



Un Verbenet du français

Laurence Danlos, Quentin Pradet, Lucie Barque, Takuya Nakamura, Mathieu Constant

► To cite this version:

Laurence Danlos, Quentin Pradet, Lucie Barque, Takuya Nakamura, Mathieu Constant. Un Verbenet du français. Traitement Automatique des Langues, ATALA, 2016, 57 (1), pp.25. hal-01392817

HAL Id: hal-01392817

<https://hal.inria.fr/hal-01392817>

Submitted on 4 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Un Verbønet du français

Laurence Danlos* — **Quentin Pradet**[◊] — **Lucie Barque**[◊] — **Takuya Nakamura**[•] — **Mathieu Constant**^{*}

* *Université Paris Diderot, ALPAGE, IUF*
laurence.danlos@linguist.univ-paris-diderot.fr

◊ *Université Paris Diderot, ALPAGE, CEA*
quentin@pradet.me

◊ *Université Paris 13, ALPAGE*
lucie.barque@univ-paris13.fr

• *LIGM (UMR 8049), Université Paris-Est*
takuya.nakamura@univ-mlv.fr

* *Université Paris-Est, ALPAGE*
Mathieu.Constant@u-pem.fr

RÉSUMÉ. VerbNet est une ressource lexicale pour les verbes anglais qui est largement utilisée en TAL du fait de sa bonne couverture lexicale et syntaxique et de son encodage systématique des rôles thématiques. Aucune ressource équivalente n'existe pour le français, ce qui nous a motivés pour développer un Verbønet du français. Nous présentons comment nous avons développé Verbønet à partir de VerbNet tout en utilisant au maximum les ressources lexicales existantes du français, et comment sont encodées les différentes alternances du français en mettant l'accent sur les différences avec l'anglais (l'existence de formes pronominales, par exemple). Cet article devrait permettre à un chercheur en TAL une utilisation simple et efficace de Verbønet pour une tâche comme l'annotation en rôles sémantiques.

ABSTRACT. VerbNet is a lexical resource for English verbs that has proven useful for NLP thanks to its high lexical and syntactic coverage and its systematic coding of thematic roles. Such a resource doesn't exist for French. This has motivated us to develop a Verbønet for French. We present how we have developed Verbønet from VerbNet while using as far as possible the available lexical resources for French, and how the various French alternations are coded, focusing on differences with English (existence of pronominal forms, for example). This paper should allow an NLP researcher to use Verbønet in a simple and efficient way for a task such as semantic role labeling.

MOTS-CLÉS : VerbNet, rôle thématique, cadre de sous-catégorisation, alternance syntaxique.

KEYWORDS: VerbNet, thematic role, sub-categorization frame, syntactic alternation.

1. Introduction

VerbNet (Kipper-Schuler, 2005) est un lexique électronique pour les verbes anglais qui encode leurs différents sens avec, pour chaque sens, le cadre de sous-catégorisation (nombre et forme des arguments), les rôles thématiques des arguments et les alternances syntaxiques. De plus, les différents sens des verbes sont organisés dans une hiérarchie arborescente permettant de regrouper des verbes (ou emplois de verbes) partageant une sémantique et un comportement syntaxique communs. La couverture de VerbNet, son organisation hiérarchique et la richesse des informations syntaxiques et sémantiques qui y sont codées font que cette ressource – manuellement construite et librement disponible¹ – est largement utilisée pour l’annotation en rôles sémantiques (Semantic Role Labelling, SRL). Cette tâche est utile pour de nombreuses applications du TAL, citons l’extraction d’information (Surdeanu *et al.*, 2003), les systèmes de questions-réponses (Shen et Lapata, 2007), la traduction automatique (Bazrafshan et Gildea, 2013), la détection de plagiat (Osman *et al.*, 2012), la prédiction des cours de Bourse (Xie *et al.*, 2013), l’interprétation de recettes de cuisine (Malmaud *et al.*, 2014) ou la génération de scènes 3D (Chang *et al.*, 2014).

Il nous semble donc important que le français soit doté d’une ressource équivalente, ce qui n’est pas le cas actuellement vu qu’aucun lexique de verbes français n’encode les rôles thématiques (section 3). Nous avons ainsi entrepris de développer Verbønet, ressource décrite dans cet article. Signalons qu’il y a déjà eu des ébauches d’élaboration de ressources pour le français comparables à VerbNet, soit de manière manuelle (Saint-Dizier, 1996) soit avec des techniques d’acquisition automatique (Messiant *et al.*, 2010 ; Falk *et al.*, 2012), mais elles n’ont été que partielles et sans suite. À l’inverse, la ressource Verbønet que nous présentons ici a une couverture quasiment équivalente à celle de VerbNet et encode les mêmes informations syntaxiques et sémantiques². Elle est amenée à évoluer comme VerbNet évolue au fil des ans. Elle a déjà été utilisée pour l’annotation en rôles sémantiques (Pradet, 2015) avec des résultats prometteurs même dans un état moins abouti qu’aujourd’hui.

Après une présentation succincte de VerbNet (section 2) et des ressources existantes pour le français (section 3), nous décrivons à la section 4 les étapes de développement de Verbønet en montrant comment nous avons évité de refaire le travail déjà fait pour l’anglais pour nous concentrer sur les particularités du français par rapport à l’anglais. Ces particularités du français sont développées dans les sections ultérieures consacrées successivement aux frames, rôles thématiques³, alternances et clitiques.

1. Voir le site <http://verbs.colorado.edu/verb-index>

2. Cette ressource est consultable sur <https://verbenet.inria.fr> et librement disponible sur <https://github.com/aymara/verbenet>.

3. Les termes « rôles thématiques » et « rôles sémantiques » sont employés indifféremment.

2. Présentation succincte de VerbNet

VerbNet a été élaboré à partir du livre de Levin (1993) qui présente (i) l'ensemble des alternances syntaxiques de l'anglais (*e.g.* la *dative alternation* où un complément introduit par la préposition *to* devient le premier objet direct, comme dans *Luc gave a bike to Max* et *Luc gave Max a bike*), (ii) un classement sémantique des verbes de l'anglais qui vise à tester l'hypothèse qu'une classe sémantique regroupe des verbes dont le comportement syntaxique est similaire.

Dans son livre, Levin ne s'est pas vraiment intéressée aux compléments phrastiques ou infinitivaux, se concentrant sur les compléments nominaux. VerbNet a étendu le travail de Levin dans deux directions (Kipper *et al.*, 2006) : d'une part, en décrivant les différentes réalisations non nominales d'un complément, d'autre part, en prolongeant la classification avec extension de la couverture lexicale.

VerbNet repose donc sur un classement syntaxico-sémantique des verbes anglais en 270 classes, organisées en 109 superclasses⁴. Chacune de ces 270 classes est potentiellement divisée en sous-classes, ce qui forme une hiérarchie arborescente. Pour chaque classe ou sous-classe, ce lexique comprend la liste des verbes de la classe, la liste de leurs rôles thématiques éventuellement assortis de restrictions de sélection, et la liste des frames syntaxico-sémantiques partagés par les verbes de la classe (la composition précise d'un frame est donnée à la section 5).

VerbNet couvre aux alentours de 8 000 emplois de verbes « ordinaires », qui relèvent de la langue courante, à l'exclusion de tout emploi figé ou relevant d'un domaine technique et de tout emploi d'auxiliaires, de semi-auxiliaires ou de verbes supports. Les emplois métaphoriques ne sont inclus que dans la mesure où la métaphore est conventionnelle, passant généralement du concret à l'abstrait, voir *John climbed the hill/social ladder* emprunté à (Brown et Palmer, 2012). Verbønet cherche à atteindre une couverture similaire à celle de VerbNet.

3. Les ressources lexicales françaises

Les deux principales ressources lexicographiques pour les verbes français, manuellement construites et organisées en classes, sont le LVF (*Les verbes français*, (Dubois et Dubois-Charlier, 1997)) et le LG (Lexique-grammaire, (Gross, 1975 ; Leclère, 1990)). Ces ressources lexicales n'encodent pas les rôles thématiques, et ne sont donc pas utilisables pour une tâche comme l'annotation en rôles sémantiques. De plus, elles ont cherché à atteindre l'exhaustivité (avec 25 000 entrées pour le LVF et 14 000 pour le LG) et par là même à décrire tout emploi verbal : en plus des emplois « ordinaires » du français courant, elles encodent des emplois rares ou relevant de domaines techniques (*e.g.* le verbe *biqueter* dans le LVF ou *bovaryser* dans le LG),

4. Les superclasses n'ont souvent aucune réalité matérielle autre qu'un numéro. Ainsi la super-classe 11: *Verbes of Sending and Carrying* ne fait que regrouper cinq classes notées 11.i. qui sont elles décrites en détail.

des emplois de verbes supports (*e.g.* l'emploi de *caresser* illustré dans *Luc caresse l'espoir de partir aux US* est codé dans LVF), et des emplois argotiques ou métaphoriques non conventionnels (*e.g.* l'emploi de *galoper* illustré dans *Il a galopé dans son esprit que Marie allait venir* est codé dans le LG). Du coup, les régularités syntactico-sémantiques pour les verbes ordinaires ont tendance à être noyées dans la masse et toute tentative d'enrichir ces ressources en affectant aux arguments d'un verbe un rôle thématique serait des plus laborieuses. Il n'empêche que le LVF et le LG ont été fort utiles dans la première étape du développement de Verbønet : en effet, il a généralement été possible de faire correspondre à une des 270 classes de VerbNet une des sous-classes du LVF et une sous-table du LG⁵, ce qui a permis un premier peuplement des classes de Verbønet (section 4).

Il existe deux autres ressources lexicographiques couvrantes pour les verbes français, mais qui ne proposent pas d'organisation syntactico-sémantique des verbes : Dicovalence (Van den Eynde et Mertens, 2003) dont la couverture est celle visée dans VerbNet, et *Lefff* (lexique des formes fléchies du français, (Sagot, 2010)) fréquemment utilisé en TAL. On peut envisager d'établir des liens entre Verbønet et ces deux ressources en vue de les enrichir mutuellement (ajout de formes pronominales dans Verbønet, ajout des rôles thématiques dans Dicovalence et *Lefff*, etc.).

Signalons également deux ressources actuellement en cours de développement qui proposent une description sémantique des verbes et des autres parties du discours : (i) une ressource lexicale inspirée de FrameNet (Baker *et al.*, 1998), actuellement en cours de construction dans le cadre de l'ANR Asfalda (Candito *et al.*, 2014a) – il est prévu d'établir des liens entre cette nouvelle ressource et Verbønet, comme ont été établis des liens pour l'anglais entre FrameNet et VerbNet (Palmer, 2009) – ; (ii) le Réseau lexical du français ou RL-Fr (Polguère, 2014), reposant sur l'approche explicative et combinatoire (Mel'čuk et Polguère, 2007), dans lequel chaque sens d'un verbe se voit attribuer une étiquette sémantique provenant d'une hiérarchie.

En bref, il nous a semblé que les ressources verbales existantes ne permettaient pas de développer facilement un lexique codant les rôles thématiques, soit parce que trop complexes pour être aisément enrichies (LVF et LG), soit parce que non structurées sémantiquement (Dicovalence et *Lefff*), soit parce qu'en cours de développement (ASFALDA et RL-Fr). C'est ce qui nous a motivés pour développer Verbønet à partir de VerbNet, selon une procédure décrite à la section suivante. Bien sûr, la construction d'une ressource du français à partir d'une ressource de l'anglais ne va pas sans poser de problème. Néanmoins, compte tenu des moyens à notre disposition, l'économie que représente cette méthode de développement l'a emporté sur ses inconvénients. D'autant que, conscients de ces limites, nous nous sommes attachés à contrôler la supposée équivalence des classes sémantiques entre anglais et français en établissant des liens

5. Une table du LG comporte des colonnes qui encodent toutes sortes de restrictions (morphologiques, syntaxiques, sémantiques...). Chaque restriction s'applique à un sous-ensemble des verbes de la table, formant par là même une sous-table.

avec les classes sémantiques des ressources existantes pour le français, comme nous allons l'expliquer à la section suivante.

Précisons avant cela en quoi une ressource encodant les rôles thématiques des arguments verbaux nous paraît intéressante. Une telle ressource est tout d'abord utile au TAL, en l'occurrence pour la tâche d'étiquetage en rôles sémantiques qui constitue actuellement l'un des principaux modes d'analyse sémantique automatique (cf. *supra* l'introduction). Elle joue également un rôle important à l'interface syntaxe-sémantique puisque l'on dispose d'un appariement entre fonctions syntaxiques et rôles thématiques : Verbønet code en effet implicitement les fonctions syntaxiques, si l'on définit celles-ci par des critères de position (sujet *versus* complément), de forme (complément direct *versus* prépositionnel), et de rôles thématiques. Enfin, l'utilisation de rôles thématiques permet d'établir des regroupements sémantiques originaux. On peut par exemple avoir accès à l'ensemble des verbes agentifs ou à des classes de verbes partageant un même ensemble de rôles thématiques. Dans VerbNet (et Verbønet) les rôles thématiques permettent de regrouper, au sein de superclasses, des classes de verbes sémantiquement proches, relativisant ainsi l'hypothèse de Levin selon laquelle le comportement syntaxique d'un verbe reflète sa sémantique, hypothèse mise à mal par l'existence de verbes sémantiquement proches qui n'ont pourtant pas le même comportement syntaxique.

4. Développement de Verbønet à partir de VerbNet

Le développement de Verbønet s'est fait en respectant deux principes : (i) garder le premier niveau de la hiérarchie de VerbNet avec ses 270 classes de verbes⁶ (dont le nom anglais a été temporairement conservé), (ii) garder autant que faire se peut les informations sémantiques (attribution des rôles thématiques et décomposition sémantique des éventualités). Ces principes ont permis d'éviter de concevoir une hiérarchie et de limiter le travail aux particularités syntaxiques du français (cadres de sous-catégorisation, réalisation syntaxique des arguments, alternances syntaxiques). Avec ces principes de base, la construction de Verbønet s'est faite en deux étapes.

La première étape semi-automatique a consisté à déterminer les verbes français appartenant à chacune des 270 classes de VerbNet en suivant la procédure suivante :

1) pour une classe VerbNet notée C_e , nous avons assigné manuellement la ou les classes LVF C_{lvf} et LG C_{lg} correspondant à la définition sémantique de C_e avec invariance des rôles thématiques (il peut s'agir d'une classe ou d'une conjonction de classes, les degrés de granularité variant d'une ressource à l'autre). Par exemple, la classe **put-9.1**, qui regroupe des verbes dénotant l'action de mettre quelque chose quelque part (e.g. *poser, mettre, ranger*), est associée aux classes sémantiquement équivalentes de LVF et LG, i.e. **L3b** et **38LD** ;

6. Il n'empêche que certaines classes ont été supprimées, par exemple la classe **weekend-56** qui contient des verbes comme *weekend* ou *december* sans équivalent lexicalisé en français.

2) la liste L_{trad} des traductions françaises des verbes anglais de la classe C_e a été obtenue grâce à deux dictionnaires bilingues librement disponibles (SCI-FRAN-EURADIC et le Wiktionnaire) ;

3) la liste des verbes français de la classe C_e est *a priori* composée des verbes de L_{trad} appartenant à l'intersection de C_{vf} et C_{lg} ⁷.

Cette étape a été réalisée rapidement et a donné des résultats encourageants : les verbes français à l'intersection de L_{trad} , C_{vf} et C_{lg} sont souvent cohérents syntaxiquement et sémantiquement. Soulignons que l'assignation manuelle de classes C_{vf} et C_{lg} correspondant à C_e n'est pas assise sur la traduction de l'anglais vers le français mais sur la définition sémantique de C_e et sur l'invariance des rôles thématiques dans la même position syntaxique. Pour illustrer ce point, considérons la superclasse 31:Psychological des verbes psychologiques qui mettent en jeu deux rôles thématiques, *Experiencer* et *Stimulus*, liés par la formule sémantique *emotional-state(E, Emotion, Experiencer)&in-reaction-to(E, Stimulus)* ; autrement dit, l'*Experiencer* (qui dénote un être animé) est dans un certain état émotionnel en réaction au *Stimulus*. En anglais, la superclasse 31 comporte quatre classes déterminées selon deux critères, le rôle thématique du sujet et la transitivité du verbe, comme présenté ci-dessous :

Eg.31.1 This movie amuses Mary = Stimulus V Experiencer

Eg.31.2 Mary likes this movie = Experiencer V Stimulus

Eg.31.3 Mary grooved on this movie = Experiencer V {Prép} Stimulus

Eg.31.4 This movie appeals to Mary = Stimulus V {Prép} Experiencer

Ces quatre classes sont directement transposables au français :

Fr.31.1 Ce film amuse Marie = Stimulus V Experiencer

Fr.31.2 Marie aime ce film = Experiencer V Stimulus

Fr.31.3 Marie se pâme devant ce film = Experiencer V {Prép} Stimulus

Fr.31.4 Ce film plaît à Marie = Stimulus V {Prép} Experiencer

Ceci ne signifie pas pour autant que les verbes des classes anglaises et françaises soient forcément en relation de traduction : ainsi, un verbe peut être transitif en anglais et sa traduction littérale française intransitive, voir *please* qui appartient à Eng.31.1 et *plaire* à Fr.31.4. Par ailleurs, si la classe Eng.31.1 contient près de 200 verbes, la classe Eng.31.4 ne contient que 5 verbes (*appeal, grate, jar, matter, niggle*), ce qui signifie qu'en anglais les verbes psychologiques dont le sujet dénote le *Stimulus* sont très majoritairement transitifs. Mais ceci est moins le cas en français : les verbes transitifs de Fr.31.1 sont majoritaires (plus de 200 verbes) mais les verbes intransitifs de Fr.31.4 (*agréer, aller, botter, convenir, déplaire, désagréer, importer, peser, plaire, répugner, seoir*), même s'ils sont moins nombreux, sont relativement fréquents et ne sont pas des traductions des verbes de Eng.31.4.

7. Quand cette intersection est vide, l'intersection des verbes de L_{trad} avec l'union de C_{vf} et C_{lg} est choisie.

La seconde étape, nettement plus laborieuse et totalement manuelle, a consisté à adapter les frames anglais au français, à redéfinir des sous-classes éventuelles, à vérifier les rôles thématiques, à préciser la réalisation syntaxique des compléments, à revoir la distribution des verbes dans les classes, etc. Cette étape repose sur des principes linguistiques qui font l’objet des sections suivantes.

Pour mener à bien ces deux étapes, nous avons développé une interface en ligne qui permet d’éditer collaborativement Verbønet en travaillant directement sur un site Web, sans avoir à modifier des fichiers XML⁸. Une copie d’écran de cette interface se trouve à la figure 1. Cette interface s’accompagne d’un certain nombre d’onglets qui donnent l’index des verbes de Verbønet, l’index des correspondances entre tables de LG et classes Verbønet⁹, la liste des champs syntaxiques avec liens sur les frames concernés, etc. Ces onglets sont fort utiles pour vérifier la cohérence de la ressource (*e.g.* contrôler l’utilisation d’un rôle thématique) et disposer de statistiques.

5. Les frames

Comme on peut le voir à la figure 1, la classe `settle-89`, qui regroupe des verbes dénotant un accord entre deux personnes (*s’entendre*, *s’accorder*), comprend quatre frames : (i) deux frames donnent les cadres de sous-catégorisation de ces verbes qui se construisent soit avec un complément introduit par la préposition *pour* et une infinitive soit avec un complément introduit par *sur* et un groupe nominal, (ii) deux autres frames donnent les constructions dérivées en appliquant l’alternance symétrique (décrite à la section 7.2). Plus généralement, si un rôle thématique se réalise de différentes façons (*e.g.* réalisation nominale, phrastique ou infinitivale), il y a autant de frames que de réalisations possibles (ce qui est décrit à la section 5.1.1). De plus, un frame correspond à une construction canonique partagée par les verbes de la classe ou à une construction dérivée par application d’une alternance syntaxique. Soulignons qu’un argument avec le rôle X dans la construction canonique aura le même rôle X dans une construction dérivée s’il y figure, voir (1) pour l’alternance locative décrite à la section 7.1.

- (1) a. Luc charge les caisses sur le camion = Agent V Theme `{+loc}` Destination
 b. Luc charge le camion de caisses = Agent V Destination `{de/avec}` Theme

Un frame comporte les éléments suivants présentés dans l’ordre d’apparition à l’écran : 1) un champ surfacique (en gras), 2) un exemple, 3) un champ syntaxique, 4) un champ sémantique. Nous allons commencer par une section sur le champ syn-

8. Cette interface, consultable sur <https://verbenet.inria.fr> et disponible sur <https://github.com/aymara/verbenet-editor>, a été développée à l’aide du framework web Django qui manipule une base de données PostgreSQL. Cette base stocke la ressource et conserve l’historique entier des modifications.

9. Une table de LG est souvent liée à plusieurs classes de Verbønet. Par exemple, la table 6 définie par un critère syntaxique (verbe transitif dont l’objet direct peut être phrastique) correspond à quatorze classes de Verbønet.

taxique, central à la description, puis nous décrivons les champs surfacique et sémantique. La composition des frames est la même dans VerbNet et Verbønet mais les sections suivantes se concentrent sur le français.

5.1. *Champ syntaxique*

Le champ syntaxique comporte un élément verbal qui est soit un verbe simple noté *V* soit une forme pronominale notée *se V* (les formes pronominales sont décrites à la section 7.3).

Avant l'élément verbal se trouve naturellement le sujet qui est représenté par un rôle thématique (voir section 6) sauf pour les constructions impersonnelles où le sujet est noté *IL* (sans rôle thématique). Entre le sujet et un verbe simple, se trouve parfois le clitique *LUI* (représentant *me, te, lui, nous, vous, leur, se*) utilisé avec certains verbes en présence d'un complément qui désigne un nom de partie du corps (section 7.4.2).

Après l'élément verbal, se trouvent le ou les compléments sous-catégorisés (compléments aussi dits essentiels ou appartenant à la réaction verbale) représentés par leurs rôles thématiques. Les compléments indirects sont précédés d'une préposition qui se trouve entre accolades¹⁰ : par exemple, l'emploi du verbe *donner* de la classe [give-13.1](#) est décrit dans un frame illustré par l'exemple *Luc a donné une bicyclette à Max* et dont le champ syntaxique est *Agent V Theme {à} Recipient*. En dehors de ces compléments portant un rôle thématique, ne peut figurer en position postverbale que le symbole *ADV*, abréviation pour adverbe ou groupe prépositionnel adverbial, qui apparaît principalement avec les formes pronominales moyennes (section 7.3).

5.1.1. *Les restrictions*

Chaque élément du champ syntaxique est susceptible d'être assorti d'une restriction qui figure entre chevrons après l'élément concerné. Les restrictions sont :

- <+plural> accolée à un rôle thématique qui indique que le groupe nominal réalisant le rôle thématique doit obligatoirement être au pluriel, ce que l'on observe surtout dans des frames de constructions symétriques (section 7.2) ;

- <+middle>, <+neutre>, <+reciproque> ou <+reflexive> accolée à *se V* qui indique comment la forme pronominale a été obtenue à partir de la forme canonique (section 7.3) ;

- <Patient.poss> ou <+Theme.poss> accolée à *LUI* qui indique le rôle thématique du clitique, voir (section 7.4.2) ;

- <+genitive(X.poss)>, avec *X = Patient, Expérencier, Theme* ou *Stimulus*, utilisée dans les restructurations (section 7.4) ;

- une réalisation syntaxique accolée à un rôle thématique qui indique que celui-ci peut être réalisé autrement que par un groupe nominal (NP), valeur par défaut.

¹⁰. Les prépositions locatives figurent entre doubles accolades comme expliqué à la section 8.

Il peut alors être réalisé (i) sous forme de clitique, ce qui est traité à la section 8, (ii) sous forme d'adjectif, de groupe adjectival ou d'équivalent fonctionnel¹¹ ; ils ont alors le rôle *Attribute* auquel est accolé la restriction <+Adj>, (iii) sous forme de complément phrastique, infinitival ou d'interrogative indirecte, ce qui est détaillé ci-dessous.

a) Les compléments phrastiques sont codés par la restriction <+Qu Pind> ou <+Qu Psubj> qui indique le mode du complément phrastique (indicatif ou subjonctif) dans la forme affirmative du verbe ; l'induction du subjonctif dans un énoncé négatif, interrogatif ou impératif (Gross, 1975) n'a pas encore été codée.

On rappelle qu'un complément phrastique introduit par les prépositions *à* et *de* est obligatoirement précédé de *ce* qui apparaît entre la préposition et le complémenteur *que*, voir *Luc tient à ce que Max parte* et **Luc tient à que Marie parte*. Vu le caractère obligatoire de *ce*, cet élément n'a pas été codé dans le champ syntaxique mais il apparaît dans le champ surfacique (section 5.2). L'éventuel effacement conjoint de la préposition *de* et de *ce*, voir *Luc se vante (de ce) qu'il est le premier de la classe*, noté [pc z] dans (Gross, 1975), n'a pas encore été codé.

b) Les compléments infinitivaux sont repérés par une restriction de la forme <+V-inf> mais cette restriction est en fait plus complexe à cause des phénomènes de contrôle et d'éventuels complémenteurs apparaissant devant l'infinitive.

Le contrôleur d'un complément infinitival a été codé en spécifiant son rôle thématique (vis-à-vis du verbe principal) dans la restriction. Ainsi, pour le verbe *habituer* de la classe [adjust-26.9](#) dans sa construction avec une infinitive illustrée par *Max a habitué Luc à se lever tôt*, le sujet vide de l'infinitive est contrôlé par *Luc* qui a le rôle *Patient* vis-à-vis de *habituer*. L'infinitive, qui a le rôle *Goal*, reçoit donc la restriction <+VPatient-inf> indiquant le contrôleur de l'infinitive. Au total, le schéma syntaxique de *Max a habitué Luc à se lever tôt* est Agent V Patient {à} Goal<+VPatient-inf>.

Certains compléments infinitivaux sont directs mais introduits par un complémenteur qui est généralement *de* (appelé « faux *de* ») ou plus rarement *à* (appelé « faux *à* »). La nature directe du complément se teste par la pronominalisation en *cela* : une infinitive introduite par un faux *de* se pronominalise en *cela*, tandis qu'une infinitive introduite par la « vraie » préposition *de* se pronominalise en *de cela* : *Luc a projeté de partir aux US/cela* vs. *Luc doute de pouvoir partir aux US/de cela*¹². Les complémenteurs apparaissent dans la restriction : ainsi le schéma syntaxique de *Luc a projeté de partir aux US* est Agent V Theme<+de VAgent-inf>, alors que celui de *Luc doute de pouvoir partir aux US* est Agent V {de} Theme<+VAgent-inf>.

11. Nous pensons aux compléments prépositionnels comme *de bonne humeur (Marie est/semble de bonne humeur)* qui ont un fonctionnement adjectival (Danlos, 1980).

12. Pour un petit nombre de verbes (e.g. *se dépêcher/se hâter de*), aucune pronominalisation en *cela* ou *de cela* n'est possible, un complément nominal non plus. Nous avons alors opté (plus ou moins arbitrairement) pour un « vrai » *de*.

c) Les interrogatives indirectes qui ont été codées sont <+si P> (*J'ai regardé s'il avait vidé les poubelles*), <+comment P> (*Luc a raconté à Léa comment cela s'était passé*) et <+combien P> (*L'inspecteur estime combien ça a coûté*).

Pour une classe de verbes donnée, un rôle thématique peut se réaliser de différentes façons : il y a alors autant de frames que de réalisations possibles. Ainsi, pour le verbe *se fiche*, il y a quatre frames correspondant aux exemples de (2).

- (2) a. Luc se fiche de la santé de Marie = *Expérencier V {de} Stimulus*
 b. Luc se fiche de ce que Marie soit malade = *Expérencier V {de} Stimulus<+Que Psubj>*
 c. Luc se fiche de si Marie vient ou non = *Expérencier V {de} Stimulus<+si P>*
 d. Luc se fiche de comment Marie vient = *Expérencier V {de} Stimulus<+comment P>*

Soulignons toutefois que ceci n'est vrai que pour les rôles thématiques correspondant à un complément : pour le sujet, nous n'avons pas codé les réalisations phrastiques ou infinitivales car, si celles-ci sont théoriquement possibles, (3a-b), elles ne sont guère observées en corpus, cédant le pas aux pronoms ou groupes nominaux, (3c). Par conséquent, tous les sujets codés dans Verbønet sont supposés être réalisés par des groupes nominaux (sauf pour les constructions impersonnelles).

- (3) a. Que Marie soit malade a dissuadé Max de partir aux US
 b. Apprendre que Marie est malade a dissuadé Max de partir aux US
 c. Cela/La maladie de Marie a dissuadé Max de partir aux US

5.1.2. Optionnalité et ordre des compléments

Le caractère optionnel de certains compléments n'est pas encodé dans Verbønet. Ceci signifie pour un verbe comme *donner* avec deux compléments, un objet direct et un complément indirect, que ses trois sous-structures avec zéro ou un complément, voir *Luc a donné (à Max) (une bicyclette)* où un élément entre parenthèses est optionnel, ne sont pas représentées : un frame correspond donc à une entrée canonique d'un emploi de verbe dans son expansion maximale (c'est-à-dire avec tous ses compléments) ou à une alternance de cette entrée canonique. Le choix de ne pas coder les sous-structures est dû au fait que leur éventuelle acceptabilité correspond à des situations variées : parfois un complément est juste non spécifié (*Luc a mangé une pizza vs. Luc a mangé*), parfois l'absence de complément résulte de l'aspect générique de la phrase décrivant une propriété du sujet (*Ce chien a mordu Luc vs. Ce chien mord*, ou *Luc a bu un thé vs. Luc boit*), parfois un complément manquant est une « anaphore zéro » dans la mesure où le complément manquant se trouve dans le contexte gauche de la phrase (c'est le cas du Récipient de *donner* dans *Luc a été très généreux avec Amnesty International. Il a donné 1 000 euros.*). Il est donc peu informatif de simplement coder qu'une sous-structure est acceptable sans indiquer à quelle situation d'optionnalité elle correspond. Nous proposons qu'un système de TAL utilisant Verbønet dans son état actuel considère que toutes les sous-structures sont acceptables¹³.

13. En anglais, certaines sous-structures ont été codées dans VerbNet.

L'ordre des compléments n'est pas non plus encodé dans Verbønet. Ce choix vient de ce que l'ordre des compléments en français ne dépend pas du verbe mais du « poids » des compléments, poids que l'on peut calculer en fonction du nombre de mots (Gross, 1975), ou d'autres facteurs décrits dans (Thuilier, 2012)¹⁴. Ceci signifie que le champ syntaxique Agent V Theme {à} Recipient est supposé couvrir aussi bien les phrases où le Theme apparaît avant le Recipient (*Luc a donné une bicyclette à Max*) que l'ordre inverse (*Luc a donné à Max une bicyclette*). Plus généralement, un système de TAL utilisant Verbønet peut considérer comme acceptables toutes les permutations des compléments d'un frame donné¹⁵.

5.2. Champ surfacique

Le champ surfacique correspond à l'analyse syntaxique surfacique de l'exemple donné dans le frame. Il se calcule de façon déterministe à partir du champ syntaxique en respectant un ensemble de règles. Nous nous contentons ici d'illustrer ces règles par le traitement des rôles thématiques réalisés par une infinitive :

– un rôle thématique suivi de <+VX-inf> (où X est un rôle thématique) se traduit dans le champ surfacique par *Vinf* s'il n'est pas introduit par une préposition sinon par *Prép Vinf* ;

– un rôle thématique suivi de <+de VX-inf> ou <+à VX-inf> se traduit par *de Vinf* ou *à Vinf*.

Les « vrais » et « faux » *de* ou *à* introduisant une infinitive ne sont donc pas distingués dans le champ surfacique, comme ils ne sont pas distingués dans le résultat d'une analyse syntaxique surfacique sans connaissance lexicale sur la complémentation des verbes. Une analyse surfacique peut être la première étape dans un système de SRL utilisant Verbønet.

5.3. Champ sémantique

Le champ sémantique d'un frame donne une décomposition sémantique de l'éventualité décrite dans le frame. Il est composé d'une formule coordonnant des prédicats sémantiques décrivant l'action et l'état des participants avant, pendant et après l'éventualité, e.g. *emotional-state(E, Emotion, Experiencer)&in-reaction-to(E, Stimulus)* pour la classe 31 des verbes psychologiques. Les champs sémantiques de Verbønet sont toujours identiques à ceux de VerbNet : nous n'avons

14. En anglais, il arrive que deux frames d'une même classe de VerbNet ne diffèrent que par l'ordre des compléments, ce qui suggère que l'ordre des compléments dépend du verbe.

15. L'inversion sujet verbe qui dépend tant de conditions syntaxiques (*Luc se demande quand arriveront les enfants*) que de conditions lexicales (*Sur le mât flotte un drapeau*) n'a pas encore été codée dans Verbønet.

pas travaillé sur ce sujet, la sémantique pour une classe de verbes donnée étant supposément identique en anglais et en français (voir section 4).

6. Rôles thématiques

Dans VerbNet, les rôles thématiques sont définis et organisés dans une hiérarchie, présentée en annexe, qui a fait l'objet d'une norme ISO (Bonial *et al.*, 2011) et à laquelle nous invitons le lecteur à se référer. Dans Verbønet, nous avons adopté cette hiérarchie à quelques exceptions près. Par exemple, nous n'avons jamais utilisé le rôle *Predicate*, qui est employé pour l'anglais dans certains champs syntaxiques mais qui n'est pas défini dans la hiérarchie.

En ce qui concerne l'attribution des rôles thématiques, rappelons que tous les frames d'une classe partagent les mêmes rôles. Dans Verbønet nous avons attribué les mêmes rôles thématiques qu'en anglais à quelques rares exceptions près, par exemple un rôle *Patient* changé en *Theme*. Soulignons toutefois que certains rôles ont été ponctuellement supprimés car non réalisables en français¹⁶. C'est le cas de *Result* dans les constructions résultatives (Levin, 1993, p. 99) qui n'existent jamais en français, (4).

- (4) a. Carla sponged the table clean = Agent V Initial_Location Result
 b. Carla a épongé la table (*propre) = Agent V Initial_Location

7. Les alternances

Un certain nombre d'alternances anglaises décrites dans (Levin, 1993) et/ou codées dans VerbNet n'ont pas d'équivalent en français : par exemple, la *conative alternation*, la *dative/benefactive alternation* ou la *search alternation*. Pour de telles alternances, les frames dérivés à partir de la forme canonique codés dans VerbNet ont été systématiquement supprimés de Verbønet. À l'opposé, certaines alternances anglaises ont un strict équivalent en français : c'est le cas des alternances locatives et symétriques décrites respectivement à la section 7.1 et 7.2. Entre les deux, se trouvent des alternances pour lesquelles la correspondance entre le français et l'anglais n'est que partielle, par exemple la *middle alternation* illustrée pour l'anglais en (5) correspond systématiquement à une forme pronominale en français (6).

- (5) a. The butcher cut the rumsteak = Agent V Patient
 b. The rumsteak cuts easily = Patient V ADV

- (6) a. Le boucher a coupé le rumsteak = Agent V Patient
 b. Le rumsteak se coupe facilement = Patient se V<+middle> ADV

16. Nous n'avons jamais rencontré la situation inverse, c'est-à-dire où un rôle thématique est réalisable en français mais pas en anglais.

Plus difficile, la *causative/inchoative alternation* illustrée pour l’anglais en (7) met en jeu en français une forme pronominale et/ou non pronominale, (8)-(10).

- (7) a. Luc broke the carafe = Agent V Patient
 b. The carafe broke = Patient V
- (8) a. Luc a cassé la carafe = Agent V Patient
 b. La carafe a cassé = Patient V
 c. La carafe s’est cassée = Patient se V<+neutre>
- (9) a. Luc sèchent les draps = Agent V Patient
 b. Les draps sèchent = Patient V
 c. *Les draps se sèchent
- (10) a. Le soleil amollit les cerises = Agent V Patient
 b. *Les cerises amollissent
 c. Les cerises s’amollissent = Patient se V<+neutre>

La *middle alternation* et la *causative/inchoative alternation* indiquent qu’il est nécessaire de décrire en détail le traitement dans Verbønet des formes pronominales (inexistantes en anglais), ce que nous faisons à la section 7.3.

Les opérations de restructuration, qui mettent en jeu une propriété ou une partie du corps d’un humain (ou animé), se comportent aussi de façon différente des opérations équivalentes anglaises et sont décrites à la section 7.4.

L’alternance passive n’est jamais codée dans Verbønet (ni dans VerbNet d’ailleurs). Rappelons qu’en français seuls les verbes (non pronominaux) se construisant avec un objet direct peuvent être passivés¹⁷, l’objet direct devenant sujet de la construction passive. Toutefois, il faut faire attention au fait que certains verbes se construisent avec un complément direct (non prépositionnel) qui ne reçoit pas la fonction objet direct (le verbe n’étant alors pas considéré comme transitif). C’est le cas pour les compléments directs qui reçoivent le rôle Value, Asset ou Attribute, comme illustré en (11).

- (11) a. Cette valise pèse 4 kilos = Theme V Value
 b. Luc a vendu sa bicyclette 200\$ à Max = Agent V Theme Asset {à} Recipient
 c. Nous avons nommé Max président = Agent V Theme Attribute

Les rôles thématiques permettent donc de distinguer les objets directs des compléments directs et par là même d’identifier les verbes transitifs. Nous pouvons affirmer que pratiquement tous les verbes transitifs codés dans Verbønet permettent le passif¹⁸,

17. À l’exception du verbe *obéir* : *Luc obéit à Marie* et *Marie est obéie par/de Luc*.

18. Rappelons que les constructions figées ou à verbe support ne sont pas codées dans Verbønet. Certains verbes avec un complément direct entrant dans ces constructions ne permettent pas le passif, même s’ils le permettent dans une construction « ordinaire », voir le contraste entre *Beaucoup de gens collectionnent les porte-clefs/ennuis* et *Les porte-clefs/*ennuis sont collectionnés par beaucoup de gens*, emprunté à (Leclère, 1993).

à quelques exceptions notoires près, par exemple le verbe *avoir* de la classe *own-100* qui ne permet pas le passif contrairement au verbe *posséder*, ou les verbes olfactifs (*embaumer, puer*) de la classe *smell-emission-43.3*.

Verbønet permet donc (sauf cas rares) de savoir quels verbes admettent le passif. en revanche Verbønet ne permet pas de savoir par quelle préposition le complément d'agent est introduit : c'est généralement *par* mais parfois *de* (*Luc est aimé par/de tous les étudiants*), ce qui n'a pas encore été codé. De même, les participes passés à valeur adjectivale (statique) comme *Le stand est démonté* ou *Le service est doté d'un fax* n'ont pas encore été codés.

7.1. Les alternances locatives

Les alternances locatives mettent en jeu une entité (le *locatum*) dont la position change ou se situe par rapport à un lieu (Guillet et Leclère, 1992). Dans le frame canonique, le lieu apparaît dans un complément prépositionnel introduit par une préposition locative ; dans le frame dérivé, le lieu prend soit la fonction sujet – voir (12b) avec le rôle *Location* qui indique le lieu où se situe le locatum – soit la fonction objet direct – voir (13b) avec le rôle *Destination* où le locatum change de lieu.

- (12) a. Des abeilles grouillent dans le jardin = Theme V $\{ \{ +loc \} \}$ Location
 b. Le jardin grouille d'abeilles = Location V $\{ de \}$ Theme
- (13) a. Luc charge les caisses sur le camion = Agent V Theme $\{ \{ +loc \} \}$ Destination
 b. Luc charge le camion de caisses = Agent V Destination $\{ de/avec \}$ Theme

Lorsque l'on quitte les constructions locatives pour les constructions datives dans lesquelles le rôle *Destination* est remplacé par son sous-type *Recipient*[+animate] on observe aussi une alternance où le complément locatif/datif devient objet direct (Guillet et Leclère, 1992), voir (14).

- (14) a. Luc fournit du lait à Max = Agent V Theme $\{ \hat{a} \}$ Recipient
 b. Luc fournit Max en lait = Agent V Recipient $\{ en \}$ Theme

7.2. Les alternances symétriques

Dans le cas des alternances symétriques, le frame canonique comporte les deux rôles thématiques *X* et *Co-X* (où *X* = *Agent*, *Patient* ou *Theme*) ; le frame dérivé comporte uniquement le rôle *X* assorti de la restriction $\langle +plural \rangle$, ce qui correspond à la coordination des groupes nominaux liés à *X* et *Co-X* ou à un simple groupe nominal sémantiquement pluriel. Le rôle *X* peut occuper la position sujet, (15) avec *X* = *Agent*, ou objet direct, (16) avec *X* = *Patient*, le rôle *Co-X* étant toujours réalisé comme complément indirect (introduit par diverses prépositions).

- (15) a. Luc conspire avec Marie pour prendre le pouvoir = Agent V $\{ avec \}$ Co-Agent $\{ pour \}$ Theme $\langle +VAgent-inf \rangle$

- b. Luc et Marie conspirent pour prendre le pouvoir = Agent<+plural> V {pour}
Theme<+VAgent-inf>
- (16) a. Luc sépare les dés noirs des dés blancs = Agent V Patient {de} Co-Patient
b. Luc sépare les dés noirs et les dés blancs = Agent V Patient<+plural>

De fait, les rôles Co-X ne se trouvent pratiquement qu'avec des verbes permettant une alternance symétrique, à quelques rares exceptions près. Par exemple, le rôle Co-Theme est utilisé pour le verbe *soustraire* (Agent V Theme {à} Co-Theme) bien qu'il ne permette pas l'alternance symétrique (contrairement à *additionner* qui décrit une opération commutative).

7.3. Les alternances pronominales

Il existe en français un certain nombre de verbes intrinsèquement pronominaux, c'est-à-dire des formes *se V* qui ne peuvent pas être mises en relation avec une forme non pronominale de *V* soit parce que *V* n'existe pas (*s'acharner* (à), *se dégrouiller* (*de/pour*), *se souvenir* (*de*)) soit parce que *V* existe mais avec un tout autre sens (*se barrer* (*Loc*), *se lâcher*, *se piquer* (*de*), *se solder* (*par*)). Ces verbes sont alors enregistrés dans Verbænet sous l'entrée composée *se V*, les champs syntaxique et surfacique décrivant la complémentation de ces verbes étant construits autour du seul symbole *V*. À titre d'illustration, la classe [escape-51.1](#) comporte les verbes *V = se barrer* et *partir* illustrés dans *Luc se barre/part à Paris* = Theme V {+loc} Destination.

À l'opposé, toutes les formes non intrinsèquement pronominales *se V* sont enregistrées sous l'entrée *V* dans un frame dérivé d'une forme canonique non pronominale. Dans le frame dérivé, la forme pronominale *se V* est assortie d'une restriction <+middle>, <+neutre>, <+reciproque> ou <+reflexive> qui précise quelle est la relation avec la forme canonique¹⁹. Les restrictions <+middle> et <+neutre> accompagnent des formes pronominales dérivées de constructions transitives pour lesquelles l'objet direct devient sujet, <+reciproque> et <+reflexive> accompagnent des formes pronominales où *se* peut être vu comme une pronominalisation de l'objet direct.

Le trait <+middle> marque les constructions « moyennes », aussi dites à « agent fantôme » (Boons *et al.*, 1976 ; Ruwet, 1972), illustrées en (6) et (17). Le trait <+middle> implique la présence de ADV dans les champ syntaxique et surfacique. Le procès décrit est d'aspect générique.

- (17) a. Le boucher a coupé mon rumsteak = Agent V Patient
b. Le rumsteak se coupe facilement = Patient se V<+middle> ADV

Le trait <+neutre> marque les formes pronominales à diathèse neutre, d'aspect non générique. Rappelons qu'en français la diathèse neutre est pronominale ou pas, (8) et (18).

19. Sauf dans un cas illustré dans la classe [dedicate-7](#).

- (18) a. Luc a cassé la carafe = Agent V Patient
 b. La carafe a cassé = Patient V
 c. La carafe s'est cassée = Patient se V<+neutre>

Pour un verbe transitif, on peut donc s'attendre *a priori* à quatre cas de figure, le premier sans aucune diathèse neutre, par exemple *donner* (*Luc a donné une fleur à Marie*, **Une fleur a donné à Marie*, **Une fleur s'est donnée à Marie*), le deuxième et le troisième avec une seule forme de diathèse neutre (pronominale ou pas, par exemple *sécher* et *amollir* en (9) et (10)), le quatrième avec deux formes de diathèse neutre, par exemple *casser* en (18). Certes, il existe certaines régularités sémantiques, surtout pour le premier cas de figure puisque, par exemple, aucun verbe comme *donner* de la superclasse [13:Change of Possession](#) n'admet de diathèse neutre tant en anglais qu'en français (Levin, 1993 ; Boons *et al.*, 1976). Mais, sinon, il faut éventuellement répartir les verbes transitifs d'une classe en quatre sous-classes et ce travail conséquent n'a pas été fait systématiquement dans la version actuelle de Verbønet.

Le trait <+reciproque> est porté par des verbes pronominaux dont le sujet est obligatoirement pluriel. Le verbe peut être transitif, (19), ou intransitif avec un complément introduit par la préposition *à*, (20). Les ajouts pronominaux de forme *l'un (Prép) l'autre* sont répertoriés dans Dicovalence (qui repose sur l'approche pronominale (Blanche-Benveniste *et al.*, 1984)) alors qu'ils sont absents de Verbønet.

- (19) a. Luc caresse Max = Agent V Theme
 b. Luc et Max se caressent (l'un l'autre) = Agent<+plural> se V<+reciproque>
- (20) a. Luc ment à Max = Agent V {à} Theme
 b. Luc et Max se mentent (l'un à l'autre) = Agent<+plural> se V<+reciproque>

Le trait <+reflexive> est porté par des verbes dont l'objet direct ou indirect introduit par la préposition *à* est coréférent au sujet, *Marie se caresse (elle-même)*, *Marie se ment (à elle-même)*²⁰. Les constructions réflexives n'ont pas été codées dans la version actuelle de Verbønet sauf dans la classe [concealment-16](#) pour illustrer le contraste entre *cache* et *séquestrer*, (21) et (22), qui montre qu'une forme réfléchie peut être interdite même si le sujet et l'objet direct ont les mêmes restrictions de sélection (l'Agent et le Patient de *séquestrer* doivent tous deux être animés).

- (21) a. Luc cache Max dans le placard = Agent V Patient {{+loc}} Location
 b. Luc se cache dans le placard = Agent se V<+reflexive> {{+loc}} Location
- (22) a. Luc a séquestré Max dans le placard = Agent V Patient {{+loc}} Location
 b. *Luc s'est séquestré dans le placard

Bilan. Il est parfois difficile de déterminer la relation entre une forme pronominale non intrinsèque et la forme non pronominale canonique (Candito *et al.*, 2014b). Les rôles thématiques peuvent éventuellement aider, par exemple pour un verbe tel que

20. Les formes pronominales *se caresser* ou *se mentir* construites avec un sujet pluriel (*Les enfants se caressent/mentent*) sont ambiguës, présentant une interprétation réciproque et réflexive.

noyer, dont l'emploi transitif est illustré en (23a). Pour la forme pronominale, on peut hésiter entre une forme neutre (23b) ou réfléchie (23c). Le fait qu'il est difficile d'attribuer le rôle Agent au sujet de *se noyer* amène à poser que c'est la forme neutre qui est correcte (et exclure donc l'analyse réfléchie en *refnoyer*).

- (23) a. Luc a noyé un chaton = Agent V Patient
 b. Un chaton s'est noyé = Patient se V<+neutre>
 c. Un chaton s'est noyé = ~~Agent se~~ V<+reflexive>

7.4. Les restructurations

Les restructurations sont des alternances qui mettent en jeu un « possesseur » et une propriété ou une partie du corps de ce possesseur.

7.4.1. Restructuration avec une propriété du possesseur

Cette restructuration est appelée *possessor-attribute factoring alternation* chez Levin et décrite pour le français dans (Guillet et Leclère, 1981). Dans la forme canonique, le possesseur (noté *poss*) et la propriété possédée (notée *prop*) apparaissent dans un groupe nominal complexe (de fonction objet direct ou sujet) dont la propriété est la tête, le possesseur étant réalisé sous forme d'un génitif, c'est-à-dire soit un complément de nom introduit par *de* soit un déterminant possessif, (24a) ou (25a). Dans la forme dérivée, ces deux éléments apparaissent sous une forme éclatée : le possesseur prend la position (objet direct ou sujet) du groupe nominal complexe et la propriété apparaît sous forme de complément prépositionnel, (24b) ou (25b).

- (24) a. Luc adore l'humour de Max =
 Expérencier V Stimulus.prop<+genitive(Stimulus.poss)>
 b. Luc adore Max pour son humour = Expérencier V Stimulus.poss {pour}
 Stimulus.prop
 c. Luc adore cette table = Expérencier V Stimulus
- (25) a. La gentillesse de Marie étonne tout le monde =
 Stimulus.prop<+genitive(Stimulus.poss)> V Expérencier
 b. Marie étonne tout le monde par sa gentillesse = Stimulus.poss V Expérencier
 {par} Stimulus.prop
 c. Ce mariage étonne tout le monde = Stimulus V Expérencier

Une seule fonction syntaxique est donc restructurée en deux fonctions syntaxiques, ce qui est largement ignoré dans les lexiques syntaxico-sémantiques qui ne traitent que des changements de diathèse avec redistribution ou suppression (et non ajout) de fonctions syntaxiques.

Mais que faire pour les rôles thématiques ? De fait, il faut innover car les restructurations ne sont pas traitées ou mal traitées tant dans VerbNet que dans les discussions sur la hiérarchie des rôles (Bonial *et al.*, 2011). Nous proposons la solution suivante. Dans la forme canonique de la restructuration, le groupe nominal complexe reçoit le

rôle thématique X selon les critères usuels, c'est-à-dire le même rôle X attribué à un groupe nominal simple occupant la même position (24c) ou (25c). Toutefois, pour indiquer sa forme complexe et introduire le codage de la restructuration, il est noté X.prop avec la restriction <+genitive(X.poss)>. Dans la forme restructurée, le rôle X apparaît sous deux formes éclatées, X.prop et X.poss, ce qui traduit son éclatement syntaxique.

7.4.2. Restructuration avec un nom de partie du corps du possesseur

Dans l'exemple (26a), l'objet direct est un groupe nominal complexe dont la tête est un nom de partie du corps (Npc) suivi d'un génitif introduisant le possesseur du Npc (humain ou animé). En (26b), ce groupe nominal est éclaté, le possesseur prenant la position du groupe nominal complexe et le Npc apparaissant dans un groupe prépositionnel – cette restructuration est appelée *body-part possessor ascension alternation* chez Levin. Enfin, en (26c), l'humain apparaît sous forme du clitique LUI bien que la forme éclatée *Luc a agrippé le bras à Max* soit inacceptable ou d'un niveau de langue relâché (cette restructuration n'a pas d'équivalent en anglais).

- (26) a. Luc a agrippé le bras de Max =
Agent V Theme.prop<+genitive(Theme.poss)>
b. Luc a agrippé Max par le bras = Agent V Theme.poss {{loc}} Theme.prop
c. Luc lui a agrippé le bras = Agent LUI<Theme.poss> V Theme.prop

Pour certains verbes, seule la restructuration avec le clitique LUI est autorisée : (27b) est inacceptable quelle que soit la préposition introduisant le Npc. Si Agent et Patient sont coréférents, on observe une structure pronominale dite en « *se* partie du corps » dans (Boons *et al.*, 1976, p. 127) : *Luc s'est lavé les cheveux*. Cette possibilité de coréférence n'a pas été codée.

- (27) a. Luc a lavé les cheveux de Max =
Agent V Patient.prop<+genitive(Patient.poss)>
b. *Luc a lavé Max par/à/sur les cheveux
c. Luc lui a lavé les cheveux = Agent LUI<Patient.poss> V Patient.prop

Soulignons qu'il ne suffit pas que le nom tête d'un objet direct soit un Npc pour que la restructuration avec le clitique LUI soit autorisée : cette restructuration dépend du verbe, par exemple elle est interdite avec des verbes psychologiques, voir *Luc adore les cheveux de Max* et **Luc lui adore les cheveux*.

8. Les clitiqes

Les clitiqes n'ont pas été codés dans Verbønet sauf LUI dans les constructions avec un nom de partie du corps (Npc) décrites à la section précédente et *se* étudié dans

les formes pronominales (section 7.3). Cependant, on peut déduire la réalisation des autres clitiques (non sujets) grâce aux règles suivantes²¹.

Les objets directs, qui, rappelons-le, doivent être distingués des compléments directs (section 7), se pronominalisent en LE (= *me, te, le, la, nous, vous, les* et la forme élidée *l'*) si l'objet direct est défini (*Luc a donné sa bicyclette à Max, Luc l'a donnée à Max*) sinon sous la forme *en ... Dét-indéfini* (*Luc a donné deux bicyclettes à Max, Luc en a donné deux à Max*).

Les compléments indirects introduits par la préposition à se pronominalisent :

- en LUI (= *me, te, lui, nous, vous, leur*) si le complément porte le rôle Beneficiary, Co-Patient, Experiencer, Recipient ou Source et que c'est un NP avec la restriction [+animate] (et que le verbe n'est pas à une forme pronominale) ;
- en y si le complément porte le rôle Goal, Stimulus, Theme, Topic ou Result à l'exception de verbes aspectuels comme *continuer* ou *entraîner* (*Luc continue/entraîne à rédiger sa thèse* = Agent V {à} Theme<+VAgent-inf>) qui ne permettent pas la formation d'un clitique.

Les compléments indirects introduits par la préposition de se pronominalisent en *en* quel que soit le rôle du complément, modulo les exceptions suivantes :

- si le complément porte le rôle Initial-Location, il est pronominalisable en *en* s'il est employé seul (e.g. *Luc est venu de Dax* = Agent V {de} Initial-Location) mais pas s'il est accompagné d'un complément de destination (e.g. *Luc est venu de Dax à Pau* = Agent V {de} Initial-Location {{Loc-Dest}} Destination, où le symbole Loc-Dest est décrit ci-dessous), ou si la préposition *de* alterne avec *depuis* (*Luc a expédié un cadeau de/depuis Dax*) ;
- si le complément porte le rôle Theme avec des verbes aspectuels comme *se hâter* (*Luc se hâte de rédiger sa thèse*).

Les compléments locatifs (non initiaux) sont introduits par des prépositions qui peuvent dépendre du nom introduit : ainsi les exemples *Luc est parti à Paris/en France/dans le Pacifique* mettent en jeu la préposition *à, de* ou *dans* selon le nom géographique introduit, le sens du verbe et les rôles thématiques restant identiques. Ces exemples sont donc tous représentés par le même champ syntaxique, à savoir Agent V {{Loc-Dest}} Destination où le symbole Loc-Dest mis entre deux accolades représente une préposition locative introduisant la destination finale du déplacement de l'Agent. Dans le même ordre d'idées, le symbole +loc mis entre deux accolades représente une préposition locative quelconque : les exemples *Luc stocke les journaux dans/sous/sur/derrière/devant le placard* sont tous représentés par le champ syntaxique Agent V Theme {{+loc}} Location. Les compléments locatifs se prono-

21. Ces règles permettent heureusement d'éviter la multiplication des frames pour une classe donnée. En effet, sans de telles règles, il aurait fallu coder quatre frames pour une classe de verbes avec deux compléments (*donner* par exemple), un sans aucun clitique, deux avec un seul clitique, et un avec deux clitiques.

minimalisent tous selon le clitique *y* sauf en la présence d'un complément avec le rôle Initial-Location (*Luc est venu de Dax à Pau*).

Les autres compléments indirects (dits obliques) introduits par une préposition qui n'est ni *à*, ni *de*, ni une préposition locative – mais par exemple *avec*, *pour*, ou *sur* dans un sens non locatif – ne permettent pas la formation de clitique : ils se pronominalisent uniquement avec un pronom fort qui est catégorisé comme NP (*Luc compte sur Max/lui pour cette tâche/cela*).

En résumé, les règles que nous venons de poser permettent de calculer l'éventuelle réalisation clitique d'un complément, entre autres grâce au rôle thématique du complément. On peut aussi retourner le problème, c'est-à-dire attribuer un rôle thématique à un complément en fonction de ses éventuelles réalisations clittiques. Rappelons néanmoins qu'il n'existe pas de clitique en anglais et que les rôles thématiques ne diffèrent guère entre l'anglais et le français (section 6), ce qui signifie que la définition sémantique des rôles thématiques est plus prépondérante que la réalisation clitique pour leur attribution.

9. Conclusion et perspectives

Les statistiques de Verbønet disent que cette ressource compte à l'heure actuelle plus de 5 000 emplois de verbes pour environ 3 000 lemmes verbaux. Nous ne sommes donc pas loin de la couverture visée. La plupart des alternances syntaxiques du français ont été codées ou sont déductibles par règles en fonction des rôles thématiques (*e.g.* passif, réalisations clittiques).

La meilleure évaluation d'une ressource étant son utilisation, il faudrait refaire des expériences telles que celles menées dans (Pradet, 2015) pour SRL ou mener de nouvelles expériences. Pour les approches supervisées et l'évaluation d'outils de TAL, il faut souligner qu'il n'existe pas de corpus français annoté à la Verbønet avec désambiguation des verbes et attribution de rôles thématiques à leurs arguments. Cependant un corpus français est annoté à la FrameNet dans le cadre du projet Asfalda (section 3). Nous envisageons donc non seulement d'établir des liens entre les ressources françaises Verbønet et FrameNet mais aussi de se servir de ces liens pour adapter le corpus annoté à la FrameNet à Verbønet. Rappelons que de tels liens entre ressources ou corpus annotés ont été largement établis pour l'anglais.

Verbønet est amené à évoluer, et d'ailleurs est en cours d'évolution. Un défaut de VerbNet est que les frames d'une classe sont présentés « en vrac ». À charge pour l'utilisateur de deviner à quoi correspond un frame : à la construction canonique avec réalisations nominales, à la construction canonique avec une réalisation non nominale ou à une alternance syntaxique ? Une nouvelle interface de visualisation de Verbønet est en cours de développement : une copie d'écran de cette interface se trouve à la figure 2, qui concerne la classe `settle-89` comme c'est le cas de la figure 1. Cette nouvelle interface, librement disponible sur le site <https://verbenet.inria.fr/public/alternances/>, identifie explicitement à quoi correspond un frame avec,

settle-89

Classe settle-89

- Membres de VerbNet: **accorder s' arranger s' entendre s' allier s'**
- Roles :Agent[+organization][+animate], Co-Agent[+organization][+animate], Goal

Canonique

Exemple	Luc s'est entendu avec Marie sur le prix.
Surfacique	NP V PP.co-agent PP.goal
Syntaxe	Agent V {avec} Co-Agent {sur} Goal
Sémantique	agree(result(E), Agent, Co-Agent, Goal)

Canonique avec Goal[+V-inf]

Exemple	Luc s'est entendu avec Marie pour proposer un projet
Surfacique	NP V PP.co-agent PP.goal
Syntaxe	Agent V {avec} Co-Agent {pour} Goal[+VAgent-inf]
Sémantique	agree(result(E), Agent, Co-Agent, Goal)

Alt. Symetrique

Exemple	Luc et Marie se sont entendus sur le prix
Surfacique	NP V PP.goal
Syntaxe	Agent<+plural> V {sur} Goal
Sémantique	agree(result(E), Agent, Goal)

Alt. Symetrique avec Goal[+V-inf]

Exemple	Luc et Marie se sont entendus pour proposer un projet
Surfacique	NP V PP.goal
Syntaxe	Agent<+plural> V {pour} Goal[+VAgent-inf]
Sémantique	agree(result(E), Agent, Goal)

Figure 2. Nouvelle interface Web pour visualiser VerbNet

entre autres, identification automatique des alternances syntaxiques, ce qui permet de mettre en évidence des incohérences de la base lexicale. Par exemple, l'explicitation de la nature des frames (canonique ou d'alternance) a permis d'améliorer le traitement des restructurations (section 7.4).

Les autres améliorations prévues concernent la couverture lexicale et la répartition des verbes dans les classes. Rappelons en effet (section 4) que le peuplement des classes a été fait dans une première étape semi-automatique : les résultats de cette étape ont été vérifiés mais peut-être pas assez soigneusement. Par contre, nous pensons qu'il n'y aura pas lieu de modifier les principes linguistiques concernant la réalisation syntaxique des arguments, les alternances syntaxiques et l'attribution de rôles thématiques, ces principes ayant été largement développés dans cet article et devant servir de guide pour tout enrichissement futur de Verbønet.

Enfin, nous envisageons à terme d'étendre cette ressource aux nominalisations dérivées des verbes d'une part, et aux expressions polylexicales verbales, d'autre part.

10. Bibliographie

- Baker C., Fillmore C., Lowe J., « The Berkeley FrameNet project », *Proceedings of the ACL-COLING conference*, 1998.
- Bazrafshan M., Gildea D., « Semantic Roles for String to Tree Machine Translation », *Proceedings of ACL conference*, 2013.
- Blanche-Benveniste C., Delofeu J., Stefanini J., van den Eynde K., *Pronom et syntaxe. L'approche pronominale et son application au français*, SELAF, Paris, 1984.
- Bonial C., Corvey W., Palmer M., Petukhova V., Bunt H., « A Hierarchical Unification of LIRICS and VerbNet Semantic Roles », *Proceedings of the ICSC Workshop on Semantic Annotation for Computational Linguistic Resources (SACL-ICSC 2011)*, 2011.
- Boons J. P., Guillet A., Leclère C., *La structure des phrases simples en français : constructions intransitives*, Droz, Genève, 1976.
- Brown S. W., Palmer M., « Semantic Annotation of Metaphorical Verbs with VerbNet : A Case Study of Climb and Poison », *Proceedings of isa-8 : Eighth Joint ACL - ISO Workshop on Interoperable Semantic Annotation*, 2012.
- Candito M., Amsili P., Barque L., Benamara F., De Chalendar G., Djemaa M., Haas P., Huyghe R., Mathieu Y. Y., Muller P., Sagot B., Vieu L., « Developing a French FrameNet : Methodology and First results », *Proceedings of the International Conference on Language Resources and Evaluation Conference (LREC)*, Reykjavik, Iceland, 2014a.
- Candito M., Perrier G., Guillaume B., Ribeyre C., Fort K., Seddah D., Villemonte De La Clergerie É., « Deep Syntax Annotation of the Sequoia French Treebank », *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Iceland, 2014b.
- Chang A. X., Savva M., Manning C. D., « Semantic Parsing for Text to 3D Scene Generation », *Proceedings of the ACL Workshop on Semantic Parsing*, 2014.
- Danlos L., Représentation d'informations linguistiques : constructions *N être Prép X*, PhD thesis, Université Paris 7, 1980.

- Dubois J., Dubois-Charlier F., *Les verbes français*, Larousse, 1997.
- Falk I., Gardent C., Lamirel J.-C., « Classifying French Verbs Using French and English Lexical Resources », *Proceedings of the ACL conference*, 2012.
- Gross M., *Méthodes en syntaxe*, Hermann, Paris, 1975.
- Guillet A., Leclère C., « Restructuration du groupe nominal », *Langages*, vol. 63, p. 99-12, 1981.
- Guillet A., Leclère C., *La structure des phrases simples en français : Les constructions transitives locatives*, Droz, Genève, 1992.
- Kipper K., Korhonen A., Ryant N., Palmer M., « Extending VerbNet with novel verb classes », *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC)*, Genoa, Italy, 2006.
- Kipper-Schuler K., VerbNet : A broad-coverage, comprehensive verb lexicon, PhD thesis, University of Pennsylvania, 2005.
- Leclère C., « Organisation du lexique-grammaire des verbes français », *Langue Française*, 1990.
- Leclère C., « Classes de constructions directes sans passif », *Langages*, vol. 109, p. 7-31, 1993.
- Levin B., *English verb classes and alternations : a preliminary investigation*, University Of Chicago Press, 1993.
- Malmaud J., Wagner E. J., Chang N., Murphy K., « Cooking with Semantics », *Proceedings of the ACL Workshop on Semantic Parsing*, 2014.
- Messiant C., Gábor K., Poibeau T. et al., « Acquisition de connaissances lexicales à partir de corpus : la sous-catégorisation verbale en français », *Traitement automatique des langues*, vol. 51, p. 65-96, 2010.
- Meřćuk I., Polguère A., *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*, De Boeck et Larcier, Bruxelles, 2007.
- Osman A. H., Salim N., Binwahlan M. S., Alteeb R., Abuobieda A., « An improved plagiarism detection scheme based on semantic role labeling », *Applied Soft Computing*, vol. 12, n° 5, p. 1493-1502, 2012.
- Palmer M., « Semlink : Linking PropBank, VerbNet and FrameNet », *Proceedings of the Conference on Generative Approaches to the Lexicon (GL2009)*, Pisa, Italy, 2009.
- Polguère A., « Principes de modélisation systémique des réseaux lexicaux », *Actes de la 20ème conférence sur le Traitement Automatique du langage (TALN 2014)*, Marseille, France, p. 79-90, 2014.
- Pradet Q., Annotation en rôles sémantiques du français en domaine spécifique, PhD thesis, Université Paris-Diderot, 2015.
- Ruwet N., *Théorie syntaxique et syntaxe du français*, Le Seuil, Paris, France, 1972.
- Sagot B., « The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French », *Proceedings of the 7th edition on Language Resources and Evaluation Conference (LREC)*, Valletta, Malta, 2010.
- Saint-Dizier P., « Constructing Verb Semantic Classes for French : Methods and Evaluation », *Proceedings of COLING*, 1996.
- Shen D., Lapata M., « Using Semantic Roles to Improve Question Answering », *Proceedings of EMNLP-CoNLL*, 2007.

- Surdeanu M., Harabagiu S., Williams J., Aarseth P., « Using predicate-argument structures for information extraction », *Proceedings of the ACL Conference*, p. 8-15, 2003.
- Thuilier J., Contraintes préférentielles et ordre des mots en français, PhD thesis, Université Paris-Diderot, 2012.
- Van den Eynde K., Mertens P., « La valence : l'approche pronominale et son application au lexique verbal », *Journal of French Language Studies*, vol. 13, p. 63-104, 2003.
- Xie B., Passonneau R. J., Wu L., Creamer G. G., « Semantic Frames to Predict Stock Price Movement », *Proceedings of the ACL conference*, 2013.

Annexe. Hiérarchie des rôles thématiques de VerbNet (Bonial *et al.*, 2011)

