

David Kovács :
COLLOCATIONS EN FRANÇAIS ET EN HONGROIS
Thèses

Directeurs de thèse : Jean-Pierre Goudaillier et Péter Balogh

Introduction

L'étude du phénomène de collocation constitue un domaine important de la lexicologie et de la lexicographie depuis quelques dizaines d'années. Les recherches concernant le sujet se divisent essentiellement en deux blocs : certaines s'appuient sur des considérations fréquentielles (approche majoritairement anglo-saxonne), d'autres sur des observations de nature sémantique (approche majoritairement francophone). En ce qui concerne notre recherche, elle ne prend en considération aucun facteur fréquentiel, mais elle s'intègre entièrement dans la tradition linguistique francophone.

En étant opposé à l'approche sémantique traditionnelle du phénomène en question, nous avons jugé indispensable de reconsidérer la question du phénomène collocatif dans le but d'établir une classification plus précise et plus rigoureuse qui peut servir à donner une définition (au moins partiellement) différente de celles que la littérature a proposées jusqu'à présent. Nos arguments contre la classification traditionnelle des cooccurrences lexicales sont les suivants :

1. manque de l'analyse de sens précis des termes dans la collocation ;
2. substituabilité restreinte qui ne s'explique pas seulement par les habitudes langagières mais aussi par les propriétés inhérentes des éléments du lexique et des rapports lexicaux qui les relie ;
3. illégitimité de l'aspect bilingue dans la classification ;
4. insuffisance des considérations sur la compositionnalité de l'expression ;
5. caractère non discriminatoire de la description du statut sémantique différencié et des propriétés syntaxiques des éléments d'une collocation.

Notre classification a été conçue dans le cadre théorique de la théorie Sens-Texte. Selon la TST, toute langue naturelle est une sorte de système de correspondance qui, au moyen de différentes composantes et procédures, lie un sens donné à un énoncé (sens → texte) et inversement, un énoncé à un sens (texte → sens). Bien que le passage Sens → Texte soit plus essentiel selon le deuxième postulat de la théorie, nous nous sommes appuyés dans notre recherche sur le passage de la compréhension puisque notre tâche était de déterminer le type

d'une expression qui a été préalablement produite. Dans notre travail, nous avons procédé à cette analyse à l'aide du paraphrasage et de la décomposition sémantique de l'énoncé afin d'obtenir ainsi les sémantèmes qui le constituent au niveau de la représentation sémantique, et qui établissent ainsi sa structure sémantique.

A part ce modèle linguistique, nous nous sommes également appuyés dans notre travail sur la théorie des classes d'objets. Ce cadre méthodologique vise à catégoriser le lexique selon les cooccurrences des mots avec des prédicats avant tout verbaux, mais adjectivaux et nominaux aussi. Selon la TCO, ce dernier peut être général ou approprié selon qu'il se combine avec le nom à cause de l'appartenance de ce nom à une hyperclasse (opérateur général) ou à une classe d'objet (opérateur approprié).

Le point de départ de nos observations était toujours une phrase élémentaire que nous avons obtenue majoritairement des textes de journaux (*HVG* pour le hongrois et *Libération* pour le français). Etant donné que les limites de nos phrases tirées d'articles ont généralement largement dépassé le cadre d'une phrase élémentaire, nous avons recouru chaque fois à la simplification de ces phrases complexes ou simples étendues.

Thèses

1. Afin de donner une définition plus juste du phénomène des collocations, nous avons d'abord conçu une nouvelle méthode d'analyse qui consiste en une sorte d'algorithme séquentiel dans la mesure où celui-ci, dans le sens très général et populaire du terme, est un énoncé d'une suite d'opérations permettant de donner la réponse à un problème que pose, dans notre cas, la classification sémantique des séquences verbales. A la différence des catégorisations traditionnelles qui appliquent leurs critères classificatoires d'une manière linéaire et simultanée, notre méthode de classification parvient à cette fin d'une manière séquentielle.
2. Notre algorithme se constitue de deux blocs, pose six questions et range les séquences dans sept catégories selon les réponses négatives ou affirmatives :

PREMIER BLOC DE L'ALGORITHME : ANALYSE DE LA COMPOSITIONNALITE

- Est-il possible d'isoler le sens du nom du sens intégral de la séquence ? → Réponse négative → catégorie °1 : SNC A NOM NON COMPOSITIONNEL
- Est-il possible d'isoler le sens du verbe du sens intégral de la séquence ? → Réponse négative → catégorie °2 : SNC A PREDICAT NON COMPOSITIONNEL

- Le sens intégral de la séquence est-il égal à la somme des sens de ses constituants ? → Réponse négative → catégorie °3 : SNC A SURPLUS DE SENS

SECOND BLOC DE L'ALGORITHME : ANALYSE DES OPERATEURS

- Le prédicat impose-t-il de restriction sémantique au complément d'objet ? → Réponse négative → catégorie °4 : SC SANS SPECIFICATION
 - Le prédicat est-il général de l'hyperclasse à laquelle le complément d'objet appartient ? → Réponse affirmative → catégorie °5 : SC A OPERATEUR GENERAL
 - Le prédicat est-il approprié de la classe d'objet à laquelle le complément d'objet appartient ? → Réponse affirmative → catégorie °6 : SC A OPERATEUR APPROPRIE / → Réponse négative → catégorie °7 : SC A OPERATEUR SPECIFIQUE
3. Comme ce schéma le montre, la catégorie collocationnelle proprement dite ne figure pas dans notre classification. La raison pour laquelle nous refusons finalement d'introduire dans notre classification cette catégorie en tant que catégorie à part est que nous considérons que le terme de collocation ne peut être appliqué qu'à un sous-ensemble des séquences verbales qui appartiennent toutes à la catégorie des séquences compositionnelles à opérateur spécifique.
 4. Nous définissons la compositionnalité de la façon suivante : une séquence est compositionnelle si chacun de ses constituants possède un sens autonome, faisant partie du sens de la séquence, et ce dernier sens ne se constitue que des sens des lexèmes qu'elle comprend, ni plus, ni moins. Quant à la notion du sens autonome, nous le définissons ainsi : un sens est autonome s'il est possible d'une part de l'isoler du sens intégral de la séquence et de lui attribuer un lexème constitutif, et d'autre part si ce sens, sous forme du lexème qui lui a été attribué, apparaît dans d'autres séquences de sens différent.
 5. Si un élément de l'expression a un sens autonome (relativement) proche de celui qu'on pourrait isoler dans l'expression, nous proposons l'analyse suivante :
 - Définition de la structure sémantique préliminaire de la séquence, avec les deux sémantèmes complexes à décomposer.
 - Paraphrasage de l'énoncé.
 - Définition des sémantèmes plus simples qui constituent la séquence.
 - Définition de la structure sémantique complète de la séquence.
 - Examen de l'intégration des sémantèmes plus simples dans les sémantèmes complexes à décomposer.

- Recherche des notions (« sous-acceptations ») de l'acceptation qui s'apparente le plus avec le sens du nom dans la séquence.
 - Examen de l'équivalence entre ce sens et une de ces notions.
6. Bien que les démarches que nécessite l'analyse de la compositionnalité du prédicat soient en principe identiques à celles de l'analyse du nom, le caractère non compositionnel du verbe diffère en quelque sorte de celui du nom : soit le verbe n'a pas de sens véritable, car il s'agit d'un verbe support sémantiquement vide, soit son sens est isolable, mais il ne cooccure qu'avec le nom de la séquence analysée. Dans ce dernier cas, son sens est souvent transparent tout comme celui de la séquence intégrale. En réalité, il ne s'agit dans ce cas que d'une acceptation très spécifique, tout en étant transparente. Quant aux expressions ayant pour prédicat un verbe support sémantiquement vide, elles forment une sous-catégorie parmi les séquences non compositionnelles à prédicat non compositionnel encore que le fonctionnement de l'algorithme ne permet pas de les distinguer des autres SNC à prédicat non compositionnel.
 7. Pour ce qui est des questions appartenant au second bloc de l'algorithme, elles portent toutes sur l'appartenance du complément d'objet à une classe donnée, soit à une hyperclasse, soit à une classe d'objet d'après son sens dans la séquence. Soulignons donc que dans notre perspective, l'appartenance d'un terme à une classe est toujours déterminée en fonction d'une de ses acceptations. Il est ainsi possible qu'un lexème puisse appartenir à plusieurs classes.
 8. La manière dont l'algorithme range une séquence dans la catégorie des SC sans spécification est la recherche du contre-exemple. Cela signifie que l'algorithme substitue le complément d'objet à une suite de mots ayant un trait sémantique différent de celui que possède le complément d'objet. Si l'algorithme en trouve un qui forme une phrase agrammaticale, cela signifie que la séquence ne fait pas partie de cette catégorie, et donc elle peut être analysée davantage. Dans le cas contraire, l'algorithme identifie la séquence en tant que SC sans spécification.
 9. Cependant, lors de la substitution, se pose inévitablement la question de l'étendue nécessaire des termes qu'il faut substituer. En effet, pour que l'appartenance d'une séquence à cette catégorie soit sans aucune conteste prouvée, il faudrait en principe vérifier tous les termes de chaque hyperclasse, ce qui est évidemment impossible. C'est pour cette raison que nous postulons que si cette substitution est possible avec un terme aléatoirement choisi, elle serait également possible avec un autre.

10. Cependant, si on a affaire à une séquence dont le constituant verbal est l'opérateur général de son constituant nominal, la méthode de « contre-exemple » s'avère irréalisable. Dès lors, deux autres méthodes peuvent servir à donner une réponse à la deuxième question. D'une part, on peut recourir à la définition du nom, proposée par le dictionnaire, qui fournit souvent d'informations relatives à l'appartenance du terme défini à une hyperclasse. D'autre part, si la méthode précédente s'avère inefficace, notamment dans le cas où le dictionnaire se réfère au complément d'objet au moyen de l'indication « quelque chose », on peut, en principe, recourir à un corpus extrêmement large qui peut démontrer l'ensemble des termes cooccurrent avec l'opérateur dont on souhaite savoir le type. Mais, pour des raisons économiques, nous rejetons cette méthode et nous recourons plutôt à notre intuition extralinguistique au cas où nous aurions affaire à cette problématique, et nous postulons la même chose que dans le cas de la catégorie précédente : si la substitution est possible avec un terme aléatoirement choisi, elle serait également possible avec un autre.
11. Pour que la séquence appartienne à la catégorie des SC à opérateur approprié, une des trois conditions doit être remplie :
- Le verbe est approprié à la classe donnée
 - Le verbe est seulement hérité d'une classe hypéronymique
 - Le verbe est transmis par héritage multiple.
12. En effet, du point de vue des classes d'objet et des prédicats appropriés, chaque sens peut être décomposé en cinq types de fragment de sens :
- Fragment de sens 'A' ('A_X', 'A_Y', 'A_Z', etc.) : sens partiel commun dans le sens de tous les termes qui constituent la / les classe(s) d'objet hypéronymique(s) d'une classe d'objet à laquelle un terme donné appartient. (s'il n'est pas possible d'intégrer une classe d'objet dans une autre classe hypéronymique, ce fragment peut être considéré comme vide) → héritage
 - Fragment de sens 'B' : sens partiel commun dans le sens de tous les termes- mais seulement ceux-ci - qui constituent la classe d'objet à laquelle un terme donné appartient.
 - Fragment de sens 'C' : sens partiel commun avec tous les termes qui constituent une autre classe d'objet que celle à laquelle un terme donné appartient. → héritage multiple
 - Fragment de sens 'D' : sens partiel commun avec certains termes qui ne constituent pas nécessairement une classe d'objet complète, mais qui se trouvent souvent (mais

pas nécessairement) dans de diverses classes d'objet. Ce fragment de sens est ainsi « occasionnel », et son identification ne se révèle que lors de la comparaison avec d'autres termes.

- Fragment de sens 'E' ('E_X', 'E_Y', 'E_Z', etc.) : autre sens partiel non commun, unique à moins que le terme n'ait de synonyme.

Cela veut dire qu'un verbe est l'opérateur approprié du nom, et ainsi la séquence est une séquence compositionnelle à opérateur approprié si ce verbe se combine avec le nom à cause du fragment de sens 'A', 'B' ou 'C' de ce nom.

13. Cependant, la construction du réseau de classe d'objet soulève la question suivante : à part l'existence d'un fragment de sens commun et d'un prédicat approprié, y a-t-il d'autres critères de la définition d'une classe d'objet ? Nous sommes d'avis qu'il est nécessaire d'introduire deux autres critères :

- Le fragment de sens commun qui relie les termes d'une classe et leur opérateur approprié doit être également interprétable indépendamment du contexte, et il doit correspondre plus au moins à nos connaissances extralinguistiques préalables. Cela veut dire que ce fragment de sens ne peut être un fragment 'D', une partie de sens seulement contextuel.
- Chaque classe d'objet doit se constituer au moins de deux termes qui ne sont ni synonymes ni quasi-synonymes.

14. Si un verbe est l'opérateur spécifique d'un nom avec lequel il apparaît et forme une séquence compositionnelle, la raison de cette cooccurrence peut avoir ainsi en principe trois raisons :

- Le verbe se combine avec le nom à cause du fragment de sens 'D' de ce nom : le nom a un sens partiel commun avec certains termes qui appartiennent à la même classe d'objet, mais qui ne constituent pas de classe subalterne.
- Le verbe se combine avec le nom à cause du fragment de sens 'D' de ce nom : le nom a un sens partiel commun avec certains termes qui se trouvent dans de diverses classes d'objet, mais qui ne constituent pas nécessairement des classes d'objet complètes.
- Le verbe se combine avec le nom à cause du fragment de sens 'E' de ce nom : le point commun qui relie verbe et nom est alors un fragment de sens unique, non commun avec d'autres éléments.

La troisième raison pour laquelle le verbe pourrait se combiner avec le mot en formant ainsi une séquence compositionnelle à opérateur spécifique n'est que théorique, car ce

type de combinaison est identique en réalité à celui où l'algorithme déduit que le prédicat a un sens unique qui ne s'emploie qu'avec un seul nom, et où, par conséquent l'algorithme considère la séquence comme non compositionnelle à prédicat non compositionnel.

15. Ainsi, une séquence est vue comme compositionnelle à opérateur spécifique si et seulement si les trois critères suivants sont accomplis :
 - se constitue d'éléments compositionnels (de constituants libres, d'un complément d'objet non datif compositionnel et d'un prédicat compositionnel)
 - renferme un prédicat qui impose des restrictions sémantiques à son complément d'objet non datif
 - renferme un prédicat qui n'est ni l'opérateur général, ni l'opérateur approprié de la classe d'objet (construite en tenant également compte de nos deux critères supplémentaires) à laquelle appartient le nom de la séquence.
16. Nous postulons que parmi les séquences remplissant ces trois critères complexes, les collocations ne sont que celles qui subissent des restrictions au niveau de la substitution synonymique. Quant à cette dernière notion, nous avons donné des précisions suivantes :
 - Premièrement, nous considérons que la synonymie, comme notion intuitive de départ, ne peut être interprétée qu'en prenant en compte une acception donnée. Par conséquent, un nom polysémique appartient à plusieurs classes d'objet puisque son appartenance est déterminée par son sens.
 - Deuxièmement, une classe d'objet se constitue des termes qui possèdent chacun un fragment de sens 'B', justifiant leur appartenance à une classe d'objet donnée, en revanche, leur sens complet ne se constitue pas seulement de ce fragment de sens, mais des fragments supplémentaires aussi. Si la différence entre l'ensemble des fragments de sens supplémentaires de deux termes appartenant à la même classe d'objet est considérable, ces deux termes ne sont pas synonymes, quoiqu'ils appartiennent à la même classe d'objet.
 - Troisièmement, il s'ensuit de ces deux dernières précisions que le synonyme d'un terme – s'il en a – doit se trouver au niveau le plus inférieur possible dans la même structure de classe hiérarchisée que se trouve ce terme et il ne peut avoir d'autres fragments de sens ('E_{X, Y, etc.}') que ce terme.
 - Enfin, quatrièmement, une acception donnée est déterminée par un fragment de sens autonome, et non par un fragment de sens « occasionnel » qui se révèle

uniquement dans certaines séquences. Cela a une conséquence importante à l'égard de la synonymie. Quoique deux termes puissent être liés par un fragment de sens occasionnel ('D'), ils ne seront pas tout de même synonymes par ce sens partiel commun, sauf s'ils sont synonymes indépendamment du fragment de sens 'D'.

17. Nous avons vu qu'une séquence compositionnelle remplissant les trois critères du point 15 se constitue d'un opérateur spécifique seulement dans deux cas :
 - soit le terme ne s'intègre dans aucune classe d'objet spécifique mais tout de même valable (selon nos critères supplémentaires), mais seulement dans une classe d'objet hypéronymique,
 - soit le verbe cooccurre avec le nom à cause de son fragment de sens spécial ('D') que ce nom ne partage qu'avec certains termes qui appartiennent à d'autres classes d'objets.
18. Par conséquent, la substitution dans une séquence compositionnelle à opérateur spécifique n'est possible que dans les cas suivants :
 - L'opérateur spécifique dans son acception donnée dans la séquence a un synonyme. La substitution est théoriquement possible.
 - Les noms qui peuvent apparaître avec le verbe appartiennent à la même classe d'objet, mais ils ne constituent pas de classe d'objet spécifique à eux. Ces noms sont synonymes (dans ce cas, c'est ce qui explique pourquoi ils ne constituent pas de classe d'objet subalterne), la substitution est donc possible.
 - Les noms qui peuvent apparaître avec le verbe appartiennent à la même classe d'objet, mais ils ne constituent pas de classe d'objet spécifique à eux. Ces noms ont des synonymes auxquels la substitution est théoriquement possible.

Ainsi, la contrainte de la substitution synonymique signifie que même si la substitution serait possible d'après ces trois cas précédents, il n'est pourtant pas possible de substituer un élément à un de ses synonymes, notamment à cause des habitudes langagières. Cependant, nous sommes d'avis que ce phénomène caractérise beaucoup moins le lexique que cela ne puisse justifier l'introduction de la catégorie distincte des collocations. C'est pour cette raison que nous considérons tout simplement ces expressions comme des séquences verbales compositionnelles à opérateur spécifique dans lesquelles la substitution synonymique est contrainte.