
Notes de lecture

Rubrique préparée par Denis Maurel

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

Wiltrud MIHATSCH, Catherine SCHNEDECKER. Les noms d'humains : une catégorie à part ? Franz Steiner Verlag. 2015. 203 pages. ISBN 978-3-515-11157-7.

Lu par **Thierry POIBEAU**

Laboratoire Lattice (ENS/CNRS)

Ce livre rassemble six contributions autour de la notion de « noms d'humains », c'est-à-dire de noms qui désignent des êtres humains. Les auteurs montrent que cette catégorie de noms a été largement délaissée jusque-là : les rares articles existant sur le sujet présentent essentiellement des listes utiles, mais non fondées sur une étude linguistique fouillée et systématique. Les contributions rassemblées ici visent au contraire à essayer d'identifier les propriétés linguistiques intrinsèques qui motiveraient l'existence de cette classe de noms.

La première contribution est intitulée « Les (noms d') humains sont-ils à part ? » et offre un panorama très complet de la question, en présentant de manière relativement exhaustive les études précédentes, leurs qualités et leurs limites. L'autrice aborde ensuite de manière très fouillée certaines particularités des noms d'humains, comme les patrons de suffixation, les propriétés syntaxiques des noms de qualité et d'insulte ainsi que la détermination des noms d'humains dans les constructions attributives. L'autrice montre que cette étude implique de se pencher sur les adjectifs, les adverbes (comme « *personnellement* »), les verbes et les rôles thématiques associés. « Dès lors, l'étude lexicale des noms d'humains concerne le lexique dans son ensemble » et présente des enjeux importants : ces noms ne sont pas *à part*, mais sont au contraire *au cœur* de l'analyse lexicale.

Dans la contribution suivante, l'auteur s'intéresse aux noms qui réfèrent aux êtres humains en général, c'est-à-dire aux « noms au sommet de la hiérarchie des noms d'humains, comme (*être*) *humain*, *homme* (au sens général), *personne*, *individu* et *gens* » (p. 55). Divers aspects de la question sont abordés comme la préférence assez marquée de cette famille de noms pour le pluriel (par exemple, le pluriel « *individus* » est en général plus présent en corpus que le singulier. « *individu* ») et surtout les relations complexes entre « le côté physique, le côté immatériel individuel, l'importance de l'aspect social et la dimension relationnelle » de ces noms. Le même auteur s'intéresse ensuite à « la position taxinomique et [aux] réseaux méronymiques des noms généraux "être humain" français et allemands ». Il s'intéresse en particulier à certaines propriétés de ces noms, comme leur statut plutôt collectif (« *les gens* ») et dégage deux groupes de noms complémentaires, l'un étant

plutôt collectif et l'autre plutôt distributif (cf. « *la moitié des personnes* » vs « *la moitié des voisins* »).

Dans le chapitre suivant sont abordés les noms qui désignent des parties physiques (les parties du corps notamment), mais aussi des noms plus abstraits qui désignent des qualités ou des facultés. L'article décrit de façon très détaillée ces différentes familles de noms et montre les liens plus ou moins forts entre le tout et la partie, selon qu'il s'agit de désigner une partie du corps ou une qualité (par exemple, la phrase « **Paul est d'une forte tête* » n'est pas acceptable tandis que « *Paul est d'un caractère difficile* » est tout à fait correcte, ce qui traduit une différence de statut des attributs considérés). L'auteur conclut en soulignant « l'existence d'un spectre qui va de la dépendance à l'indépendance pour ce qui concerne la relation partie-tout telle qu'elle se présente chez l'humain ».

L'autrice de la contribution suivante montre que la définition de « *femme* » dans différentes ressources lexicographiques s'appuie fortement sur le sexe, voire sur des stéréotypes liés au sexe, alors que pour le mot « *homme* », les traits liés au sexe sont repoussés à l'arrière-plan, au profit de traits plus génériques référant plutôt à la simple notion d'être humain. De ce fait, la notion de « *femme* » comme "être humain féminin" n'est en aucun cas le symétrique du nom « *homme* » comme "être humain masculin" : pour l'autrice, le concept FEMME est subalterne au concept HOMME dans une ressource comme EuroWordNet. « À l'inverse, [...] l'entité *homme* subordonne toute la hiérarchie que ce soit dans le sens générique ou dans le sens spécifique. »

La dernière contribution étudie en particulier les liens entre noms d'idéalité (terme emprunté à Husserl et désignant des noms comme « *sonate* », « *roman* » ou « *théorème* ») et les liens avec les agents créateurs de ces entités (« *romancier* ») ou bénéficiaires de ces entités (« *allocutaire* », « *auditeur* »). Les liens entre ces différents types de noms sont étudiés en détail, ainsi que l'existence de noms de créateurs sans nom d'idéalité morphologique simple correspondante (compositeur et pièce musicale). Les auteurs proposent des hypothèses concernant ces lacunes lexicales, en ayant notamment recours à différents paramètres d'ordre linguistique et extralinguistique.

On le voit, il s'agit avant tout d'un livre de linguistique, même si plusieurs études sont fondées sur l'interrogation minutieuse de corpus représentatifs. Ce livre fournit un complément bienvenu aux nombreux recueils sur la notion d'entité nommée, et permet de renouveler ce domaine d'étude, en l'élargissant à une classe de noms homogène et complémentaire, bien que largement délaissée jusque-là. Il va sans dire que ce livre ouvre des perspectives fort intéressantes, et gageons qu'il donnera bientôt lieu à des travaux importants dans le domaine du traitement automatique des langues.

Amr Helmy IBRAHÍM. L'analyse matricielle définitoire : un modèle pour la description et la comparaison des langues. Éditions de la CRL. 2015. 554 pages. ISBN 2-9526027-6-X.

Lu par **Christian BOITET**

GETALP - Université Joseph Fourier

Ce livre, de 554 pages très denses, est une sorte de somme reprenant et approfondissant de nombreux travaux précédents de son auteur. La première impression est qu'il est très difficile à lire, aussi bien à cause de sa typographie très défectueuse (abondance de mots collés) qu'à cause de l'amoncellement de notes et d'incises. La seconde est qu'il s'agit d'un travail de linguistique très détaillé et profond, très intéressant par ses nombreuses comparaisons entre langues, mais très éloigné du TAL. Amr Helmy Ibrahim est parti dans les années 70 des axes méthodologiques de Maurice Gross et de son école, pour ensuite tracer et approfondir une voie originale, celle de l'AMD (analyse matricielle définitoire). Cependant, seule une faible proportion de l'ouvrage peut être considérée comme de la linguistique computationnelle, et à peu près aucune comme de la linguistique formelle (ou mathématique).

Ce livre m'a énormément intéressé, comme m'ont intéressé les travaux de l'école russe (dès 1970 Apresyan, Mel'tchuk, Zholkovskij, Rosenzweig, plus tard Boguslavsky et Iomdine avec le développement en parallèle des meilleurs analyseurs du russe et de l'anglais dans ETAP-3), ceux de Zemb (sur la « triade statutaire », et sur l'analyse comparative en trois mille pages chez Duden de l'allemand et du français), et tant de travaux de formalisation linguistique visant plus ou moins à étayer la linguistique computationnelle et par là le TAL, comme ceux de l'école de Prague et de l'école chomskyenne.

Mais qu'en dire en bref ? Amr Helmy Ibrahim affirme à plusieurs endroits de l'ouvrage que sa modélisation pourrait faciliter l'analyse automatique. Cependant, aucune indication n'est donnée sur ce qui pourrait précisément être fait, et encore moins sur la façon de le faire. D'autre part, la grande majorité des chercheurs en TAL est familière de notions linguistiques bien moins nombreuses et profondes que celles nécessaires pour suivre tout l'exposé avec profit, et n'est pas non plus à même de comprendre les nombreux exemples, très intéressants au demeurant, de comparaison de plusieurs langues (il y a des exemples dans dix-neuf langues) sur tel ou tel phénomène, sans se limiter à l'éternel couple français-anglais, très réducteur. Enfin, cette analyse linguistique, certes très approfondie, est présentée sans faire le lien avec d'autres théories linguistiques contemporaines, nettement plus pertinentes du point de vue du TAL, comme la théorie sens-texte de Mel'tchuk ou la théorie de la dépendance de l'école de Prague (Sgall, Hajičova...).

Pour encourager les lecteurs de la revue *TAL* à aborder cet ouvrage imposant, on pourrait essayer de donner une sorte de « guide de lecture », résumant les six parties principales et leurs vingt-cinq chapitres ou grandes sections, mais cela ne semble pas possible en moins d'une trentaine de pages. Nous nous bornerons ici à reprendre un exemple de ce qu'est une AMD, pour illustrer cette approche et ses buts, et à présenter une toute petite partie de l'analyse contrastive des constructions à verbe

support, dont les traitement, comme plus généralement celui des expressions polylexicales plus ou moins figées, est redevenu un domaine de recherche très actif en TAL.

1. *Exemple de matrice définitoire (citation)*

Que faut-il comprendre pour que l'énoncé suivant, produit au cours d'un dialogue oral, ne tombe pas dans l'oreille d'un sourd ?

[T2] *Quand je l'ai appelé, il dormait.*

Il faut comprendre que l'enchaînement syntaxique des données grammaticales suivantes :

- la liaison qu'établit la conjonction *quand* entre le verbe *appeler* au passé composé et le verbe *dormir* à l'imparfait ;
- la forme de *passé composé* faite de l'*auxiliaire avoir* et du *participe passé* ;
- la forme de *l'imparfait* ;
- la valeur et la fonction de référence et d'anaphore des pronoms personnels *je*, *l'* et *il* ;

correspond au discours sous-jacent suivant :

[T2'] *À un moment qui est avant le moment de mon énonciation, il se produit l'événement suivant : la personne qui est moi et que je désigne par le mot je passe un appel téléphonique à une personne de sexe masculin dont je ne précise pas le nom, que mon interlocuteur connaît ou attend de connaître, que je désigne par le mot abrégé le et qui est plongé dans un état de sommeil, c'est-à-dire dans l'état de quelqu'un qui dort.*

Ou, si l'énoncé avait été lu avant l'invention de Graham Bell :

[T2''] *À un moment qui est avant le moment de mon énonciation, il se produit l'événement suivant : la personne qui est moi et que je désigne par le mot je adresse un appel à une personne de sexe masculin dont je ne précise pas le nom, que le lecteur connaît ou attend de connaître, et que l'auteur signale par le mot abrégé le au moment où cette personne est plongée dans un état de sommeil, c'est-à-dire dans l'état de quelqu'un qui dort.*

[T2'] ou [T2''] formulent le minimum de ce que nous devons savoir pour vraiment comprendre [T2] dans le contexte où il est produit. Ce minimum correspond aussi à ce que serait une formulation totalement explicite et non ambiguë de [T2]. Ces formulations explicites sont redondantes au sens non pas qu'elles répètent à proprement parler ce qui a déjà été dit, mais au sens où elles donnent explicitement des informations dont la formulation abrégée n'est qu'un déclencheur formel.

L'ouvrage explique en détail comment, grâce aux mécanismes grammaticaux, le français et d'autres langues construisent des énoncés portant l'essentiel des

informations explicitées dans la matrice définitoire, et, en tout cas, tels que les destinataires puissent les comprendre, étant donné le contexte partagé entre eux et l'émetteur de l'énoncé.

2. Exemple d'analyse de certaines constructions à verbe support (citations)

Voici d'abord la deuxième des treize « constructions types » étudiées en détail dans l'ouvrage, après une étude encore plus détaillée des constructions à support *avoir*. On voit bien le mécanisme de génération à partir de la matrice.

[II] Construction étendue à support réflexif réversible

$$N_{i-hum} V_{-sup-avoir} Det N_{-pred} Prep_{-pour/à} l'égard de/vis-à-vis de (N_{j-hum} + N_{j-nr})$$

$$\longleftrightarrow$$

$$(N_{j-hum} + N_{j-nr}) V_{-sup-avoir} (Det + Det_{i-pos}) N_{-pred} (à N_{i-hum} + de N_{i-hum})$$

Cette construction est caractéristique des prédicats introduits dans la matrice par un nom classifieur de sentiment $N_{-class-sentiment}$. Ces prédicats sont plus précisément des *causatifs de sentiment*. Leur matrice a la forme suivante :

Il se fait que (Det) N_{-hum} (a + éprouve + ressent) Det $N_{-class-sentiment}$ Prep_{-de} N_{-pred}
Pro_{-relatif} ($N_{-hum} + N_{-nr}$) lui (inspire + fait ressentir) Prep_{-pour/à} l'égard de/vis-à-vis de/au sujet de
($N_{j-hum} + N_{j-nr}$)

Soit par exemple :

(1) Noé a [un sentiment d' → de l'] *admiration pour* Zoé

(1') **Il se fait que** Zoé a l'*admiration de* Noé → Zoé **fait** l'*admiration de* Noé

(2) Noé a des inquiétudes **au sujet de** Zoé

(2') Zoé **inspire** des inquiétudes à Noé

(3) Noé a beaucoup de méfiance **vis-à-vis de** Zoé

(3') Zoé **inspire** beaucoup de méfiance à Noé

(4) Noé a de la tendresse **pour** Zoé

(4') Zoé **inspire** de la tendresse à Noé

Cette construction montre la relation étroite que le verbe support entretient avec son sujet et l'aptitude qu'ont certains supports comme *avoir* d'établir un lien entre la relation d'appropriation qu'ils ont avec le prédicat qu'ils actualisent et leur sujet grammatical.

Les deux schémas au début, présentés comme équivalents, sont des « index structuraux »¹, pas des matrices définitoires. En revanche, le schéma commençant par « Il se fait que... » est une matrice définitoire.

La construction type suivante est déduite de la précédente par un figement presque complet.

[III] Construction figée

N_{-hum} V_{-avoir} Det_{-une} N_{-dent} Prep_{-contre} N_{-hum}

Par rapport à [II], l'index structural [III] se définit par une réduction drastique de la combinatoire et une disparition du caractère paradigmatique des constituants de la structure : on n'a plus affaire à un type de verbe, mais à un verbe spécifique, la détermination a perdu sa fonction grammaticale puisqu'il n'y a plus qu'un seul déterminant possible, on n'a plus affaire à la catégorie des noms, mais au nom *dent*, aucune variation n'est plus possible sur la préposition, et la sous-catégorisation stricte des noms est limitée au trait *humain* :

*Noé (a + garde) (une + *E + *la + *les + *des) dent contre Zoé.*

On comprend ici la différence de « variabilité » des expressions composées, et il est bien possible que ce soit une très bonne idée de s'appuyer sur ce type d'analyse pour modéliser toutes les variantes des expressions polylexicales plus ou moins variables « en discours ». C'est un problème auquel le TAL s'attaque depuis quelques années. On trouve dans l'ouvrage une très intéressante liste de quatre-vingt-onze verbes supports en français. On pourrait la compléter (par exemple, elle contient « infliger » mais pas « subir », cf. *infliger/subir une défaite*) et préciser les fonctions lexico-sémantiques associées, et on aurait tout de suite de bonnes bases pour le traitement de ce type de constructions, tant en analyse qu'en génération.

¹ On pourrait dire seulement « schémas syntaxiques », car aucune structure, par exemple dépendancielle, n'est contenue dans ces expressions.