

# Conjonctions de subordination, verbes de dire et d'attitude propositionnelle : une modélisation STAG pour le discours

Timothée Bernard  
Université Paris Diderot - ALPAGE  
timothee.bernard@inria.fr

## RÉSUMÉ

---

Nous proposons une nouvelle modélisation en grammaire d'arbres adjoints synchrone (STAG) syntaxe/sémantique pour les conjonctions de subordination (ConjSub) et les verbes de dire et d'attitude propositionnelle (VAP; *dire, penser, croire*, etc.). Cette modélisation, plus riche que les modélisations traditionnelles, est conçue pour l'analyse du discours et fondée sur l'observation que ces deux catégories sont loin d'être homogènes. En effet, des travaux antérieurs ont montré d'une part que les occurrences de ConjSub pouvaient être divisées en deux classes aux propriétés syntaxiques et sémantiques différentes, d'autre part que les VAP présentaient en discours deux usages distincts : *évidentiel* et *intentionnel*. Notre proposition vise donc à rendre compte précisément de ces différences tout en modélisant les interactions entre VAP et ConjSub.

## ABSTRACT

---

**Modelling Subordinate Conjunctions, Attitude Verbs and Reporting Verbs in STAG: a Discourse Perspective**

We propose a new model for subordinate conjunctions (SConjs) and attitude verbs and reporting verbs (AVs) in STAG syntax and semantics. This model is designed for discourse parsing, and is based on the observation that SConjs and AVs are far from homogeneous categories. Indeed, previous work has shown that SConjs can be divided into two classes according to their syntactic and semantic properties. Similarly, AVs have two different uses in discourse: *evidential* and *intentional*. Our proposition aims at representing these distinctions and modelling their discourse-related interactions.

**MOTS-CLÉS** : discours, STAG, conjonction de subordination, verbe d'attitude propositionnelle, verbe de dire, interface syntaxe/sémantique.

**KEYWORDS** : discourse, STAG, subordinate conjunction, attitude verb, reporting verb, syntax-semantics interface.

---

## 1 Introduction

Sont observées entre les différentes phrases d'un texte d'une part des marques de cohésion, comme les anaphores, et d'autre part des liens de cohérence – les *relations discursives* (RD) –, qui permettent au discours d'être interprété. Il y a débat sur le nombre et la nature des RD, il est cependant possible de nommer parmi les plus classiques *Contrast*, *Consequence* ou *Explanation* (Asher & Lascarides, 2003). Nous considérons ici que les RD sont des prédicats à deux arguments, structurant les phrases à

l'échelle discursive (textuelle ; voir (1a)) mais intervenant aussi à l'échelle phrastique (1b). Dans ces deux exemples, la relation *Consequence* est *explicite*, c'est-à-dire signalée lexicalement, mais une RD peut aussi être *implicite*, c'est-à-dire inférée sémantiquement. C'est ce que nous observons par exemple lorsque nous supprimons *donc* de (1b) pour obtenir (1c)<sup>1</sup>.

- (1) a. *Fred était malade. Par conséquent, il est resté chez lui.*
- b. *Fred était malade, il est donc resté chez lui.*
- c. *Fred était malade, il est resté chez lui.*

*Donc* ou *par conséquent* sont des *connecteurs du discours*, une classe d'éléments lexicaux qui ont justement pour fonction de signaler la présence d'une RD entre deux parties de texte<sup>2</sup>. Comme le relève (Roze *et al.*, 2012), les connecteurs discursifs du français regroupent différentes catégories syntaxiques. Parmi celles-ci, ce sont les conjonctions de subordination (ConjSub) qui vont nous intéresser plus particulièrement. Les ConjSub sont généralement considérées comme une catégorie homogène, alors que l'on peut en distinguer deux classes aux propriétés syntaxiques et sémantiques bien distinctes, comme la possibilité ou non de clivage illustrée en (2) ou les différences de portée observées en (3).

- (2) a. C'est quand **il avait vingt ans** que *Fred est allé au Brésil*.
- b. #C'est bien qu'il n'avait pas soif que Fred a repris un verre.
- (3) a. Il n'est pas venu parce qu'il avait faim : il est venu parce qu'il avait soif.
- b. #Il n'est pas venu bien qu'il ait du travail : il est venu bien qu'il soit fatigué.

Outre les ConjSub, nous nous intéressons aux *verbes de dire* et d'*attitude propositionnelle* (VAP) – ces verbes comme *dire*, *affirmer* ou *croire* et *savoir* qui peuvent indiquer une action, mais surtout décrivent la position d'un agent sur une proposition donnée. Les VAP présentent des interactions particulières avec les connecteurs discursifs en général et les ConjSub en particulier. Le but de ce travail est une modélisation de ces différents éléments dans une grammaire d'arbres adjoints synchrones (STAG, (Shieber & Schabes, 1990)). En effet, les différentes modélisations en STAG proposées jusqu'ici (par exemple (Nesson & Shieber, 2006), ou (Danlos, 2009) qui s'intéresse plus spécifiquement à l'analyse du discours) n'intègrent pas les propriétés mentionnées dans cet article. La portée de ce travail, qui contribue à l'étude des formalismes nécessaires à l'analyse discursive, est essentiellement théorique. Néanmoins, la capacité à traiter de manière précise la sémantique des connecteurs discursifs et des VAP s'avère indispensable à tout potentiel analyseur discursif opérationnel<sup>3</sup>.

Nous commençons par présenter à la section 2 différents travaux mettant en évidence les aspects des ConjSub et des VAP que nous souhaitons modéliser. À la section 3, nous utilisons des tests linguistiques pour déterminer plus précisément les interactions des ConjSub avec les VAP ainsi que d'autres opérateurs à portée. Enfin, la section 4 décrit notre proposition en STAG. La section 5 conclut cet article et indique certaines pistes de recherche future.

1. Nous suivons dans ce texte des conventions inspirées du Penn Discourse Treebank (PDTB, (Prasad *et al.*, 2007)) : dans les exemples les expressions lexicalisant les RD seront soulignés, l'un de leur argument, Arg<sub>1</sub>, sera présenté en italique et l'autre, Arg<sub>2</sub>, en gras.

2. Il existe d'autres marqueurs plus complexes formant une catégorie ouverte appelée « AltLex » (pour « alternative lexicalization » (Prasad *et al.*, 2007)).

3. Notons à cet égard que près d'une phrase sur deux dans le PDTB contient un connecteur discursif et qu'environ un tiers des RD (en incluant les implicites) sont attribuées au moins partiellement à un agent autre que l'auteur (Prasad *et al.*, 2008).

## 2 Travaux pertinents et prérequis

### 2.1 Problèmes à l'interface syntaxe-sémantique

Pour (Dinesh *et al.*, 2005) qui travaille en anglais et (Danlos, 2013) en français, dans une phrase comme (4a) l'auteur entend exprimer une opposition entre sa croyance et celle de Sabine, d'où l'inclusion de *Sabine pense qu'* dans l'Arg<sub>2</sub> de la relation *Concession* lexicalisée par *même si*. À l'inverse, dans (4b) l'auteur ne cherche pas essentiellement à opposer sa croyance à celle de Sabine, mais plutôt le voyage de Fred au Pérou à son absence (supposée) de séjour à Lima, d'où l'exclusion de *Sabine pense qu'* de l'argument de la RD. Dans ce dernier cas, on observe donc que l'argument sémantique de la RD est inclus sans être égal au contenu propositionnel de l'argument syntaxique du connecteur.

Ce type de problèmes à l'interface syntaxe-sémantique (que nous appellerons dans la suite de cet article des « disparités ») apparaît régulièrement en présence de verbes de dire ou d'attitude propositionnelle (VAP) : quand un VAP constitue, avec la complétive qu'il introduit, l'argument syntaxique d'un connecteur, il peut (4a) ou non (4b) être inclus dans l'argument de la relation correspondante. En reprenant la terminologie de (Asher *et al.*, 2006), nous disons que le VAP est (en usage) *intentionnel* dans le premier cas et *évidentiel* dans le second. Remarquons que le cas évidentiel – lorsqu'il est possible – est équivalent à l'expression *d'après Sabine*.

- (4) a. *Fred est allé au Pérou même si Sabine pense qu'il n'a jamais mis les pieds en Amérique.*  
 b. *Fred est allé au Pérou même si Sabine pense qu'il n'est jamais passé par Lima.*

Il apparaît cependant, comme le fait remarquer (Hunter & Danlos, 2014), que tous les connecteurs ne peuvent pas donner lieu à ces disparités. La ConjSub *parce que*, par exemple, n'est pas compatible avec l'usage évidentiel des VAP (5). (Hunter & Danlos, 2014), reprenant la taxinomie du PDTB, propose l'hypothèse que seuls les connecteurs portant une relation appartenant aux classes COMPARISON (comme *même si*) et EXPANSION (comme *de plus*) autorisent l'usage évidentiel en plus de l'usage intentionnel, contrairement à ceux portant une relation de type TEMPORAL (comme *ensuite*) et CONTINGENCY (comme *parce que*) qui n'autorisent que l'usage intentionnel.

- (5) a. *Fred ne peut pas venir parce qu'il n'est pas en ville.*  
 b. #*Fred ne peut pas venir parce que Sabine pense qu'il n'est pas en ville.*

### 2.2 Deux types de propositions adverbiales

Une proposition adverbiale est une proposition subordonnée qui, comme un adverbe, a une fonction de « complément circonstanciel ». C'est le cas d'*après qu'il a mangé* (rôle temporel) et *bien qu'il n'ait pas mangé* (rôle concessif) dans (6).

- (6) a. *Fred est parti après qu'il a mangé.*  
 b. *Fred est parti bien qu'il n'ait pas mangé.*

En se basant sur un grand nombre de phénomènes – coordination, portée, ellipse, ou liés à la prosodie, la typographie ou encore la présence d'ambiguïtés – (Haegeman, 2004) propose une

partition des propositions adverbiales en deux classes : les *propositions centrales* (PropC) d’une part et les *propositions périphériques* (PropP) d’autre part, ces dernières étant syntaxiquement moins intégrées à leur principale et à rapprocher des coordinations. D’un point de vue logico-sémantique, les PropC apportent une information (de temps, de causalité, etc.) sur l’éventualité décrite par la proposition matrice alors que les PropP ont une fonction plus pragmatique de structuration du discours (concession, information préliminaire, etc.). Par exemple en (7), *quand il avait vingt ans* et *s’il fait beau* sont des propositions centrales, elles peuvent donner lieu à un clivage (voir (2)) contrairement à *quand Sabine n’a jamais pu y aller* et *s’il fait beau* en (8) qui sont des propositions périphériques.

- (7) a. *Fred est allé au Brésil quand il avait vingt ans.*  
 b. *S’il fait beau, j’irai dehors.*
- (8) a. *Fred est allé au Brésil quand Sabine n’a jamais pu y aller.*  
 b. *S’il fait beau, pourquoi ne vas-tu pas dehors ?*

Indiquons dès à présent que (Haegeman, 2004) observe qu’une négation ou une interrogation contenue dans une principale peut avoir portée sur une proposition centrale mais pas sur une proposition périphérique. Il est aussi pertinent de remarquer qu’une proposition centrale ne peut pas contenir de modal épistémique *speaker-oriented* (c’est-à-dire dont le point de référence est l’auteur, comme ce serait le cas en (9a) mais pas en (9c) où *may* est surtout « John-oriented »), alors que cela est tout à fait possible pour une proposition périphérique (9b)<sup>4</sup>.

- (9) a. #Mary accepted the invitation without hesitation after John may have accepted it. (Haegeman, 2004)  
 b. *The ferry will be fairly cheap, while/whereas the plane may/will probably be too expensive.* (Haegeman, 2004)  
 c. *John is worried because he may be ill.*

### 3 Tests de projections et conjonctions de subordination

Afin de modéliser les ConjSub, nous cherchons à mettre en évidence leurs interactions avec les autres composants de la phrase. C’est dans ce but que nous les étudions dans le cadre de cinq schémas<sup>5</sup> qui vont nous permettre de déterminer la portée de différents opérateurs en présence de ConjSub. Notons que certaines traductions des schémas originaux (en anglais) forcent l’utilisation du subjonctif ; les formulations choisies ici tentent de limiter ce phénomène, qui n’apparaît pas en anglais. Ainsi, « It is not the case that A » devient « Non,  $\neg A$  » et « It is possible that A » devient « Peut-être que A » :

- Négation : Non,  $\neg A$ .  
 Interrogation : Est-ce que A ?  
 Conditionnel : Si A, B.  
 Épistémique : Peut-être que A.  
 VAP : Sabine pense que A.

4. Ci-dessous la traduction des exemples en (9) :  
 #Mary accepta l’invitation sans hésitation après que John l’a peut-être acceptée.  
 Le ferry sera plutôt abordable, alors que l’avion sera probablement trop onéreux.  
 John est inquiet parce qu’il est peut-être malade.

5. Ces schémas sont notamment utilisés pour tester les propriétés de projection (Chierchia & McConnell-Ginet, 1990).

Dans ces schémas, nous remplaçons A par un énoncé de la forme «  $A_1$  CONJ  $A_2$  », où CONJ est une conjonction de subordination lexicalisant une RD  $R$  et nous nous demandons si  $\text{Arg}_1$ ,  $\text{Arg}_2$  et  $R(\text{Arg}_1, \text{Arg}_2)$  sont logiquement impliqués par la phrase. Notons que nous n'imposons pas de contrainte sur la structure syntaxique des énoncés ainsi produits ; nous ne savons donc *a priori* pas si A forme un unique constituant ou bien plusieurs constituants diversement rattachés au reste de la phrase. Pour des raisons de concision, nous n'illustrons dans ce travail que les résultats pour *parce que* et *bien que* – introduisant respectivement une PropC et une PropP – mais les résultats présentés peuvent facilement être étendus aux autres ConjSub.

**Parce que (Explanation) :** Les énoncés en (10) sont le résultat des schémas de négation et d'interrogation appliqués à une occurrence de *parce que* lexicalisant *Explanation*.

- (10) a. Non, *Fred n'était pas absent* parce qu'il était malade.  
 b. Est-ce que Fred était absent parce qu'il était malade ?

Une interprétation locale de la négation en (10a) (ayant seulement portée sur *Fred était absent*) ne serait pas logiquement cohérente ; elle doit donc avoir une portée globale, c'est-à-dire sur la principale et sa subordonnée. Cette analyse est compatible avec (Haegeman, 2004), *parce que* introduisant une PropC. Étudions alors quelles continuations débouchent sur un discours cohérent<sup>6</sup>. Toutes les continuations en (11) sont possibles et décrivent des situations différentes<sup>7</sup> :

- avec (11a), ni  $\text{Arg}_1$  (*Fred est absent*), ni  $\text{Arg}_2$  (*Fred est malade*), ni la relation (*Explanation*) ne sont vérifiés ;
  - avec (11b), seul  $\text{Arg}_1$  est vérifié ;
  - avec (11c), seul  $\text{Arg}_2$  est vérifié ;
  - avec (11d),  $\text{Arg}_1$  et  $\text{Arg}_2$  sont vérifiés, mais pas la relation *Explanation*.
- (11) a. Il était là et en pleine forme.  
 b. C'est seulement qu'il avait raté son train.  
 c. Il est venu quand même, bien que souffrant.  
 d. Il était effectivement malade, mais il serait quand même venu s'il n'avait pas dû réceptionner un colis important.

Nous observons donc que la négation a portée sur la totalité de  $\text{Arg}_1 \wedge \text{Arg}_2 \wedge R(\text{Arg}_1, \text{Arg}_2)$  et qu'aucun de ces trois éléments n'est sémantiquement impliqué par l'utilisation de *parce que*. Remarquons aussi que toutes ces continuations, si introduites par *Non*, constituent des réponses possibles pour la forme interrogative (10b). L'opérateur interrogatif partage donc avec la négation les mêmes conclusions.

Dans le même ordre d'idées, nous instancions le schéma conditionnel par la phrase (12), que nous pouvons faire suivre par chacune des phrases en (13). Nous retrouvons les quatre configurations précédentes : ni  $\text{Arg}_1$  ni  $\text{Arg}_2$  ni  $R$  n'est vérifié avec (13a), seul  $\text{Arg}_1$  est vrai en (13b), seul  $\text{Arg}_2$  en (13c),  $\text{Arg}_1$  et  $\text{Arg}_2$  sont vrais mais non  $R$  avec (13d).

6. Nous nous autorisons tout type de continuation, dialogue inclus, tant qu'aucune relation de correction ou d'expression de type *Hey, wait a minute* (Shanon, 1976) n'intervient.

7. Pour certains relecteurs, les continuations (11a) et (11c) ne sont pas acceptables. La négation et l'interrogation ignoreraient donc pour eux l' $\text{Arg}_1$  de *parce que*. Je pense que ces personnes accepteraient par contre ces continuations pour *Il est faux de dire que  $A_1$  parce que  $A_2$ . et On m'a dit que  $A_1$  parce que  $A_2$  ; est-ce vrai ?* Il s'agit à mon avis d'un phénomène de présupposition que je n'ai pas encore cherché à modéliser.

- (12) Si *Fred s'est vexé* parce que Sabine l'a taquiné, alors c'est qu'il en est secrètement amoureux.
- (13) a. Mais Fred est de très bonne humeur et je connais Sabine, elle n'est pas du genre à l'embêter.  
 b. Mais je n'ai pas entendu dire qu'elle l'ait taquiné.  
 c. Mais je n'ai pas entendu dire qu'il soit vexé.  
 d. Mais je ne pense pas que là soit la raison.

Le lecteur pourra vérifier que les mêmes conclusions s'appliquent à tous les autres schémas : l'opérateur correspondant peut avoir portée globale sur la totalité «  $A_1$  parce que  $A_2$  ».

**Bien que (Concession) :** La sémantique de *although* (*bien que* en français) a été l'objet de nombreuses discussions depuis (Frege, 1948). Pour (Potts, 2015), la relation *Concession* lexicalisée par la conjonction n'est pas *at-issue*, c'est-à-dire qu'elle ne constitue pas le message principal du locuteur mais plutôt une information qui doit être intégrée (si elle ne l'est pas déjà) par les interlocuteurs, sans avoir vocation à être débattue. Il apparaît notamment que cette relation est difficilement niable. Il est aussi intéressant de remarquer que la *Concession* est liée au locuteur : celui-ci ne peut employer *bien que* que s'il adhère lui-même à la relation, même si le connecteur est sous la portée d'un VAP (qui sont des *presupposition plugs*). Ainsi, *although* est souvent cité comme un déclencheur d'*implicature conventionnelle* depuis (Grice, 1975).

Remarquons tout d'abord que comme illustré en (14), il n'est pas possible d'appliquer le schéma négatif avec un énoncé «  $A_1$  bien que  $A_2$  » déjà sémantiquement cohérent. Pour produire un discours acceptable comme en (14c), il est nécessaire que la négation soit incluse dans  $Arg_1$ . Le fait que les discours tels que (14b) soient incohérents montrent qu'avec *bien que* – contrairement à ce que nous venons de voir pour *parce que* – une négation présente dans la principale a toujours une portée seulement locale. Ceci est conforme aux observations de (Haegeman, 2004), *bien que* introduisant une PropP.

- (14) a. *Fred a mangé de la viande* bien qu'il soit végétarien.  
 b. #Non, Fred n'a pas mangé de viande bien qu'il soit végétarien.  
 c. Non, *Fred n'a pas refusé de manger de la viande* bien qu'il soit végétarien.

Considérons maintenant (15a) suivant le schéma interrogatif. Alors que la phrase en (15b) constitue une réponse parfaite, celle en (15c), niant  $Arg_2$  au lieu de  $Arg_1$ , n'est pas acceptable<sup>8</sup>. Ceci tend à montrer que l' $Arg_2$  de *bien que*, en plus de la relation *Concession* elle-même, n'est pas *at-issue* et que l'interrogation porte uniquement sur le contenu de la principale.

- (15) a. A-t-il mangé de la viande bien qu'il se dise végétarien ?  
 b. Non, il a refusé.  
 c. #Non, il n'est plus végétarien.

Ce serait cependant faire un raccourci que d'affirmer que l' $Arg_2$  est toujours non *at-issue* et hors de la portée des opérateurs introduits dans les tests étudiés ici. La phrase (16a), construite sur le schéma avec VAP, peut être continuée par (16b) où l' $Arg_2$  est nié par l'auteur lui-même : c'est donc que l' $Arg_2$  est interprété comme sous la portée sémantique de *Sabine pense que*.

8. Il est effectivement possible de répondre à (15a) par *Il n'est plus végétarien*, mais il s'agit plus d'une remarque que d'une réponse : le dialogue pourrait se continuer par *Tu n'as pas répondu à ma question*.

- (16) a. Sabine pense que *Fred est venu travailler* bien qu'il soit malade.  
 b. Mais elle se trompe : il était déjà guéri depuis plusieurs jours.  
 c. Bien qu'il soit malade, Sabine pense que *Fred est venu travailler*.

Toutefois, il est vrai que hors contexte, (16a) semble intuitivement affirmer la maladie de Fred ; ce serait donc là une lecture par défaut. C'est pourquoi nous pensons qu'une telle phrase est ambiguë, l'Arg<sub>2</sub> pouvant être sous la portée sémantique du VAP ou non ; cette dernière interprétation étant la plus probable et choisie par défaut. Remarquons que seulement dans ce cas la subordonnée adverbiale est antéposable (sans changement de sens) comme en (16c).

**Résumé :** La table 1 compile ces résultats avec ceux concernant les ConjSub *après que* et *alors que*, non discutés ici par manque de place. Beaucoup de connecteurs sont ambigus entre différentes RD (voir *quand* dans les exemples (7) et (8), ou (Prasad *et al.*, 2008) pour une étude des ambiguïtés des connecteurs discursifs dans le PDTB), or nous savons déjà que le sens d'une ConjSub – selon qu'elle introduit une PropC ou une PropP – en modifie ses propriétés syntaxiques et sémantiques, c'est pourquoi nous précisons à chaque fois quel sens est considéré.

Il semble que les ConjSub introduisant une PropP (que nous appellerons dans la suite « conjonctions périphériques », ou ConjP) ont toutes le même comportement : elles sont compatibles avec les disparités entre syntaxe et sémantique concernant Arg<sub>2</sub>, l'auteur ne peut pas les utiliser sans accepter la relation qu'elles lexicalisent et dans une lecture par défaut (fortement préférée), l'auteur s'engage aussi sur Arg<sub>2</sub>. Les ConjSub introduisant une PropC (« conjonctions centrales », ConjC) partagent aussi un certain nombre de propriétés : elles ne sont pas compatibles avec les disparités sur leur Arg<sub>2</sub> et l'interprétation de la relation lexicalisée est toujours soumise aux opérateurs présents dans les schémas utilisés dans notre étude ; le statut de Arg<sub>2</sub> est lui variable.

<i>R</i> (type)	CONJ	Arg <sub>2</sub>	<i>R</i> (Arg <sub>1</sub> , Arg <sub>2</sub> )	disparité sur Arg <sub>2</sub>
<i>Concession</i> (périphérique)	bien que	(+)	+	+
<i>Contrast</i> (périphérique)	alors que	(+)	+	+
<i>Explanation</i> (centrale)	parce que	-	-	-
<i>Narration</i> (centrale)	après que	(+)	-	-

TABLE 1 – Les colonnes Arg<sub>2</sub> et *R*(Arg<sub>1</sub>, Arg<sub>2</sub>) expriment si ces propositions sont toujours impliqués par l'usage de CONJ dans les schémas étudiés : « - » signifie « non », « + » signifie « oui » et « (+) » signifie « oui dans la lecture par défaut » ; « + »/« - » dans la dernière colonne indique si CONJ est compatible ou non avec une disparité sur son Arg<sub>2</sub>. Il n'y a pas de colonne pour Arg<sub>1</sub> car l'interprétation de cet élément est toujours soumise aux opérateurs utilisés dans les différents schémas.

Avant d'exposer notre modélisation en STAG, nous tenons à insister sur le fait que c'est la relation lexicalisée qui détermine le caractère central ou périphérique d'une occurrence de ConjSub. Ceci nous amène à classer les RD en deux catégories : les *relations centrales* d'une part et les *relations périphériques* de l'autre. Cette classification est très proche de celle ébauchée par (Hunter & Danlos, 2014) mais repose sur des critères syntaxiques et sémantiques, et non plus sur la taxinomie du PDTB (établie dans un objectif d'annotation).

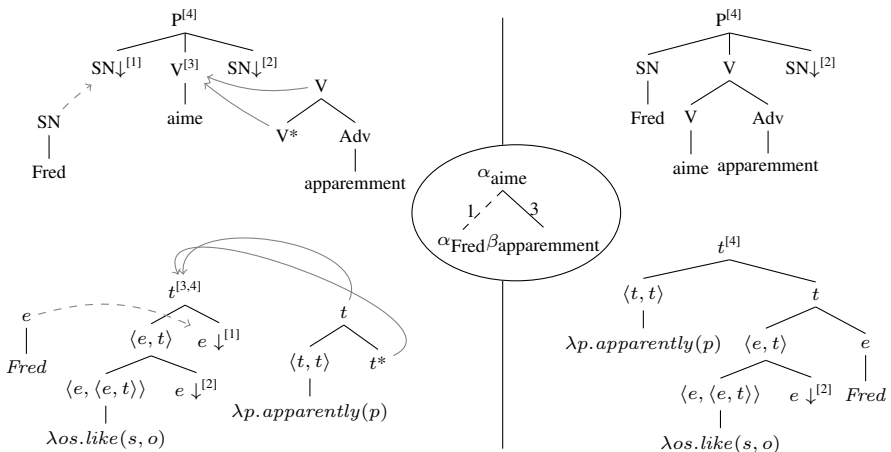


FIGURE 1 – STAG : à gauche, substitution de  $\alpha_{\text{Fred}}$  au lien [1] de  $\alpha_{\text{aime}}$  et adjonction de  $\beta_{\text{apparemment}}$  au lien [3]; entourée au milieu est l’arbre de dérivation ; à droite, les arbres dérivés syntaxiques et sémantiques.

## 4 Modélisation en STAG

Nous proposons dans cette section de nouvelles structures STAG pour représenter les VAP et les ConjSub, ainsi qu’une légère variation des structures phrastiques traditionnelles. Ces modifications permettent de rendre compte des disparités syntaxe/sémantique ainsi que des phénomènes de portée mis en lumière dans les sections précédentes.

### 4.1 Rappels sur STAG

Nous supposons ici que le formalisme des grammaires d’arbres adjoints (TAG) présenté dans (Joshi, 1987) est connu du lecteur. Les TAG synchrones (STAG) en sont une extension introduite par (Shieber & Schabes, 1990) qui utilise pour structures élémentaires des triplets  $(t_L, t_R, l)$ , où  $t_L$  et  $t_R$  sont deux arbres TAG et  $l$  est un ensemble de liens (notés [entre crochets]) entre un nœud de  $t_L$  et un nœud de  $t_R$ . La substitution et l’adjonction sont définies à partir de leur équivalent en TAG, avec la contrainte que chaque opération doit être effectuée de manière synchrone sur les deux arbres de chaque paire et sur des nœuds appariés. La figure 1 présente une substitution et une adjonction sur des structures STAG syntaxe/sémantique, qui permettent de calculer simultanément la syntaxe et la sémantique d’un énoncé.

Remarquons dès à présent que si un arbre est substitué dans un autre, alors toute adjonction sur le premier ne pourrait modifier que localement sa sémantique, sans avoir portée sur le second. À l’inverse, si un arbre est modifié par un arbre auxiliaire, alors une adjonction sur le premier peut très bien avoir une portée globale incluant le second, comme se serait le cas en figure 1 d’une adjonction supplémentaire sur le lien [4] de  $\alpha_{\text{aime}}$ , qui aurait  $\beta_{\text{apparemment}}$  dans sa portée. C’est justement cette asymétrie entre substitution et adjonction que nous allons exploiter pour modéliser les ConjSub, après avoir exposé notre proposition pour les VAP.



## 4.2 Verbes d’attitude propositionnelle et verbes de dire

En TAG, les VAP – comme les autres verbes ponts – sont usuellement représentés par un arbre auxiliaire venant s’adjoindre sur le nœud P qui enrachine la complétive introduite (Joshi, 1987). Une telle représentation permet de rendre compte naturellement des phénomènes d’extraction non bornée qui apparaissent dans des phrases comme *C’est l’homme que Jean croit que [...] Paul a dit que Marie aime* et confère à *SN V que* le même statut que d’après *SN*. Mais cette équivalence n’a pas lieu d’être pour un VAP en usage intentionnel : un tel VAP décrit l’état ou l’action (de penser, de dire, etc.) qui est argument de la RD, la complétive précisant alors un élément central de cette éventualité mais ne constituant pas l’éventualité elle-même. Le VAP intentionnel n’est pas sémantiquement un modifieur de sa complétive, contrairement au VAP évidentiel.

C’est pourquoi nous proposons une nouvelle paire d’arbres initiaux en addition de la paire auxiliaire traditionnelle (figure 2). La sémantique des deux structures est légèrement différente : un VAP évidentiel utilise un prédicat – noté ici sans apostrophe contrairement au prédicat intentionnel – qui est « effacé » lorsqu’il apparaît dans une RD. Ceci s’effectue *via* un ensemble de règles de réécriture de la forme  $\text{Contrast}(p, \text{think}(a, q)) \rightarrow \text{Contrast}(p, q)$ . Inversement, les disparités non naturelles sont évitées en supprimant toute analyse dans laquelle un prédicat évidentiel est argument d’une RD centrale :  $\text{Explanation}(p, \text{think}(a, q)) \rightarrow \perp$ . Ce sont ces règles qui permettent d’obtenir les analyses sémantiques correctes tout en rendant compte des possibilités de disparité.

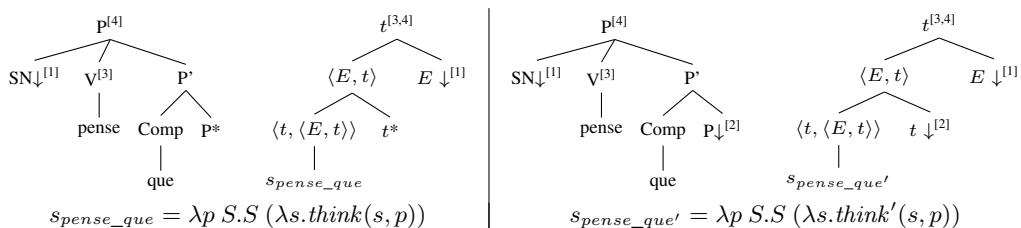
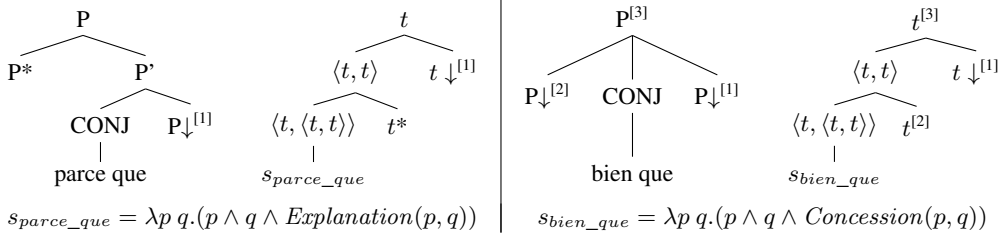


FIGURE 2 – VAP :  $\beta_{\text{pense que}}$  (évidentiel ; à gauche) et  $\alpha_{\text{pense que}}$  (intentionnel ; à droite)

## 4.3 Conjonctions de subordination

Similairement, les différences de syntaxe et de sémantique entre propositions centrales et propositions périphériques peuvent être expliquées à l’aide de structures différentes pour les ConjC et les ConjP comme en figure 3<sup>9</sup>. L’utilisation chez les ConjP d’un nœud à substitution même pour l’argument gauche, à la place du nœud d’adjonction présent chez les ConjC, permet de garantir que tout modifieur (tel qu’une négation) ne peut être que local, c’est-à-dire ne peut avoir portée sur la totalité de la conjonction  $\text{Arg}_1 \wedge \text{Arg}_2 \wedge R(\text{Arg}_1, \text{Arg}_2)$ , alors que cela est possible avec les ConjC. Un nœud d’adjonction, au contraire, laisse ouverte cette possibilité, comme nous l’avons rappelé précédemment.

9. La présence du nœud P’ pour les ConjC est due à la possibilité de clivage (*C’est parce que A que B*), soulignant l’existence d’un tel constituant. De telles phrases clivées n’apparaissent pas avec les ConjP.


 FIGURE 3 –  $\beta_{parce\ que}$  (ConjC; à gauche) et  $\alpha_{bien\ que}$  (ConjP; à droite)

## 4.4 Structures phrastiques

En nous inspirant de (Nesson & Shieber, 2006), nous considérons des structures verbales possédant deux sites d’adjonction différents pour les modifieurs verbaux (comme la négation ou les adverbes) et les verbes ponts (typiquement les VAP). Dans un formalisme où les adjonctions multiples sont licites et utilisées pour représenter certaines ambiguïtés de portée, ceci permet d’éviter les interprétations (non naturelles) où un modifieur verbal aurait portée sur un verbe pont. Cependant, alors que les ConjC se présentent comme des modifieurs phrastiques, en ce point similaires aux verbes ponts, nous avons vu que leur portée rentrait en concurrence avec celle des modifieurs verbaux. C’est pourquoi sur une structure comme celle de la figure 4.a, nous distinguons un autre lien (le lien [3]) attaché syntaxiquement à la racine P mais sémantiquement au même niveau que les modifieurs verbaux (et pouvant de ce fait générer une ambiguïté). Un système de traits permet d’interdire ce nouveau lien [3] aux VAP et de leur réserver le précédent (le lien [2]).

La figure 4.b montre les arbres de dérivations obtenus par l’adjonction d’une négation (ou de n’importe quel modifieur verbal) sur la principale d’une ConjSub. On observe alors que la négation peut avoir portée locale *ou* globale dans le cas d’une ConjC, mais seulement locale dans le cas d’une ConjP.

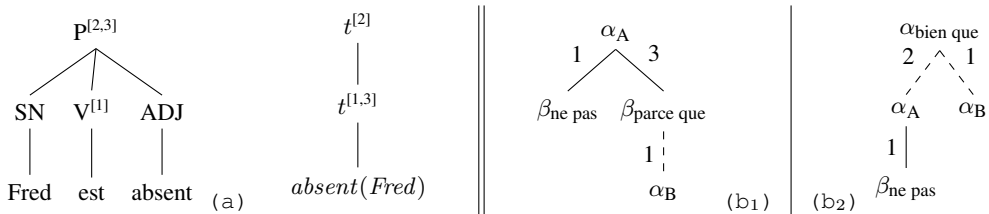


FIGURE 4 – Une structure phrastique (a), accompagnée des arbres de dérivations pour  $\neg A\ CONJ\ B$  avec une ConjC (b<sub>1</sub>) ou une ConjP (b<sub>2</sub>). Les liens [1] et [3] de  $\alpha_A$  étant situé au même nœud sémantique, l’arbre de gauche est une représentation sous-spécifiée qui, suivant l’ordre des adjonctions, conduit à deux arbres dérivés différents en sémantique mais un unique arbre pour la syntaxe.

## 4.5 Discussion

La table 2 montre les arbres de dérivation obtenus pour les phrases de la forme *A CONJ Sabine pense que B* comme en (4). Nous laissons au lecteur le soin de vérifier que les arbres dérivés sont bien ceux attendus ; c’est-à-dire que les arbres syntaxiques ne dépendent pas de l’usage du VAP contrairement

aux arbres sémantiques. Comme indiqué ci-dessus, nous éliminons l'analyse faisant usage d'un VAP évidentiel avec une ConjC, mais il existe pour les ConjP une ambiguïté que seule la sémantique des différents éléments peut permettre de résoudre.

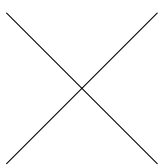
	VAP intentionnel	VAP évidentiel
ConjP :	$  \begin{array}{c}  \alpha_{\text{bien que}} \\  \swarrow \quad \searrow \\  2 \quad 1 \\  \swarrow \quad \searrow \\  \alpha_A \quad \alpha_{\text{Sabine pense que}} \\  \quad \quad \quad \downarrow \\  \quad \quad \quad 2 \\  \quad \quad \quad \downarrow \\  \quad \quad \quad \alpha_B  \end{array}  $	$  \begin{array}{c}  \alpha_{\text{bien que}} \\  \swarrow \quad \searrow \\  2 \quad 1 \\  \swarrow \quad \searrow \\  \alpha_A \quad \alpha_B \\  \quad \quad \quad \downarrow \\  \quad \quad \quad 2 \\  \quad \quad \quad \downarrow \\  \quad \quad \quad \beta_{\text{Sabine pense que}}  \end{array}  $
ConjC :	$  \begin{array}{c}  \alpha_A \\  3 \downarrow \\  \beta_{\text{parce que}} \\  1 \downarrow \\  \alpha_{\text{Sabine pense que}} \\  2 \downarrow \\  \alpha_B  \end{array}  $	

TABLE 2 – Arbres de dérivation pour les phrases de la forme *A CONJ Sabine pense que B*.

La table 3 montre les arbres de dérivation obtenus pour les phrases de la forme *Sabine pense que A CONJ B*, c'est-à-dire cette fois avec le VAP situé avant la conjonction comme en (16a). Dans cette configuration, *A CONJ B* peut former un constituant phrastique introduit par le VAP ; ce dernier ne fait alors pas partie des arguments de la relation discursive lexicalisée par CONJ et est qualifié ici d'« externe ». Notons que dans ce cas et en l'absence de contexte, le statut intentionnel ou évidentiel du VAP est indéterminé ; nous avons choisi d'utiliser alors la paire traditionnelle  $\beta_{\text{pense que}}$ . Comme précédemment, cette configuration présente des ambiguïtés qui ne peuvent être résolues que par la sémantique particulière des RD.

La figure 5 montre les deux arbres syntaxiques obtenus dans cette configuration pour la ConjP *bien que*, suivant le rôle du VAP (à gauche : intentionnel ou évidentiel ; à droite : externe). Remarquons que nous analysons le cas évidentiel avec une disparité syntaxe/sémantique sur  $\text{Arg}_1$ . En effet, on voit dans l'arbre de gauche que *Sabine pense que A* est un argument syntaxique de *bien que*, alors que si le VAP est évidentiel, seul le contenu propositionnel de *A* constitue l' $\text{Arg}_1$  de la *Concession*. Cette analyse est appuyée par la possibilité d'antéposition de la subordonnée adverbiale dans les cas intentionnels ou évidentiels illustrée en (16c) : l'antéposition de *bien que B* apparaît naturel à partir de l'arbre de gauche et non de l'arbre de droite.

## 5 Conclusion et perspectives

Après avoir exposé la notion de disparité entre syntaxe et sémantique pour les arguments des connecteurs discursifs – liée à la distinction entre VAP intentionnel et VAP évidentiel – nous avons rappelé l'existence de deux classes de conjonctions de subordination, aux propriétés syntaxiques et sémantiques bien distinctes. Nous avons alors étudié quelles interactions de portées pouvaient exister entre ConjSub et d'autres opérateurs, notamment les VAP. Ceci nous a conduit à enrichir les modélisations existantes en STAG en proposant de nouvelles structures pour les ConjSub ainsi que les VAP encodant

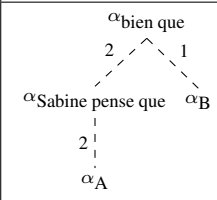
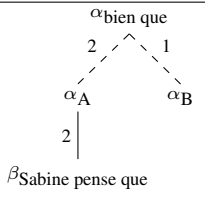
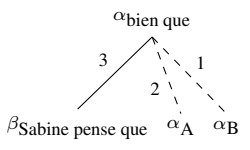
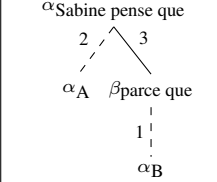
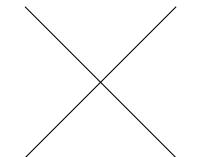
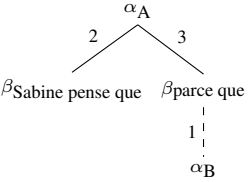
	VAP intentionnel	VAP évidentiel	VAP externe à CONJ
ConjP :			
ConjC :			

TABLE 3 – Arbres de dérivation pour les phrases de la forme *Sabine pense que A CONJ B*.

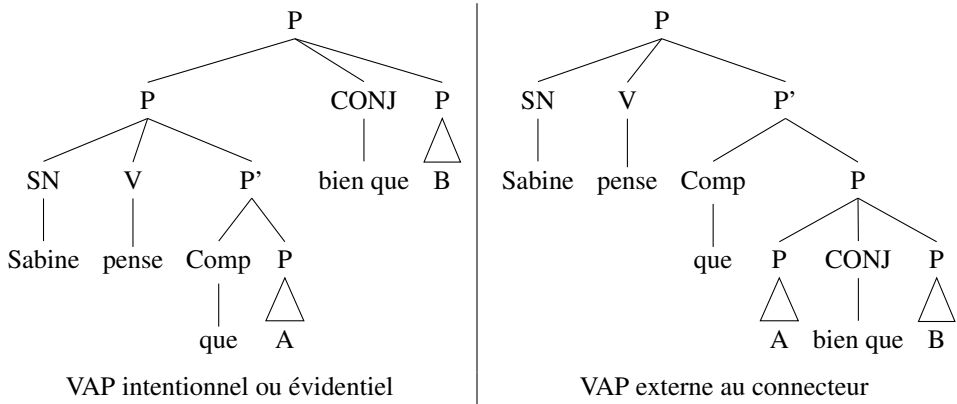


FIGURE 5 – Arbres dérivés syntaxiques pour les phrases de la forme *Sabine pense que A bien que B*. Notons qu’avec *bien que* (une ConjP), l’arbre de droite correspond à une lecture moins probable devant être forcée par le contexte.

les distinctions précédemment mises en lumière. Ces structures permettent bien de rendre compte de la plupart des phénomènes observés. Néanmoins, notre modélisation doit être affinée : elle ne rend pas encore compte de la projection des RD périphériques qui apparaît en (16b), ni de la projection de l’Arg<sub>2</sub> dans la lecture par défaut de certaines ConjC telles que *après que* (non discutée ici). Nous pensons qu’une étude plus précise de la sémantique des connecteurs discursifs, notamment en termes de présupposition et d’implicature conventionnelle s’avère nécessaire. Nous avons de plus l’intention d’étendre notre approche à d’autres connecteurs dans l’objectif de définir une grammaire STAG du discours.

Références

ASHER N., HUNTER J., DENIS P. & REESE B. (2006). Evidentiality and intensionality : Two

- uses of reportative constructions in discourse. In *Workshop on Constraints in Discourse Structure*, Maynooth, Ireland.
- ASHER N. & LASCARIDES A. (2003). *Logics of Conversation*. Cambridge University Press.
- CHIERCHIA G. & MCCONNELL-GINET S. (1990). *Meaning and grammar : an introduction to semantics*. MIT Press.
- DANLOS L. (2009). D-STAG : un formalisme d'analyse automatique de discours basé sur les TAG synchrones. *Revue TAL*, **50**(1), 111–143.
- DANLOS L. (2013). Connecteurs de discours adverbiaux : Problèmes à l'interface syntaxe-sémantique. *Linguisticae Investigationes*, **36**(2), 261–275.
- DINESH N., LEE A., MILTSAKAKI E., PRASAD R., JOSHI A. & WEBBER B. (2005). Attribution and the (Non-)Alignment of Syntactic and Discourse Arguments of Connectives. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II : Pie in the Sky*, p. 29–36, Ann Arbor, Michigan : ACL.
- FREGE G. (1948). Sense and Reference. *The Philosophical Review*, **57**(3), 209–230.
- GRICE H. (1975). Logic and conversation. In P. COLE & L. JERRY, Eds., *Syntax and semantics 3 : Speech acts*, p. 41–58. Academic Press, San Diego, CA.
- HAEGEMAN L. (2004). The syntax of adverbial clauses and its consequences for topicalisation. In M. COENE, G. DE CUYPER & Y. D'HULST, Eds., *Current Studies in Comparative Romance Linguistics*, number 107 in APiL, p. 61–90. Antwerp University.
- HUNTER J. & DANLOS L. (2014). Because We Say So. In *Proceedings of the EACL 2014 Workshop on Computational Approaches to Causality in Language*, CAtOCL, p. 1–9, Gothenburg, Sweden : ACL.
- JOSHI A. (1987). An introduction to Tree Adjoining Grammars. *Mathematics of Language*, **1**, 87–115.
- NESSON R. & SHIEBER S. (2006). Simpler TAG semantics through synchronization. In *Proceedings of FG 2006*, p. 129–142, Malaga, Spain.
- POTTS C. (2015). Presupposition and Implicature. In S. LAPPIN & C. FOX, Eds., *The Handbook of Contemporary Semantic Theory*, p. 168–202. Wiley-Blackwell, 2 edition.
- PRASAD R., DINESH N., LEE A., MILTSAKAKI E., ROBALDO L., JOSHI A. K. & WEBBER B. (2008). The Penn Discourse TreeBank 2.0. In *Proceedings of LREC-2008*, Marrakech, Morocco.
- PRASAD R., MILTSAKAKI E., DINESH N., LEE A., JOSHI A., ROBALDO L. & WEBBER B. (2007). The Penn Discourse Treebank 2.0 Annotation Manual. *IRCS Technical Reports Series*.
- ROZE C., DANLOS L. & MULLER P. (2012). LEXCONN : A French Lexicon of Discourse Connectives. *Discours. Revue de linguistique, psycholinguistique et informatique*, (10).
- SHANON B. (1976). On the Two Kinds of Presuppositions in Natural Language. *Foundations of Language*, **14**(2), 247–249.
- SHIEBER S. & SCHABES Y. (1990). Synchronous Tree-adjoining Grammars. In *Proceedings of the 13th Conference on Computational Linguistics - Volume 3, COLING '90*, p. 253–258, Stroudsburg, PA, USA : ACL.