

Un "langage pivot" pour articuler description et formalisation : l'exemple des verbes introducteurs de "que-phrases"¹

Christophe Benzitoun

Equipe Delic – Université de Provence
29, Av. Robert Schuman, 13100 Aix-en-Provence
christophebenzitoun@yahoo.fr
1^{ère} année de Doctorat

Mots-clefs – Keywords

Langage "pivot", que-phrase, description, formalisation
"pivot" language, Wh-sentence, description, formalization

Résumé – Abstract

Nous présentons dans cet article une réflexion en vue de la modélisation d'une partie du patrimoine descriptif du français finalement peu utilisé en TALN. Pour ce faire, nous utilisons le concept de langage "pivot" qui permet d'articuler la description et la présentation formalisée.

This paper presents insights to rewrite in a "one level" formal language a part of French descriptive patrimony finally not much employed in NLP. We make use the concept of "pivot" language that allows to combine the description of a language and his formalized presentation.

1 Introduction

Le présent article présente un projet de recherche et a comme objectifs de montrer les avantages qu'apporterait un langage de traduction entre la linguistique "descriptive" et la linguistique "formelle" ainsi que de proposer une méthodologie apte à atteindre cet objectif (d'où l'aspect provisoirement théorique). La linguistique descriptive désigne une discipline reposant sur des critères et une méthodologie qui lui sont propres, par opposition à la linguistique formelle dérivant d'autres disciplines telles que la logique ou les mathématiques. La linguistique descriptive s'appuie, selon (Dixon, 1997 : 128-131), sur une "Théorie Linguistique de Base" qui constitue *"the fundamental theoretical concepts that underlie all work in language description..."* (Dixon, 1997 : 128), Théorie que l'on ne retrouve pas forcément dans les formalismes ou de manière fragmentaire.

La description et la formalisation ne s'excluent pas mais se complètent et, selon nous, elles doivent constituer deux étapes nécessairement distinctes car *"En toute rigueur théorique, il convient donc de distinguer la description des connaissances (notamment linguistiques), l'expression de ces connaissances dans un formalisme, et l'élaboration de techniques et de*

¹ Nous employons le terme de "que-phrases" (désormais que-P) pour ne pas présumer de la nature ni de la fonction du constituant. Le phénomène que nous allons étudier ici se rapproche de ce que l'on appelle traditionnellement les "complétives".

stratégies informatiques de traitement effectif. Dans la pratique toutefois [...], les traitements automatiques ne mettent pas toujours en œuvre ces trois types d'outils" (Fuchs et al., 1993 : 12). Cette citation met en évidence un paradoxe intéressant qui se trouve au cœur du TALN, en tant que discipline au confluent de la description et de la formalisation. En effet, le nombre de descriptions linguistiques ne cesse de grandir, sans qu'il y ait vraiment de compilations faites par les descriptivistes ni de récupération systématique des descriptions par les formalistes². Or, le recueil et le filtrage des descriptions sont des travaux qu'il est important de faire dans un but de cumul des connaissances. Une traduction intermédiaire dans un langage que nous qualifions de "langage pivot" est une première étape vers la formalisation des descriptions. Un double bénéfice résulterait de l'élaboration et de l'utilisation de ce langage : pour la linguistique, l'évaluation sur de vastes données authentiques des propriétés dégagées et pour les formalistes, l'existence d'une base de données pour la constitution de grammaires formelles.

Nous présenterons brièvement (partie 2) la méthodologie proposée. Puis, nous illustrerons (partie 3) son utilisation par l'intermédiaire d'une étude sur les comportements des verbes dominant des que-P en position objet et les incises. Cette étude concerne notamment les verbes de "dire" ainsi que *croire, falloir, penser, savoir, trouver, voir, vouloir...*

Nous avons choisi les que-P car cela recoupe un problème largement présent dans la littérature et qui a été traité à l'aide de cadres formels. Cela nous permettra d'effectuer une étude comparative. Nous montrerons que notre approche permet de tenir compte de distinctions empiriquement motivées qui peuvent être masquées dans le formalisme utilisé, celui-ci ne disposant pas forcément d'outils aptes à révéler de telles propriétés.

2 Méthodologie

Notre méthodologie se propose de faire l'interface entre deux approches. En effet, le fait que les formalistes utilisent des formalismes de plus en plus contraignants et sophistiqués n'impose-t-il pas de refaire partiellement ou entièrement une description à chaque formalisme créé ? De leur côté, les descriptivistes se préoccupent finalement peu de la formalisation et de l'implémentation qui pourraient être faites de leurs analyses et s'en tiennent le plus souvent à une description "morcelée" de la langue, sans forcément avoir une visée d'uniformité. Cet état de fait vient sans doute de la nature même de ces disciplines. Pour éviter un tel paradoxe et garder la spécificité de chacune des disciplines, nous proposons de formaliser les propriétés mises en évidence par des descriptions. Nous pouvons résumer notre méthode globale comme suit (nous nous centrerons seulement sur les deux premières étapes).

² Nous utilisons les raccourcis de "descriptiviste" pour linguiste travaillant sur la description des langues en dehors d'un formalisme et "formaliste" pour linguiste travaillant à partir d'un formalisme.

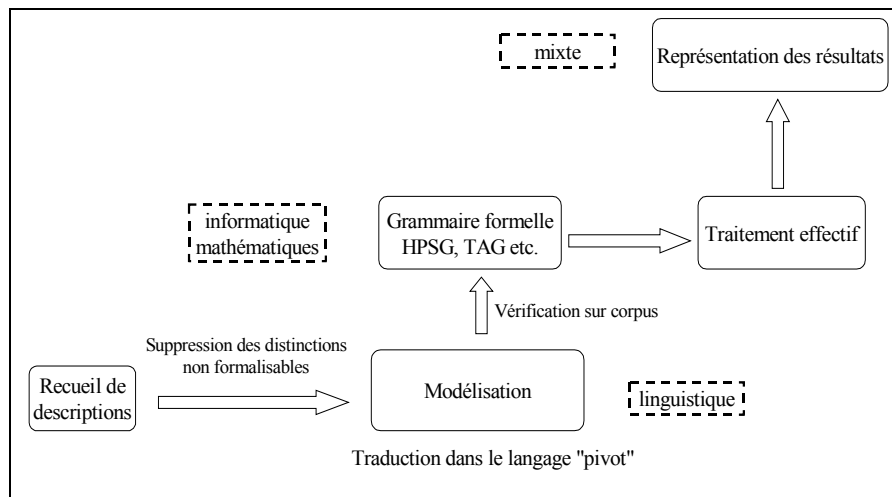


Figure 1 : Etapes successives de la description à la formalisation

3 Application de la méthode

En (3.1), nous évoquons brièvement deux analyses concernant les "complétives" : (Abeillé, 2002) et (Blanche-Benveniste, 1989). En (3.2) et (3.3), nous décrivons les propriétés traduites dans le langage pivot.

3.1 Présentation des analyses

(Abeillé, 2002 : 128) considère "les complétives comme des phrases de base et les "principales" comme des ajouts", ce qu'illustre l'arbre de dérivation suivant³ (p.117) :

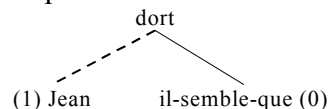


Figure 2 : arbre de dérivation de l'énoncé *Il semble que Jean dort*

Une alternative à cette analyse uniforme des complétives est celle que propose notamment (Blanche-Benveniste, 1989) qui distingue deux fonctionnements. En effet, une série de tests révèle l'existence d'au moins deux types de verbes introducteurs de que-P : les verbes "recteurs faibles" (non constructeurs comme ceux de (Abeillé, 2002)) et "recteurs forts" (constructeurs). Plus un autre fonctionnement associé aux recteurs faibles : les incises. Ces trois constructions sont illustrées par les exemples suivants :

- 1) Recteur fort : *il a prouvé qu'il était brave.* => *il l'a prouvé, qu'il était brave.*
- 2) Recteur faible : *il paraît que tu es parti hier.* => **il le paraît, que tu es parti.*
- 3) Incise : *"je reviendrai demain", a-t-il dit.* => **"je reviendrai demain", l'a-t-il dit.*

Certains verbes peuvent avoir les deux fonctionnements (*croire*), d'autres ont uniquement le statut de recteur fort (*souhaiter, prouver*) ou faible (*paraître*). Les verbes ne pouvant être que recteurs forts régissent notamment une que-P proportionnelle à un pronom clitique (ex. 1) et ils ne peuvent pas apparaître en incise.

- 4) *Il a prouvé qu'il était brave.* => *?Il était brave, il a prouvé.*

Ceci montre bien que certains verbes peuvent construire des que-P et ne sont donc pas des "ajouts", ce qui oblige à nuancer l'analyse de (Abeillé, 2002) qui pourtant énumère de telles propriétés.

Les recteurs faibles, quant à eux, sont caractérisés par une absence de proportionnalité entre la que-P et un clitique. Cela induit le fait que le complément "apparent" du verbe est grammati-

³ Pour une présentation des arbres dérivés et de dérivation, voir (Abeillé, 2002).

calement "informe" : il n'a donc pas les propriétés d'un complément. Cette propriété se retrouve dans les incises qui, de plus, ont leur verbe bloqué dans la modalité assertive :

- 5) *Il est parti, je pense.* => **il est parti, je ne pense pas.* => ? *il est parti, je le pense.*

Dans ce cas, il est clair que le verbe n'est pas recteur. Il y a donc bien au moins deux fonctionnements distincts pour les verbes introducteurs de que-P. Mais si l'on entre dans les détails, le problème se révèle encore plus ardu. En effet, Blanche-Benveniste pense que les "recteurs faibles" ont de nombreuses propriétés qu'il serait très difficile de hiérarchiser. Ainsi, une étude descriptive plus précise serait nécessaire, ce qu'illustre la partie suivante.

3.2 Dérivation des propriétés linguistiques

Nous allons maintenant évoquer le passage de la description linguistique au langage pivot. Une propriété importante est illustrée par le subjonctif. En effet, les verbes contraignant le verbe présent dans la que-P à être au subjonctif sont rarement recteurs faibles et n'ont pas la possibilité d'être incise. D'où la règle : *Subjonctif dans la que-P = indice de rection forte.*

Des logiciels d'étiquetage morpho-syntaxique tel Cordial 8⁴ permettent de repérer ce type de propriété lié à la flexion du verbe, ce qui en fait un critère apte à un traitement automatique. De même, le blocage du pronom sujet sur une personne en particulier est un indice opératoire.

- 6) *Je crois bien qu'il va neiger.* => **Tu crois bien qu'il va neiger.*

En revanche, le blocage des modalités n'est pas une propriété facile à traduire dans le langage pivot. En effet, une modalité particulière n'est pas l'indice d'une rection faible.

- 7) *Il n'empêche pas que la fenêtre soit ouverte.*

- 8) *Il n'empêche que la fenêtre est ouverte.*

Ce qui fait du verbe dans l'exemple (8) un recteur faible, c'est l'absence de la particule de négation *pas* et l'absence du subjonctif dans la que-P. On remarquera aussi le sujet bloqué à l'impersonnel. En revanche, l'absence de la modalité négative est l'indice d'une rection forte dans le cas du verbe *empêcher*. Pour traduire le blocage des modalités, il faut donc constituer des entrées répertoriant un certain nombre de propriétés des verbes. Par exemple, deux entrées simplifiées se présenteraient sous la forme :

<i>empêcher</i> Vnc	mod. nég ! <i>pas</i>	<i>empêcher</i> Vc	si mod. nég. alors <i>pas</i>
	mode = ind.		mode = subj.
	sujet = <i>il</i> impers.		sujet = ! <i>il</i> impers.

On peut donc traduire cette propriété dans le langage pivot moyennant la constitution d'entrées. De plus, certaines propriétés étant inhibées à cause de l'interaction avec la construction à objet de type SN, ces entrées nous permettront de donner des restrictions sur la bonne application des propriétés. Du reste, c'est la solution envisagée par la plupart des formalismes contemporains avec le recours aux structures de traits et à la lexicalisation des grammaires. Une dérivation est donc facile à proposer.

3.3 Modélisation dans le langage pivot

Une correspondance bi-univoque entre une description et un cadre formel étant difficile à obtenir, il semble nécessaire de passer par une sorte de dictionnaire de traduction : ce dictionnaire est ce que nous avons appelé le "langage pivot". Il doit tendre vers une neutralité terminologique et être apte à modéliser le plus de propriétés linguistiques possibles. Dans le cas qui nous occupe ici, une partie du langage pivot est de la forme⁵ : soit des énoncés du type CV₁-que-CV₂ (que-P) ou CV₂-CV₁ (incise).

⁴ Logiciel développé par la société Synapse Développement.

⁵ Comme nous l'avons signalé précédemment, il serait nécessaire de faire des entrées pour chacun des verbes concernés. Ne disposant pas de l'espace nécessaire, nous ne pouvons les reproduire ici.

construction	propriétés
Vc	$V \rightarrow \text{bien}$ $\text{mode}(V2) = \text{subj.}$ $O(V1) = \text{que-P}$ $V1 > \text{que-P}$ $\text{mode}(V2) = \text{sél.}$ $\text{aspect}(V2) = \text{sél.}$ $\text{tps}(V2) = \text{sél.}$
Vnc	$V1 > \text{que-P}$ $\text{mode}(V2) \neq \text{subj.}$ $\text{mode}(V2) = \text{indic.}$ $\text{mod.}(V1) = \text{const.}$ $\text{subj}(V1) = \text{bloc.}$ $\text{tps}(V1) = \text{bloc.}$ $O(V1) = \text{que-P}$ $V > \text{bien}$
Vi	$CV2 > CV1 \mid CV2 > CV1 > CV2$ $\text{subj.} > V1 \mid V > \text{subj.}$ $\text{mod}(V1) = \text{assert.}$ $!O(V1)$ $\text{tps}(V1) \neq \text{subj.}$

Figure 3 : traduction des propriétés linguistiques dans le langage pivot

On peut donc traiter des exemples du type :

- 9) *Je ne regrette pas qu'il soit parti* avec $\text{mode}(V2) = \text{subj.}$; $O(V1) = \text{que-P}$;
 $\text{mod.}(V1) = \text{nég.}$ d'où *regrette* = Vc.

4 Conclusion et perspectives

En confrontant deux analyses, nous avons pu mettre en évidence l'intérêt d'une approche descriptive indépendante précédant l'approche formelle. On pourrait appliquer les distinctions opérées dans un formalisme (en TAG, par exemple) :

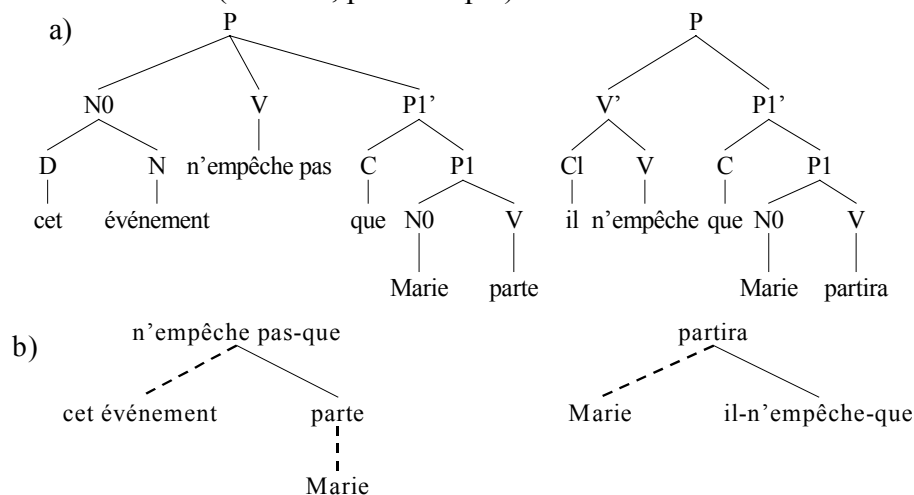


Figure 4 : arbres dérivés (a) et de dérivation (b) pour deux constructions du verbe *empêcher-que*

Dans ce contexte, les propriétés mises en évidence par une approche descriptive peuvent donc être systématisées, réfutées, reproductibles et passées à l'épreuve de larges corpus. Cette méthode serait donc un bon moyen pour l'évaluation des descriptions et leur intégration dans des grammaires formelles. Mais c'est seulement lorsque nous aurons pris en compte plus de phénomènes, étendu le langage pivot et implémenté effectivement les propriétés que nous pourrions conclure quant à la fécondité d'une telle approche.

Références

- Abeillé A. (2002), *Une grammaire électronique du français*, Paris, CNRS éditions.
Andersen H.L. (2000), Discours rapporté en français parlé, Actes du colloque international Université de Copenhague 1998, Etudes Romanes N°47, pp. 143-155.
Blanche-Benveniste Cl. (1989), Constructions verbales "en incise" et rection faible des verbes, *Recherches sur le français parlé*, Vol. 9, pp. 53-73.

- Blanche-Benveniste Cl., Deulofeu J., Stéfani J., van den Eynde K. (1984), *Pronom et syntaxe. L'approche pronominale et son application à la langue française*, Paris, SELAF.
- Boone A. (1996), *Les complétives et la modalisation*, Muller (éd.), pp. 45-51.
- Dixon R.M.W. (1997), *The rise and Fall of languages*, Cambridge University Press.
- Fuchs C., Danlos L., Lacheret-Dujour A., Luzzati D., Victorri B. (1993), *Linguistique et traitements automatiques des langues*, Hachette Université.
- Lebarbé T. (2002), *Hiérarchie Inclusive des Unités Linguistiques en Analyse Syntaxique Coopérative*, Thèse de Doctorat, Université de Caen.
- Vergne J. (2001), *Analyse syntaxique automatique de langue : du combinatoire au calculatoire*, Actes de TALN 2001, pp. 15-29.

Annexe

Lexique non exhaustif du langage pivot

Vc	verbe constructeur	!	non	mod.	modalité
Vnc	verbe non constructeur		ou	que-P	que-phrasé
Vi	verbe en incise	O(V1)	objet de V1	>	suivi de

Liste des 10 verbes les plus fréquents ayant une que-P en position objet

Ce dénombrement a été entrepris sur le corpus de français parlé CORPAIX⁶ (1.228.645 occ.)

dire	924	trouver	288
penser	839	voir	191
croire	609	vouloir	101
savoir	440	apercevoir	52
falloir	381	comprendre	51

⁶ CORPAIX est un corpus de français parlé constitué par le GARS à l'Université de Provence.