

Le lexique syntaxique Proton: une base de données de la valence verbale en français

Karel van den Eynde & Piet Mertens
Université de Leuven

Résumé

Cet exposé présente d'abord le dictionnaire de valence Proton des verbes pleins en français, son contenu, les propriétés répertoriées. Ensuite il situe le cadre de description syntaxique adopté (l'approche pronominale), l'utilité des paradigmes de pronoms, les types de compléments essentiels identifiés et la notion de groupe de formulations. Afin d'illustrer les applications du lexique dans le domaine du TAL, on décrit une procédure permettant de passer des paradigmes de pronoms à des structures de traits représentant des propriétés syntaxiques.

1. Le dictionnaire de valence Proton

La base de données Proton a été réalisée entre 1986 et 1992, au Département de linguistique de l'université de Leuven, dans le cadre d'un projet de recherche dirigé par Karel van den Eynde.

Couverture

La base de données répertorie les propriétés de construction valentielles des verbes du français. Il s'agit des verbes simples qui figurent dans le Petit Robert, à l'exclusion des acceptions vieilles ou techniques. En tout, 3700 infinitifs figurent dans les 8600 entrées représentant chacune un emploi syntactico-sémantique différent. Les locutions verbales (*porter plainte, retrousser les manches, ne pas y aller par quatre chemins*) ne sont pas reprises dans le lexique.

Propriétés syntaxiques inventoriées

Chaque entrée décrit les propriétés syntaxiques d'un emploi donné. Elle énumère les « compléments essentiels » (ou « termes de valence », « actants », « paradigmes ») figurant dans le cadre de sous-catégorisation (« schéma valentiel », « schéma actantiel »). Pour chaque complément, elle indique la nature (cf. infra), le caractère obligatoire ou facultatif, le paradigme de pronoms (cf. infra), les réalisations syntagmatiques possibles (pronom, groupe nominal, complétive, infinitif, etc.). L'entrée indique également les reformulations éventuelles (cf. infra), les « groupes de formulations liés » (constructions apparentées avec le même infinitif) et la traduction en néerlandais du verbe. L'entrée est complétée par une phrase exemple qui illustre l'emploi.

Les entrées ne mentionnent que les termes de valence (compléments essentiels, actants), écartant ainsi les « compléments accessoires » (« circonstants ») qui s'emploient avec tous les verbes.

Constitution du dictionnaire

Les schémas valentiels ont été constitués à partir de sources lexicographiques existantes : les dictionnaires Petit Robert et Larousse, le dictionnaire de valence de Busse & Dubost (1983). Ces données étaient complétées par les observations des membres de l'équipe, et réinterprétées en fonction de l'approche syntaxique adoptée. Le dictionnaire a été validé par plusieurs locuteurs natifs.

2. Cadre syntaxique : l'approche pronominale

Paradigmes de pronoms

Dans la perspective de l'approche pronominale (Blanche-Benveniste *et al.*, 1984 ; Melis *et al.*, 1994 ; Eynde *et al.*, 2003), les constructions verbales à compléments pronominaux fonctionnent comme des grilles classificatoires pour les constructions comportant des éléments lexicaux. Aussi le lexique Proton énumère-t-il les pronoms pouvant figurer dans chaque position du cadre de sous-catégorisation. Cependant, la classe des formes citées ne se limite pas aux pronoms clitiques et non clitiques de la grammaire traditionnelle. Elle inclut les pronoms dits interrogatifs (appelés "suspensifs"), certains "adverbes" dont le comportement est analogue à celui des pronoms, les pronoms proportionnels à une complétive (que-Phrase, infinitive, etc.) et les indéfinis.

Intérêt des paradigmes de pronoms

L'utilisation de paradigmes de pronoms pour la caractérisation des compléments présente plusieurs avantages.

- Elle permet de distinguer des compléments à première vue similaires sur le plan syntagmatique, sans besoin de définir des traits sémantiques ou des fonctions.

- parler à <i>qn</i> [objet indirect]	à <i>qn</i> /lui/*y/*là
s'adresser à <i>qn</i> [complément prép.]	à <i>qn</i> /*lui/y/*là
habiter à la gare [compl. locatif]	*à <i>qn</i> /*lui/y/là
- parler de <i>qc</i> , se libérer de <i>qc</i>	de <i>qc</i> /de <i>lui</i> /en/*de <i>là</i>
venir de <i>qp</i>	*de <i>qc</i> /*de <i>lui</i> /en/de <i>là</i>
caractériser son style d'impressionniste	ainsi/*de <i>lui</i> /*en/*de <i>là</i>
- Elle permet également d'identifier des compléments apparents (par exemple le « sujet » des verbes impersonnels, morphèmes discontinus...).

il/*elle/*ça neige
il en arrive; *il m'est arrivé ça*
- Enfin elle permet d'expliciter certaines restrictions sélectionnelles (p.ex. trait de « complexité » ou pluriel obligatoire).

Jean et Anne se séparent; *ils se séparent*; *nous nous séparons*; *je les sépare*
**Jean se sépare*; **il se sépare*; **je me sépare*
elle s'est séparée de son mari; *elle se sépare de lui*; *je me sépare de lui*

Typologie des compléments essentiels

La constitution des paradigmes pronominaux débouche sur une typologie des compléments, plus riche que dans la tradition grammaticale.

- P0 (sujet),

nous déménageons
- P1 (objet),

il les déchire, ces lettres
- P2 (objet indirect) y/lui/à *lui*/à *ça* (avec restrictions supplémentaires selon le verbe),

j' y pense; *je pense à lui*; *je lui parle*
- P3 (objet indirect) en/lui/de *ça* (avec restrictions supplémentaires),

il le lui enlève; *il l' en enlève*; *il en diffère par son caractère*
- PQ (complément de quantité)

ça te coûtera autant; *il mesure combien?*
- PL (locatif) y/là,

il va quelque part
- PD (délocatif) en/de *là*,

- *j' en reviens*
- PT (temps),
*le médecin ne consulte qu' à ce moment-là
la pharmacie ferme quand?*
- PM (manière)
*ils l'ont nommé comment; on l'appelle ainsi
sa mère l'attifait ainsi (comme un sapin de Noël)*
- PP (complément prépositionnel à préposition fixe) avec mention de la (des) préposition(s)
*je vous désapprouve dans votre comportement
adosser qc contre; aligner qc sur; s'acharner contre/sur;
s'accompagner de; accéder à qc par qp
céder devant les menaces*

Les compléments valenciels locatif, délocatif, temporel et de manière sont à distinguer des compléments accessoires (« rectionnels ») correspondants.

Là (Rec), à l'aéroport, tu prends le métro et tu y (Val) arriveras, à la Gare du Nord.

Le 15 janvier (Rec), le magasin fermait exceptionnellement à 15 heures (Val).

La tour déverse légalement (Rec) dans ce sens-là (Val).

Formulation, reformulation et groupe de formulations.

Chaque entrée décrit donc les compléments essentiels (ou paradigmes) d'un verbe dans un emploi donné, indépendamment de l'ordre des éléments dans la chaîne, ordre qui dépend de facteurs indépendants comme la réalisation pronominale, nominale ou verbale du complément, etc. Une telle entité sera appelée une « *formulation* ».

Un infinitif donné peut ainsi présenter plusieurs formulations:

*je retrouve la clé
la clé a été retrouvée par moi*

Les deux constructions ci-dessus font intervenir un même verbe et les mêmes référents, bien que dans des fonctions différentes. Elles seront appelées des « reformulations », si l'une implique l'autre et vice versa : si je retrouve la clé, la clé a été retrouvée par moi.

Les types de reformulations recensées dans Proton sont les suivantes:

- reformulation "passif être"
(le conducteur charge le camion ↔) le camion est chargé par le conducteur
- reformulation "se passif"
le camion se charge [la veille de la livraison] par le conducteur
- reformulation "se faire passif"
*(il a été acheté par un des partis ↔) il s'est fait acheter par un des partis
(je paie le propriétaire [autant] ↔) le propriétaire se fait payer [autant]*

Les possibilités de reformulation varient selon l'emploi du verbe :

*ça ne le regarde pas * il n'est pas regardé par ça
tu nous étonnes * nous sommes étonnés par toi*

Pour cette raison, le « groupe de formulations » correspond à l'ensemble formé par une formulation donnée et ses reformulations éventuelles.

Réalisations verbales d'un complément

Le lexique Proton précise également les différents types de complétives admises pour un complément verbal:

- infinitif
*écouter de la musique délasse l'esprit
réussir suppose qu'on a essayé*

- de + infinitif
 - il lui téléphonait de rentrer immédiatement (P1)*
 - il tente/accepte/arrête/affecte de parler au voisin (P1)*
 - il feint d'éprouver un malaise (P1)*
 - il m'est parfois advenu/arrivé de me tromper (P0, il..de)*
 - il appartient aux parents d'élever leurs enfants (P0, il..de)*
 - ça me choque/amuse d'être traité ainsi (P0, ça ÷ de+inf)*
 - ça l'arrange vraiment de devoir partir (P0, ça ÷ de+inf)*
- à + infinitif
 - il s'acharne à obtenir cette promotion (P2)*
 - elle apprend à parler le chinois (P1)*
- que-phrase à l'indicatif
 - j'indiquerai que la première page manque (P1)*
 - il s'est avéré que toute son histoire n'était qu'un bluff (P0, il..que-phrase)*
- que-phrase au subjonctif
 - il me déplait que Max soit parti sans me prévenir (P0, il..que-phrase)*
 - il attend de moi que je vienne lui rendre visite (P1)*
 - à ce que-phrase à l'indicatif/au subjonctif*
 - il s'est habitué à ce qu'elle parle peu*
 - ses parents ont consenti à ce qu'il prenne des vacances*
- de ce que-phrase à l'indicatif/au subjonctif
 - elle se plaint de ce qu'il ne s'occupe plus de ça*
- si-phrase à l'indicatif
 - elle demande si tu viens*
- question indirecte (= avec pronom suspensif régi par le verbe subordonné)
 - évaluer combien de mois il faudrait pour y arriver leur serait impossible (P0)*
 - t'as maintenant pigé pourquoi je t'ai dit ça! (P1)*

Quelques entrées du lexique Proton seront examinées en détail dans la présentation.

La forme des entrées

Une description détaillée du lexique Proton a été publiée en 2003 dans la revue *Journal of French Language Studies* (vol. 13, pp. 63-104).

Le lexique peut être consulté sur le site suivant:

<http://bach.arts.kuleuven.ac.be/PA/protonfr.html>

3. Des pronoms aux traits syntactico-sémantiques

La majorité écrasante des analyseurs syntaxiques qui prennent en compte le cadre de sous-catégorisation et les propriétés syntaxiques et sémantiques (restrictions sélectionnelles) des compléments représentent ces propriétés sous forme de traits syntaxiques, le plus souvent organisés en structures de traits. Celles-ci sont associées aux positions dans le cadre de sous-catégorisation. Comme il a été dit plus haut, bon nombre de ces informations sont présentes dans le dictionnaire Proton sous forme de paradigmes de pronoms. Au niveau de la description linguistique l'utilisation des paradigmes de pronoms présente un intérêt certain. En revanche, ce type de représentation convient peu au fonctionnement des analyseurs à unification. Nous avons donc tenté de passer des paradigmes de pronoms aux structures de traits qu'ils impliquent.

Le principe de cette transformation est simple. Un terme valentiel se voit attribuer un trait *T* avec une valeur donnée sous certaines conditions. 1. Le paradigme de pronoms correspondant présente un pronom donné, une classe de pronoms, une disjonction ou une conjonction de

pronoms donnée. 2. Le paradigme ne présente pas tel pronom, telle classe, disjonction ou conjonction de pronoms. 3. L'attribution d'un trait peut dépendre de la présence d'un autre trait (avec une valeur donnée). Les traits s'organisent donc dans une hiérarchie.

L'approche a été mise en oeuvre pour attribuer les traits [\pm HUMAIN], [\pm CONCRET], [EVENT], [\pm COMPLEXE] (pluriel obligatoire), [\pm INDIVIDUEL]. Elle a aussi permis de préciser (a) la nature syntagmatique du complément : exclusivement pronominale (verbes pronominaux), nominale, verbale (proposition infinitive ou complétive), ou à la fois nominale et verbale, (b) certaines caractéristiques de la forme verbale du complément : infinitif, mode indicatif ou subjonctif.

Dans la présentation, nous illustrerons la procédure d'attribution des traits.

Exploitation en traitement automatique des langues

La base de données Proton a été transformée en vue de son intégration dans le système de traduction automatique METAL, conçu à l'université du Texas et développé dans les années 80 et 90 à Leuven, avec le soutien de la société Siemens.

L'analyse en traits à partir des paradigmes de pronoms s'inscrit dans une recherche qui vise l'intégration du lexique valentiel Proton dans un analyseur pour grammaires de dépendance (Mertens, 2002). Les entrées lexicales y prennent la forme d'arbres élémentaires, dans lesquels les arcs représentent des relations de dépendance et les nœuds des éléments lexicaux ou des positions (de dépendant) à saturer. A chaque nœud est associée une structure de traits (FS) qui explicite les propriétés syntaxiques et sélectionnelles. Comme dans les grammaires d'arbres adjoints (TAG), l'analyseur procède par adjonction d'arbres. Cependant, dans notre cas, toutes les données, y compris les arbres, sont représentées comme des FSs, et la combinaison des arbres revient à l'unification du FS du dépendant avec un nœud du régissant.

Références

- Blanche-Benveniste, C.; Delofeu, J.; Stefanini, J. & Eynde, K. van den (1984). *Pronom et syntaxe. L'approche pronominale et son application au français*. Paris : SELAF, pp. 245.
- Busse, Winfried & Dubost, Jean-Pierre (1983) *Französisches Verblexikon. Die Konstruktion der Verben im Französischen*. Stuttgart : Ernst Klett, 2. überarbeitete Auflage, pp. 422.
- Dehaspe, Luc & Karel Van den Eynde (1991) The pronominal approach to verbal valency: a formal description of *speak, say, tell and talk*. In: Klein E. et al. (eds.) *Betriebslinguistik und Linguistikbetrieb*. Tübingen: Niemeyer, vol. 1, pp. 273-280.
- Eggermont, C. & Eynde, K. van den (1990) A Pronominal Basis for Computer Assisted Translation. The Proton Project. in Thelen, M. & Lewandowska, B. (eds) (1990) *Translation and Meaning, Part 1*. (Proceedings of the Maastricht Session of the 1990 Maastricht-Lodz Duo Colloquium).
- Eynde, K. van den ; Mertens, P. (2003) La valence : l'approche pronominale et son application au lexique verbal. *French Language Studies* 13 (1), 63-104. (Cambridge)
- Eynde, K. van den; Broeders, E.; Eggermont, C. & Vangilbergen, L. (1988) The pronominal approach in NLP: a pronominal feature analysis of coordination in French. *Computers and Translation* 3, 177-213.
- Eynde, Karel van den; Sabine Kirchmeier-Andersen; Piet Mertens; Lene Schøsler (2002) Distributional Syntactic Analysis and Valency : Basic Notions, Procedures and Applications of the Pronominal Approach. in Nevin, Bruce (ed) (2002) *The Legacy of*

Zellig Harris. Language and information into the 21st century. Volume 2. Mathematics and computability of Language. John Benjamins Publishing. 163-202.

Melis, Ludo & Eggermont, Carmen (1994) L'approche pronominale de la valence. *International Journal of Lexicography* 7(2), 79-89.

Mertens, Piet (2002) Parsing Dependency Grammar using ALE. *Proceedings COLING 2002*, vol. 2, p. 653-659 (19th Int. Conf. on Comp. Ling., August 26-30, 2002, Taipei, Taiwan).