

DESCRIPTIF SUCCINCT DU PROJET

Le projet vise la description d'une catégorie lexicale de noms encore en marge des études dans le domaine de la sémantique nominale : **les noms qui dénotent des « humains »** (*personne, individu, touriste, linguiste, banlieusard, président, socialiste,...*) (désormais NH).

De prime abord, ces N ne semblent pas poser de problème majeur, du fait qu'ils satisfont aux tests utilisés ordinairement pour classer les N d'entités concrètes (on peut « voir/toucher/interagir (avec) » un « enfant/piéton/médecin »), comptables (on peut « compter/dénombrer » « un/deux/quelques » « enfant(s)/piéton(s)/médecin(s) » et hétérogènes (les « parties » d'un « enfant/piéton/médecin » sont d'un autre type que leur « tout »). Par ailleurs, leur « évidence » sociologique ou socioprofessionnelle (en un mot extralinguistique) telle que la question des modalités et des enjeux des classifications en usage ne semble pas devoir être questionnée d'un point de vue linguistique.

Pourtant, un examen attentif de la littérature et des faits révèle un certain nombre de problèmes :

- la bipartition humain vs non humain n'aurait rien d'évident selon Lakoff (1986) qui montre que dans certaines langues aborigènes d'Australie, les entités humaines mâles vs femelles sont dissociées et rangées avec les animaux d'une part et l'eau, le feu, la nourriture, d'autre part ;
- le principe de la « hiérarchie être », utilisé pour classer les entités non animées ou animées non humaines (p.ex. *un singe est un animal, une pomme est un fruit*) et catalysant la structuration lexicale, s'applique difficilement aux entités humaines (**un plombier est un homme ; *une actrice est une femme*) (cf. Cruse, 1986, 2002) ;
- des expérimentations menées en psycholinguistique (Medin *et al.* 2000) appuient l'idée que les catégories sociales, renvoyant aux humains, seraient basées sur la race, l'âge, le sexe et les occupations (p.ex. *adolescente de 10 ans, asiatique, transporteur de courrier*) et représenteraient une « classification croisée » (cross-classification) davantage qu'une taxonomie.

De là, l'idée que “different cognitive domains may have different types of semantics, and that, *in particular, the semantics of living things is different from the semantics of artifact*” (Wierzbicka, 1996, 370 ; nous soulignons).

C'est cette idée que le présent projet va défendre en :

- proposant, tout d'abord, un état de l'art substantiel sur la question des NH (les deux disciplines linguistiques ayant abordé la question étant, d'une part, celle de la morphologie, du fait que les NH usent de procédés de construction très nombreux et diversifiés (p.ex. : conversion : adjectif > nom : *riche > les riches, pauvre > les pauvres* ; dérivation en -ard : *banlieusard* ; -ant : *fabricant* ; -eur : *chercheur* ; -ien : *musicien* ; -iste : *communiste* ; -ais : *français, polonais* ; -in : *messin, ...*) et, d'autre part, celle de la syntaxe qui a montré que les N dits de fonction ou de profession génèrent des interprétations distinctes selon qu'ils s'emploient avec ou sans l'article indéfini comme dans *Paul est un médecin vs Paul est médecin*)¹ ;
- inventoriant les NH et en les classant suivant des principes linguistiques sémasiologiques (en partant des caractéristiques formelles des NH) dans le but de compléter et améliorer les classements existants (cf. Gross, 1995, Bonnard, 1984) ;
- en mettant au jour **leurs spécificités linguistiques**, par contraste avec d'autres N plus couramment étudiés, notamment :
 - le principe de structuration qui gouverne ce lexique particulier, avec la question des N génériques (*personne, individu, ...*) et celle d'une organisation taxonomique vs multi-critériée (cf. *supra*) ;
 - **leur importance ontologique** : la personne humaine est par nature fondamentale et participe (avec l'espace, le temps et la manière) des grandes catégories ontologiques qui structurent notre expérience et notre cognition ;
 - **leur rapport au temps** : certains N s'appliquent aux individus humains pour toute leur vie si l'on peut dire : on naît blanc ou noir, français ou américain, homme ou femme (d'où l'irrecevabilité de **Paul est devenu noir/asiatique* ; d'autres sur une partie seulement : on est considéré comme un enfant jusqu'à environ 15 ans (d'où des irrecevabilités comme : **Paul est né adulte*) ; d'autres, enfin, le font épisodiquement ou sporadiquement (*piéton,*

1. Voir les références bibliographiques dans la partie détaillée du projet.

- touriste*, voire *communiste*); d'où l'intérêt de creuser ce point en rapport avec les temps grammaticaux, l'aspect et les questions d'ontologies sortales (individual vs stage level) de Carlson (1977) ;
- le **caractère intrinsèque vs extrinsèque de certains attributs** (cf. Anscombe, 2001) à corrélérer avec la possibilité pour certains N d'apparaître dans des constructions manifestant l'agentivité (*il a tout fait pour devenir enfant vs il a tout fait pour devenir pompier/*piéton).

Outre le **caractère inédit du thème retenu**, l'originalité du projet tient à sa **dimension pluridisciplinaire et plurilinguistique**. En effet, l'équipe constituée pour mener à bien cette étude se compose de spécialistes de sous-domaines linguistiques : morphologues, sémanticiens, linguistes informaticiens, traductologues et diachroniciens, réunissant, qui plus est, des langues diversifiées (langues romanes : espagnol, italien, roumain ; langues germaniques : allemand, anglais ; langues slaves : serbe, bulgare).

Retombées et résultats attendus

- ♦ **Avancées en recherche fondamentale**, grâce aux inventaires et typologies des NH, à l'élucidation de la structuration lexicale de ce lexique particulier (détermination des relations sémantiques entre ces NH et mise au point d'une ou de taxinomies qui manifesteront l'originalité de ces N) ; Une autre avancée intéresse particulièrement la frontière entre morphologie et lexique. Elle concerne le croisement des points de vue, souvent difficiles à concilier, de la sémasiologie et de l'onomasologie. Comme le souligne Roché (2011), l'étude du lexique construit consiste soit en démarches comme celle qui vise à répondre à la question « quels sont les caractéristiques sémantiques des dérivés nominaux morphologiques en *-eur* ? en *-ier* ? en *-iste* ? », soit en des approches tournées vers la caractérisation des unités produites, plutôt que vers les modes de formation de ces unités, et qui s'occupent de questions comme : « Parmi les noms humains, comment forme-t-on les noms collectifs ? singulatifs ? les noms d'idéologues ? de métier ? de relation ? etc. ». Le projet offre, de par la spécialité des membres qui y participent, le double regard nécessaire pour permettre le développement d'une méthodologie mixte d'analyse du lexique : sémasiologique *et* onomasologique. Nous pourrions à terme proposer une cartographie du lexique des NH mesurant, pour chaque type sémantique de nom, la participation de telle règle constructionnelle (singulatif : *équipe* > *équipier*, sportif : *basket* > *basketteur*, disque > *discobole*, musicien : *flute* > *flutiste*) ou l'activation d'une locution 'N prep N' (singulatif : *club* > *membre d'un club*, sportif : *poids* > *lanceur de poids*, musicien : *pipeau* > *joueur de pipeau*) ou autre.

- ♦ **Avancées dans le domaine de la recherche appliquée** : la progression des connaissances sur le fonctionnement de ces N devrait en effet fournir un jeu d'étiquettes morphosémantiques susceptibles d'améliorer et d'affiner l'étiquetage des bases de données catégorisées (qui est souvent à « gros grain »), et par là, faciliter leur exploitation ainsi que celle de la constitution de corpus ; par ailleurs, une meilleure connaissance du fonctionnement des NH devrait aider à l'indexation des chaînes de référence dans les types de textes centrés sur les humains (textes narratifs, informatifs, etc.) et se révéler utile dans des applications de fouille de texte, d'extraction d'informations ou de recherche d'informations ou de traduction automatique ou assistée par ordinateur ; enfin, la constitution de bases de données bimodales (linguistiques / iconiques) permet d'envisager certaines applications didactiques, p.ex. en matière de communication scientifique (texte documentaire, vulgarisation, etc.) ou pédagogique (ouvrages scolaires, méthodes de langues, dictionnaires visuels, etc.).

♦ Produits délivrables

- « **Module** » de **dictionnaire électronique sur les NH** (contacts avec les équipes de lexicographes de Paris-7 (B. Sagot qui élabore un Wordnet français gratuit) et de Paris-13 (équipe de J. Pruvost) pour proposer d'intégrer « nos » traits descriptifs dans les dictionnaires électroniques à disposition) ;
- **Jeux d'étiquettes morpho-sémantiques** (à intégrer à des outils d'annotations, type ANALEC, Paris-3, équipe avec laquelle nous collaborons, ainsi que les modules d'analyse morphologique comme celui de DériF élaboré par F. Namer, ATILF, U. de Lorraine) ;
- **2 numéros de revues thématiques** (revue *Langages*, numéro spécial de *Journal of French Language Studies*) ;
- **1 colloque international** (Strasbourg, fin 2013-début 2014) et les actes du colloque ;
- **1 site Web**.