HAL
open science

Improvement of VerbNet-like resources by frame typing

Laurence Danlos, Mathieu Constant, Lucie Barque

To cite this version:

Laurence Danlos, Mathieu Constant, Lucie Barque. Improvement of VerbNet-like resources by frame typing. Workshop on Grammar and Lexicon: interactions and interfaces (GramLex), The COLING 2016 Organizing Committee, Dec 2016, Osaka, Japan. pp.61-70. hal-01392822

HAL Id: hal-01392822

https://inria.hal.science/hal-01392822v1

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.

Improvement of VerbNet-like resources by frame typing

Laurence Danlos
Université Paris Diderot
ALPAGE INRIA
Laurence.Danlos@inria.fr

Mathieu Constant
Université de Lorraine
ATILF CNRS
Mathieu.Constant@univ-lorraine.fr

Lucie Barque
Université Paris 13
ALPAGE INRIA
Barque@univ-paris13.fr

Abstract

Verbənet is a French lexicon developed by "translation" of its English counterpart — VerbNet (Kipper-Schuler, 2005) — and treatment of the specificities of French syntax (Pradet et al., 2014; Danlos et al., 2016). One difficulty encountered in its development springs from the fact that the list of (potentially numerous) frames has no internal organization. This paper proposes a type system for frames that shows whether two frames are variants of a given alternation. Frame typing facilitates coherence checking of the resource in a "virtuous circle". We present the principles underlying a program we developed and used to automatically type frames in Verbənet. We also show that our system is portable to other languages.

1 Introduction

VerbNet is a broad-coverage resource for English verbs in which verbs are grouped according to shared syntactic behaviors, namely surface realizations of their argument structure and syntactic alternations they are involved in (Kipper-Schuler, 2005; Kipper et al., 2006). Based on the hypothesis that verbs with similar semantics share similar syntactic properties (Levin, 1993), VerbNet extends Levin's classification up to 270 hierarchically organized verb classes. VerbNet has proven useful for NLP thanks to its high coverage (more than five thousand distinct verbs), useful verb separation and systematic coding of thematic roles. In particular, VerbNet is widely used in Semantic Role Labelling (SRL), a task that has grown steadily in importance: it serves as an aid to information extraction (Surdeanu et al., 2003), question-answering (Shen and Lapata, 2007), event extraction (Exner and Nugues, 2011), plagiarism detection (Osman et al., 2012), machine translation (Bazrafshan and Gildea, 2013), or stock market prediction (Xie et al., 2013).

Given the success of VerbNet, equivalent ressources for other languages have been developed, including Italian (Busso and Lenci, 2016), Portuguese (Scarton et al., 2014) and others. For French, a French Verbənet was first automatically derived from English VerbNet (Pradet et al., 2014); next the focus turned to accounting for the syntactic specificities of French – for example the existence of pronominal forms (noted as *se V*) which are illustrated in (1) for the middle alternation (Danlos et al., 2016).

(1) a. Le boucher a coupé le rumsteak = `Agent V Patient` *(The butcher cut the rump steak)*
b. Le rumsteak se coupe facilement = `Patient se V<+middle> ADV` *(The rump steak cuts easily)*

During the development of French Verbənet, problems were encountered owing to the lack of structure in the list of frames for a given class in the original English resource. The aim of this paper is to propose a solution to these problems. We first explain why the current organization of the information is detrimental to the resource quality (Section 2). Then, after highlighting differences between English VerbNet and its French counterpart in terms of optionality and order of complements (Section 3), we explain how frames are automatically typed for French and how this typing helps to enhance the resource (Section 4). Finally, Section 5 discusses the portability of the typing program to other languages.

This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: http://creativecommons.org/licenses/by/4.0/

2 Problems with frame structuring in VerbNet

In VerbNet, verb classes are organized in a hierarchy in which sub-classes inherit properties of parent classes. The components of a class are 1) **Members**, a list of verbs – considered in one of their senses, if polysemous – belonging to the class; 2) **Roles**, a list of thematic roles shared by the members, with each thematic role optionally further characterized by certain selectional restrictions;[1] and 3) **Frames** which is a list of frames characterizing the syntactico-semantic behavior of the members. Each frame consists of a **syntactic surface construction**, an `EXAMPLE` sentence, a `SYNTAX` field in which thematic roles are mapped to syntactic complements, and a `SEMANTICS` field that indicates how the participants are involved in the event. As an illustration, Figure 1 shows the list of the seven frames that describe the get-13.5.1 class.[2] This class includes verbs that denote an action consisting of obtaining something (e.g *buy, catch, order, reach*). This kind of action generally implies the following thematic roles: `Agent`, `Theme`, `Source`, `Beneficiary` and `Asset`.

FRAMES

NP V NP

EXAMPLE	"Carmen bought a dress."
SYNTAX	AGENT V THEME
SEMANTICS	HAS_POSSESSION(START(E), ?SOURCE, THEME) TRANSFER(DURING(E), THEME) HAS_POSSESSION(END(E), AGENT, THEME) CAUSE(AGENT, E)

NP V NP PP.SOURCE

EXAMPLE	"Carmen bought a dress from Diana."
SYNTAX	AGENT V THEME {FROM} SOURCE
SEMANTICS	HAS_POSSESSION(START(E), SOURCE, THEME) TRANSFER(DURING(E), THEME) HAS_POSSESSION(END(E), AGENT, THEME) CAUSE(AGENT, E)

NP V NP PP.BENEFICIARY

EXAMPLE	"Carmen bought a dress for Mary."
SYNTAX	AGENT V THEME {FOR} BENEFICIARY
SEMANTICS	HAS_POSSESSION(START(E), ?SOURCE, THEME) TRANSFER(DURING(E), THEME) HAS_POSSESSION(END(E), AGENT, THEME) CAUSE(AGENT, E) BE

NP V NP.BENEFICIARY NP

EXAMPLE	"Carmen bought Mary a dress."
SYNTAX	AGENT V BENEFICIARY THEME
SEMANTICS	HAS_POSSESSION(START(E), ?SOURCE, THEME) TRANSFER(DURING(E), THEME) HAS_POSSESSION(END(E), AGENT, THEME) CAUSE(AGENT, E) BE

NP V NP PP.ASSET

EXAMPLE	"Carmen bought a dress for $50."
SYNTAX	AGENT V THEME {FOR} ASSET
SEMANTICS	HAS_POSSESSION(START(E), ?SOURCE, THEME) TRANSFER(DURING(E), THEME) HAS_POSSESSION(END(E), AGENT, THEME) CAUSE(AGENT, E) CO

NP.ASSET V NP

EXAMPLE	"$50 won't even buy a dress."
SYNTAX	ASSET V THEME
SEMANTICS	HAS_POSSESSION(START(E), ?SOURCE, THEME) TRANSFER(DURING(E), THEME) HAS_POSSESSION(END(E), ?AGENT, THEME) CAUSE(?AGENT, E)

NP V NP PP.SOURCE NP.ASSET

EXAMPLE	"FMC has bought 565,000 shares from Nortek Inc. at $23.50 a share."
SYNTAX	AGENT V THEME {FROM} SOURCE {AT FOR} ASSET
SEMANTICS	HAS_POSSESSION(START(E), SOURCE, THEME) HAS_POSSESSION(END(E), AGENT, THEME) TRANSFER(DURING(E), THEME) COST(E, ASSET)

Figure 1: List of frames for the get-13.5.1 class.

One can see in Figure 1 that frames differ in the number of complements (e.g. one complement in the 1st and 6th frames, two complements in the 2nd and 3rd frame, etc.), in their syntactic nature (e.g. `Beneficiary` is realized as a prepositional complement in the 3rd frame, as a direct complement in the 4th frame), in their syntactic function (e.g. `Asset` is realized as an oblique complement in the 5th frame, as the subject in the 6th frame), etc. In addition to these cases, it must be noted that if a given argument can be realized in different ways (nominal, infinitival, sentential, etc.), then there is one frame by type of realization. This is illustrated in Figure 2 showing the list of frames for the urge-58-1 class with two frames according to the realization of `Topic`. The non-nominal syntactic realizations of a thematic role are specified in syntactic restrictions, which are written between (angle) brackets.

In summary, there can be quite a number of frames in English VerbNet for a given class. The problem we want to highlight is the absence of organization and typing in the list of frames. First, alternation variants are not explicitly related, which is a loss of information for any NLP system using VerbNet.

[1] The terms thematic and semantic roles refer to the very same notion.

[2] Because of lack of room, the `SEMANTICS` field is cut.

FRAMES	
NP V NP	
EXAMPLE	"I advised him."
SYNTAX	AGENT V RECIPIENT
SEMANTICS	URGE(DURING(E), AGENT, RECIPIENT, ?TOPIC)
NP V NP S_ING	
EXAMPLE	"I advised him to come."
SYNTAX	AGENT V RECIPIENT TOPIC <+OC_TO_INF>
SEMANTICS	URGE(DURING(E), AGENT, RECIPIENT, TOPIC)
NP V NP THAT S	
EXAMPLE	"I advised him that he should come."
SYNTAX	AGENT V RECIPIENT TOPIC <+THAT_COMP>
SEMANTICS	URGE(DURING(E), AGENT, RECIPIENT, TOPIC)

Figure 2: List of frames for the urge-58-1 class.

For instance, the 3rd and 4th frames in Figure 1 are not explicitly related as variants in the "Benefactive Alternation" (Levin, 1993, pp 48-49). The information that they describe exactly the same situation is thereby lost. By the same way, the alternations that induce a change of meaning — e.g. the "locative alternation" (Levin, 1993, pp 49-55) — are not identified and so the change in meaning is lost. Next, it is difficult to know whether a combination of complements is either impossible or possible but not coded. For example, the sentence *Carmen bought a dress from Diana for Mary* seems to be correct but no frame explicitly encodes this configuration. Similarly, the sentence in the 4th frame *Carmen bought Mary a dress* can be extended with a `Source` complement (*Carmen bought Mary a dress from Diana*) or an `Asset` complement (*Carmen bought Mary a dress for $50*) but there is no frame for such extensions.

Our point is not to criticize the coding of the get-13.5.1 class or any class but to emphasize that the absence of organization and/or typing in the list of frames for a given class can lead to errors, incoherencies and oversights, because the linguist is not guided in her work when creating the list of frames. In Section 4, we propose a solution to overcome this problem. We first underscore the differences in frame coding between the English and French resources.

3 Optionality and order of complements in French Verbənet

As stated previously, the French Verbənet was initially created by adapting the English resource to French, which means that the structure of this French resource is nearly identical to that of the English one (Pradet et al., 2014). It was developed using a web interface available at `https://verbenet.inria.fr` and illustrated in Figure 3 for the Settle-89 class.

However, there are two points where the French Verbənet differs from the English VerbNet.[3] The first one is that sub-structures, i.e. structures in which an optional complement is unexpressed, are never coded for French while they are sometimes coded for English. The second one is that the order of complements is not coded in French Verbənet. A prime example of the different coding choices between English VerbNet and French Verbənet is given in the class send-11.1. It has five frames in English given in Figure 4, while it has only one frame in French, which corresponds to the fourth one in Figure 4 (*Nora a envoyé le livre de Paris à Londres / Nora sent the book from Paris to London*). We will see that these different coding choices have implications for the automatic frame typing program (Section 5).

The choice not to encode sub-structures in French Verbənet is due to the fact that an unexpressed complement may lead to different interpretations. Considering only optional objects, (Levin, 1993, pp 33-40) identified eight unexpressed object alternations such as "Unspecified object Alternation" (*Mike ate the cake* → *Mike ate*) or "Characteristic property of Agent Alternation" (*That dog bites people* → *That dog bites*). In conclusion, it's not informative to simply encode a sub-structure as acceptable without stating to which situation it corresponds. For example, the frame for *That dog bites* should be typed as a variant in the "Characteristic property of Agent Alternation". For French Verbənet, such a work has not

[3] We remind the reader (Section 1) that the differences between French and English VerbNet due to the discrepancies between the two languages are discussed in (Danlos et al., 2016) and left aside here.

settle-89

Classe 89 ✖ Cacher

U2a ou F4b ou X1a

14 ou 16

Valider les verbes : ✔ LADL ✔ LVF ✔ Les deux

une classe LVF par verbe, s'accorder à, s'arranger pour

- Paragon : s'harmoniser
- VerbNet members: bargain cohere concur consent contract covenant harmonize jibe resolve settle square
- Membres de Verb∋Net: **accorder s'** **arranger s'** ~~concourir~~ **entendre s'** ~~réduire~~ ~~correspondre~~ ~~harmoniser~~ **allier s'** ~~arrêter~~ ~~concilier~~ ~~décider se~~ ~~dissiper~~ ~~marier~~ ~~marier se~~ ~~régler~~ ~~résoudre se~~ ~~tenir se~~ ~~trancher~~ [montrer] ✚
- Roles : Agent [+animate | +organization] Co-Agent [+animate | +organization] Goal ✚

NP V PP.co-agent pour V-Inf ✖

Exemple	Luc s'est entendu avec Marie pour proposer un projet
Syntaxe	Agent V {avec} Co-Agent {pour} Goal<+VAgent-inf>
Sémantique	agree(result(E), Agent, Co-Agent, Goal)

NP V PP.co-agent PP ✖

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

NP V pour V-inf ✖

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

NP V PP ✖

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

Frames supprimées :

- NP V (Eventually, they settled.) ↻ Réintégrer
- NP V PP.goal what S_INF (They settled on what to do.) ↻ Réintégrer
- NP V PP.co-agent PP.goal (I settled with them on a novel proposal.) ↻ Réintégrer
- NP V PP.co-agent PP.goal <+interrog> (Luc s'est entendu avec Marie sur ce qu'ils doivent faire.) ↻ Réintégrer

✚ Ajouter une frame

❯ Ajouter une sous-classe ↻ Envoyer les verbes

Figure 3: Web interface for editing and viewing Verb∋net.

FRAMES	
NP V NP	
EXAMPLE	"Nora sent the book."
SYNTAX	AGENT V THEME
SEMANTICS	MOTION(DURING(E), THEME) CAUSE(AGENT, E)
NP V NP PP.INITIAL_LOCATION	
EXAMPLE	"Nora sent the book from Paris."
SYNTAX	AGENT V THEME {{+SRC}} INITIAL_LOCATION
SEMANTICS	MOTION(DURING(E), THEME) LOCATION(START(E), THEME, INITIAL_LOCATION) CAUSE(AGENT, E)
NP V NP PP.DESTINATION	
EXAMPLE	"Nora sent the book to London."
SYNTAX	AGENT V THEME {TO} DESTINATION
SEMANTICS	MOTION(DURING(E), THEME) LOCATION(END(E), THEME, DESTINATION) CAUSE(AGENT, E)
NP V NP PP.INITIAL_LOCATION PP.DESTINATION	
EXAMPLE	"Nora sent the book from Paris to London."
SYNTAX	AGENT V THEME {{+SRC}} INITIAL_LOCATION {TO} DESTINATION
SEMANTICS	MOTION(DURING(E), THEME) LOCATION(START(E), THEME, INITIAL_LOCATION) LOCATION(END(E), THEME, DESTINATION)
NP V NP PP.DESTINATION PP.INITIAL_LOCATION	
EXAMPLE	"TransCanada is shifting its HQ to Calgary from Toronto."
SYNTAX	AGENT V THEME {TO} DESTINATION {{+SRC}} INITIAL_LOCATION
SEMANTICS	MOTION(DURING(E), THEME) LOCATION(START(E), THEME, INITIAL_LOCATION) LOCATION(END(E), THEME, DESTINATION)

Figure 4: List of frames for the send-11.1 class.

yet been done, and so no sub-structure is coded.[4]

The choice not to encode the order of complements is due to the fact that the order of the complements of a verb in French does not depend on the verb itself but on the "weight" of the complements, the weight being computed according to the number of words (Gross, 1975) or other factors described in (Thuilier, 2012). [5] In English VerbNet, it happens that two frames in a class differ only by the order of the complements. This is the case in the class send-11.1, where the last two frames in Figure 4 differ only by the order of `Initial-Location` and `Destination` complements.

4 French Verbənet: version with typed frames

The typed version of Verbənet from the English-like version has been created automatically [6]. In this typed version, any frame is typed so as to show explicitly whether it is "canonical" or obtained by an alternation. Before going into detail, let us provide an illustrative example. The settle-89 class, which was shown in Figure 3 in the English-like version, is shown in the typed version in Figure 5. The type of the first frame is **Canonique** (canonical). The type of the second one is **Canonique avec Goal[+V-inf]** which states that the `Goal` complement is realized as an infinitival phrase, while it is understood that is is realized as a noun phrase — the default value — in the canonical frame. The type of the third frame is **Alt. Symetrique** which states that this frame is obtained from the canonical one by the alternation named "Simple reciprocal Alternation (intransitive)" in (Levin, 1993, pp 62-63). The type of the last frame is **Alt. Symetrique avec Goal[+V-inf]** which states that the `Goal` complement is realized as an infinitival phrase in the symmetrical (reciprocal) alternated form.

The automatic conversion of the untyped version of Verbənet into the typed one led us to discover and correct incoherencies for the enhancement of the resource in a virtuous circle: the untyped version is changed until the typed version is satisfactory.

The conversion program is going to be explained step by step, first in parent classes (section 4.2), second in sub-classes (section 4.3). Next we will illustrate how the frame typing allows the enhancement of the original resource by discovering and correcting incoherences (section 4.4). Before that, we need to discuss the notion of canonical frame.

[4] To use French Verbənet for a task such as SRL, as an initial approximation one may consider any complement to be optional.

[5] To use French Verbənet for a task such as SRL, one may consider any permutation of complements to be acceptable.

[6] The input of this program is the XML version of Verbənet available at `https://github.com/aymara/verbenet`.

settle-89

Classe settle-89

- Membres de Verbɔ̃Net: 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 5: New Web interface for viewing Verbɔ̃net

4.1 What is a canonical frame ?

In Levin's description of English alternations, this notion is absent: there exists only the notion of two variants of an alternation which are on the same footing. It is not our intention here to enter a theoretical discussion on the topic, however, from a lexicographic and NLP perspective, it seems justified to state that verbs such as *break, cut, hit* and *touch* are "canonically" transitive and participate in various diathesis alternations (middle, conative, causative/inchoative alternations) which makes them intransitive — although some authors, e.g. (Dubois and Dubois-Charlier, 1997) for French, consider intransitive forms such as *The carafe broke* as canonical compared to the transitive causative form *Fred broke the carafe*.

One of the outcomes of this work — which is still in progress — will be to determine classes of verbs for which a canonical form can be identified and to understand why the other classes do not exhibit a canonical frame. As a prime illustration, our frame typing program gives poor results and doesn't identify a canonical frame in class 55 of aspectual verbs *(commencer (begin), continuer (continue))*, but this is not a surprise: these verbs are included in VerbNet for the sake of coverage, but they could have been excluded, much as modal and light verbs are excluded.

A second question is: should there exist only a unique canonical frame? Consider the class banish-

10.2 in Figure 6. In (Levin, 1993, p 123) it is stated that all banish verbs "allow *to* phrases as well as *from* phrases, though not simultaneously", so the second or third frame in Figure 6 cannot be considered as a sub-structure of a frame including both `Source` and `Destination` complements. Thus, there is apparently no reason to consider one as more canonical than the other, except that banish verbs are verbs "which relate to the removal of an entity, typically a person, from a location" (Levin, 1993, p 123), which seems to promote the frame with a `Source` as canonical on semantic grounds. So the idea of two canonical frames could be accepted and one of the outcomes of this work will be to understand what are the classes of verbs with two potential canonical forms and whether one of these forms can be promoted as canonical on semantic grounds.

FRAMES	
NP V NP	
EXAMPLE	"The king banished the general."
SYNTAX	AGENT V THEME
SEMANTICS	CAUSE(AGENT, E) LOCATION(START(E), THEME, ?SOURCE) LOCATION(END(E), THEME, ?DESTINATION)
NP V NP PP.SOURCE	
EXAMPLE	"The king banished the general from the army."
SYNTAX	AGENT V THEME {{+SRC}} SOURCE
SEMANTICS	CAUSE(AGENT, E) LOCATION(START(E), THEME, SOURCE) NOT(LOCATION(END(E), THEME, SOURCE))
NP V NP PP.DESTINATION	
EXAMPLE	"The king deported the general to the isle."
SYNTAX	AGENT V THEME {TO} DESTINATION
SEMANTICS	CAUSE(AGENT, E) LOCATION(START(E), THEME, ?SOURCE) LOCATION(END(E), THEME, DESTINATION)

Figure 6: List of frames for the banish-10.2 class.

4.2 Frame typing in parent classes

The typing program identifies first the canonical frame, then the alternated frames and finally the frames with syntactic restrictions, as described below.

Canonical frame. The program which converts the untyped version into the typed one first requires the canonical frame to be identified, which raises problems discussed above. Currently, the strategy we use to automatically identify the canonical form is to spot the frame that includes all the thematic roles listed in the component **Roles** of the class (Section 2) — and which doesn't include any syntactic restrictions (written between (angle) brackets). This strategy relies on the fact that the order of complements is not coded in French Verbɔnet (Section 3).

Alterned frame. The program detects an alternated frame thanks to a set of rules designed for the set of coded alternations. For example, a rule can type a frame as **Alt. Instrument subject** (Levin, 1993, pp 80) when the subject of the canonical frame, i.e. the thematic role on the left of the symbol *V*, is replaced by the thematic role `Instrument`, see `Agent V Patient {avec} Instrument` → `Instrument V Patient`. Another rule types a frame as **Alt. Symetrique** when a `Co-Agent` (resp. `Co-Patient`) in the canonical frame is replaced by an `Agent` (resp. `Patient`) marked as `<+plural>`, see `Agent V {avec} Co-Agent {sur} Goal` → `Agent<+plural> V {sur} Goal`.

One of the main difficulties encountered in typing alternated frames is observed with cascades of alternations. This is illustrated in the paradigm in (2) from the almagate-22.2 class: (2a) gives the canonical frame, (2b) the symmetrical alternated frame, (2c) the neutral alternated frame. The difficulty in typing is for (2d) which gives the alternated form when both the symmetrical and neutral alternations apply in any order from the canonical frame.

(2) a. **Canonique**: `Agent V Patient {avec} Co-Patient`
Fred a associé la compagnie α avec la compagnie β (Fred associated company α with company β)
b. **Alt. Symetrique**: `Agent V Patient<+plural>`
Fred a associé les compagnies α et β (Fred associated companies α and β)
c. **Alt. Neutre**: `Patient se V<+neutre> {avec} Co-Patient`

La compagnie α s'est associée avec la compagnie β (Company α associated itself with company β)

d. **Alt. Symetrique & Neutre**: `Patient<+plural> se V<+neutre>`

Les compagnies α et β se sont associées (Companies α and β associated themselves with one another)

Frame with syntactic restriction Finally, the program has to identify two frames which are identical except that in one of the frames a thematic role `X` has a syntactic restriction [synt], which states that the realization of `X` is not nominal but infinitival, for example. The two frames are then easily typed **T** and **T with X[synt]**.

4.3 Frame typing in sub-classes

For sub-classes, the frame typing program relies on the idea that the canonical frame is in the parent class. Consider the French sub-class bend-45.2.1 which includes verbs for which the non-pronominal inchoative form — named **Alternation inchoative** — is possible on top of the pronominal form *(La tige a plié / La tige s'est pliée (The rod bent))*, whereas the parent class bend-45.2 includes verbs for which only the pronominal inchoative form — named **Alternation neutre** — is possible *(La tige s'est incurvée / *La tige a incurvé (The rod curved))*.[7] As a consequence, the unique frame in the sub-class bend-45.2.1 receives the type **Alternation inchoative** which is to be understood as an alternated form of the canonical form in the parent class. More generally, if a sub-class was created because of an alternation A which is possible for only some verbs of the parent class, there is no problem with frame typing: the unique frame in the sub-class receives type **Alternation A**.

However, a sub-class may also be created for other reasons, one of them being variants in prepositions introducing complements. In French, the class correspond-36.1 includes communication verbs for which the `Theme` is introduced by the preposition *sur*, as illustrated in *Le comité a délibéré sur ce point (The committee deliberated on this issue)*. A sub-class of correspond-36.1 includes verbs that also allow the `Theme` to be introduced by the preposition *de*, as illustrated in *Le comité a discuté de ce point (The committee discussed this issue)*. The frame in the sub-class, which does not fall within alternation variants nor alternative syntactic realization (in the sense given above), has not been typed yet.

4.4 Discovering incoherencies

The automatic conversion of the untyped Verbənet version into the typed one led us to discover incoherencies. As an illustration, it has been discovered that the coding of "possessor-attribute factoring alternations" (Levin, 1993, pp 72-78) called "restructurations" in French was not satisfactory (it is not satisfactory in English either). These alternations arise because a possessor and a possessed attribute may be expressed in two different ways. As one option, they may be expressed in a single complex noun phrase whose head is the attribute modified by a genitive for the possessor *(Fred adores Jane's humour)*. Alternatively, they may be expressed separately, one as a direct argument (subject or object), and the other via a prepositional phrase *(Fred adores Jane for her humour)*. A unique syntactic function which is expressed in a complex noun phrase is thus restructured into two syntactic functions. For thematic roles, this is unusual since the complex noun phrase receives a unique thematic role while there must be two roles for the two syntactic functions in the restructured variant.

As a consequence, we chose the following solution: we use two distinct frames to code that a given thematic role, for example `Stimulus` for the object of *adorer (adore)*, is either a simple noun phrase *(Fred adores this paint)* or a complex noun phrase *(Fred adores Jane's humour)*, see (3a) and (3b). The frame with the complex noun phrase is identified with the restriction `Stimulus[+genitive]`, which requires to divide the role `Stimulus` into two parts: `Stimulus.prop` for the attribute which is the head noun and `Stimulus.poss` for the possessor expressed in a genitive phrase. These two parts are naturally used in the restructured form, see (3c) with two thematic roles assigned to the two complements. The coding for a verb such as *adorer* in the admire-31.2 class is schematized in (3), in which the non-nominal realizations of `Stimulus` are left aside. [8]

[7] In English, there is no sub-class bend-45.2.1 since the inchoative alternated form is compulsorily a non-pronominal form while it is pronominal and/or non-pronominal in French.

[8] In VerbNet, the role `Stimulus.prop` is named `Attribute` in the frame equivalent to (3c), and (3b) is not coded in a specific frame.

(3) a. **Canonique**: `Experiencer V Stimulus`
Fred adore cette peinture (Fred adores this paint)
b. **Canonique with Stimulus[+genitive]** : `Experiencer V Stimulus.prop<+genitive(Stimulus.poss)>`
Fred adore l'humour de Jane (Fred adores Jane's humour)
c. **Restructuration**: `Experiencer V Stimulus.poss {pour} Stimulus.prop`
Fred adore Jane pour son humour (Fred adores Jane for her humour)

We have not yet finished our virtuous circle between the un-typed and typed versions of Verbənet but the first results we get are promising. As an illustration, 16 classes have been corrected for the coding of restructuration alternations.

5 Methods and portability for typing VerbNet-like resources in other languages

Typing frames can be done in two ways: manually or automatically. Manual typing can be a good solution for a language for which no VerbNet-like ressource exists. When starting the work from scratch, the linguist has every reason to type each frame right away so as to be guided in her work. Manual typing for a language for which a VerbNet-like ressource already exists can be time-consuming (and tiresome). This is why we choose automatic typing for French with the existing Verbənet. This automatic typing is effective in improving the resource as typing errors are more often attributable to errors in the resource itself than to the typing program.

What is the feasibility of porting a typing program from one language to another? We cannot currently answer this question since, as far as we know, French is the only language for which there is a typing program, however we do have some indication. We focus on the identification of the canonical frame (in a parent class) since it should be clear from the description of the French typing program in Section 4 that this is the most difficult point. The canonical frame in a parent class can generally be identified thanks to the fact that it is this frame that includes all the thematic roles pertinent for the class with no restriction on any thematic role (which means that all thematic roles are realized as simple noun phrases) and no role such as `Stimulus.poss` or `Stimulus.prop` (which means that no restructuration is involved, Section 4). In French, there are only a few exceptions to this principle. For example, in class appoint-29.1 *(nommer (nominate))*, the two frames include all the thematic roles and differ only by the presence of *comme (as)* in the "*as* Alternation" (Levin, 1993, p 78). For this class, we typed (somewhat arbitrarily) **Canonique** the frame without *comme (as)* and **Alt. comme** the other frame. In English, there are more exceptions to this principle. First, the "Dative and Benefactive alternations" (Levin, 1993, pp 45-49), which don't exist in French, don't change the number of thematic roles. So specific rules must be designed to spot these alternated frames. Second, two frames may differ only by the order of complements (Section 3). In this case, it is not clear what should be done.

Finally, let us examine sub-structures. As we explained in Section 3, sub-structures have not yet been coded in French Verbənet because it is not informative to code a sub-structure as acceptable without stating to which situation it corresponds. This means that when they are coded they will be simultaneously typed, e.g. `Agent V` = *Ce chien mord (This dog bites)* typed as **Alternation Characteristic property of Agent** variant of the canonical frame `Agent V Patient` = *Ce chien a mordu Jane (This dog bit Jane).* For English, an automatic typing program from the existing VerbNet can only type *This dog bites* as a sub-structure, without any other information.

6 Conclusion

We have shown that the lack of structure/typing in the list of frames for a VerbNet class make the coding task of the linguist difficult and it can lead to incoherencies or oversights. We have proposed frame typing as a method to overcome this problem. The types are: (i) canonical, (ii) canonical with a non-nominal or complex (i.e with a genitive) nominal realization of a thematic role, (iii) alternated form of another frame where the other frame can iteratively be canonical, canonical with a non-nominal simple realization or alternated form, (iv) untyped when the typing program gives no result.

We have presented an automatic typing program for French which we believe is easily portable to other languages. The automatic conversion of the untyped Verbənet version into the typed one enabled

us to discover and correct incoherencies, thus enhancing the resource in a "virtuous circle". We have not yet evaluated the effects of this enhancement, but other VerbNet-like resources could be enhanced as we have done for French.

Acknowledgements

We thank André Bittar for editing our English.

References

Marzieh Bazrafshan and Daniel Gildea. 2013. Semantic Roles for String to Tree Machine Translation. In *Proceedings of ACL conference*, pages 419–423, Sofia, Bulgaria.

Lucia Busso and Alessandro Lenci. 2016. Italian-Verbnet: A construction-based approach to Italian verb classification. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portorož, Slovenia.

Laurence Danlos, Quentin Pradet, Lucie Barque, Takuya Nakamura, and Matthieu Constant. 2016. Un Verbenet du français. *Traitement Automatique des Langues*, 57(1):33–58.

Jean Dubois and Françoise Dubois-Charlier. 1997. *Les verbes français*. Larousse-Bordas, Paris, France.

Peter Exner and Pierre Nugues. 2011. Using semantic role labeling to extract events from Wikipedia. In *Proceedings of the Workshop on Detection, Representation, and Exploitation of Events in the Semantic Web (DeRiVE 2011). Workshop in conjunction with the 10th International Semantic Web Conference (ISWC 2011)*, pages 38–47, Bonn, Germany.

Maurice Gross. 1975. *Méthodes en syntaxe*. Hermann, Paris.

Karin Kipper, Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extending VerbNet with novel verb classes. In *Proceedings of LREC*, Genoa, Italy. Citeseer.

Karin Kipper-Schuler. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. Ph.D. thesis, University of Pennsylvania.

Beth Levin. 1993. *English verb classes and alternations: a preliminary investigation*. University Of Chicago Press.

Ahmed Hamza Osman, Naomie Salim, Mohammed Salem Binwahlan, Rihab Alteeb, and Albaraa Abuobieda. 2012. An improved plagiarism detection scheme based on semantic role labeling. *Applied Soft Computing*, 12(5):1493–1502.

Quentin Pradet, Laurence Danlos, and Gaël De Chalendar. 2014. Adapting VerbNet to French using existing resources. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 2014)*, Reykjavík, Iceland.

Carolina Scarton, Magali Sanches Duran, and Sandra Maria Alusio. 2014. Using cross-linguistic knowledge to build VerbNet-style lexicons: Results for a (Brazilian) Portuguese VerbNet. In Jorge Baptista, Nuno Mamede, Sara Candeias, Ivandré Paraboni, Thiago A. S. Pardo, and Maria das Graças Volpe Nunes, editors, *Computational Processing of the Portuguese Language*. Springer International Publishing.

Dan Shen and Mirella Lapata. 2007. Using Semantic Roles to Improve Question Answering. In *Proceedings of EMNLP-CoNLL*, pages 12–21, Prague, Czech Republic.

Mihai Surdeanu, Sanda Harabagiu, Johns Williams, and Paul Aarseth. 2003. Using predicate-argument structures for information extraction. In *Proceedings of the ACL Conference*, pages 8–15, Sapporo, Japan.

Juliette Thuilier. 2012. *Contraintes préférentielles et ordre des mots en français*. Ph.D. thesis, Université Paris-Diderot.

Boyi Xie, Rebecca J. Passonneau, Leon Wu, and Germán G. Creamer. 2013. Semantic Frames to Predict Stock Price Movement. In *Proceedings of the ACL conference*, pages 873–883, Sofia, Bulgaria.