40(4), 606-618. doi: doi:10.1111/j.1467-8535.2008.00872.x
- Linacre, J. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328.
- Linacre, J. M. (1998). Detecting multidimensionality: which residual data-type works best? *Journal of outcome measurement*, 2(3), 266-283. Repéré à <http://www.winsteps.com/a/Linacre-multidimensionality-residuals.pdf>
- Linacre, J. M. (2004). Optimizing rating scale category effectiveness. Dans E. V. Smith et R. M. Smith (dir.), *Introduction to rasch measurement* (p. 258-278). Maple Grove: JAM Press.
- Linacre, J. M. (2017) Winsteps® Rasch measurement computer program. (Version 3.93.0). Beaverton, Oregon. Repéré à Winsteps.com
- Liu, J. (2010). *The assessment agent system: assessing comprehensive understanding based on concept maps*. (Thèse de doctorat inédite, Virginia Polytechnic Institute and State University, Blacksburg, Virginia). Repéré à <http://scholar.lib.vt.edu/theses/available/etd-10092010-152756/>
- Liu, X. (1994). *The validity and reliability of concept mapping as an alternative science assessment when item response theory is used for scoring*. Communication présentée à American Educational Research Association, New Orleans, LA. Repéré à <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=eric3&AN=ED370992>
- Liu, X. et Hinchey, M. (1996). The internal consistency of a concept mapping scoring scheme and its effect on prediction validity. *International Journal of Science Education*, 18(8), 921-937. doi: 10.1080/0950069960180805
- Lord, F. M. et Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wisley.
- Luckie, D., Harrison, S. H. et Ebert-May, D. (2011). Model-based reasoning: using visual tools to reveal student learning. *Advances in Physiology Education*, 35(1), 59-67. doi: <https://doi.org/10.1152/advan.00016.2010>
- Lutz, J. A. (1987). A study of professional and experienced writers revising and editing at the computer and with pen and paper. *Research in the Teaching of English*, 398-421. Repéré à <https://www.jstor.org/stable/40171125>
- Mahler, D., Großschedl, J. et Harms, U. (2017). Using doubly latent multilevel analysis to elucidate relationships between science teachers' professional knowledge and students'

- performance. *International Journal of Science Education*, 39(2), 213-237. doi: <https://doi.org/10.1080/09500693.2016.1276641>
- Marchand, C. et d'Ivernois, J.-F. (2004). Les cartes conceptuelles dans les formations en santé. *Pédagogie Médicale*, 5(4), 230-240. doi: <https://doi.org/10.1051/pmed:2004031>
- Marchand, C., D'Ivernois, J. F., Assal, J. P., Slama, G. et Hivon, R. (2002). An analysis, using concept mapping, of diabetic patients' knowledge, before and after patient education. *Medical Teacher*, 24(1), 90-99. doi: 10.1080/01421590120091087
- Markham, K. M., Mintzes, J. J. et Jones, M. G. (1994). The concept map as a research and evaluation tool: further evidence of validity. *Journal of Research in Science Teaching*, 31(1), 91-101. doi: <https://doi.org/10.1002/tea.3660310109>
- Markus, K. A. et Borsboom, D. (2013). *Frontiers of test validity theory: measurement, causation, and meaning*. New York: Routledge.
- Masters, G. N. (1982). A rasch model for partial credit scoring. *Psychometrika*, 47(2), 149-174. doi: 10.1007/bf02296272
- McClure, J. R., Sonak, B. et Suen, H. K. (1999). Concept map assessment of classroom learning: reliability, validity and logistical practicality. *Journal of Research in Science Teaching*, 36(4), 475-492. doi: [http://dx.doi.org/10.1002/\(SICI\)1098-2736\(199904\)36:4<475::AID-TEA5>3.0.CO;2-O](http://dx.doi.org/10.1002/(SICI)1098-2736(199904)36:4<475::AID-TEA5>3.0.CO;2-O)
- McDonald, A. S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education*, 39(3), 299-312. doi: [https://doi.org/10.1016/S0360-1315\(02\)00032-5](https://doi.org/10.1016/S0360-1315(02)00032-5)
- McNamara, D. S., Allen, L. K., Crossley, S. A., Dascalu, M. et Perret, C. A. (2017). Natural language processing and learning analytics. Dans C. Lang, G. Siemens, A. Wise et D. Gasevic (dir.), *Handbook of learning analytics* (p. 93). doi: 10.18608/hla17.
- Messick, S. (1989). Validity. Dans R. L. Linn (dir.), *Educational measurement* (3^e éd.). New York: McMillan.
- Michael, R. S. (1995). *The validity of concept maps for assessing cognitive structure*. (Thèse de doctorat, Indiana University, US). Accessible par Proquest. (9507650)
- Millard, D. E., Bailey, C., Davis, H. C., Gilbert, L., Howard, Y. et Wills, G. (2006). *The e-Learning assessment landscape*. Communication présentée à Proceedings - Sixth International Conference on Advanced Learning Technologies, ICALT 2006. Repéré à <https://www.scopus.com/inward/record.uri?eid=2-s2.0-34247123218&partnerID=40&md5=79fd5298e585dc581135134c5a65cca5>
- Mintzes, J. J., Wandersee, J. H. et Novak, J. D. (1998). *Teaching science for understanding: a human constructivist view*. Amsterdam: Academic Press.
- Mintzes, J. J., Wandersee, J. H. et Novak, J. D. (2001). Assessing understanding in biology. *Journal of Biological Education*, 35(3), 118-124. doi: <https://doi.org/10.1080/00219266.2001.9655759>
- Mintzes, J. J., Wandersee, J. H. et Novak, J. D. (dir.). (2005). *Assessing science understanding: a human constructivist view*. Burlington: Elsevier Academic Press.
- Mislevy, R. J., Almond, R. G. et Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, 2003(1). doi: <https://doi.org/10.1002/j.2333-8504.2003.tb01908.x>
- Mislevy, R. J., Steinberg, L. S., Almond, R. G. et Lukas, J. F. (2006). Concepts, terminology, and basic models of evidence-centered design. Dans D. M. Williamson, R. J. Mislevy et

- I. I. Bejar (dir.), *Automated scoring of complex tasks in computer-based testing* (p. 15-47). New York, NY: Routledge.
- Moher, D., Liberati, A., Tetzlaff, J. et Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*, 339. doi: 10.1136/bmj.b2535
- Molaison, E. F., Taylor, K. A., Erickson, D. et Connell, C. L. (2009). The use and perceptions of concept mapping as a learning tool by dietetic internship students and preceptors. *Journal of allied health*, 38(3), e97-103.
- Morin, M. et Paquette, F. (2014). La cartographie conceptuelle: un exemple d'utilisation en science et technologie. Dans M.-J. Durand et N. Loyer (dir.), *L'Instrumentation pour l'évaluation : la boîte à outils pour l'enseignant évaluateur*. Montréal: Marcel Didier.
- Mühling, A. (2016). Aggregating concept map data to investigate the knowledge of beginning CS students. *Computer Science Education*, 26(2), 16. doi: <https://doi.org/10.1080/08993408.2016.1241340>
- Mühling, A. (2017). Concept landscapes: aggregating concept maps for analysis. *Journal of Educational Data Mining*, 9(2), 1-30. Repéré à <http://jedm.educationaldatamining.org/index.php/JEDM/article/view/138>
- Nesbit, J. C. et Adesope, O. O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research*, 76(3), 413-448. doi: <https://doi.org/10.3102/00346543076003413>
- Nottis, K. E. K. (1996). *The effective use of analogies in earth science*. (Thèse de doctorat, State University of New York at Buffalo, New York, US). Accessible par Proquest. (9617897)
- Novak, J. D. (1990a). Concept mapping: a useful tool for science education. *Journal of Research in Science Teaching*, 27(10), 937-949. doi: <https://doi.org/10.1002/tea.3660271003>
- Novak, J. D. (1990b). Concept maps and Vee diagrams: two metacognitive tools to facilitate meaningful learning. *Instructional Science*, 19(1), 29-52. doi: <https://doi.org/10.1007/BF00377984>
- Novak, J. D. (2002). Meaningful learning: the essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86(4), 548-571. doi: <https://doi.org/10.1002/sce.10032>
- Novak, J. D. (2003). The promise of new ideas and new technology for improving teaching and learning. *Cell Biology Education*, 2(2), 122-132. doi: <https://doi.org/10.1187/cbe.02-11-0059>
- Novak, J. D. et Canas, A. J. (2008). *The theory underlying concept maps and how to construct and use them*. Pensacola, US: IHMC CmapTools. Repéré à <http://eprint.ihmc.us/id/eprint/5>
- Novak, J. D. et Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization*, 5(3), 175-184. doi: 10.1057/palgrave.ivs.9500126
- Novak, J. D. et Gowin, D. B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.
- Novak, J. D., Gowin, D. B. et Johansen, G. T. (1983). The use of concept mapping and knowledge Vee mapping with junior high school science students. *Science Education*, 67(5), 625-645. doi: <http://dx.doi.org/10.1002/sce.3730670511>

- Novak, J. D. et Musonda, D. (1991). A twelve-year longitudinal study of science concept learning. *American Educational Research Journal*, 28(1), 117-153. doi: 10.3102/00028312028001117
- Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology*, 3(1), 1-18. doi: [http://dx.doi.org/10.1016/0022-2496\(66\)90002-2](http://dx.doi.org/10.1016/0022-2496(66)90002-2)
- NVivo qualitative data analysis software. (Version 10). (2012) Doncaster, Australia: QSR International Pty Ltd.
- Okoli, C. (2009). *A brief review of studies of Wikipedia in peer-reviewed journals*. Communication présentée à Digital Society, 2009. ICDS'09.
- Parshall, C. G., Harmes, C. J., Davey, T. et Pashley, P. J. (2010). Innovative items for computerized testing. Dans W. J. van der Linden et C. A. W. Glas (dir.), *Elements of adaptive testing* (p. 215-230). doi: <http://dx.doi.org/10.1007/978-0-387-85461-8>. Repéré à <http://dx.doi.org/10.1007/978-0-387-85461-8>
- Pellegrino, J. W., Chudowsky, N. et Glaser, R. (2001). *Knowing what students know: the science and design of educational assessment*. Washington, DC: National Academy Press.
- Piolat, A. (2004). Approche cognitive de l'activité rédactionnelle et de son acquisition. Le rôle de la mémoire de travail. *Linx*, 51, 55-74. doi: 10.4000/linx.174
- Pudelko, B., Young, M., Vincent-Lamarre, P. et Charlin, B. (2012). Mapping as a learning strategy in health professions education: A critical analysis. *Medical Education*, 46(12), 1215-1225. doi: 10.1111/medu.12032
- Qasim, I., Jeong, J.-W., Heu, J.-U. et Lee, D.-H. (2013). Concept map construction from text documents using affinity propagation. *Journal of Information Science*, 39(6), 719-736. doi: <https://doi.org/10.1177/0165551513494645>
- Quillian, M. R. (1969). The teachable language comprehender: a simulation program and theory of language. *Commun. ACM*, 12(8), 459-476. doi: 10.1145/363196.363214
- R Development Core Team. (2017) R: A language and environment for statistical computing. (Version 3.4.0). Vienna, Austria: R Foundation for Statistical Computing. Repéré à <http://www.R-project.org>
- Raikes, N. et Harding, R. (2003). The Horseless Carriage Stage: replacing conventional measures. *Assessment in Education: Principles, Policy & Practice*, 10(3), 267-277. doi: 10.1080/0969594032000148136
- Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Oxford, England: Nielsen & Lydiche.
- Reavley, N. J., Mackinnon, A. J., Morgan, A. J., Alvarez-Jimenez, M., Hetrick, S. E., Killackey, E., . . . Jorm, A. F. (2012). Quality of information sources about mental disorders: a comparison of Wikipedia with centrally controlled web and printed sources. *Psychological Medicine*, 42(8), 1753-1762. doi: 10.1017/S003329171100287X
- Rice, D. C., Ryan, J. M. et Samson, S. M. (1998). Using concept maps to assess student learning in the science classroom: Must different methods compete? *Journal of Research in Science Teaching*, 35(10), 1103-1127. doi: 10.1002/(SICI)1098-2736(199812)35:10<1103::AID-TEA4>3.0.CO;2-P
- Rochette, A., Bélisle, M., Laflamme, A., Doucet, M., Chaput, M. et Fillion, B. (2010). Étude descriptive de l'utilisation des cartes conceptuelles comme stratégie pédagogique en sciences de la santé. *Pédagogie Médicale*, 11(2), 97-109. doi: <https://doi.org/10.1051/pmed/2010011>

- Roegiers, X. (2010). *L'école et l'évaluation: des situations complexes pour évaluer les acquis des élèves*. (2^e éd.). Bruxelles: De Boeck.
- Roth, M. et Bernhardt, K. L. S. (2016). *Using concept maps for assessment and improvement of a multi-section introduction to engineering course*. Communication présentée à ASEE Annual Conference and Exposition, Conference Proceedings. Repéré à <https://www.scopus.com/inward/record.uri?eid=2-s2.0-84983317356&partnerID=40&md5=01d353c9233408b82e36ebdaa0b5ae02>
- Royer, R. et Royer, J. (2004). Comparing hand drawn and computer generated concept mapping. *Journal of Computers in Mathematics and Science Teaching*, 23(1), 67-81.
- Ruiz-Primo, M. A. (2014). [Challenges of studying concept map assessment]. Communication personnelle.
- Ruiz-Primo, M. A., Iverson, H. et Yin, Y. (2009, Avril 17-21). *Towards the use of concept maps in large-scale assessments: Exploring the efficiency of two scoring methods*. Communication présentée à National Association of Research in Science Teaching (NARST) Annual International Conference, Garden Grove, CA.
- Ruiz-Primo, M. A., Schultz, S. E., Li, M. et Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Teaching*, 38(2), 260-278.
- Ruiz-Primo, M. A. et Shavelson, R. J. (1996). Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6), 569-600. doi: [http://dx.doi.org/10.1002/\(SICI\)1098-2736\(199608\)33:6<569::AID-TEA1>3.0.CO;2-M](http://dx.doi.org/10.1002/(SICI)1098-2736(199608)33:6<569::AID-TEA1>3.0.CO;2-M)
- Ruiz-Primo, M. A., Shavelson, R. J., Li, M. et Schultz, S. E. (2001). On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educational Assessment*, 7(2), 99-141. doi: https://doi.org/10.1207/S15326977EA0702_2
- Rumelhart, D. E. (1980). Schemata : the building blocks of cognition. Dans R. J. Spiro, B. C. Bruce et W. F. Brewer (dir.), *Theoretical issues in reading comprehension : perspectives from cognitive psychology, linguistics, artificial intelligence, and education*. Hillsdale, N.J.: L. Erlbaum Associates.
- Rumelhart, D. E. et Norman, D. A. (1978). Accretion, tuning, and restructuring : three modes of learning. Dans J. W. Cotton et R. L. Klatzky (dir.), *Semantic factors in cognition*. New Jersey, US: Lawrence Erlbaum Associates.
- Rumelhart, D. E. et Ortony, A. (1977). The representation of knowledge in memory. Dans R. C. Anderson, R. J. Spiro et W. E. Montague (dir.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rye, J. A. et Rubba, P. A. (1998). An exploration of the concept map as an interview tool to facilitate the externalization of students' understandings about global atmospheric change. *Journal of Research in Science Teaching*, 35(5), 521-546. doi: [https://doi.org/10.1002/\(SICI\)1098-2736\(199805\)35:5<521::AID-TEA4>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1098-2736(199805)35:5<521::AID-TEA4>3.0.CO;2-R)
- Rye, J. A. et Rubba, P. A. (2002). Scoring concept maps: an expert map-based scheme weighted for relationships. *School Science and Mathematics*, 102(1), 33-44. doi: 10.1111/j.1949-8594.2002.tb18194.x
- Safayeni, F., Derbentseva, N. et Cañas, A. J. (2005). A theoretical note on concepts and the need for cyclic concept maps. *Journal of Research in Science Teaching*, 42(7), 741-766. doi: <http://dx.doi.org/10.1002/tea.20074>

- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika*, 34(1), 1-97. doi: 10.1007/bf03372160
- Scalise, K. et Gifford, B. (2006). Computer-based assessment in e-learning: A framework for constructing "intermediate constraint" questions and tasks for technology platforms. *The Journal of Technology, Learning and Assessment*, 4(6). Repéré à <https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1653>
- Scallon, G. (2004). *L'évaluation des apprentissages dans une approche par compétences*. Québec: Éditions du Renouveau Pédagogique.
- Schau, C. et Mattern, N. (1997). Use of map techniques in teaching applied statistics courses. *The American Statistician*, 51(2), 171-175. doi: 10.1080/00031305.1997.10473955
- Schau, C., Mattern, N., Zeilik, M., Teague, K. W. et Weber, R. J. (2001). Select-and-fill-in concept map scores as a measure of students' connected understanding of science. *Educational and Psychological Measurement*, 61(1), 136-158. doi: 10.1177/00131640121971112
- Schwendimann, B. A. (2013). *Mapping biological ideas: Concept maps as knowledge integration tools for evolution education*. (ProQuest Information & Learning, US).
- Shallcross, D. C. (2016). Concept Maps for Evaluating Learning of Sustainable Development. *Journal of Education for Sustainable Development*, 10(1), 160-177.
- Shavelson, R. J., Baxter, G. P. et Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215-232. doi: <https://doi.org/10.1111/j.1745-3984.1993.tb00424.x>
- Shemwell, J., Fu, A., Figueroa, M., Davis, R. et Shavelson, R. (2010). Assessment in schools – secondary science. Dans P. Peterson, E. Baker et B. McGaw (dir.), *International Encyclopedia of Education* (3^e éd., p. 300-310). Oxford: Elsevier.
- Shudong, W., Hong, J., Young, M. J., Brooks, T. et Olson, J. (2007). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: A meta-analysis of testing mode effects. *Educational and Psychological Measurement*, 68(1), 5-24. doi: 10.1177/0013164407305592
- Shymansky, J. A., Woodworth, G., Norman, O., Dunkhase, J., Matthews, C. et Liu, C. T. (1993). A study of changes in middle school teachers' understanding of selected ideas in science as a function of an in-service program focusing on student preconceptions. *Journal of Research in Science Teaching*, 30(7), 737-755. doi: 10.1002/tea.3660300711
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. doi: 10.1007/s11336-008-9101-0
- Sireci, S., G. et Zenisky, A. L. (2006). Innovative item formats in computer-based testing: in pursuit of improved construct representation. Dans T. M. Haladyna et S. M. Downing (dir.), *Handbook of test development*. Mahwah, NJ: Erlbaum.
- Sireci, S. G. (2009). Packing and unpacking sources of validity evidence. Dans R. W. Lissitz (dir.), *The concept of validity : Revisions, new directions and applications*. Charlotte, NC: Information Age Publishing.
- Sireci, S. G., Thissen, D. et Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28(3), 237-247. doi: 10.1111/j.1745-3984.1991.tb00356.x
- Sireci, S. G. et Zenisky, A. L. (2015). Computerized innovative item formats: Achievement and credentialing. Dans S. Lane, M. R. Raymond et T. M. Haladyna (dir.), *Handbook of test development* (2^e éd., p. 329-350). New York, NY: Routledge.

- Smith, E. E. (1989). Concepts and induction. Dans M. I. Posner (dir.), *The foundations of cognitive science*. Cambridge, MA: MIT Press.
- Smith, E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. Dans E. V. Smith et R. M. Smith (dir.), *Introduction to rasch measurement* (p. 93-122). Maple Grove: JAM Press.
- Smith Jr, E. V. (2001). Evidence for the reliability of measures and validity of measure interpretation: a Rasch measurement perspective. *Journal of Applied Measurement*, 2(3), 281-311.
- Smith, R. M. (1999). Fit analysis in latent trait measurement models. *Journal of Applied Measurement*, 1(2), 199-218.
- Sowa, J. F. (2000). *Knowledge representation: logical, philosophical, and computational foundations*. Pacific Grove: Brooks/Cole.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1), 72-101. doi: 10.2307/1412159
- Srinivasan, M., McElvany, M., Shay, J. M., Shavelson, R. J. et West, D. C. (2008). Measuring knowledge structure: reliability of concept mapping assessment in medical education. *Academic Medicine*, 83(12). doi: <https://doi.org/10.1097/ACM.0b013e31818c6e84>
- Tabachnick, B. G. et Fidell, L. S. (2013). *Using multivariate statistics*. (6^e éd.). New York: Pearson.
- Tardif, J. (1992). *Pour un enseignement stratégique: L'apport de la psychologie cognitive*. Montréal: Logiques.
- Tardif, J. (2006). *L'évaluation des compétences : documenter le parcours de développement*. Montréal: Chenelière Éducation.
- Taricani, E. M. et Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development*, 54(1), 65-82. doi: <https://doi.org/10.1007/s11423-006-6497-z>
- Temple, N. J. et Fraser, J. (2014). How accurate are Wikipedia articles in health, nutrition, and medicine?/Les articles de Wikipédia dans les domaines de la santé, de la nutrition et de la médecine sont-ils exacts? *Canadian Journal of Information and Library Science*, 38(1), 37-52. doi: 10.1353/ils.2014.0000
- Tennant, A., McKenna, S. P. et Hagell, P. (2004). Application of Rasch analysis in the development and application of quality of life instruments. *Value in Health*, 7, S22-S26. doi: <https://doi.org/10.1111/j.1524-4733.2004.7s106.x>
- The International Test Commission. (2006). International guidelines on computer-based and internet-delivered testing. *International Journal of Testing*, 6(2), 143-171. doi: 10.1207/s15327574ijt0602_4
- Thissen, D. (1991). *MULTILOG user's guide: Multiple, categorical item analysis and test scoring using item response theory*. : Scientific Software International.
- Thomas, A. (2016). *Evaluating the validity of technology-enhanced educational assessment items and tasks: an empirical approach to studying item features and scoring rubrics*. (Thèse de doctorat inédite, City University of New York, New York).
- Trickett, S. et Trafton, J. G. (2009). A primer on verbal protocol analysis. Dans D. Schmorow, J. Cohn et D. Nicholson (dir.), *The PSI handbook of virtual environments for training and education* (Vol. 1, p. 332-346). Westport, CT: Praeger Security International.
- Trochim, W. M. (1989). Concept Mapping for Evaluation and Planning. *Evaluation and Program Planning*, 12(1).

- Van Someren, M. W., Barnard, Y. F. et Sandberg, J. A. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. : Academic Press London.
- Van Waes, L. et Schellens, P. J. (2003). Writing profiles: the effect of the writing mode on pausing and revision patterns of experienced writers. *Journal of Pragmatics*, 35(6), 829-853. doi: [http://dx.doi.org/10.1016/S0378-2166\(02\)00121-2](http://dx.doi.org/10.1016/S0378-2166(02)00121-2)
- Venkatesan, D., Bharathi Priya, C. et Rajarajan, S. (2017). Concept map based learning and evaluation-a review. *International Journal of Civil Engineering and Technology*, 8(8), 947-954.
- Vikaros, L. S. (2017). *New methods for analyzing semantic graph based assessments in science education*. (Thèse de doctorat, Teachers College, Columbia University, US). Accessible par Proquest. (10126376)
- Vitale, M. R. et Romance, N. R. (2005). Portfolios in science assessment: a knowledge-based model for classroom practice. Dans J. J. Mintzes, J. H. Wandersee et J. D. Novak (dir.), *Assessing Science understanding: a human constructivist view* (p. 167-196). Burlington: Elsevier Academic Press.
- von der Heide, T. (2015). Concept maps for assessing change in learning: A study of undergraduate business students in first-year marketing in China. *Assessment & Evaluation in Higher Education*, 40(2), 286-308. doi: <https://doi.org/10.1080/02602938.2014.910637>
- Walker, J. M. T., Cordray, D. S., King, P. H. et Fries, R. C. (2005). Expert and student conceptions of the design process: Developmental differences with implications for educators. *International Journal of Engineering Education*, 21(3 PART 1), 467-479.
- Wallace, J. D. et Mintzes, J. J. (1990). The concept map as a research tool: exploring conceptual change in biology. *Journal of Research in Science Teaching*, 27(10), 1033-1052. doi: <https://doi.org/10.1002/tea.3660271010>
- Watson, M. K., Pelkey, J., Noyes, C. R. et Rodgers, M. O. (2016). Assessing conceptual knowledge using three concept map scoring methods. *Journal of Engineering Education*, 105(1), 118-146. doi: 10.1002/jee.20111
- Weinerth, K., Koenig, V., Brunner, M. et Martin, R. (2014). Concept maps: A useful and usable tool for computer-based knowledge assessment? A literature review with a focus on usability. *Computers & Education*, 78, 201-209. doi: 10.1016/j.compedu.2014.06.002
- West, D. C., Park, J. K., Pomeroy, R. J. et Sandoval, J. (2002). Concept mapping assessment in medical education: a comparison of two scoring systems. *Medical Education*, 36(9), 820-826. doi: <https://doi.org/10.1046/j.1365-2923.2002.01292.x>
- West, D. C., Pomeroy, J. R., Park, J. K., Gerstenberger, E. A. et Sandoval, J. (2000). Critical thinking in graduate medical education. *JAMA: The Journal of the American Medical Association*, 284(9), 1105-1110. doi: 10.1001/jama.284.9.1105
- Wiggins, G. (1993). *Assessing student performance : exploring the purpose and limits of testing*. San Francisco, CA: Jossey-Bass Publishers.
- Williamson, D. M., Bejar, I. I. et Mislevy, R. J. (2006). Automated scoring of complex tasks in computer-based testing: An introduction. Dans D. M. Williamson, R. J. Mislevy et I. I. Bejar (dir.), *Automated scoring of complex tasks in computer-based testing* (p. 1-13). New York, NY: Routledge.
- Wills, S. W. (2008). *The effect of embedded prompts in graphic organizer instruction on students' relational understanding of secondary content material*. (Thèse de doctorat inédite, University of Alabama, US).

- Wilson, M. et Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181-208. doi: 10.1207/S15324818AME1302_4
- Wright, B. D., Linacre, J. M., Gustafson, J. et Martin-Lof, P. (1994). Reasonable mean-square fit values. *Rasch measurement transactions*, 8(3), 370.
- Wu, M. et Adams, R. (2013). Properties of Rasch residual fit statistics. *Journal of Applied Measurement*, 14(4), 339.
- Wunnasri, W., Pailai, J., Hayashi, Y. et Hirashima, T. (2017). *Reliability investigation of automatic assessment of learner-build concept map with Kit-Build method by comparing with manual methods*. Communication présentée à Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). doi: https://doi.org/10.1007/978-3-319-61425-0_35 Repéré à https://www.scopus.com/inward/record.uri?eid=2-s2.0-85022212647&doi=10.1007%2f978-3-319-61425-0_35&partnerID=40&md5=fe2f02d963caff76f9b149f01ff91a65
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125-145. doi: <https://doi.org/10.1177/014662168400800201>
- Yin, Y. et Shavelson, R. J. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education*, 21(3), 273-291. doi: <http://dx.doi.org/10.1080/08957340802161840>
- Yin, Y., Vanides, J., Ruiz-Primo, M. A., Ayala, C. C. et Shavelson, R. J. (2005). Comparison of two concept-mapping techniques: implications for scoring, interpretation and use. *Journal of Research in Science Teaching*, 42(2), 166-184. doi: <http://dx.doi.org/10.1002/tea.20049>
- Yue, M., Zhang, M., Zhang, C. et Jin, C. (2017). The effectiveness of concept mapping on development of critical thinking in nursing education: A systematic review and meta-analysis. *Nurse Education Today*, 52, 87-94. doi: 10.1016/j.nedt.2017.02.018
- Zwick, W. R. et Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, 99(3), 432-442. doi: <http://dx.doi.org/10.1037/0033-2909.99.3.432>

Annexe A. Formulaire de consentement de l'enseignant



Faculté des sciences de l'éducation
Département d'administration et fondements de l'éducation

FORMULAIRE DE CONSENTEMENT à l'attention des enseignants

Titre de la recherche : Expérimentation de la cartographie conceptuelle en vue de l'évaluation des apprentissages

Chercheur : Maxim Morin, étudiant au doctorat, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal

Directeur de recherche : Jean-Guy Blais, professeur titulaire, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal

A) RENSEIGNEMENTS AUX PARTICIPANTS

1. Objectifs de la recherche

Ce projet vise à tester des tâches de cartographie conceptuelle en science et technologie au secondaire. Ces tâches seront réalisées en mode papier-crayon ou en mode informatique. La recherche permettra d'explorer des méthodes pour analyser les cartes produites par les élèves et elle tentera d'identifier les actions des élèves lors des tâches de construction de cartes conceptuelles. De telles tâches pourront à l'avenir s'ajouter à la boîte à outils des enseignants. La recherche permettra aussi de s'interroger sur l'utilisation de la technologie informatique dans la réalisation de tâches complexes.

2. Participation à la recherche

L'enseignant participant à cette recherche accepte d'introduire et d'expérimenter des tâches de cartographie conceptuelle avec des élèves du deuxième cycle en science et technologie. Dans l'étude, les élèves participants à la recherche auront à répondre à un questionnaire d'items à choix multiples et à réaliser deux tâches de cartographie conceptuelle au cours des activités en salle de classe. Les élèves remettront leur cahier de réponses et leurs cartes à la fin des tâches. Parmi l'ensemble des élèves, un sous-groupe sera convié à être filmé lors de la construction des cartes conceptuelles afin de garder des traces des processus cognitifs utilisés par les élèves.

La participation de l'enseignant implique qu'il accepte de faire travailler individuellement les élèves avec du matériel papier-crayon ou avec un ordinateur portable. L'enseignant participera au choix du contenu disciplinaire des tâches de cartographie conceptuelle, de façon à ce qu'elles soient intégrées dans les activités habituelles de la salle de classe. Chaque tâche sera réalisée dans une séance de cours, au moment qui convient à l'enseignant.

Au cours de la recherche, l'enseignant participant devra identifier des élèves ayant des niveaux d'habiletés différents dans le but d'observer différents processus de construction des cartes conceptuelles. L'enseignant participant, avec l'aide du chercheur, sera responsable de la distribution et de la récolte des formulaires de consentements destinés aux élèves. Il assistera aussi le chercheur dans la gestion de la remise des travaux et dans la planification des rencontres avec les élèves qui effectueront la tâche devant la caméra vidéo. Une fois la collecte terminée, l'enseignant participant sera convié à une entrevue avec le chercheur dans le but de faire un retour au sujet de l'introduction de la cartographie conceptuelle en salle de classe. Après la collecte des données, il aura le choix ou non de participer à l'analyse de l'ensemble des données recueillies pour la recherche.

2. Confidentialité

Les données récoltées au moyen des entrevues demeureront confidentielles. Chaque enseignant participant à la recherche se verra attribuer un numéro et seuls le chercheur principal et/ou la personne mandatée à cet effet auront la liste des participants et des numéros qui leur auront été attribués. De plus, les renseignements seront conservés dans un classeur sous clé situé dans un bureau fermé. Aucune information permettant de vous identifier d'une façon ou d'une autre ne sera publiée. Ces renseignements personnels seront détruits 7 ans après la fin du projet. Seules les données ne permettant pas de vous identifier seront conservées après cette date, le temps nécessaire à leur utilisation.

4. Avantages et inconvénients

En participant à cette recherche, vous pourrez contribuer à l'avancement des connaissances et à l'amélioration des instruments d'évaluation des apprentissages. Votre participation à la recherche pourra également vous donner l'occasion d'explorer un dispositif favorisant l'acquisition de connaissances en science et technologie. Par contre, votre participation implique de consacrer du temps en classe pour les activités prévues dans la recherche. Les données récoltées au cours de cette recherche ne pourront influencer en aucun cas le dossier scolaire des élèves.

5. Droit de retrait

Votre participation est entièrement volontaire. Vous êtes libre de vous retirer en tout temps sur simple avis verbal, sans préjudice et sans devoir justifier votre décision. Si vous décidez de vous retirer de la recherche, vous pouvez communiquer avec le chercheur, dont les coordonnées sont indiquées ci-dessous. Si vous vous retirez de la recherche, les renseignements qui auront été recueillis au moment de votre retrait seront détruits.

6. Compensation

Les participants ne recevront aucune compensation financière pour leur participation à la recherche.

7. Diffusion des résultats

Une séance d'information sera offerte à l'ensemble des enseignants de l'institution participante afin de leur communiquer les conclusions générales de la recherche. Cette séance se déroulera l'année suivant la recherche, une fois que les données seront analysées.

B) CONSENTEMENT

Je déclare avoir pris connaissance des informations ci-dessus, avoir obtenu les réponses à mes questions sur ma participation à la recherche et comprendre le but, la nature, les avantages, les risques et les inconvénients de cette recherche.

Après réflexion et un délai raisonnable, je consens librement à prendre part à cette recherche. Je sais que je peux me retirer en tout temps sans aucun préjudice, sur simple avis verbal et sans devoir justifier ma décision.

Signature : _____ Date : _____

Nom : _____ Prénom : _____

Je suis intéressé(e) à participer à l'analyse des données recueillies dans le cadre de ce projet. Je comprends que le fait de cocher la case OUI m'engage à respecter la confidentialité des informations avec lesquelles j'entrerai en contact dans le cadre de ces analyses	Oui <input type="checkbox"/>	Non <input type="checkbox"/>
---	------------------------------	------------------------------

Je déclare avoir expliqué le but, la nature, les avantages, les risques et les inconvénients de l'étude et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature du chercheur : _____ Date : _____

Nom : _____ Morin _____ Prénom : _____ Maxim _____

Pour toute question relative à la recherche ou pour vous retirer du projet, vous pouvez communiquer avec Maxim Morin, étudiant au doctorant, au numéro de téléphone : (514) 343-6156 ou à l'adresse courriel : maxim.morin@umontreal.ca.

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal, au numéro de téléphone (514) 343-2100 ou à l'adresse courriel suivante: ombudsman@umontreal.ca (l'ombudsman accepte les appels à frais virés).

Pour toute information d'ordre éthique concernant les conditions dans lesquelles se déroule votre participation à ce projet, vous pouvez contacter le coordonnateur Comité plurifacultaire d'éthique de la recherche (CPER) par courriel : cper@umontreal.ca ou par téléphone au (514) 343-6111 poste 1896.

Pour plus d'information sur vos droits comme participants, vous pouvez consulter le portail des participants de l'Université de Montréal à l'adresse suivante : <http://recherche.umontreal.ca/participants>.

Un exemplaire du formulaire d'information et de consentement signé doit être remis au participant

Annexe B. Formulaire de consentement des élèves



Faculté des sciences de l'éducation
Département d'administration et fondements de l'éducation

FORMULAIRE DE CONSENTEMENT à l'attention des élèves

Titre de la recherche : Expérimentation de la cartographie conceptuelle en vue de l'évaluation des apprentissages

Chercheur : Maxim Morin, étudiant au doctorat, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal

Directeur de recherche : Jean-Guy Blais, professeur titulaire, Département d'administration et fondements de l'éducation, Faculté des sciences de l'éducation, Université de Montréal

A) RENSEIGNEMENTS AUX PARTICIPANTS

1. Objectifs de la recherche

Ce projet de cette recherche vise à expérimenter des tâches d'évaluation en science et technologie au secondaire. Ces tâches seront réalisées à l'écrit ou à l'ordinateur. La recherche tentera de déterminer les avantages et les limites de telles tâches pour l'évaluation des apprentissages en sciences. La recherche permettra aussi de s'interroger sur l'utilisation de l'informatique lors de la réalisation de ces tâches.

2. Participation à la recherche

Votre participation à la recherche consiste à répondre par écrit à un questionnaire d'items à choix multiples et à réaliser deux tâches. Ces tâches pourront être faites avec du matériel papier-crayon ou avec un ordinateur portable. Chaque activité de la recherche sera effectuée en salle de classe. Les tâches seront réparties en 3 séances et elles se dérouleront au moment choisi avec l'enseignant. À la fin de chaque tâche, vous remettrez une copie, papier ou électronique, du travail effectué aux chercheurs.

Parmi les élèves participants, huit élèves seront invités à participer à une entrevue. Si vous participez à cette entrevue, vous aurez à effectuer une tâche en exprimant votre pensée à haute voix. Cette entrevue sera réalisée hors des heures de classe et elle devrait durer entre 60 et 75 minutes. Vos actions seront filmées lors de la tâche et les ordinateurs utilisés seront programmés de façon à garder des captures d'écran vidéo de vos actions.

Notez que seulement 8 élèves pourront participer à l'entrevue. Votre enseignant choisira les élèves participants parmi ceux qui auront rempli le formulaire de consentement. Une clause spécifique à la fin de ce document vous permet d'accepter ou non l'invitation à participer à cette entrevue.

3. Confidentialité

Les renseignements que vous donnerez au cours de la recherche demeureront confidentiels. Chaque participant à la recherche se verra attribuer un numéro et seuls le chercheur principal et/ou la personne mandatée à cet effet auront la liste des participants et des numéros qui leur auront été attribués. De plus, les renseignements seront conservés dans un classeur sous clé situé dans un bureau fermé. Aucune information permettant de vous identifier ne sera publiée. Ces renseignements personnels seront détruits 7 ans après la fin du projet. Seules les données ne permettant pas de vous identifier seront conservées après cette date, le temps nécessaire à leur utilisation. Les vidéos ne seront utilisées qu'à des fins d'analyse des données. Toutefois une clause spécifique à la fin de ce document vous permet d'autoriser ou non l'utilisation de certains extraits à des fins didactiques (par exemple, dans le cadre d'une présentation dans un cours ou dans un congrès). Seuls les chercheurs auront accès aux extraits vidéos, qui ne seront aucunement diffusés à travers les réseaux sociaux.

4. Avantages et inconvénients

En participant à cette recherche, vous ne courez pas de risques ou d'inconvénients particuliers. Votre participation, ou non, ne pourra pas entraîner de conséquence dans votre dossier scolaire. Votre

participation pourra contribuer à l'avancement des connaissances et à l'amélioration des instruments d'évaluation des apprentissages.

5. Droit de retrait

Votre participation est entièrement volontaire. Vous êtes libre de vous retirer en tout temps sur simple avis verbal, sans préjudice et sans devoir justifier votre décision. Si vous décidez de vous retirer de la recherche, vous pouvez communiquer avec le chercheur, au numéro de téléphone indiqué ci-dessous. Si vous vous retirez de la recherche, les renseignements qui auront été recueillis au moment de votre retrait seront détruits.

6. Compensation

Vous ne recevrez aucune compensation financière pour votre participation à la recherche.

7. Diffusion des résultats

Les publications en rapport avec le projet de recherche porteront uniquement sur les résultats généraux de façon à garantir la confidentialité des renseignements et l'anonymat des participants.

Une séance d'information sera offerte à l'ensemble des enseignants de l'institution participante afin de leur communiquer les conclusions générales de la recherche lorsque les analyses seront effectuées. Les élèves qui en manifesteront le désir pourront aussi participer à cette séance d'information. Si vous souhaitez y participer, vous pouvez inscrire votre adresse électronique dans la partie consentement de ce formulaire. Ces coordonnées resteront strictement confidentielles et ne serviront qu'à vous inviter à la séance d'information.

B) CONSENTEMENT

Je déclare avoir pris connaissance des informations ci-dessus, avoir obtenu les réponses à mes questions sur ma participation à la recherche et comprendre le but, la nature, les avantages, les risques et les inconvénients de cette recherche.

Après réflexion et un délai raisonnable, je consens librement à prendre part à cette recherche. Je sais que je peux me retirer en tout temps sans aucun préjudice, sur simple avis verbal et sans devoir justifier ma décision.

Je consens à ce que les données anonymisées recueillies dans le cadre de cette étude soient utilisées pour des projets de recherche subséquents de même nature, conditionnellement à leur approbation éthique et dans le respect des mêmes principes de confidentialité et de protection des informations Oui Non

Je suis intéressé à participer à l'entrevue et je consens à être filmé : Oui Non

Si vous avez répondu OUI à la question précédente :
Je consens à ce que des extraits des vidéos soient utilisés à des fins didactiques Oui Non

Signature de l'élève : _____ Date : _____
Nom : _____ Prénom : _____
Courriel (si vous désirez être invité à la séance d'information) _____

Je déclare avoir expliqué le but, la nature, les avantages, les risques et les inconvénients de l'étude et avoir répondu au meilleur de ma connaissance aux questions posées.

Signature du chercheur : _____ Date : _____
Nom : Morin Prénom : Maxim

Pour toute question relative à la recherche ou pour vous retirer du projet, vous pouvez communiquer avec Maxim Morin, étudiant au doctorant, au numéro de téléphone : (514) 343-6156 ou à l'adresse courriel : maxim.morin@umontreal.ca.

Toute plainte relative à votre participation à cette recherche peut être adressée à l'ombudsman de l'Université de Montréal, au numéro de téléphone (514) 343-2100 ou à l'adresse courriel suivante: ombudsman@umontreal.ca (l'ombudsman accepte les appels à frais virés).

Pour toute information d'ordre éthique concernant les conditions dans lesquelles se déroule votre participation à ce projet, vous pouvez contacter le coordonnateur Comité plurifacultaire d'éthique de la recherche (CPER) par courriel : cper@umontreal.ca ou par téléphone au (514) 343-6111 poste 1896.

Pour plus d'information sur vos droits comme participants, vous pouvez consulter le portail des participants de l'Université de Montréal à l'adresse suivante : <http://recherche.umontreal.ca/participants>.

Un exemplaire du formulaire d'information et de consentement signé doit être remis au participant

Annexe C. Lettre d'information aux parents



Faculté des sciences de l'éducation
Département d'administration et fondements de l'éducation

Objet : Participation de votre enfant à un projet de recherche

Bonjour,

Ce printemps, votre enfant sera invité à participer à une recherche dans son cours en science et technologie. Le projet de cette recherche vise à expérimenter des tâches d'évaluation en science et technologie au secondaire. Ces tâches seront réalisées à l'écrit ou à l'ordinateur. La recherche tentera de déterminer les avantages et les limites de telles tâches pour l'évaluation des apprentissages en sciences. La recherche permettra aussi de s'interroger sur l'utilisation de l'informatique lors de la réalisation de ces tâches.

La participation de votre enfant à la recherche consistera à répondre par écrit à un questionnaire d'items à choix multiples et à réaliser deux tâches. Chaque activité de la recherche sera effectuée en salle de classe. Les tâches seront réparties en 3 séances et elles se dérouleront au moment choisi avec l'enseignant. De plus, l'enseignant participera au choix du contenu disciplinaire des tâches, de façon à ce qu'elles soient intégrées dans les activités habituelles de la salle de classe.

En complément à la recherche, certains élèves seront invités à participer à une entrevue au cours de laquelle ils devront effectuer une tâche en exprimant leur pensée à haute voix. Cette tâche sera réalisée hors des heures de classe et elle devrait durer entre 60 et 75 minutes. Les élèves en action seront filmés lors de la tâche et les ordinateurs utilisés dans la version informatisée seront programmés de façon à garder des captures d'écran vidéo des actions des élèves. Notez que seulement 8 élèves pourront participer à la recherche. L'enseignant identifiera les élèves participants parmi ceux dont le formulaire de consentement aura été rempli.

La participation de votre enfant est volontaire. Toutes les données récoltées dans le cadre de la recherche demeureront confidentielles. Aucune information ne permettra d'identifier votre enfant. La participation de votre enfant est sans risque et la participation ou la non-participation de votre enfant au projet n'entraînera aucune pénalité. Les élèves qui ne participeront pas à la recherche pourront exécuter une activité pédagogique équivalente, menée par l'enseignant.

Nous vous remercions grandement de votre précieuse collaboration,

Maxim Morin
Étudiant au doctorat

Jean-Guy Blais
Professeur titulaire,
Directeur de recherche

Annexe D. Description de la tâche d'entraînement

Nom : _____

Groupe : _____

Carte conceptuelle sur l'effet de serre

Consignes

Examine attentivement la liste de concepts ci-dessous. Ils te serviront à construire une carte conceptuelle pour faire la tâche suivante :

Expliquer le mécanisme de l'effet de serre.

Ta carte doit inclure obligatoirement tous les concepts de la liste. N'oublie pas, les relations entre deux concepts doivent toujours comporter un mot-lien.



Concepts obligatoires

Inscris un P à côté des concepts principaux et un S à côté des concepts secondaires.

Chaleur du Soleil	_____
Couche d'ozone	_____
Dioxyde de carbone	_____
Énergie thermique de la Terre	_____
Gaz à effet de serre	_____
Lumière du Soleil	_____
Processus naturel	_____
Rayons infrarouges	_____
Rayons ultraviolets	_____
Surface de la Terre	_____
Température de la Terre	_____

Aide-mémoire

Remplis l'aide-mémoire avant de remettre ta carte finale.

- Tous les concepts de la liste sont présents.
- Les concepts sont utilisés une seule fois.
- Toutes les relations contiennent un mot-lien.
- Le sens des mots-liens est le plus précis possible.
- Les éléments de ta carte sont organisés clairement.

Annexe E. Description de la tâche principale

Les relations trophiques

Consignes

Examine attentivement la liste de concepts ci-dessous. Ils te serviront à construire une carte conceptuelle au sujet des :

Relations trophiques

Ta carte doit obligatoirement inclure tous les concepts de la liste. Tu peux aussi ajouter d'autres concepts pour améliorer ta carte.

N'oublie pas, les relations entre deux concepts doivent toujours comporter un mot-lien.



Concepts obligatoires (17)

Animaux	Décomposeurs
Autotrophes	Hétérotrophes
Cadavres et déchets d'autres organismes	Niveaux trophiques
Chaînes alimentaires	Photosynthèse
Champignons et bactéries	Producteurs
Consommateurs	Relations trophiques
Consommateurs primaires	Réseaux trophiques
Consommateurs secondaires	Végétaux
Consommateurs tertiaires	

Aide-mémoire

Remplis l'aide-mémoire avant de remettre ta carte finale.

- Tous les concepts de la liste sont présents.
- Les concepts sont utilisés une seule fois.
- Toutes les relations contiennent un mot-lien.
- Le sens des mots-liens est le plus précis possible.
- Les éléments de ta carte sont organisés clairement.

Annexe F. Présentation des consignes au sujet de la cartographie conceptuelle

La carte conceptuelle: un outil pour mieux apprendre

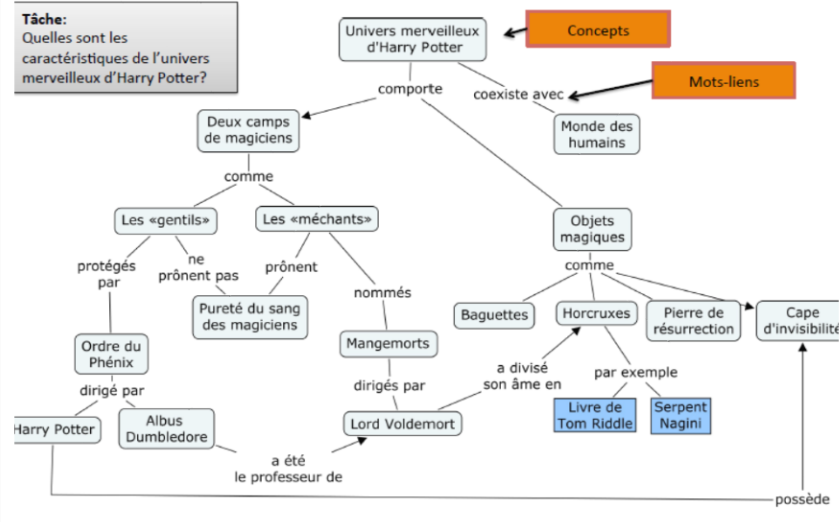
Par Maxim Morin

Doctorant, Université de Montréal

Mai 2014

La carte conceptuelle, c'est quoi?

Tâche:
Quelles sont les caractéristiques de l'univers merveilleux d'Harry Potter?



Construire une carte conceptuelle

1. Identifie clairement les concepts principaux et les concepts secondaires.
2. Établis une première relation entre deux concepts.
3. Répète l'étape 3 en établissant des relations avec les autres concepts.
4. Révise la carte de façon à respecter les règles suivantes :
 1. Le sens des mots de relations doit être le plus précis possible.
 2. Les relations doivent absolument contenir un mot lien.
 3. Chaque concept doit être utilisé UNE SEULE FOIS.
 4. Les éléments de la carte doivent être organisés clairement.

Exercice d'entraînement

Tâche

Explique le fonctionnement des barrages hydroélectriques.

Concepts

Barrage hydroélectrique
Énergie électrique
Énergie hydraulique
Mouvement de l'eau
Turbines hydrauliques

Annexe G. Protocole de verbalisation de la pensée

Protocole pour les entretiens de verbalisation avec les élèves

Paramètres

Les entretiens seront réalisés dans une salle tranquille, à l'abri de sources de distraction et de bruits extérieurs. La salle doit contenir le matériel nécessaire à la collecte de données et elle doit offrir une table de travail et une chaise adaptées à la tâche et confortables.

Voici la liste de matériel à prévoir :

Matériel général.

- Chaises
- Table de travail
- Caméra vidéo
- Eau (en bouteilles et verre)

Matériel spécifique.

- Feuille des consignes de la tâche (principale)
- Feuille des consignes de la tâche d'entraînement
- Ordinateur portable comportant le logiciel CMap et le logiciel CamStudio
- Feuille de format 11 X 17 pouces, crayons à mine, Post-its.

Déroulement de l'entretien

Objectifs

Il est nécessaire d'introduire les objectifs de l'entretien au participant. Pour ce faire, voici une procédure à suivre.

Introduction.

As-tu déjà souhaité être dans la tête de quelqu'un d'autre? D'une personne que tu admires, d'un parent, d'un ami, d'un animal? Si oui, qui?

Objectifs.

Dans les entrevues, mon objectif est d'être dans votre tête, de mieux comprendre ce que vous pensez pendant que vous construisez une carte

conceptuelle.

Déroulement de la séance

Avant d'entreprendre les activités liées à l'entretien, il convient d'en présenter le déroulement au participant.

Étapes.

- Présentation des consignes de passation de la tâche
- La séance sera enregistrée. Les paroles seront captées par enregistrement audio et les actions seront captées directement par l'ordinateur ou par la caméra vidéo.
- Exercice d'entraînement dans lequel vous devrez penser à voix haute, et débriefage.

Consignes

Le chercheur doit informer le participant des principales consignes en vue de l'entretien de verbalisation, mais il ne doit pas compromettre la réalisation de la tâche en donnant trop de consignes. Il doit alors anticiper certaines réactions des participants en les prévenant qu'il leur demandera de «continuer à parler» ou qu'il ne répondra pas nécessairement à leur question. Il importe que le participant se sente à l'aise à exécuter la tâche au moment de l'entretien.

Consignes

- Nous t'avons invité à participer à la recherche dans le but de tester la tâche de cartographie conceptuelle. Si tu sens que tu éprouves des difficultés lors de la réalisation de la tâche, c'est la faute des chercheurs et non la tienne.
- Les données seront traitées de façon confidentielle.
- Tu peux arrêter de faire la tâche en tout temps si tu es mal à l'aise;
- Tu dois savoir que tu peux poser des questions lors de l'entretien, mais il est possible que nous n'y répondions pas. Nous t'encourageons à réaliser la tâche de façon autonome;
- Tu ne dois pas être surpris si nous te demandons de «continuer à parler»
- Tu devras nous le dire lorsque tu auras terminé la tâche. Nous n'arrêterons en aucun cas ton travail. Tu devras nous aviser lorsque tu auras terminé la tâche.

Entraînement à la verbalisation de la pensée

Il est important de faire un exercice d'entraînement avec le participant avant qu'il effectue la tâche de la recherche. Voici trois exemples de tâches.

Exemple 1.

On te demande de compter le nombre de fenêtres que contient votre maison ou votre appartement. Dis tout ce que tu penses pendant que tu réalises cette tâche.

Exemple 2.

Vous avez deux jarres d'eau. L'une peut contenir 5 litres et l'autre, 3 litres. Les jarres n'ont pas de marques et il est impossible de voir la quantité d'eau qu'elles contiennent. Elles peuvent être remplies au robinet et elles peuvent être vidées dans un évier. Il est aussi possible de transvider le contenu d'une jarre à l'autre. Est-ce clair?

On vous demande de prendre en notes le contenu des jarres sur une feuille. Faites deux colonnes, une pour chaque jarre. Les deux jarres sont initialement vides. Leur contenu respectif est de zéro et zéro litre. Est-ce clair? Votre tâche est de produire 4 litres d'eau.

Exemple 3.

Voici les consignes de la tâche de cartographie conceptuelle que tu as faites lors de la dernière séance. La tâche est de créer une mini-carte à partir de quelques concepts, 3 ou 4.

Présentation de la tâche

Au départ, le chercheur a le mandat d'exposer le déroulement de la tâche, et de rappeler qu'il s'agit d'un entretien au cours duquel le participant doit verbaliser sa pensée.

Tâche générale.

Dans quelques instants, vous recevrez une tâche de cartographie conceptuelle. S'il vous plaît, dites à voix haute tout ce que vous pensez pendant que vous réalisez cette tâche.

Ensuite, le chercheur doit présenter la tâche ainsi que le matériel qui sera utilisé pour construire la carte conceptuelle.

Tâche spécifique.

Voici la tâche de cartographie conceptuelle. Nous allons lire l'énoncé ensemble une première fois, et ensuite je te laisserai commencer la tâche, sauf si tu as des questions précises.

Avant le début de la tâche

À faire avant la tâche

- Vérifier que le participant n'a plus de questions.
- Demander au participant de commencer.

Pendant la tâche

À faire pendant la tâche

- Rappeler au participant de s'exprimer (de «continuer à parler»)
- Réorienter le participant vers la réalisation de la tâche s'il commence à analyser et à interpréter son propre processus.
- Prendre des notes exhaustives

Après la tâche

À faire à la fin de l'activité

- Lorsque le participant croit avoir terminé la tâche, le remercier.
- Demander au participant s'il a d'autres remarques ou commentaires.
- Le remercier à nouveau.

Annexe H. Lexique initial de la verbalisation

Lexique des codes pour la verbalisation

- Les [] représentent les actions (précédentes ou coïncidentes)
- Les () représentent des références à certains objets (par exemple, un concept) ou paroles, qui peuvent être inférées à partir de la vidéo.

Commentaires non applicables

Code	Description	Exemple
NA	Information qui ne précise pas la pensée conceptuelle de l'apprenant. Information qui décrit les opérations techniques * À certaines occasions, l'unité verbale renvoie à la fois à des opérations techniques, à la fois à des explications de nature conceptuelle. Il faut alors codé cette unité dans les catégories des explications.	Ok. Oublions ça pour l'instant. On va juste relier trois classes ici pour.... Juste comme je fais pour mettre la flèche, eh?

Explication de nature conceptuelle

Code	Description	Exemple
Définir	L'apprenant fournit plus d'information au sujet du contenu de la carte, (d'un concept ou d'une proposition) ou des options de réponses.	Ex. 1. Consommateurs primaires sont composés de réseaux trophiques Ex. 2. Et la chaîne alimentaire c'est composé de plusieurs relations trophiques. Ah non, les relations trophiques c'est les relations entre les chaînes alimentaires, je pense. Je sais pas. Relation trophiques.
Comparer ou contraster des concepts	L'apprenant fait ressortir les similarités ou différences entre des concepts ou des propositions.	Autotrophes. Ce serait tous ceux qui dépendent. Non, hétérotrophes ce serait tous ceux qui dépendent des autres. Et autotrophes ce serait les végétaux, qui dépendent de personne d'autres
Regrouper	L'apprenant forme des regroupements entre des concepts, sans exprimer de mots-liens entre les concepts.	Ex. 1. Donc qu'ils [cadavres et déchets; champignons et bactéries] font partie des groupes des décomposeurs. Ex. 2. Pour ce qui est de autotrophes. Ça doit pas être loin de producteurs, végétaux.
Justifier	L'apprenant justifie une partie du contenu de sa carte.	Les animaux et les champignons et bactéries seraient des consommateurs, puisqu'ils consomment cadavres et déchets, des animaux et d'autres animaux.

Régulation

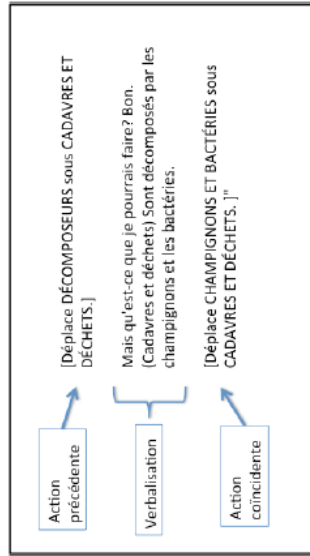
Code	Description	Exemple
Décrire / Appliquer une stratégie Planification et contrôle	L'apprenant porte son attention sur la progression de la tâche, en planifiant la prochaine étape, ou en contrôlant le contenu qui est absent.	Hum, ben au début je les disperse un peu sur le plan pour voir si je peux trouver des liens entre chaque. Pour mieux les voir. Parce que j'ai une mémoire plus visuelle. Pis ça m'aide après ça pour les placer en ordre, si on veut.
Relire	L'apprenant relit le contenu de sa carte	Les producteurs sont des végétaux qui font la photosynthèse. Les niveaux trophiques sont composés de tout ça. Les réseaux trophiques, sont composés. Ah comme ça. Les relations trophiques, parmi les autotrophes et les hétérotrophes.
Réfléchir sur le contenu de la carte	L'apprenant vérifie le contenu (conceptuel) de sa carte, qui entraîne ou non à la modification de celle-ci.	[relisant] Je n'ai pas mis la flèche dans le bon sens. C'est les producteurs et les décomposeurs qui sont les niveaux trophiques.
Réfléchir inefficacement	L'apprenant fait ressortir la difficulté de la tâche ou donne une information ambiguë qui indique qu'il ne sait pas quoi faire. Exclusion : Lors de certaines situations, l'élève formule un lien entre des concepts ou une proposition sous une forme interrogative. Ces unités verbales doivent alors se retrouver dans l'un des codes des explications.	Eh, réseaux trophiques. C'est quoi ça déjà ?
Réviser la qualité du produit	L'apprenant fait une réflexion sur l'aspect global de la carte	Mais je crois que ça va être mélangeant. Je crois que mes lignes vont se mélanger. C'est correct? Comme si mes lignes vont se croiser.

Annexe I. Lexique initial des stratégies

Lexique des codes pour les stratégies

Les unités du verbatim pour le codage des stratégies sont divisées en trois éléments. L'action qui précède, la verbalisation (s'il y a lieu), et l'action coïncidente.

- Les [] représentent les actions (précédentes ou coïncidentes)
- Les () représentent des références à certains objets (par exemple, un concept) ou paroles, qui peuvent être inférées à partir de la vidéo.



Les stratégies représentent des mécanismes intériorisés qui guident l'action. Elles sont inférées sur la base de la séquence des actions et sur la base de la verbalisation. Ce codage représente un microcodage des transitions entre chaque action.

Les stratégies sont divisées en trois grandes catégories.

- La stratégie de démarrage représente une unité qui marque le début de la construction de la carte.
- Les stratégies de poursuite représentent les transitions entre deux actions. Elles concernent à la fois la nature de l'action coïncidente et le lien entre celle-ci et l'action précédente.
- Les stratégies de régulation représentent les processus d'évaluation de la carte, ce qui peut conduire ou non à une modification de celle-ci.

Ordre d'importance des stratégies

Il est possible qu'une action coïncidente contienne des références à différentes stratégies. Dans ces situations, il faut considérer l'action qui correspond à l'opération cognitive de niveau supérieur. Par exemple, dans la catégorie des stratégies de poursuite :

- Les stratégies de déplacement et de formulation ont préséance sur les stratégies «écrire» et «retranscrire»,
- les stratégies de formulation ont préséance sur les stratégies de déplacement,
- les stratégies «réviser le contenu» ont préséance sur les autres stratégies.

Démarrage

Code	Description.	Exemple
Démarrage	Début de la construction de la carte L'élève...	[début] J'ai déjà lu les consignes, alors concept 1, Animaux. Concept 2, autotrophes [Écrit le concept ANIMAUX. Écrit le concept AUTOTROPHES. Place les concepts à l'extérieur de la feuille]

Poursuite (1)

Code	Description. L'élève...	Exemple
Écritre	Écrit un concept sur un Post-it	[Aucune action] Donc, champignons et bactéries. Champignons et bactéries, si je me rappelle bien c'est des décomposeurs. [Écrit CHAMPIGNONS ET BACTÉRIES.]
Retranscrire	Retranscrit un concept sur la page, après avoir retiré le Post-it.	[Trace un lien entre CONSOMMATEURS et CONSOMMATEURS TERTIAIRES. Formule la relation «se divise en»] Hétérotrophes. On va continuer à écrire des mots. Champignons et bactéries. Cadavres et déchets. [Retranscrit CHAMPIGNONS ET BACTÉRIES.]

Poursuite (déplacer ou lier)

Déplacer un concept dans la page ou **Tracer** un lien entre deux concepts, sans rédiger un mot-lien.

Code	Description. L'élève...	Exemple
Sans justification	Ne donne aucune information (verbale ou observable) qui justifie une action	[Déplace HÉTÉROTROPHES.] Réseaux trophiques, [Déplace RÉSEAUX TROPHIQUES.]
En révisant la liste	Réalise une action coïncidente après avoir révisé la liste initiale des concepts.	[Écrit PHOTOSYNTHÈSE. Place le concept près de PRODUCTEURS.] Relations trophiques. Ça, ça irait en haut, pas loin de niveaux trophiques. [Écrit RELATIONS TROPHIQUES à la suite d'une révision de la liste initiale. Place le concept près de NIVEAUX TROPHIQUES]
Sans lien verbal et sans action liée	Réalise une action coïncidente sans que cette dernière ne soit liée directement à l'action précédente.	[Déplace les ANIMAUX près de CONSOMMATEURS.] Ok, mais le thème principal c'est chaînes alimentaires. [Déplace CHAINES ALIMENTAIRES.] [Écrit PRODUCTEURS.]
Avec lien verbal mais sans action liée	Réalise une action coïncidente sans que cette dernière ne soit liée directement à l'action précédente, mais pour laquelle un raisonnement a été explicité.	Donc, hum. Les végétaux sont des producteurs. Non. comment est-ce que je pourrais le mettre? Ce sont des producteurs qui font de la photosynthèse. Non on va le mettre ensemble. Les végétaux ce sont des producteurs, [Déplace VÉGÉTAUX dans la feuille. Déplace PHOTOSYNTHÈSE à gauche de VÉGÉTAUX.]

Avec lien	Réalise une action coïncidente et cette dernière est liée directement à l'action précédente.	<p>[Déplace DÉCOMPOSEURS sous CADAVRES ET DÉCHETS.]</p> <p>Mais qu'est-ce que je pourrais faire? Bon. [Cadavres et déchets] Sont décomposés par les champignons et les bactéries.</p> <p>[Déplace CHAMPIGNONS ET BACTÉRIES sous CADAVRES ET DÉCHETS.]</p>
-----------	--	---

Poursuite (formuler)

Rédige un mot-lien entre deux concepts, après (ou non) avoir déplacé des objets.

Code	Description. L'élève...	Exemple
Sans lien verbal et sans action liée	Formule un mot-lien entre deux concepts, mais ceux-ci n'ont pas de lien direct avec les éléments de l'action précédente.	[Trace un lien entre CONSOMMATEURS et CONSOMMATEURS PRIMAIRES. Formule la relation «sont divisés»]. Les végétaux ils font la photosynthèse. Font la photosynthèse pour se nourrir, dans le fond. [Déplace PHOTOSYNTHESE. Trace un lien entre VÉGÉTAUX et PHOTOSYNTHESE. Formule la relation «font».]
Avec lien verbal mais sans action liée	Réalise une action coïncidente sans que cette dernière ne soit liée directement à l'action précédente, mais pour laquelle un raisonnement a été explicité.	[relisant] (Les producteurs sont constitués par des végétaux, principalement qui eux font la photosynthèse, ce qui les rend autotrophes, dans le fond.) On pourrait aussi faire un lien ici (autotrophes) entre photosynthèse. Ce serait une cause. Voilà. [Déplace PHOTOSYNTHESE. Trace un lien entre PHOTOSYNTHESE et AUTOTROPHES. Formule la relation «cause».]
Avec lien	Formule un mot-lien entre deux concepts, et ceux-ci ont un lien direct avec les éléments de l'action précédente.	[Retranscrit DÉCOMPOSEURS. Trace un lien entre DÉCOMPOSEURS et CHAMPIGNONS et BACTÉRIES. Formule la relation «sont».] Et décomposent des cadavres et autres déchets d'organismes. [Retranscrit CADAVRES ET DÉCHETS. Trace un lien entre CHAMPIGNONS ET BACTÉRIES et CADAVRES et DÉCHETS. Formule la relation «décomposent».]

Régulation

Code	Description. L'élève...	Exemple
Ne pas se rappeler	Cherche le sens d'un concept ou d'une proposition, mais n'arrive pas à fournir une réponse.	[Déplace légèrement NIVEAUX TROPHIQUES. Trace un lien entre NIVEAUX TROPHIQUES et CONSOMMATEURS PRIMAIRES.] Ok. Hum. Quel verbe je pourrais utiliser niveaux trophiques à. Les niveaux trophiques, mais qu'est-ce que les niveaux trophiques? [Aucune action]
Relire	Relit en partie le contenu de sa carte	[Formule la relation «affectent» entre RELATIONS TROPHIQUES et NIVEAUX TROPHIQUES.] Ok on va relire une autre fois pour voir si ça fait du sens. On commence avec le réseau trophique qui est constitués de plusieurs chaines alimentaires qui sont affectées par les relations trophiques. Et qui, elles affectent les niveaux trophiques. Et les niveaux trophiques sont bien sûr les consommateurs primaires, secondaires et tertiaires. Donc, et ainsi, les niveaux trophiques constituent aussi la chaine alimentaire. Donc les consommateurs primaires consomment généralement les producteurs, qui eux sont des autotrophes. Les producteurs sont constitués par des végétaux, principalement qui eux font la photosynthèse, ce qui les rend autotrophes, dans le fond. [relisant]
Réorganiser	Déplace un ou plusieurs objets sans altérer le contenu de la carte.	[Trace un lien entre VÉGÉTAUX et AUTOTROPHES. Formule la relation «sont».] Parfait, on va juste tout sélectionner tout notre schéma et on va le descendre un peu pour laisser de la place aux autres concepts. [Crée de l'espace]

Régulation (suite)

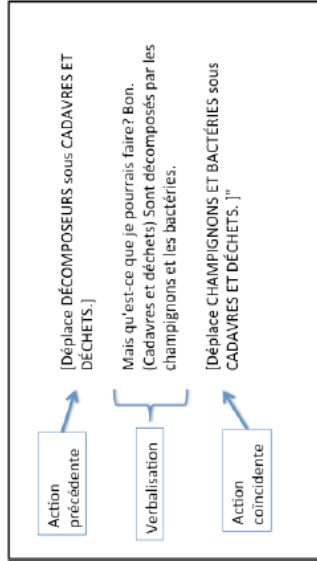
Code	Description. L'élève...	Exemple
Réviser le contenu	Réviser de façon approfondie le contenu dans la carte, ce qui conduit ou non à une modification de son contenu.	[Trace un lien entre NIVEAUX TROPHIQUES et PRODUCTEURS. Formule la relation «sont».] Je n'ai pas mis la flèche dans le bon sens. C'est les producteurs et les décomposeurs qui sont les niveaux trophiques.
Vérifier pour d'autres liens	Cherche à établir d'autres liens entre les concepts, mais cela ne conduit pas à une action .	[Inverse la flèche entre NIVEAUX TROPHIQUES et PRODUCTEURS. Inverse la flèche entre NIVEAUX TROPHIQUES et DÉCOMPOSEURS. Inverse la flèche entre NIVEAUX TROPHIQUES et CONSOMMATEURS.] [Aucune action] Relations. Hum. La relation, on pourrait dire. Sont trophiques. On pourrait dire que la relation entre producteurs décomposeurs et consommateurs c'est que seulement les producteurs ont besoin, pas besoin des autres. Mais les deux autres ont besoin. [Aucune action]
Vérifier les concepts utilisés	Détermine les concepts qui ne sont pas encore utilisés.	[Trace un lien entre CHAMPIGNONS ET BACTÉRIES et CADAVRES ET DÉCHETS D'ORGANISMES. Formule la relation «qui se nourrissent».] Là il nous reste réseaux et niveaux trophiques. [Relisant la liste]

Annexe J. Lexique final des stratégies

Lexique des codes pour les stratégies

Les unités du verbatim pour le codage des stratégies sont divisées en trois éléments. L'action qui précède, la verbalisation (s'il y a lieu), et l'action coïncidente.

- Les [] représentent les actions (précédentes ou coïncidentes)
- Les () représentent des références à certains objets (par exemple, un concept) ou paroles, qui peuvent être inférées à partir de la vidéo.



Les stratégies représentent des mécanismes intériorisés qui guident l'action. Elles sont inférées sur la base de la séquence des actions et sur la base de la verbalisation. Ce codage représente un microcodage des transitions entre chaque action.

Les stratégies sont divisées en trois grandes catégories.

- La stratégie de démarrage représente une unité qui marque le début de la construction de la carte.
- Les stratégies de poursuite représentent les transitions entre deux actions. Elles concernent à la fois la nature de l'action coïncidente et le lien entre celle-ci et l'action précédente.
- Les stratégies de régulation représentent les processus d'évaluation de la carte, ce qui peut conduire ou non à une modification de celle-ci.

Planification / Récupération d'idées

Code	Description.	Exemple
P.1 Exprimer verbalement des relations entre des concepts	Le participant lit des énoncés de concepts et exprime verbalement des relations entre des concepts sans faire quelconque action.	[Aucune action] Niveaux trophiques, divisés en consommateurs, producteurs. Végétaux sont producteurs. Végétaux sont autotrophes. (Végétaux) utilisent photosynthèse. Champignons et bactéries sont des décomposeurs. (Décomposeurs) mangent cadavres et déchets. Animaux font cadavres et déchets, représentent. Consommateurs se divisent en consommateurs primaires, secondaires, tertiaires. Animaux est un consommateur. Animaux, et champignons et bactéries sont hétérotrophes. Hum. Animaux, qui mangent végétaux. Ça me prend un lien pour relation trophique, qui exprimerait que les animaux mangent les végétaux. Les champignons et bactéries mangent les animaux morts. Là je ne sais pas où relier les relations trophiques. Et réseaux trophiques c'est l'ensemble des relations trophiques. Je peux pas y aller avec hétérotrophes ni autotrophes. Ça c'est clair. [relisant les concepts]
P.2 Définir un concept	Le participant lit un énoncé de concept et verbalise une définition de ce concept.	[Trace un lien entre VÉGÉTAUX et PHOTOSYNTHÈSE. Formule la relation «pratique».] Autotrophes c'est ceux qui se nourrissent par eux-mêmes et hétérotrophes c'est ceux qui ont besoin des autres pour se nourrir. Juste pour être bien sûr. [...] Parfait, je vais y aller à auto qui est par lui-même et hétéro qui est par les autres. [Déplace AUTOTROPHES près de PHOTOSYNTHÈSE.]

Planification / Organisation

Code	Description.	Exemple
P.3 Identifier le thème principal	Le participant désigne un concept comme étant le concept principal de la tâche.	[Déplace les ANIMAUX près de CONSOMMATEURS.] Ok, mais le thème principal c'est chaînes alimentaires. [Déplace CHAINES ALIMENTAIRES au haut de la page.]
P.4 Identifier les concepts principaux	Le participant désigne certains concepts comme étant les concepts principaux de la tâche.	[Aucune action] Décomposeurs. Hétérotrophes. Niveaux trophiques. Photosynthèse. Producteurs. Relations trophiques, réseaux trophiques. Végétaux. D'après moi, les concepts les plus importants ce seraient les relations trophiques, réseaux trophiques, consommateurs, décomposeurs et producteurs parce que c'est les trois niveaux trophiques. Je vais écrire tous les concepts sur les Post-it avant même de faire des liens.
		[Lit les concepts]

Code	Description.	Exemple
<p>P.5 Réunir deux concepts en vue de former ultérieurement une proposition</p>	<p>Le participant positionne deux concepts côte-à-côte et exprime qu'il va former une relation entre ces concepts ultérieurement.</p> <p>Le participant peut aussi tracer un lien entre ces concepts, mais il ne rédige pas d'expression de liaison.</p>	<p>Exemple 1 [Déplace DÉCOMPOSEURS près de CHAINES ALIMENTAIRES]</p> <p>et, hum. Les décomposeurs sont composés. Et bien c'est les champignons et les bactéries.</p> <p>[Déplace CHAMPIGNONS ET BACTÉRIES près de DÉCOMPOSEURS]</p> <p>Exemple 2 [relisant]</p> <p>(Les producteurs sont des végétaux, qui eux sont autotrophes, et pratiquent la photosynthèse.)</p> <p>On pourrait même relier autotrophes avec photosynthèse. Pour dire, les végétaux sont autotrophes, donc. Sont autotrophes.</p> <p>[Trace un lien entre AUTOTROPHES et PHOTOSYNTHESE.]</p>
<p>P.6 Regrouper plusieurs concepts en vue de former ultérieurement des relations</p>	<p>Le participant regroupe trois ou plusieurs concepts et exprime que ces concepts forment un groupe de concepts.</p>	<p>[Déplace PRODUCTEURS près de CONSOMMATEURS.]</p> <p>Ensuite je vais prendre consommateurs primaires, consommateurs secondaires et consommateurs tertiaires aussi. Ça (consommateurs primaires, secondaires et tertiaires) je vais tous les mettre dans le même coin parce que c'est trois, c'est des classes.</p> <p>[Déplace CONSOMMATEURS PRIMAIRES près de CONSOMMATEURS. Déplace CONSOMMATEURS SECONDAIRES. Déplace CONSOMMATEURS TERTIAIRES]</p>

Code	Description.	Exemple
P.7 Déplacer des concepts et des relations en vue d'ajouter d'autres propositions	Le participant déplace un groupe de concepts ET de relations. Il explique déplacer ces composantes en vue de former d'autres propositions.	<p>[Déplace RÉSEAUX TROPHIQUES. Trace un lien entre RÉSEAUX TROPHIQUES et CHAINES ALIMENTAIRES. Formule la relation «est basé».]</p> <p>Relations trophiques. Mais là comment je vais faire pour mettre ça là si je veux dire que ça (décomposeurs), ça pourrait être là, avec ça (champignons et bactéries et cadavres et déchets d'organismes) qui va là ici comme ça. Parfait, si je déplace comme ça ici. Les champignons ici, nourrissent, qui se nourrissent ici, avec cadavres. Parfait, j'ai mon décomposeurs ici. J'ai mon consommateurs ici. Mais ça, mon relation trophique je l'ai là. Hey boy.</p> <p>[Réorganiser: Déplace DÉCOMPOSEURS (et les liens associés). Déplace CHAMPIGNONS ET BACTÉRIES et déplace CADAVRES ET DÉCHETS D'AUTRES ORGANISMES (et les liens associés). Déplace CONSOMMATEURS. Déplace RELATIONS TROPHIQUES.]</p>

Mise en relation

Code	Description.	Exemple
<p>F.1 Formuler une relation entre des concepts</p>	<p>Le participant rédige une expression de liaison entre deux concepts. Note : si le participant ne fait que tracer un lien, il faut utiliser l'un des codes dans la catégorie « Planification ».</p>	<p>[Déplace légèrement le concept ANIMAUX. Éprouve des problèmes techniques. Trace à nouveau un lien entre HÉTÉROTROPHES et ANIMAUX. Formule la relation «constitués»] (Hétérotrophes) Constitués d'animaux qui sont des. Qui sont des consommateurs. Constitués d'animaux. Hum. Constitués d'animaux. [Déplace le concept ANIMAUX à proximité de CONSOMMATEURS. Trace un lien. Formule la relation «qui sont des».]</p>
<p>F.2 Formuler plusieurs relations contiguës</p>	<p>Le participant rédige les expressions de liaison entre plusieurs concepts contiguës. Généralement, le participant exprime les relations entre tous ces concepts avant de commencer à rédiger les expressions de liaison.</p>	<p>[Trace un lien entre CONSOMMATEURS PRIMAIRES et ANIMAUX. Formule la relation «peuvent être».] Les chaînes alimentaires ... sont composées... des... autotrophes et hétérotrophes... [délai] composée par [Retranscrit AUTOTROPHES. Retranscrit HÉTÉROTROPHES. Trace un lien entre CHAINES ALIMENTAIRES et AUTOTROPHES. Formule la relation «composées par». Trace un lien entre CHAINES ALIMENTAIRES et HÉTÉROTROPHES. Formule la relation «composées par»]</p>

Code	Description.	Exemple
F.3 Tracer une accolade	Le participant trace une accolade et exprime qu'un concept A, le concept principal, englobe tous les concepts subordonnés.	<p>[Trace un lien entre HÉTÉROTROPHES et AUTOTROPHES. Formule la relation «dépend»]</p> <p>Alors la chaine alimentaire. Je dirais que c'est ces trois là. Pas ces trois là. Je veux dire, les consommateurs primaires, secondaires et tertiaires avec les végétaux et les décomposeurs, qui sont les bactéries et les champignons. Alors. Je dirais que c'est toute la feuille. Alors je vais faire un petit truc comme cela.</p> <p>[Crée une accolade. Déplace le concept CHAINES ALIMENTAIRES au centre de l'accolade.]</p>

Révision / Relecture

Code	Description.	Exemple
R.1 Relire des propositions	Le participant relit des propositions qui sont déjà inscrites dans la carte.	<p>[Formule la relation «affectent» entre RELATIONS TROPHIQUES et NIVEAUX TROPHIQUES.]</p> <p>Ok on va relire une autre fois pour voir si ça fait du sens. On commence avec le réseau trophique qui est constitué de plusieurs chaînes alimentaires qui sont affectées par les relations trophiques. Et qui, elles affectent les niveaux trophiques. Et les niveaux trophiques sont bien sûr les consommateurs primaires, secondaires et tertiaires. Donc, et ainsi, les niveaux trophiques constituent aussi la chaîne alimentaire. Donc les consommateurs primaires consomment généralement les producteurs, qui eux sont des autotrophes. Les producteurs sont constitués par des végétaux, principalement qui eux font la photosynthèse, ce qui les rend autotrophes, dans le fond.</p> <p>[relisant]</p>

Révision / Modification et Édition

Code	Description.	Exemple
R.2 Corriger une ou plusieurs relations	Le participant change la direction des traits reliant des concepts ou modifie les expressions de liaison entre ces concepts.	[Trace un lien entre DÉCOMPOSEUR et VÉGÉTAUX. Formule la relation «consomme / décompose».] Alors, ici je vais rajouter le mot consomme. Parce que, consomme aussi. Hum. [Modifie un mot-lien entre DÉCOMPOSEURS et VÉGÉTAUX.]
R.3 Corriger l'étiquette d'un concept	Le participant change l'énoncé d'un concept (par ex., parce qu'il l'avait mal inscrit au départ).	[relisant] Oups j'ai oublié d'ajouter chimique. [Modifie le concept D'ANGENDRER LE PROCESSUS CHIMIQUE.]
R.4 Remplacer une ou plusieurs propositions par d'autres propositions	Le participant supprime les relations entre certains concepts et formule de nouvelles relations entre ces concepts ou entre ces concepts et d'autres concepts. Le participant exprime généralement son intention avant de procéder. Il explique ce qui motive ces modifications.	[Trace un lien entre AUTOTROPHES et PHOTOSYNTHÈSE.] Ensuite il faudrait que je dise que autotrophes c'est leur manière de se nourrir par eux-mêmes. Ah. On pourrait aussi «deleter» ça (la relation entre autotrophes et photosynthèse), ensuite enlever le lien entre végétaux et hétérotrophes, pour dire que les producteurs sont végétaux et pratiquent la photosynthèse. Ce qui les considère comme autotrophes. Ce qui les considère... quand on dit «les» on dit les végétaux. (Photosynthèse) Ce qui les considère comme autotrophes. Parfait. [Supprime la relation entre AUTOTROPHES et PHOTOSYNTHÈSE. Supprime la relation entre VÉGÉTAUX et AUTOTROPHES. Trace une relation entre PHOTOSYNTHÈSE et AUTOTROPHES. Formule la relation «ce qui les considèrent».]

Code	Description.	Exemple
R.5 Supprime une ou des relations	Le participant supprimer une ou des relations et exprime les raisons qui motivent cette action.	[Elle déplace le concept CONSOMMATEURS SECONDAIRES à côté du concept CONSOMMATEURS PRIMAIRES.]
		Eh. Niveaux trophiques. Peut-être faire un autre lien.
		[Supprime le lien entre NIVEAUX TROPHIQUES et CONSOMMATEURS PRIMAIRES]

Révision / Évaluation et Diagnostic

Code	Description.	Exemple
R.6 Évaluer le contenu de la carte	Le participant donne son appréciation au sujet du contenu de la carte.	<p>[Trace un lien entre CHAMPIGNONS ET BACTÉRIES et CADAVRES ET DECHETS. Efface partiellement. Formule la relation «qui décomposent»]</p> <p>C'est le seul qui soit différent des autres. Autour. Cadavres et déchets, et les champignons décomposent. Les décomposeurs sont eh. Sont tous reliés ensemble. Je suis en train de faire des liens un peu moins perturbants. Peut-être. hum.</p> <p>[Aucune action]</p>

Code	Description.	Exemple
R.7 Chercher des améliorations potentielles.	Le participant mentionne qu'il cherche des améliorations potentielles et exprime des changements qu'il pourrait apporter à la carte.	<p>[Trace un lien entre CHAINES ALIMENTAIRES et RÉSEAUX TROPHIQUES: Formule la relation «composée de».]</p> <p>Donc là si regarde pour faire d'autres liens entre les différents concepts. Hum, C'est sûr, les consommateurs et les décomposeurs sont étroitement liés mais pas nécessairement avec un mot. Et c'est plus tard dans. Animaux qui produisent ça. Champignons mangent ça. Champignons et bactéries et animaux sont hétérotrophes. Il n'y a rien d'autre d'hétérotrophe. Peut-être qu'on pourrait dire que non. On ne peut pas dire que les consommateurs et décomposeurs sont hétérotrophes. On spécifie déjà que eux ils le sont. Eh, végétaux, sont autotrophes. Ils utilisent la photosynthèse. Il n'y a rien d'autre de relier à photosynthèse ici. Hum. On pourrait peut-être dire que les cadavres et déchets aident les végétaux lorsqu'ils sont mangés. Hum, non. Eh. Consommateurs ici. Ça prend des S. Donc, animaux, ça dépend des animaux. Ils ne sont pas nécessairement primaires, secondaires ou tertiaires. Ils peuvent être les trois, dépendant du type d'animal. Consommateurs se divisent en. Ça reste comme ça. Les trois sont des niveaux trophiques. Hum. Est-ce qu'on peut relier chaînes alimentaires avec autres choses. Peut-être niveaux trophiques. Non c'est déjà relié à relations. Réseaux trophiques. Ça peut juste être associé à relations trophiques. Eh. J'ai pas d'autres liens à faire. J'ai pas mal fini.</p>
R.8 Vérifier le nombre de concepts	Le participant fait le compte des concepts utilisés dans la carte et vérifie que tous les concepts de la liste sont utilisés.	<p>[reaisant]</p> <p>[Aucune action]</p> <p>Je crois que oui. Ah attendez. Un deux trois quatre cinq six sept huit neuf dix onze douze treize quatorze quinze seize dix-sept. Ok.</p> <p>[Compte les concepts utilisés]</p>

Annexe K. Certificat d'approbation éthique



Comité plurifacultaire d'éthique de la recherche

31 janvier 2014

Monsieur Maxim Morin
Candidat au doctorat
Administration et fondements de l'éducation - Faculté des sciences de l'éducation

OBJET: Reconnaissance d'une approbation éthique

M. Maxim Morin,


Le Comité plurifacultaire d'éthique de la recherche (CPER) a étudié le projet de recherche intitulé « Expérimentation de la cartographie conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages » et a délivré le certificat d'éthique demandé suite à la satisfaction des exigences précédemment émises.

Notez qu'il y apparaît une mention relative à un suivi annuel et que le certificat comporte une date de fin de validité. En effet, afin de répondre aux exigences éthiques en vigueur au Canada et à l'Université de Montréal, nous devons exercer un suivi annuel auprès des chercheurs et étudiants-chercheurs.

De manière à rendre ce processus le plus simple possible et afin d'en tirer pour tous le plus grand profit, nous avons élaboré un court questionnaire qui vous permettra à la fois de satisfaire aux exigences du suivi et de nous faire part de vos commentaires et de vos besoins en matière d'éthique en cours de recherche. Ce questionnaire de suivi devra être rempli annuellement jusqu'à la fin du projet et pourra nous être retourné par courriel. La validité de l'approbation éthique est conditionnelle à ce suivi. Sur réception du dernier rapport de suivi en fin de projet, votre dossier sera clos.

Il est entendu que cela ne modifie en rien l'obligation pour le chercheur, tel qu'indiqué sur le certificat d'éthique, de signaler au CPER tout incident grave dès qu'il survient ou de lui faire part de tout changement anticipé au protocole de recherche.

Nous vous prions d'agréer, Monsieur, l'expression de nos sentiments les meilleurs,


Pierre Lapointe, Président
Comité plurifacultaire d'éthique de la recherche (CPER)
Université de Montréal

PL/RS/rs

c.c. Gestion des certificats, BRDV
Jean-Guy Blais, Professeur titulaire, Administration et fondements de l'éducation - Faculté des sciences de l'éducation
Lucie Lefrançois
p.j. Certificat CPER-13-134-D

3744 Jean-Brillant, B-430-8
C.P. 6128, succ. Centre-ville
Montréal QC H3C 3J7

Téléphone : 514-343-6111 poste 1896
cper@umontreal.ca
<http://www.recherche.umontreal.ca/ethique-de-la-recherche/les-comites/les-comites-sectoriels-de/ethique-de-la-recherche/cper/>

Comité plurifacultaire d'éthique de la recherche

CERTIFICAT D'APPROBATION ÉTHIQUE

Le Comité plurifacultaire d'éthique de la recherche (CPER), selon les procédures en vigueur, en vertu des documents qui lui ont été fournis, a examiné le projet de recherche suivant et conclu qu'il respecte les règles d'éthique énoncées dans la Politique sur la recherche avec des êtres humains de l'Université de Montréal.

Projet	
Titre du projet	Expérimentation de la cartographie conceptuelle comme dispositif de collecte de données en vue de l'évaluation des apprentissages
Étudiant requérant	Maxim Morin Candidat au doctorat, Administration et fondements de l'éducation - Faculté des sciences de l'éducation Université de Montréal
Financement	
Organisme	Non financé
Programme	--
Titre de l'octroi si différent	--
Numéro d'octroi	--
Chercheur principal	--
No de compte	--
Approbation reconnue	
Approbation émise par	non
Certificat:	s.o.

MODALITÉS D'APPLICATION

Tout changement anticipé au protocole de recherche doit être communiqué au CPER qui en évaluera l'impact au chapitre de l'éthique.

Toute interruption prématurée du projet ou tout incident grave doit être immédiatement signalé au CPER.

Selon les règles universitaires en vigueur, un suivi annuel est minimalement exigé pour maintenir la validité de la présente approbation éthique, et ce, jusqu'à la fin du projet. Le questionnaire de suivi est disponible sur la page web du CPER.



Pierre Lapointe, Président
Comité plurifacultaire d'éthique de la recherche
Université de Montréal

31 janvier 2014
Date de délivrance

1 février 2015
Date de fin de validité

adresse postale
3744 Jean-Brillant, B-430-8
C.P. 6128, succ. Centre-ville
Montréal QC H3C 3J7
www.cper.umontreal.ca

Téléphone : 514-343-6111 poste 1896
cper@umontreal.ca

Annexe L. Illustrations des difficultés rencontrées par les participants lors de la production d'une carte conceptuelle

Premier exemple - Illustration d'une difficulté liée à la création d'une bulle vide

VERBATIM

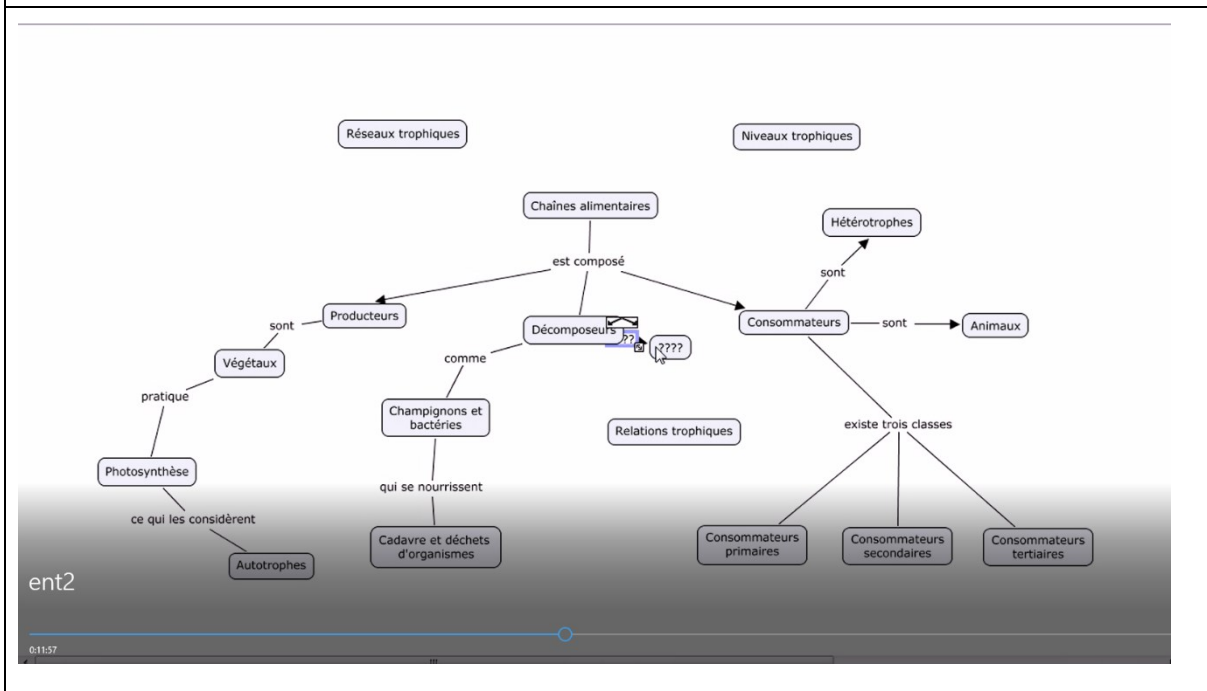
[Supprime la relation entre CONSOMMATEURS et RELATIONS TROPHIQUES.]

On va mettre que les décomposeurs, eh ben. Ok ça marche pas. On va faire les décomposeurs et les consommateurs sont en relations trophiques. Celui là (une bulle vide) on va juste l'enlever.

[Crée une bulle. Supprime la bulle. Crée une bulle. Supprime la bulle. Crée une bulle et Trace une relation avec DÉCOMPOSEURS. Trace un lien entre CONSOMMATEURS et la bulle. Formule la relation « sont ». Inscrit RELATIONS TROPHIQUES dans la bulle.]
 [Supprime l'autre bulle contenant le concept RELATIONS TROPHIQUES.]

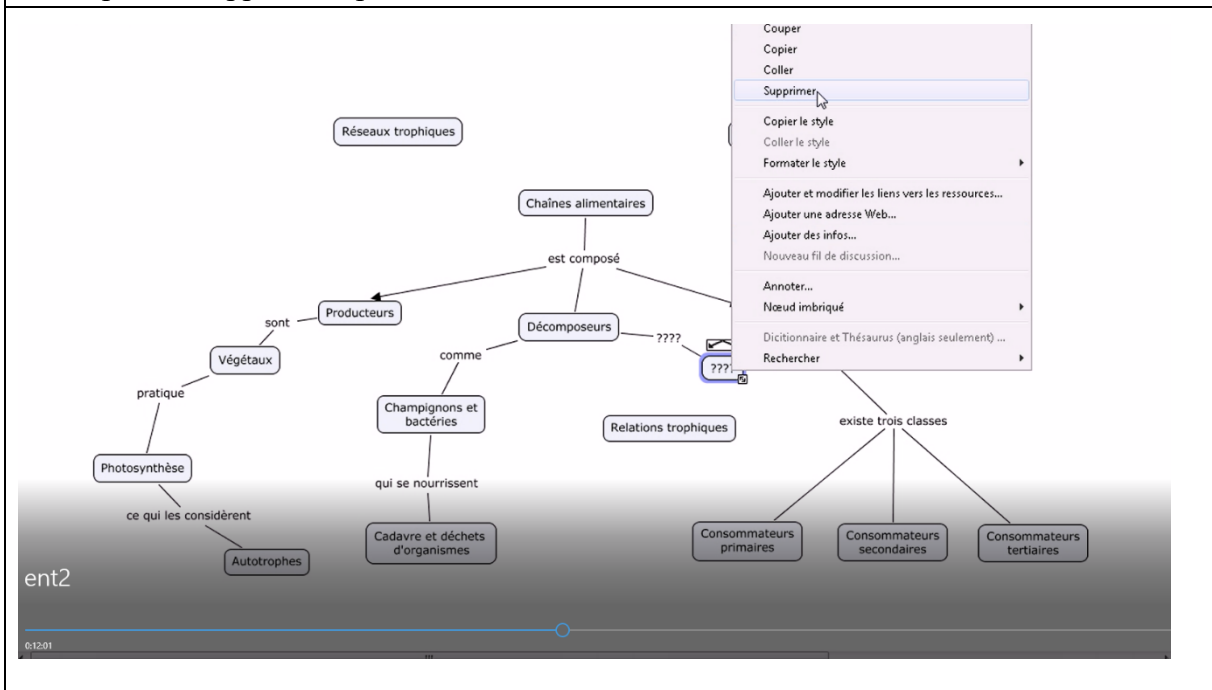
Chronomètre : 11 min 57 s

Description : Début de la séquence d'actions. Crée de la première bulle vide



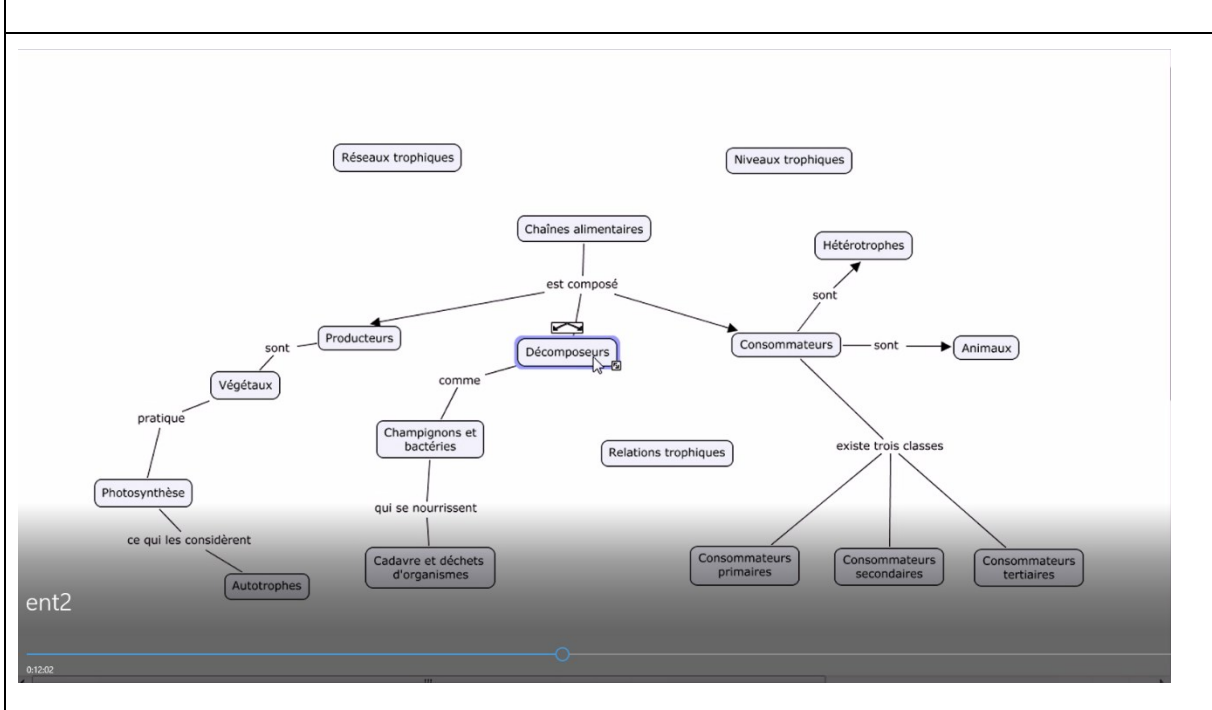
Chronomètre : 12 min 01 s

Description : Supprime la première bulle vide.



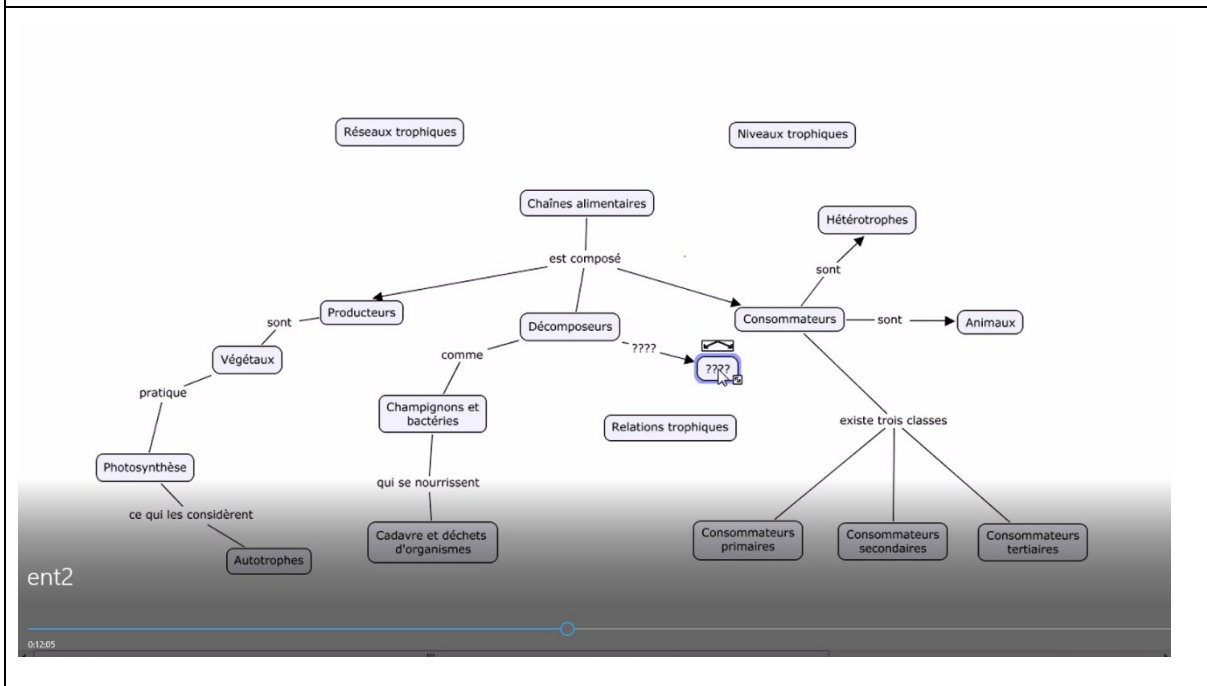
Chronomètre : 12 min 02 s

Description : Première bulle vide supprimée.



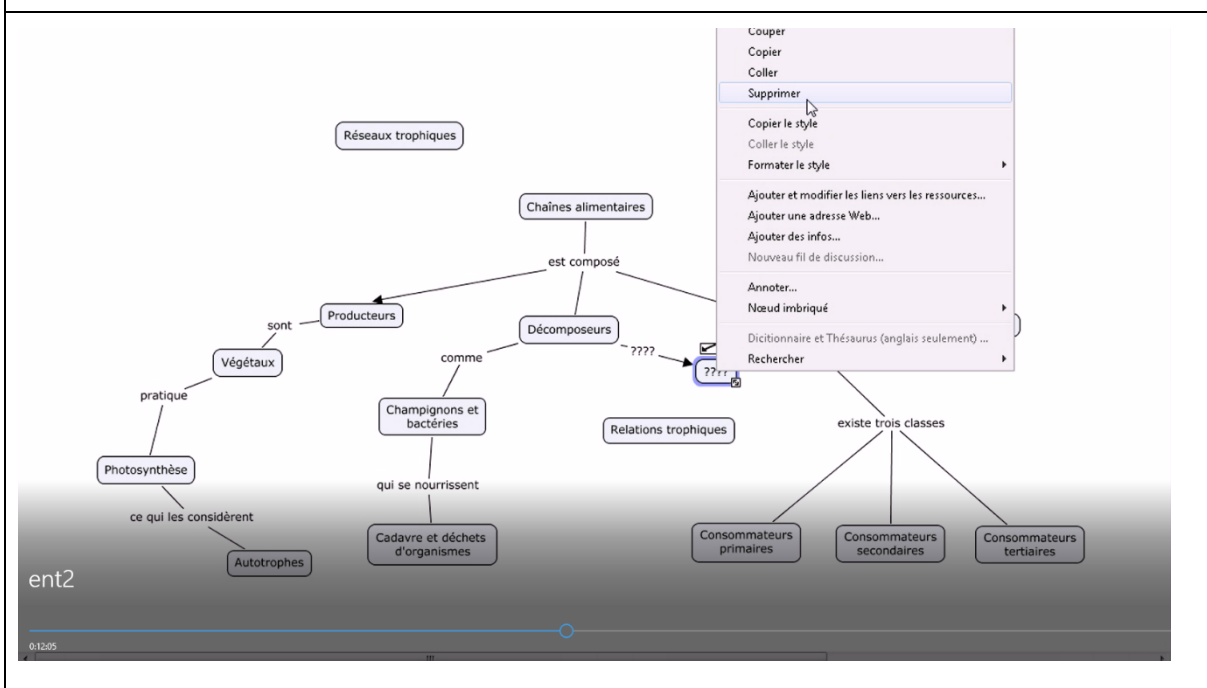
Chronomètre : 12 min 05 s

Description : Crée de la deuxième bulle vide



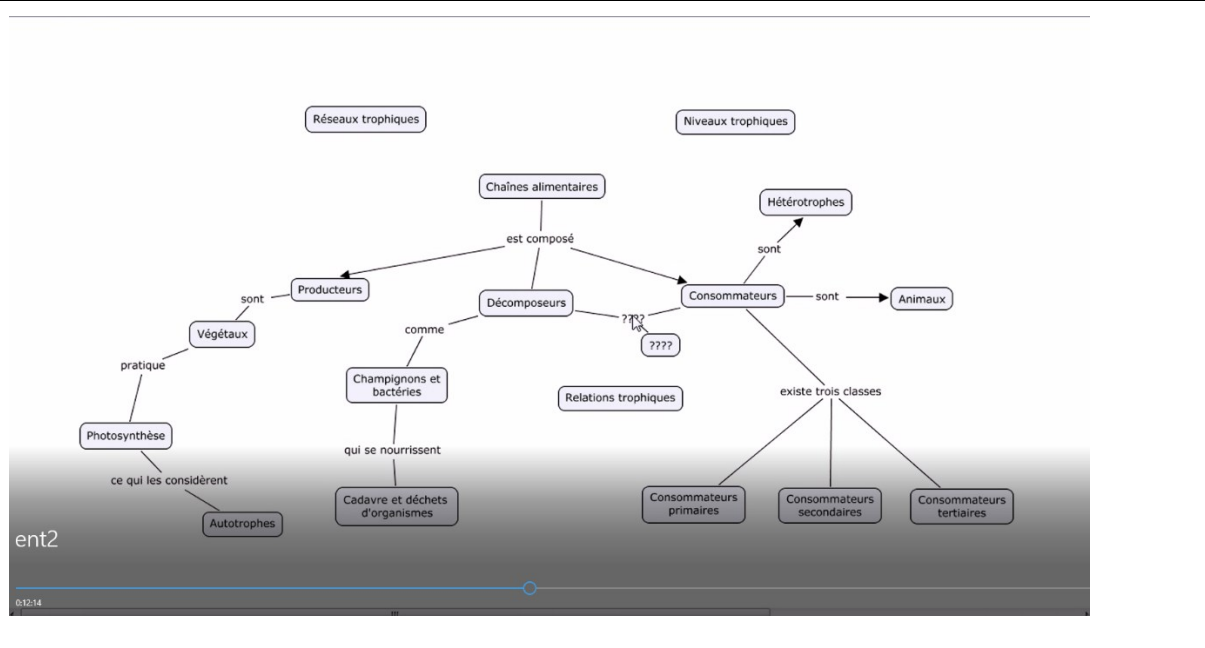
Chronomètre : 12 min 05 s

Description : Supprime la deuxième bulle vide.



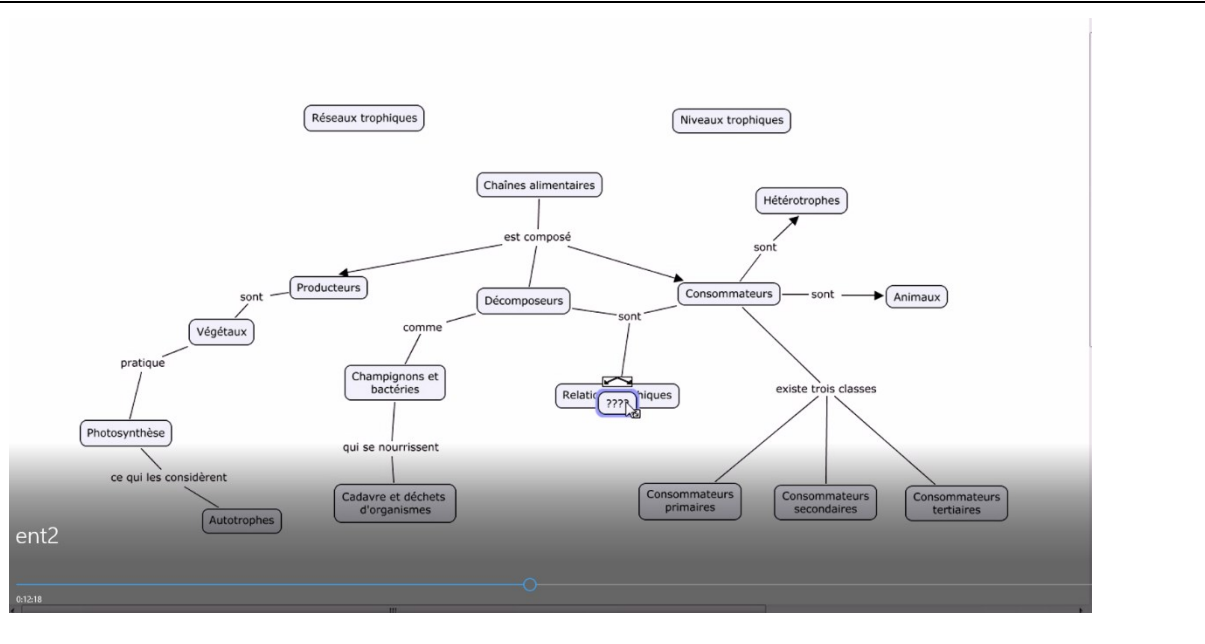
Chronomètre : 12 min 14 s

Description : Crée une bulle et Trace une lien avec DÉCOMPOSEURS. Trace un lien entre CONSOMMATEURS et la bulle.



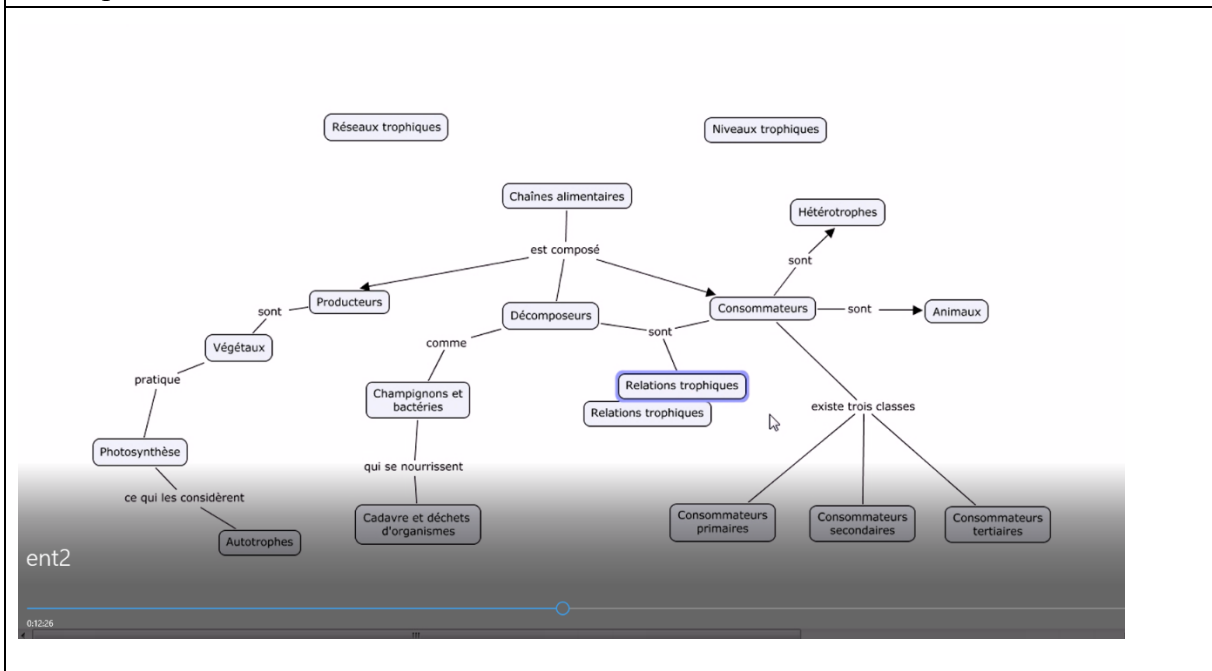
Chronomètre : 12 min 18 s

Description : Formule la relation « sont »



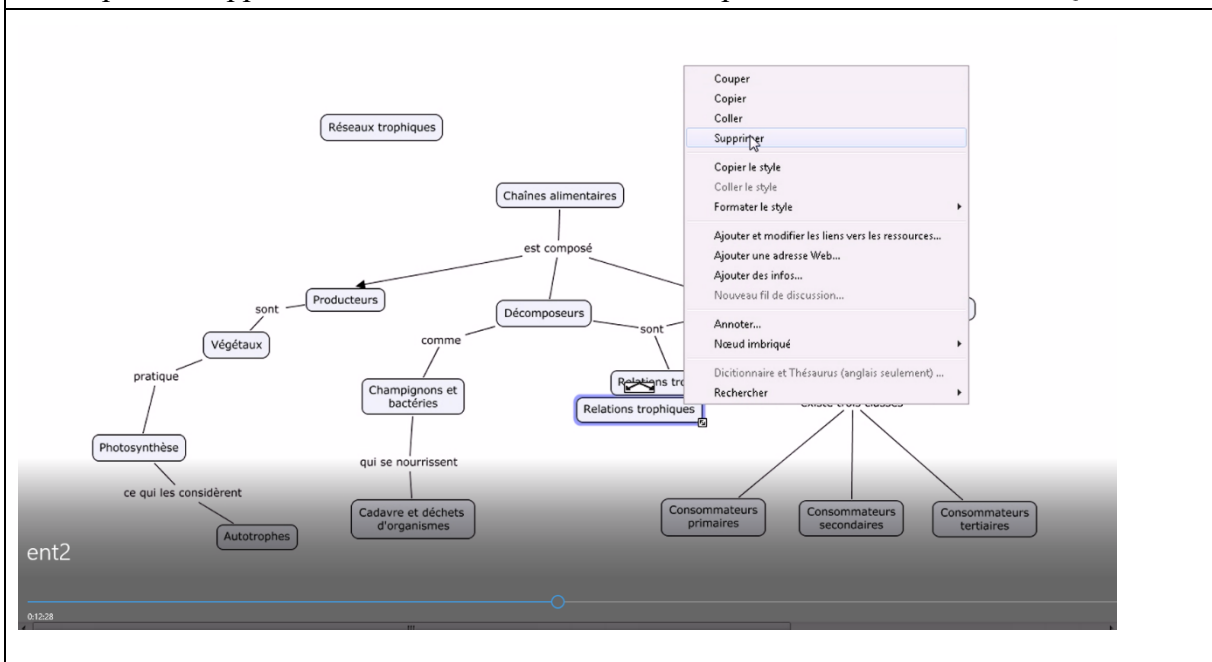
Chronomètre : 12 min 26 s

Description : Inscrit RELATIONS TROPHIQUES dans la bulle.



Chronomètre : 12 min 28 s

Description : Supprime l'autre bulle contenant le concept RELATIONS TROPHIQUES.



Deuxième exemple - Illustration d'une difficulté liée à la méconnaissance du logiciel

VERBATIM

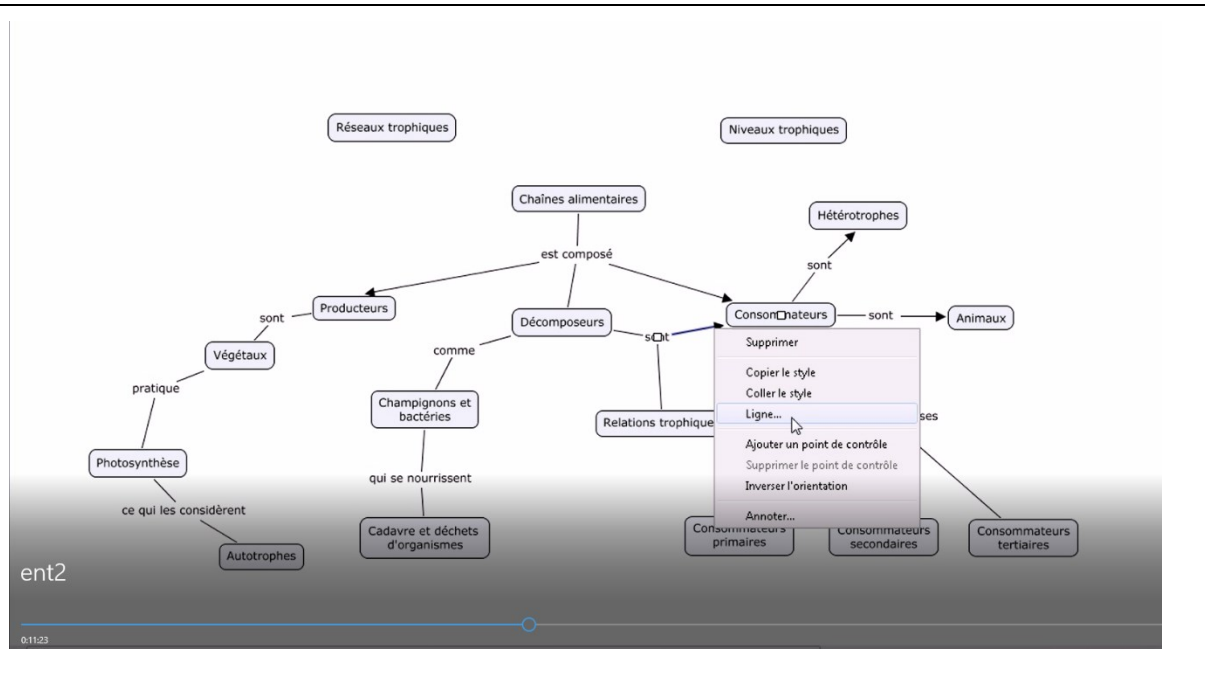
[Crée une bulle. Supprime la bulle. Trace un lien entre DÉCOMPOSEURS et CONSOMMATEURS. Formule la relation « sont ».]

on va juste avancer la flèche pour pas qu'on la voit. Ligne. Forme. Non. Style. On veut juste que la ligne soit droite et que ça soit pas une flèche. Police. Op. On va juste faire. Faut que je sois de l'autre côté. On va faire consommateurs [trace un lien entre consommateurs et relations trophiques]. Et voilà.

[Éprouve des difficultés. Cherche dans le menu de styles. Trace une relation entre CONSOMMATEURS et RELATIONS TROPHIQUES.]

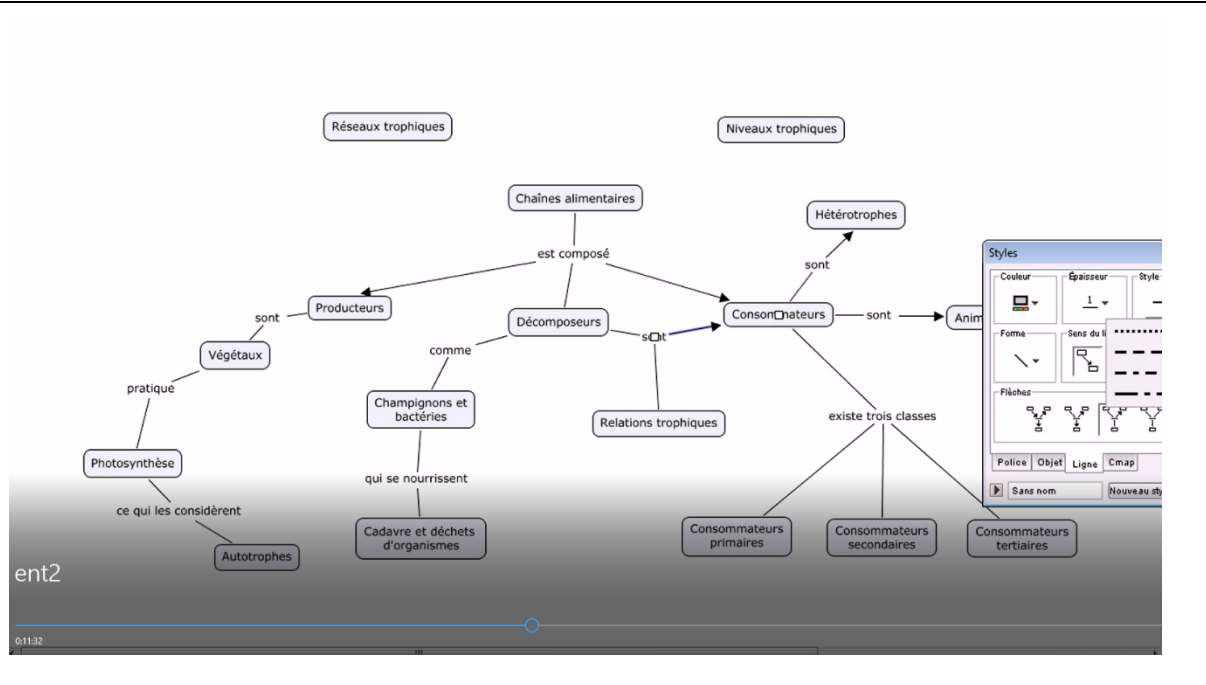
Chronomètre : 11 min 23 s

Description : Recherche dans le menu de styles



Chronomètre : 11 min 32 s

Description : Recherche toujours dans le menu de styles



Troisième exemple - Illustration d'une difficulté liée à la méconnaissance du logiciel

VERBATIM

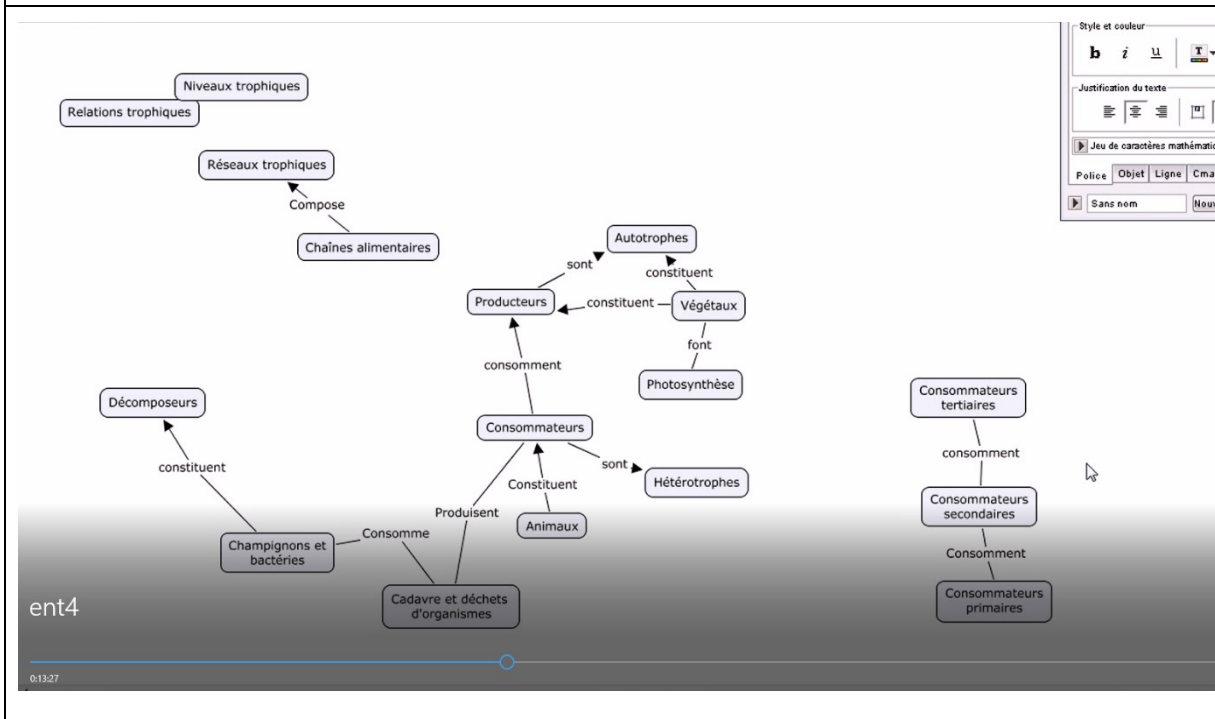
[Trace un lien entre CONSOMMATEURS TERTIAIRES et CONSOMMATEURS SECONDAIRES. Formule la relation « consomment »]

Donc, je sais que des décomposeurs peuvent être à n'importe quel niveau là-dessus. Donc à n'importe quel niveau ils (les décomposeurs) peuvent consommer les cadavres ou déchets. Ah. Ok. Je vais juste faire un peu d'espace. Donc je vais les bouger un peu vers la gauche.

[Crée de l'espace. Déplace le groupe de CONSOMMATEURS TERTIAIRES. Éprouve des difficultés techniques.]

Chronomètre : 13 min 27 s

Description : Début de la séquence.



Chronomètre : 13 min 30 s

Description : Sélectionne un groupe de concepts et de mots de liaison

The screenshot shows a concept map titled "ent4" with a timer at 0:13:30. The map illustrates trophic levels and relationships. Key concepts include "Niveaux trophiques", "Relations trophiques", "Réseaux trophiques", "Chaînes alimentaires", "Autotrophes", "Producteurs", "Végétaux", "Photosynthèse", "Consommateurs", "Hétérotrophes", "Animaux", "Décomposeurs", "Champignons et bactéries", and "Cadavre et déchets d'organismes". Relationships are indicated by arrows with labels like "sont", "constituent", "consomme", "produisent", "font", and "consomment". A vertical stack of concepts on the right is highlighted with a dashed box: "Consommateurs tertiaires", "Consommateurs secondaires", "Consommateurs primaires", with connecting arrows labeled "consomment". A text editor toolbar is visible in the top right corner.

Chronomètre : 13 min 32 s

Description : Déplace le groupe d'objets sur les autres objets positionnés au centre de la page.

This screenshot shows the same concept map as above, but with a group of concepts moved to the center. The group, highlighted with a dashed box, includes "Consommateurs tertiaires", "Consommateurs secondaires", and "Consommateurs primaires". The central placement of this group is intended to show its relationships with other central concepts like "Producteurs", "Végétaux", "Animaux", and "Hétérotrophes". The timer now shows 0:13:32.

Chronomètre : 13 min 37 s

Description : Tente de sélectionner le groupe d'objets à nouveau.

ent4

0:13:37

Chronomètre : 13 min 40 s

Description : Déplace seulement certains objets du groupe, dont le mot de liaison « constituant » qui se trouvait entre les concepts « Végétaux » et « Producteurs ».

ent4

0:13:40

Chronomètre : 13 min 44 s

Description : Remplace le mot de liaison « constituent » à sa position initiale.

ent4
0:13:44

Chronomètre : 13 min 47 s

Description : Déplace l'un des concepts (Consommateurs Tertiaires) qui n'a pas été pris dans le groupe précédent.

ent4
0:13:47

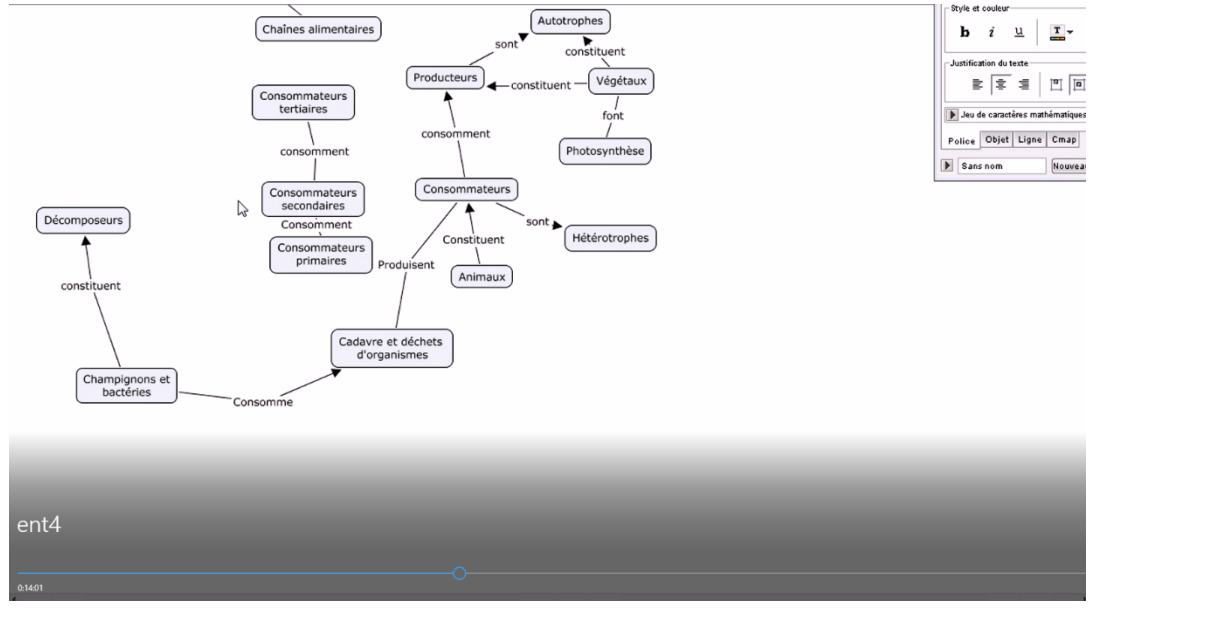
Chronomètre : 13min 49 s

Description : Déplace le mot de liaison entre « Consommateurs secondaires » et « Consommateurs tertiaires ».

The screenshot displays a software interface for a trophic network diagram. The diagram is a hierarchical flowchart showing the relationships between different trophic levels. At the top, 'Niveaux trophiques' and 'Relations trophiques' are shown. Below them, 'Réseaux trophiques' leads to 'Chaînes alimentaires'. The main part of the diagram shows 'Autotrophes' (Producers) which 'constituent' (constitute) 'Végétaux' (Plants). 'Végétaux' 'font' (perform) 'Photosynthèse' (Photosynthesis). 'Producteurs' (Producers) 'sont' (are) 'Autotrophes'. 'Producteurs' 'constituent' (constitute) 'Consommateurs' (Consumers). 'Consommateurs' 'consomment' (consume) 'Producteurs'. 'Consommateurs' are divided into 'Consommateurs primaires' (Primary consumers), 'Consommateurs secondaires' (Secondary consumers), and 'Consommateurs tertiaires' (Tertiary consumers). 'Consommateurs primaires' 'produisent' (produce) 'Cadavre et déchets d'organismes' (Organism carcasses and waste). 'Consommateurs secondaires' 'consomment' (consume) 'Consommateurs primaires'. 'Consommateurs tertiaires' 'consomment' (consume) 'Consommateurs secondaires'. 'Décomposeurs' (Decomposers) 'constituent' (constitute) 'Champignons et bactéries' (Mushrooms and bacteria). 'Champignons et bactéries' 'consomme' (consume) 'Cadavre et déchets d'organismes'. 'Hétérotrophes' (Heterotrophs) 'sont' (are) 'Consommateurs'. 'Hétérotrophes' 'constituent' (constitute) 'Consommateurs'. 'Hétérotrophes' 'produisent' (produce) 'Cadavre et déchets d'organismes'. 'Animaux' (Animals) 'constituent' (constitute) 'Consommateurs'. 'Animaux' 'produisent' (produce) 'Cadavre et déchets d'organismes'. A text editor toolbar is visible on the right side of the interface, with options for 'Style et couleur', 'Justification du texte', 'Jeu de caractères mathématiques', 'Police', 'Objet', 'Ligne', 'Cmap', 'Sans nom', and 'Nouveau'. The interface also shows a timer at the bottom left indicating 13:49.

Chronomètre : 14 min 01 s

Description : Créé de l'espace en déplaçant les concepts « Décomposeurs » et « Champignons et bactéries ».



Annexe M. Représentation des démarches mobilisées par les participants lors des séances de verbalisation de la pensée

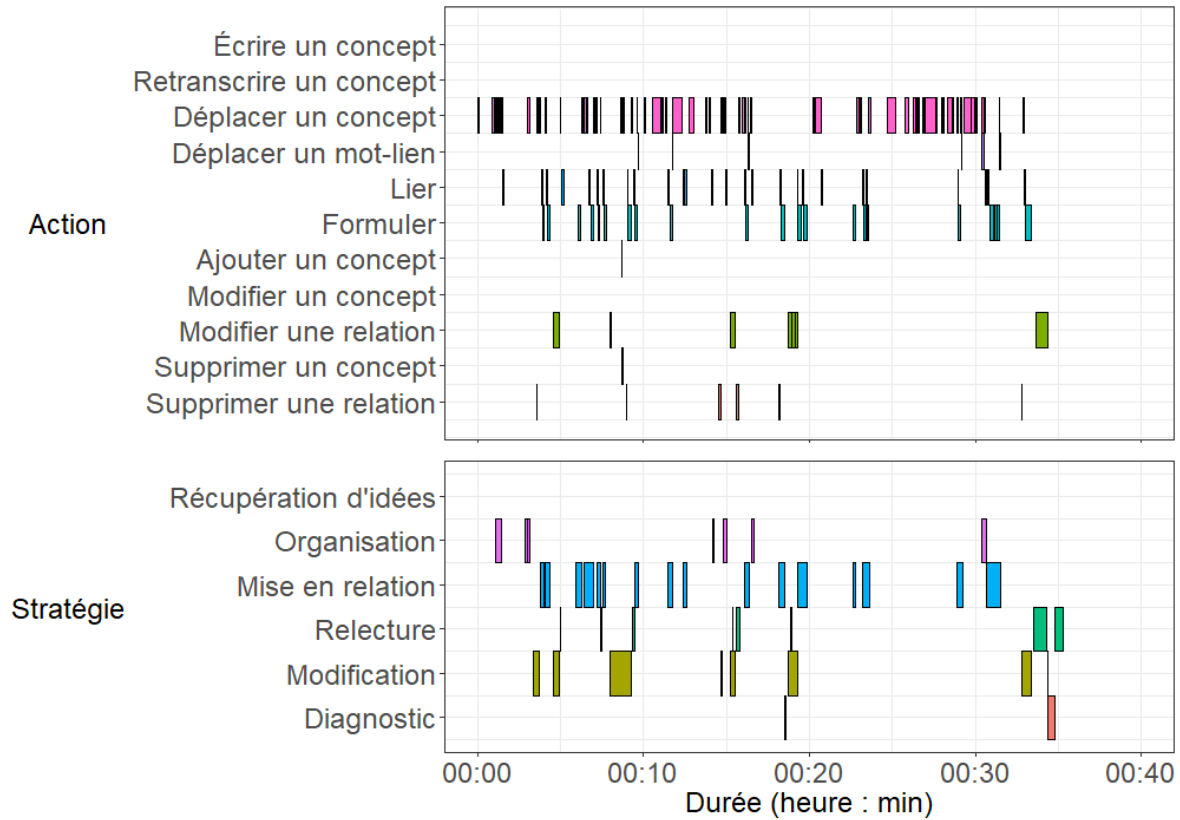


Figure 22. Démarche du participant 1

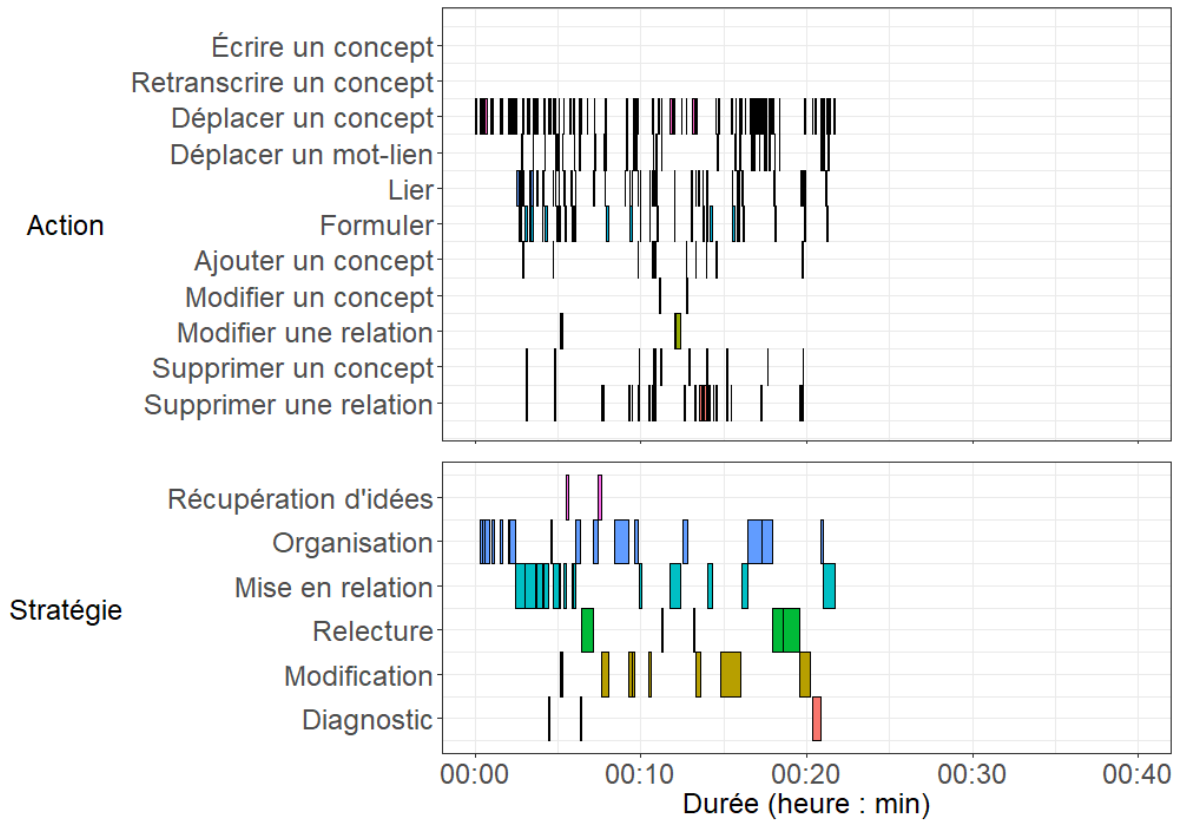


Figure 23. Démarche du participant 2

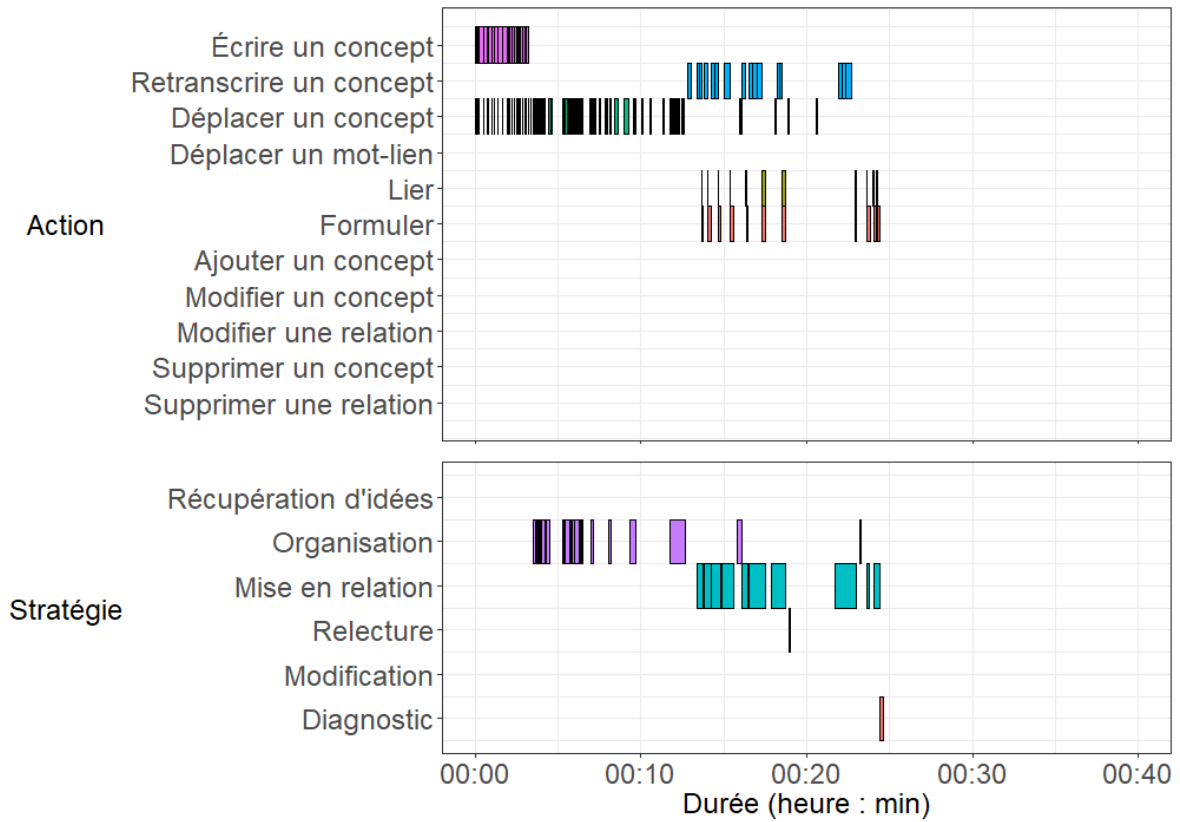


Figure 24. Démarche du participant 5

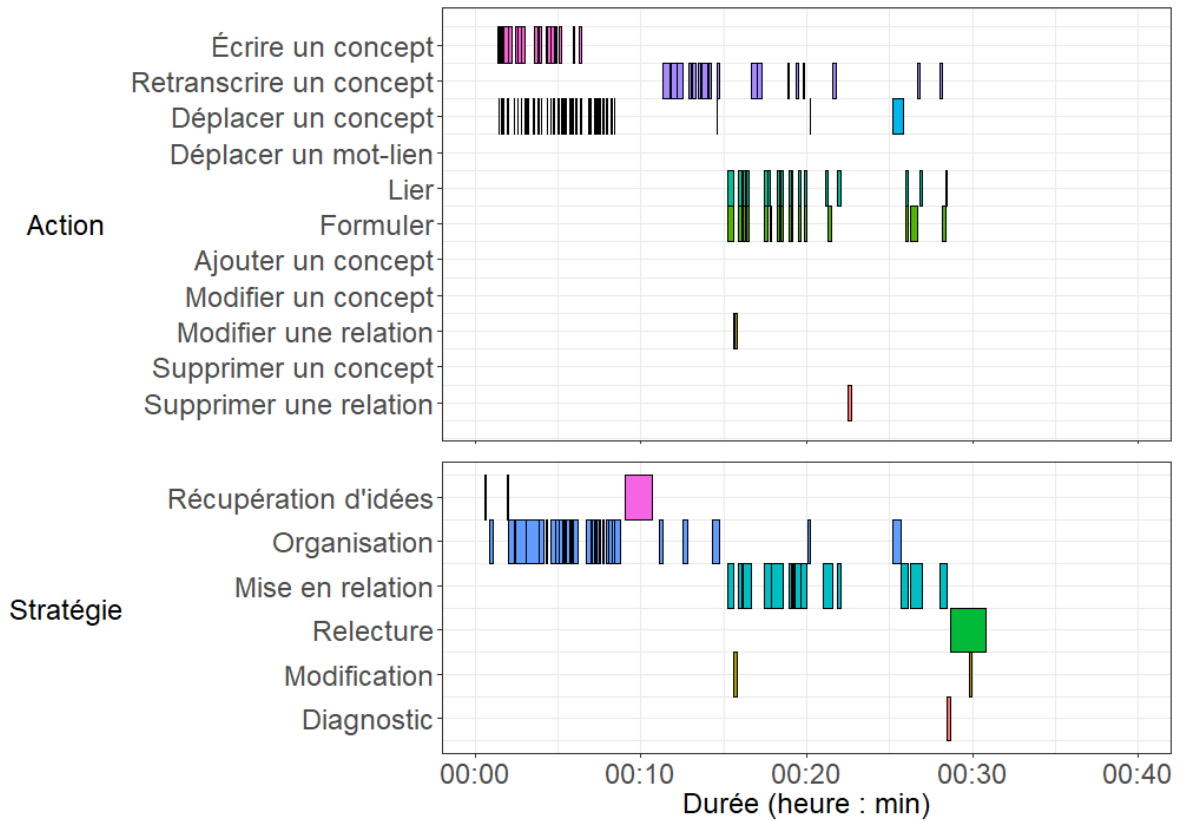


Figure 25. Démarche du participant 6

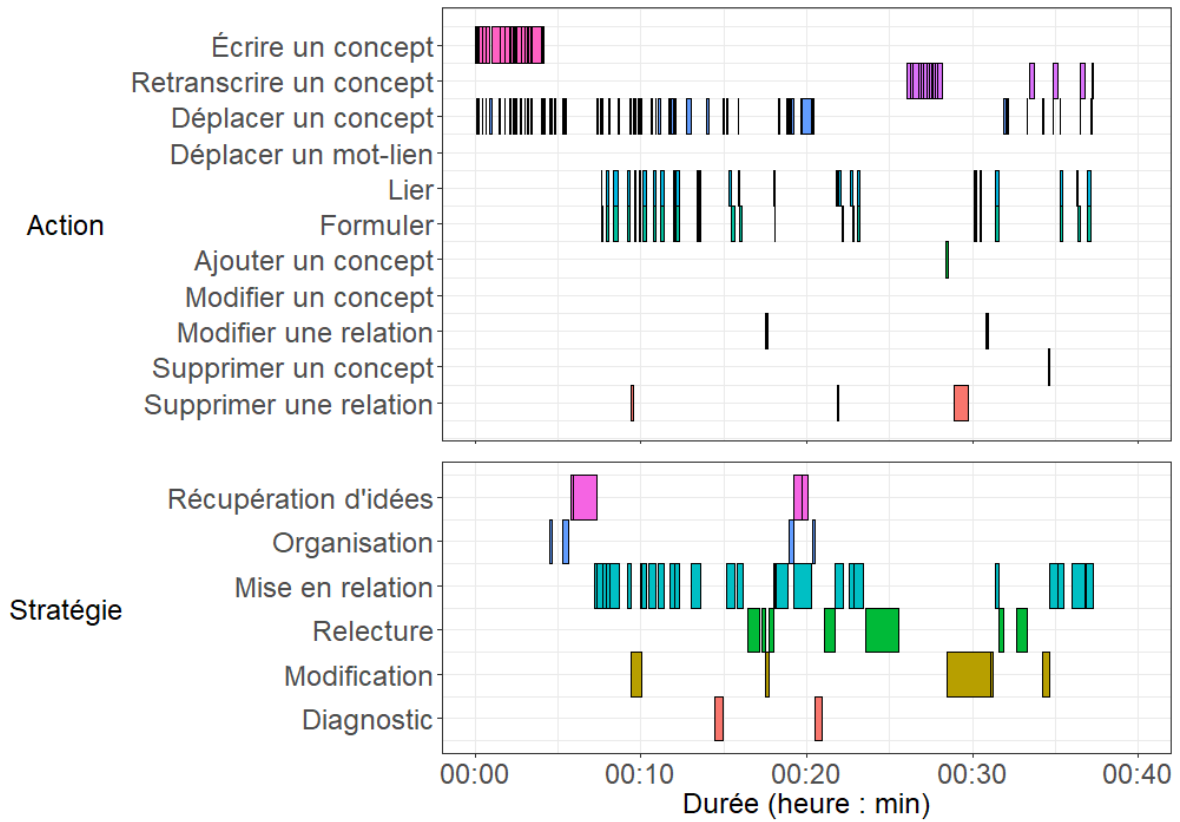


Figure 26. Démarche du participant 8

Annexe N. Concepts et synonymes acceptés lors de la production des cartes des références

Tableau XXXIX. Concepts et synonymes acceptés pour le corpus sur l'effet de serre

<i>Concepts</i>	<i>Synonymes acceptés</i>
Chaleur du Soleil	Énergie émise par le Soleil, énergie solaire
Couche d'ozone	Ozone stratosphérique
Dioxyde de carbone	CO ₂
Énergie thermique de la Terre	Chaleur de la Terre
Gaz à effet de serre	GES
Lumière du Soleil	Rayonnement solaire, rayons du Soleil, rayon solaire, lumière solaire
Processus naturel	
Rayon infrarouge	Rayonnement infrarouge, infrarouge, IR
Rayon ultraviolet	Rayonnement ultraviolet, ultraviolet, UV
Surface de la Terre	Surface terrestre, sol
Température de la Terre	Température moyenne de la Terre, température sur la Terre

Tableau XL. Concepts et synonymes acceptés pour le corpus sur les relations trophiques

<i>Concepts</i>	<i>Synonymes acceptés</i>
Animal	
Chaîne alimentaire	Chaîne trophique
Champignon et bactérie	Champignon, bactérie
Consommateur	Consommateur
Consommateur primaire	Consommateur de premier ordre
Consommateur secondaire	Consommateur de second ordre
Consommateur tertiaire	Consommateur de troisième ordre
Déchet et cadavre d'autres organismes	Détritus
Décomposeur	
Organisme autotrophe	Autotrophe
Organisme hétérotrophe	Hétérotrophe
Niveau trophique	
Photosynthèse	
Producteur	
Relation trophique	
Réseau trophique	Réseau alimentaire
Végétal	Plante