Italian VerbNet: A Construction-based Approach to Italian Verb Classification

Lucia Busso, Alessandro Lenci

CoLingLab, Department of Philology, Literature and Linguistics, University of Pisa
via S. Maria 36, 56126 Pisa, Italy
E-mail: lucia.busso90@gmail.com, alessandro.lenci@unipi.it

Abstract

This paper proposes a new method for Italian verb classification -and a preliminary example of resulting classes- inspired by Levin (1993) and VerbNet (Kipper-Schuler, 2005), yet partially independent from these resources; we achieved such a result by integrating Levin and VerbNet's models of classification with other theoretic frameworks and resources. The classification is rooted in the constructionist framework (Goldberg, 1995; 2006) and is distribution-based. It is also semantically characterized by a link to FrameNet'ssemanticframesto represent the event expressed by a class. However, the new Italian classes maintain the hierarchic "tree" structure and monotonic nature of VerbNet's classes, and, where possible, the original names (e.g.: Verbs of Killing, Verbs of Putting, etc.). We therefore propose here a taxonomy compatible with VerbNet but at the same time adapted to Italian syntax and semantics. It also addresses a number of problems intrinsic to the original classifications, such as the role of argument alternations, here regarded simply as epiphenomena, consistently with the constructionist approach.

Keywords: verb classification, construction grammar, VerbNet

1. Introduction

Verb classification and its implications have been at the center of linguistic research on the syntactic-semantic interface for almost 50 years. In fact, since the pioneer works of Charles Fillmore (Fillmore, 1968, 1970), it has been recognized that verbs that share common semantic features can be grouped on the basis of their regular syntactic patterns of argument realizations. Fillmore suggested that a verb syntactic behavior (i.e., its argument realization) and lexical semantics are linked, and that verbs tend to group accordingly to these variables. In Fillmore (1970), the author proposed the famous distinction between *break* and *hit*, which do not show similar syntactic behavior even though they share a intuitive and basic meaning. For instance, only *break* allows for the inchoative alternation:

(1)   a. *The stick broke.*
      b. **The tree hit.*

                [Fillmore, 1970: 126-128]

The two verbs actually represent larger classes that include other similar verbs that show the same syntactic realizations:

    a. Break verbs: *bend, fold, shatter, crack...*
    b. Hit verbs: *slap, strike, bump, stroke...*

                [Fillmore, 1970: 130]

A verb class is thus intended as a homogenous group of verbs, defined in terms of shared meaning components and similar syntactic behavior, or more specifically a set of "semantically related verbs sharing a range of linguistic properties, such as the possible realizations of their arguments and the particular interpretation associated with each possible argument realization." (Levin, 2013:1).

Understanding how verbs can be grouped and what is precisely the status of the resultant verb classes has been a major goal of many scholars operating in various and different research fields and from different approaches: lexical semantics, computational linguistics, cognitive sciences etc.(see among others Pinker, 1989; Jackendoff, 1990; Levin, 1993; Dorr, 1997; Dang et al., 1998; Merlo and Stevenson, 2001).

A very large literature on verb classification has appeared since the seminal works by Fillmore, because semantically coherent verb classes are recognized as a useful device for capturing many generalizations over a vast range of properties, both within a given language and cross-linguistically, and can therefore be used as a valuable means of inquiry. English has been indubitably the more studied and analyzed language. Several classifications are currently available for English verbs (e.g.: Pinker, 1989; Jackendoff, 1990; Faber, 1999). The largest and most widely renown is however Levin (1993), a pioneering study on verb classes based on argument alternations. A number of studies has sprung from Levin's work; in particular, the broad-coverage online lexicon VerbNet (Kipper et al., 2000; Kipper-Schuler, 2005), which is almost completely based on Levin (1993) and on its subsequent implementations (see below).

Research on cross-linguistic verb classifications grounded in the syntax-semantics interface, following the model of Levin (1993) and VerbNet, has gained prominence in the scientific community during the last two decades. Different versions of VerbNet-style lexicons have been developed cross-linguistically (see for example Pradet et al., 2014), but besides Merlo et al. (2002), no significant attempt in this direction has been made for Italian yet.[1]

---

[1]Other resources have been translated, adapted or created for Italian: the different semantic classifications of *Italian WordNet*(Pianta et al., 2002) and *ItalWordNet*(Roventini et al. 2000), or *Simple*'s verb classes (Lenci et al., 2000), which is partly inspired to the Generative Lexicon developed by Pustejovsky (1995).Jezek (2003) provided an independent syntactic-semantic classification for Italian verbs, limited to transitive and (unaccusative / unergative)intransitives.

This work aims to fill this gap by developing a reliable, comprehensive and coherent method for the creation of Italian verb classes based on both syntactic and semantic ground. The present classification is rooted in the constructionist framework and modeled on the system of Levin (1993) and VerbNet.

The paper is articulated as follows: in section two, we will be briefly review the Levin and VerbNet classification model; section three is dedicated to a description of the new classification, including the method we developed, the online resources that were used, and a general class layout and structure. The case study of verbs of KILLING can be found in the appendix, along with the synoptic schema of our class hierarchy and the taxonomy used for selectional preferences.

2. Levin and VerbNet Classifications

Levin classified English verbs according to the presence- or absence- in their syntactic patterns of argument alternations, that is "alternations in the expressions of arguments, sometimes accompanied by changes of meaning." (Levin, 1993:2) In fact, Levin agrees with various studies (Fillmore 1967, Guerssel et al.1985, Hale and Keyser 1986, 1987) in considering argument alterations as sensitive to particular components of verb meaning, and therefore the best indicator of differences in verb behavior.

*English verb classes and alternation*s represents a milestone work on verb classification, but it has nonetheless several weak points which led a number of scholars to try and revise the original work. First of all, given the preliminary nature of Levin's work, the scope of the classification is quite limited: it only considers verbs taking noun and prepositional phrase complements, and excludes those taking sentential arguments. Secondly, even though the author states that class membership depends on alternations, the resulting classes appear to have in some cases a "mixed character", and to be at least partially semantically motivated as well. In the following years, several extensions of Levin (1993) were proposed. The major implementations of the original classification are Dang et al. (1998) and Korhonen and Briscoe (2004), which has been carried out with a semi-automatic approach.

The original and extended versions of Levin's work have been often used by the Natural Language Processing (NLP) community as evidence for the semantic similarity of verbs (Jing &McKeown, 1998; Lapata& Brew, 1999; Kohl et al., 1998). Class-based information such as that provided by Levin (1993) has in fact proved to be extremely useful for NLP tasks, like language generation, machine learning and word sense disambiguation tasks, in order to capture important generalizations among verbs, thereby leaning data-sparseness problems (Kipper et al., 2008).

The most influential computational implementation of Levin's classes, however, is the online resource VerbNet (Kipper et al., 2000; Kipper-Schuler, 2005). VerbNet is the most important broad-coverage class-based online verb lexicon developed for English. It is hierarchically organized, giving rise to a "tree" structure, and the relationship between parent-class and child-class is strictly monotonic. Its verb classes are crucially based on Levin (1993) and its extensions, especially Korhonen and Briscoe (2004). Original classes have been extended and refined in order to achieve a higher (syntactic and semantic) coherence among members of a class (see Kipper et al. 2006a and 2006b; Kipper et al. 2008;).

Numerous attempts have been made to translate or adapt such a useful resource in other languages, the main assumption of most approaches being that the basic meaning components shared by classes can be applied cross-linguistically (Jackendoff,1990). For example, Merlo et al. (2002) have used cross-linguistic similarities to convert 20 Levin classes to Italian, obtaining high accuracy (86.3%). Recent direct translations of VerbNet are the ones of Estonian VerbNet (Jentson, 2014) and Brazilian Portuguese (Scarton & Aluısio, 2012). Other studies comparable to VerbNet were also done for Spanish (Ferrer, 2004), German (Schulte ImWalde, 2006), and Japanese (Suzuki &Fukumoto, 2009).

For French, several studies and researches have been done throughout the last 20 years. Saint-Dizier (1996) first produced a resource rather similar to VerbNet. Later work has focused on the automatic acquisition of subcategorization frames, grouped according to their syntactic and semantic similarity (Sun et al., 2010). The most complete work on the creation of a French VerbNet is however Pradet et al. (2014).

Differently from most of the researches above, our aim has not been to directly translate VerbNet in Italian, but rather to develop a new, autonomous and independent classification for Italian verbs, accounting for Italian the syntactic peculiarities, and stemming from the integration of various resources and theoretical frameworks, and yet compatible with Levin/VerbNet classes (henceforth: L/V)

3. Towards an Italian Verb Classification

The method that we propose to build an Italian verb classification is rooted in the Construction Grammar paradigm (Goldberg 2006, Hoffmann and Trousdale 2013).Moreover, we adopted a radical *distributional* [2] approach. In other words, the classes were found by examining corpus-based data and empirical distributions. As our classification is constructional in nature, we did

---

[2] One of the main divergence among different verb classification lies in the starting assumption, in particular, on the main distinction between *ontology-based* and *distribution-based* classifications. These two types of approaches differ with respect to "the extent to which the distributional properties of verbs, i.e. the set of linguistic constructions and patterns they occur with, is adopted as the main criterion for class identification and class membership." (Lenci, 2014:17)

not take argument alternations to be reliable indicators of verb meaning and behavior, but rather we considered them as mere epiphenomena. Each "alternating" form was analyzed independently, since "differences among instances of the same surface pattern are often most naturally attributed directly to the different verbs and arguments involved." (Goldberg, 2002: 327). That is, Goldberg -and the constructionist approach in general- claims that it is preferable to avoid positing derivations and to concentrate on surface forms, i.e., constructions, since there are generally more numerous, broad and powerful generalizations surrounding particular surface forms than the ones that come from transformational and derivational accounts. The constructionist approach argues that each argument structure construction specifies its semantic and information-structure properties; this greatly enhances the role of the lexicon, which is believed to include phrasal patterns with their own idiosyncratic syntactic or semantic properties. This is in striking contrast with earlier derivational accounts of argument alternations (see among others, Jackendoff, 1975; Pinker, 1989), which posited specific processes for deriving one variant of an alternation from another.

Since constructions are intrinsically defined as pairings of syntactic templates and semantic and pragmatic content, verb classes are to be characterized by mutually dependent syntactic and semantic layers. An integrated syntactic-semantic description is common to VerbNet as well; however, beside a description of the participants of the event, we also described each class associating it to the corresponding FrameNet conceptual frame (Baker et al., 1998); this overall semantic frame paired with each class is to be understood as a conceptual, schematic representation of a situation, describing the general scenario evoked by the verb. In addition, using FrameNet's frames as a semantic, cognitive and conceptual reference for the classes allowed us to integrate characteristics of the so-called *ontological* taxonomies in an otherwise distributional one. For example, the class of UCCIDERE ("to kill") verbs was associated with the frame of KILLING, in which "a Killer or Cause causes the death of the Victim with an Instrument". Therefore, we provide the following definition of verb class: *verb classes are sets distinct senses of verbs that share common argument realization patterns (i.e., they share the same constructions) and that profile the same template of event (i.e., they evoke the same conceptual frame).* Therefore, it is not a verb *per se*, but rather a particular verb sense is to be considered member of a given class; verb classes, that is, are equivalence classes of verb meanings.

Another important feature of our approach is that the classes are here conceived as having a prototype-like structure (Rosch, 1973). The greater or lesser degree of prototypicality is given by the number of typical constructions of the class the verbs participate in: each class contains a group of "core" verbs that share all the syntactic and semantic characteristics of the general class, and fuzzy boundaries of less prototypical members. In

particular, we argue that specific semantic properties of a specific verb meaning may "block" the realization of some constructions that are instead shared by most of the other verbs of a class, or instead add some idiosyncratic constructions as well. This idea is consistent with the distinction between two components in a verb meaning: one component that is shared by a verb with other members of its class, and one component which is instead specific of a single verb (cf. for instance the notions of template and root in Rappaport-Hovav& Levin, 2003; Levin & Rappaport-Hovav, 2005). We assume here that both such components determine the constructions a verb participates in.

Having described the general theoretical principles that have guided our work, we will now turn to analyze the methodology and the online resources that were used to build Italian verb classes.

3.1. Methodology and Resources

Italian semantic classes were bootstrapped from a representative sample of Italian verbs, the 1000 most frequent Italian verbs, as found in Lebani et al. (2014): these are the verbal lemmas which are marked as highly frequent in the monolingual Italian dictionary Sabatini e Coletti (henceforth: *S&C*) (Sabatini & Coletti, 2012) - the only Italian dictionary that indicates the verb valency in the lexical entry - which were then matched with the corresponding verbs in the *La Repubblica* corpus (Baroni et al., 2004).

The first step towards the building of a new class was always a corresponding Levin's class (e.g. KILL verbs), from which we tried to individuate a comparable Italian class. We then extrapolated from the S&C sample the candidate members (e.g. *uccidere, ammazzare*, "to kill"), trying to create sets of lexemes that cohesively profiled the same type of event. Other verbs were further added (despite not appearing in our original sample), if they satisfied the syntactic and semantic constraints of a particular verb class.

In a second phase, we refined the set of selected verbs by closely examining the syntactic constructions they shared, and linked each syntactic pattern to a semantic roles list and selectional preferences constraints imposed on the various fillers of the slots of the constructions [3]. For example, in the KILL verbs class the roles associated with the 3 arguments are Killer, Victim and Instrument; in the subclass of UCCIDERE verbs the Victim must be Animate, while the Killer may be an Agent or an Inanimate Cause. In the subclass of ASSASSINARE [4] verbs, instead, the Killer must strictly be Animate.

In order to determine the range of constructions associated with each class, we used various online resources. Firstly, S&C provided us with general information about valency and allowed patterns for each

---

[3]The selectional preferences of the arguments are defined with respect to the ontology of semantic types presented by Lebani and Lenci, (2013). The taxonomy of arguments is minimal but linguistically plausible.

[4] "to assassinate"

verb; however, S&C only gives very coarse-grained dictionary-like information, that is does not provide any true insight on the verb distributional properties. This is why we combined its use with the exploration of the corpus-based lexical resource on Italian argument structure LexIt (Lenci et al., 2012). LexIt was particularly useful since it organizes the allowed syntactic frames by their frequency, thus distinguishing between typical and more rare and marked uses. The third resource that was consulted is the Italian section (Cennamo& Fabrizio, 2013) of the typological database ValPal (Hartmann et al. 2013).The ValPal database is part of the Leipzig Valency Classes Project, which aims to follow up cross-linguistically the works by Levin (1993) for English and by Apresjan (1967) for Russian; the authors analyze the lexical realization of 80 verb meanings in 35 different languages. Selected meanings should exemplify a representative sample of the verbal lexicon, are cognitively salient (e.g.: EAT, PUSH, KILL),are reported to have distinctive syntactic behaviors (within and across languages).For each language, ValPaL explores "basic" (i.e.: frequent and not rare) verbs lexicalizing the various meanings (e.g.: in Italian *uccidere* is considered to be the lexicalization of the general meaning KILL). Each verb is then described by its coding frames, associated examples, a table indicating verb meaning, semantic micro-roles, coding sets and argument types, as well as valency alternations.

In order to characterize the semantic side of constructions, we associated each class with a FrameNet's frame. Using Frame Semantics as a theoretical framework for semantic representation allowed us to integrate in our otherwise distributional paradigm an "ontological" perspective (i.e.,FrameNet), so that verb classification can also take into account the properties of the event profiled by a given verb. We adopted FrameNet's roles (aka Frame Elements) as well, in order to be able to ground the description on a preexisting and well-established role list.

We did not include, at least in this firstphase, a formal description of the event and of its temporal structure (which is instead present in VerbNet's semantic description).

3.2. Class Description and Structure

As a preliminary note, it is important to remark that the classes presented here are only meant to be a preliminary showcase of the proposed method. It is clear that much additional work should be done to fully develop a full-blown, large-scale classification of Italian verbs.

Seven Italian verb classes were constructed. Specifically, these classes include verbs of KILLING, PUTTING, FILLING, REMOVING, CLEARING, SENDING AND CARRYING and of CHANGE OF POSSESSION (see appendix). In table 1 at the end of this paragraph we report the total number of verbs, argument structure constructions, classes and subclasses that were analyzed in our work.

We will describe the differences with the original classification below. In this work, we concentrated only on transitive verbs, in particular on verbs that all refer to a macro-event conceptualizing the movement of a Theme performed by an Agent or a Cause, with the sole exception of verbs of KILLING. In fact, this class was used to "tune" our method, because it is not particularly numerous, has clear-cut semantic and syntactic distinctions within subclasses, and the distance between Italian and English class structure is quite small, despite that even in this case some significant differences between the two languages could be detected. At the end of this paragraph the case-study of verbs of KILLING is displayed as an example of our classes' layout.

We maintained the original L/V English name for the general description of the classes (e.g.: verbs of KILLING, verbs of PUTTING, etc.) to underline the compatibility with the English classification, but we named each subclass with its most prototypical Italian member (e.g.: UCCIDERE verbs, METTERE verbs).For instance, our KILLING class corresponds to VerbNet 42nd class; however its subclass ASSASSINARE verbs does not precisely correspond to a VerbNet class but contains elements from the classes 42.1 and 42.2, MURDER verbs and POISON verbs. We will now turn to briefly describing the structure of our verb classes.

Each verb class is completely described by the corresponding VerbNet or Levin class (and the possible subclass number), its member verbs, a semantic frame specifying the associated FrameNet frame, a list of constructions, plus possible idiosyncratic constructions of specific verbs.

The first type of information is the alphabetically ordered list of the class members (see below). Each verb, or rather sense of the verb, is associated with a glossed example. The associated semantic frame includes an informal description of the event itself together with the FrameNet reference frame, and a table specifying the number of arguments and the roles (i.e. the Frame Elements) linked to them (e.g.: *a1* = Killer or Cause, *a2*= Victim, *a3*= Instrument), followed by semantic types specifying the selectional preferences (e.g., in the KILLING frame, the *a2*must be ANIMATE).

Classes are then specified with the constructions shared by its members. Each construction is associated with an example and the (possible) idiosyncratic verbs that block that specific pattern. As we said above, this allows us to represent verb class with a prototype structure, with more core verbs sharing all the constructions, and so more peripheral members for which some constructions are impossible. For example, the ASSASSINARE verbs subclass (subclass of UCCIDERE verbs, under verbs of KILLING) include several verbs that display the Direct Reflexive construction (1); however, other members of the class do not show this construction, since the feature [+ external_Agent] –i.e. the Agent cannot corefer with the Patient- is characteristic of most of these verbs (2).

(1) *Gianni si è impiccato* (John hanged himself)
(2) **Gianni si è assassinato* (*John assassinated himself)

The syntactic frames are represented with a formalism

adapted from ValPaL, which allowed us to link the syntactic layer of the constructions to the arguments described in the semantic frame and to specify unmarked word order. E.g.: *subj[a1]> V>obj[a2]* states that the first argument is associated with the subject and is followed by the verb. The direct object is associated with the second argument.

Classes are organized in a strictly monotonic way, as in VerbNet; e.g.: ASSASSINARE verbs adds further specifications to the UCCIDERE subclass, for example a necessarily animate and volitional agent. When possible, the internal structure of Italian classes was kept compatible with L/V: in the internal subdivision we tried to follow Levin and VerbNet subclasses system. We used such subdivision as a starting point, and departed from it only when the syntactic properties defining an English class do not apply to Italian (e.g.: subclasses based on the presence vs. absence of a syntactic construction that is not possible in Italian, like the Double Object construction), or when other solutions were regarded to better describe the verb behavior.

In other words, during our analysis we established several mismatches between the English and the Italian models, due to syntactic or semantic differences between the two languages, or discrepancies deriving from the construction methodologies. We discuss such cases in the next paragraph.

VERBS	CONSTRUCTIONS	CLASSES
Total of verbs examined and classified: **224**	Total of argument structure constructions examined: **28**	Classes: **7** Subclasses: **35**

Table 1: number of elements examined

1. VERBS of KILLING

KILL Verbs

VerbNet Classes: 42

Class Members:

Verbs	Senses
ammazzare "to kill"	Ammazzare qualcuno (To kill someone)
uccidere "to kill"	Uccidere qualcuno (To kill someone)

Semantic frame: A Killer or Cause causes the death of the Victim with an Instrument

Framenet Frame: Killing

Arguments	Roles	Semantictypes
a1	Killer, Cause	
a2	Victim	Animate
a3	Instrument	

Constructions:

Syntacticframes	Examples	Roots not allowing the construction
subj[a1] > V >obj[a2]	*Il rapitore uccide l'ostaggio* (The kidnapper kills the hostage) *L'incidente ammazzò sei persone* (the accident killed six people)	
subj[a1] > V	*Medusa uccide con lo sguardo* (Medusa kills with a glance)	
subj[a1] > V >obj[a2]> {con,a} [a3]	*Il rapitore uccise gli ostaggi con un coltello* (The kidnapper killed the hostages with a knife)	
subj[a1=a2:Animate] > si-dir_refl-V	*Cleopatra si uccise con il morso di un serpente* (Cleopatra killed herself with a snake'sbite)	
subj[a1,a2]>si-recip -V	*I due nemici si uccisero a vicenda (*The two enemies killed each other)	
ci-si-refl-impers-V	*Ci si uccide per disperazione*(One kills oneself out of despair)	
subj[a2]> si-refl_pass-V	*I nemici si uccidono con facilità* (Enemies are killed easily)	
subj[a2] > pass-V ({da}[a1])	*Il soldato fu ucciso dal nemico* (The soldier was killed by the enemy)	
si-impers-pass-V	*In guerra si viene uccisi (*At war, one gets killed)	

Idiosyncratic constructions:

Construction	Examples
subj[a1] > V >obj[a2] > {di}{botte, pugni, legnate,...}	*Gianni ha ammazzato Il ragazzo di botte* (John killed the boy by beating him → John beat the boy a lot)
subj[a1=a2] > si-refl-V > {di}{fatica, lavoro, studio,...}	*Mi ammazzo di lavoro* (I'm *killing* myself with work → I'm working to exhaustion)

3.3. Some notable differences with Levin classes

Four different groups of mismatches were found during our analysis of the data. The first group includes subclasses that had to be eliminated from the Italian classification due to syntactic or semantic differences between the two languages; these subclasses are SLIDE verbs (VerbNet class: 11.2), CONTRIBUTE verbs (VerbNet class: 13.2), and BERRY verbs (VerbNet class: 13.7).

Levin's SLIDE verbs (*bounce*, *float*, *move*, *roll*, *slide*), subclass of SENDING and CARRYING verbs, includes verbs which can be used both as intransitive and transitive, and the latter use can be roughly paraphrased as "cause *x* to *y*", where *y* is the verb in question. In Italian such a subclass is not possible, since the causative use of an intransitive verb of manner of motion is typically marked by a causative structure such as *"fare y"* ("to make y", literally), e.g.:

(3) *La palla rotola* (The ball rolls)
(4) *Il bambino fa rotolare la palla* (The boy rolls the ball)

Another excluded subclass within the CHANGE of POSSESSION class is the one of CONTRIBUTE verbs (Levin class: 13.2)(e.g.: *administer*, *contribute*, *disburse*, *distribute*, etc). Levin describes this subclass as lacking the Dative alternation, because these "Latinate" verbs do not admit the Double Object construction (Levin, 1993:139).Therefore, it is not surprising that a satisfactory Italian counterpart for CONTRIBUTE verbs cannot be found. In fact, the Double Object construction is not possible for Italian (i.e., the change of possession is always encoded with a direct object and a PP); moreover, it is clearly impossible to distinguish a subclass of "Latinate" verbs in a Romance language.

Finally, BERRY verbs (e.g.: *berry*, *birdnest*, *blackberry*, *clam*, *crab*, *fish*, etc.) were excluded for both syntactic and semantic reasons. The members of this class, in fact, are all zero-related to nominals. In Italian, however, not only the zero-relation is not possible, but there are also no semantically direct counterparts of these verbs. Italian generally realizes the zero-related English verb in two

parts (a verb indicating the action and the noun indicating the object that is collected) to maintain the original meaning:

(3) The children like to berry in the summer
I bambini amano raccogliere more [to pick berrei lit.] *durant e l'estate*

A second group of discrepancies the Italian classification and L/V concerns classes that are individuated solely based on the presence of particular alternations. Since we have assumed, consistently with Construction Grammar, that syntactic alternations are merely epiphenomena, similar classes have been excluded. An example is represented by the SPRAY/LOAD (Levin class: 9.7) and WIPE verbs (Levin class: 10.4). We chose not to create a separate group for those verbs, but rather to re-distribute them according to the different constructions they occur in. E.g., SPRAY/LOAD verbs are cross-classified between METTERE and RIEMPIRE verbs.

(5) *Gianni carica il camion di paglia* (John loads the truck with hay) → RIEMPIRE(holistic sense)
(6) *Gianni carica la paglia sul camion* (John loads hay on the truck) → METTERE (partitve sense)

We listed the "holistic" sense of *caricare* "load" in the RIEMPIRE class because, like the other verbs of this class, it expresses an event in which *a2* is totally filled with the content expressed by *a3*. Contrarily, the "partitive" use of *caricare* refers to a PUTTING semantic frame and has consistently been classified as a METTERE verb. The same cross-classification was adopted with WIPE verbs: their "holistic" sense was assigned to VUOTARE verbs, and the "partitive" one to RIMUOVERE verbs.

(7) *Il maestro cancella la lavagna* (The teacher wipes the blackboard)→VUOTARE (holistic sense)
(8) *Il maestro cancella le scritte dalla lavagna* (Theteacher wipes the writings from the blackboard) → RIMUOVERE (partitive sense)

The third type of mismatch consists in several groups of verbs whose "status" was changed (i.e., subclasses in the original model were transformed into classes in our system, or independent classes were reduced to subclasses). For instance, we created two new autonomous classes, verbs of FILLING and of CLEARING, which in the L/V classification are subclasses of - respectively- VERBS OF PUTTING and REMOVING (In Levin, FILL verbs is the 9.8 subclass and CLEAR the 10.3). We chose to re-classify those two groups of verbs following FrameNet, in which they both are treated as autonomous classes referring to independent frames (FILLING and EMPTYING).

Under the newly created FILLING class, we listed the

subclass of IMBURRARE verbs, the counterpart to Levin's BUTTER verbs (Levin's class: 9.9), which she instead lists under PUT verbs.

Finally, we also created new Italian subclasses that are not present in Levin or VerbNet classifications. Under the CHANGE OF POSSESSION class we added two further subclasses, VENDERE and COMPRARE verbs, which both imply parallel money transactions. On the one hand VENDERE verbs are listed under the DARE subclass and take the perspective of the Seller, which gives Goods to a Buyer in exchange for a sum of money. That is, they lexicalize the receiver of the action (the Buyer) as the complement of a PP headed by {a}.

(9) *Ho venduto il libro <u>a</u> Giulia* (I sold the book to Julia)

On the other hand, COMPRARE verbs reflect the Buyer's perspective, which in this case is the Agent that acquires some Goods from a Seller. These verbs lexicalize the Seller as the complement of a PP headed by {da}, and they also admit an additional argument *(a4)* –the Recipient or Beneficiary of the action of purchase- in a PP headed by {a}.

(10) *Giulia ha comprato un computer <u>da</u> un amico* (Julia bought a computer from a friend)

(11) *Ho comprato un computer <u>a</u> mio figlio* (I bought my son a computer)

Lastly, we created a new subclass, CONDIVIDERE verbs, to include verbs lexicalizing a possession that is shared rather than transferred between the participants. These verbs are constructed with the preposition {con} as head of a Beneficiary PP.

(12) *Gianni ha condiviso il pranzo <u>con</u> Giulia* (John shared his lunch with Julia)

4.   Concluding Remarks

In this paper, we have here presented the results of a preliminary investigation aiming at developing an Italian verb classification, inspired to VerbNet/Levin but differing from the original English one in a number of ways.

The present classification is not only adapted to Italian syntax and semantics, but it is also strongly rooted in a distributional and constructionist perspective. It integrates a strong distribution-based approach with ontology-based taxonomies by matching each class to a conceptual-semantic frame derived from FrameNet.

Verb classes have been defined as sets of semantically related verbs that share the same patterns and constructions. However, the differences between the two systems notwithstanding, the resulting classification remains compatible with VerbNet/Levin taxonomy.

It is clear that much additional work will be required, but we believe that the presented methodology will allow the development of a linguistically sound classification of Italian verbs, as a step to explore the complex interplay of argument structure constructions and verb semantics.

5.   Bibliographic References

Apresjan, D. (1967) *Experimental investigations of the semantics of the Russian verb*. Nauka: Moscow.

Baker, C. F., Fillmore, C. J., e Lowe, J. B. (1998). "The Berkeley FrameNet project". In *Proceedings of the 17 International Conference on Computational Linguistics*, pp.86-90.

Boas, H. C. (2002). "On constructional polysemy and verbal polysemy in Construction Grammar". In Samiian, V. (ed.), *Proceedings of the 2000 Western Conference on Linguistics*. Vol. 12, pp. 126-39.

Cennamo, M. &Fabrizio, C. (2013). "Italian Valency Patterns". In Hartmann, I., Haspelmath, M.,Taylor, B. (eds.), *Valency Patterns Leipzig*. Max Planck Institute for Evolutionary Anthropology:Leipzig. (available online at: http://valpal.info/languages/italian).

Dang, H.T., Kipper, K., Palmer, M., Rosenzweig, J. (1998). "Investigating regular sense extensions based on intersective Levin classes". In Proc. of COLING/ACL Montreal, Canada, pp. 293–99.

Dorr, B. (1997). "Large-scale dictionary construction for foreign language tutoring and interlingual machine translation". *Machine Translation*, 12(4), pp. 271–325.

Faber, P. B. &Mairal Usón, R. (1999). *Constructing a Lexicon of English Verbs*. Mouton deGruyter: Berlin/New York.

Ferrer, E. (2004). "Towards a Semantic Classification of Spanish Verbs Based on Subcategorisation Information". In ACL 2004: Student Research Workshop. Barcelona, Spain.

Fillmore, C. (1970). "The Grammar of Hitting and Breaking". In Jacobs, R. A. and Rosenbaum, P. S. (eds.), *Readings in English Transformational Grammar*. Ginn: Waltham, MA, pp. 120–33.

Goldberg, A. (1995). *Constructions*, University of Chicago Press: Chicago, IL.

Goldberg, A. (2002). "Surface generalizations: An alternative to alternations". In *Cognitive Linguistics* 13(4), pp. 327-56.

Goldberg, A. (2006). *Constructions at Work: the nature of generalization in language*. Oxford University Press: Oxford.

Hale, K.L. & Keyser, S.J. (1987). "A View from the Middle". *Lexicon Project Working Papers*10, Center for Cognitive Science, MIT, Cambridge, MA.

Hartmann, I., Haspelmath, M., Taylor, B. (eds.) (2013). *Valency Patterns Leipzig*. Max Planck Institute for Evolutionary Anthropology: Leipzig (Available online at: http://valpal.info).

Hoffmann, Th. Trousdale, G. (eds.)(2013). *The Oxford Handbookof Construction Grammar*. Oxford University Press,: Oxford.

Jackendoff, R.S. (1975). "Morphological and semantic regularities in the lexicon". In *Language* 51, pp. 639–71.

Jackendoff, R. S. (1990). *Semantic Structures*. MIT Press:

Cambridge, MA.

Jentson, I. (2014). "VerbNet Workbench". In *GWC* 2014.

Jezek, E. (2003). *Classi di verbi tra semantica e sintassi.* Edizioni ETS: Pisa.

Jing, H. & McKeown, K.R., (1998). "Combining multiple, large-scale resources in a reusable lexicon for natural language generation". In *Proceedings of the Joint 17th International Conference on Computational Linguistics 36th Annual Meeting of the Association for 232 Computational Linguistics (COLING-ACL'98).* Universite de Montreal, Quebec, Canada,

Kipper-Schuler, K. (2005). *VerbNet: A Broad -coverage, Comprehensive Verb Lexicon.* PhD dissertation, University of Pennsylvania, Philadelphia, PA.

Kipper, K., Dang, H.T. and Palmer, M. (2000). "Class-based construction of a verb lexicon". In *AAAI/IAAI,* pp. 691–96.

Kipper, K., Korhonen, A. Ryant, N. and Palmer, M., (2006a). "A Large-Scale Extension of VerbNet with Novel Verb Classes". In *Proceedings of the 12th euralex international congress (euralex 2006).* Turin, Italy, pp. 173-84.

Kipper,K., Korhonen, A., Ryant, N. and Palmer, M. (2006b). "Extending VerbNet with Novel Verb Classes". In *Proceedings of 5th international conference on Language Resources and Evaluation.* Genova, Italy, pp.1027-32.

Kipper, K. , Korhonen, A., Ryant, N. and Palmer, M. (2008.). "A large-scale classification Of English verbs". In *Language Resources and Evaluation Journal,* 42, pp. 21-40.

Kohl, K. T., Jones, D.A., Berwick, R.C. and Nomura, N. (1998). "Representing verb Alternations in WordNet". In *Christiane FellbaumWordNet: An Electronic LexicalDatabase. Language, Speech, and Communication.* MIT Press: Cambridge, MA, pp. 153-78.

Korhonen, A. & Briscoe, T. (2004). "Extended Lexical-Semantic Classification of English Verbs". In *Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics,* Boston, MA, pp. 38-45.

Lapata, M. & Brew, C. (1999). "Using Subcategorization to Resolve Verb Class Ambiguity". In *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora,* College Park, MD, pp. 153-78.

Lebani, G., Viola, V. and Lenci, A. (2014), "Bootstrapping an Italian VerbNet: Data-driven Analysis of Verb Alternations". In *Proceedings of the 9th Edition of the Language, Resources and Evaluation Conference (LREC 2014).* Reykjavik, 26-31 May 2014, pp. 1127-34.

Lenci, A., Bel, N., Busa, F., Calzolari,N., Gola, E., Monachini, M., Ogonowsky, A., Peters, I., Peters, W., Ruimy, N., Villegas, M. and Zampolli, A. (2000). "SIMPLE: A General Framework for the Development of Multilingual Lexicons". In *International Journalof Lexicography* 13:4, pp. 249-63.

Lenci, A., Lapesa, G., Bonansinga,G. (2012), "LexIt: A Computational Resource on Italian Argument Structure". In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12),* Istanbul, 23-25 May 2012, pp. 3712-18.

Lenci, A. (2014). "Carving Word classes from corpora". In Simone, R. and Masini, F. (eds.), *WordClasses: Nature, typology and representations.* Vii, pp. 17–36.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation,* University ofChicago Press: Chicago, IL.

Levin, B. (2013). "Verb Classes Within and Across Languages". In Comrie, B. and Malchukov, A.(eds.), *Valency Classes: A Comparative Handbook.* De Gruyter: Berlin, pp. 1-37.

Levin, B. & Rappaport Hovav, M. (2005). *Argument Realization.* Cambridge University Press: Cambridge, MA.

Merlo P., & Stevenson, S. (2001). "Automatic Verb Classification based on Statistical Distribution of Argument Structure". In *Computational Linguistics,* 27:3, pp.373-408.

Merlo, P., Stevenson, S., Tsang, V. and Allaria, G. (2002). "A Multilingual Paradigm for Automatic Verb Classification". In *Proceedings of 40th Annual Meeting of the Association for Computational Linguistics.* Association for Computational Linguistics: Philadelphia, PA, pp. 207–14.

Pinker, S. (1989). *Learnability and Cognition: The Acquisition of Argument Structure.* MIT press:Cambridge,MA.

Pianta, E., Bentivogli, L. and Girardi, C. (2002). "MultiWordNet: Developing an Aligned Multilingual Database". In *Proceedings of the 1st Global WordNetConference.* Mysore, India, pp. 293-302.

Pradet, Q., Danlos, L., De Chalendar, G. (2014). "Adapting VerbNet to French using existing resources". *LREC'C'14-Ninth International Conference on Language Resources and Evaluation,* May 2014, Reykjavik, Iceland, pp. 1122-26.

Pustejovsky, J. (1995). *The Generative Lexicon.* MIT Press: Cambridge, MA.

Rappaport Hovav, M. & Levin, B. (2003). "Root and template in the representation of Verb meaning". Presentation. Stanford, CA, May 2003 (available online at: web.stanford.edu/~bclevin/su03.pdf)

Rosch, E.H. (1973). "Natural categories". In *Cognitive Psychology* 4. pp. 328-350.

Roventini, A., Alonge, A., Calzolari, N., Magnini, B. &Bertagna, F. (2000). "ItalWordNet: A Large Semantic Database for Italian". In *Proceedings of LREC 2000,* Athens, pp. 783-90.

Ruppenhofer, J., et al. (2006). *FrameNet II: Extended Theory and Practice.* International Computer Science Institute: Berkeley, CA.

Sabatini,F.&Coletti,V. (2012). *Il Sabatini-Coletti: dizionario della lingua italiana.* Rizzoli-Larousse: Milano.

Saint-Dizier, P. (1996). "Constructing Verb Semantic

Classes for French: Methods and Evaluation". In *COLING 1996*, pp. 1127-30.

Scarton, C. &Aluisio, S. (2012). "Towards a cross-linguistic VerbNet-style lexicon for Brazilian Portuguese". In *Workshop on Creating Cross-language Resources for Disconnected Languages and Styles Workshop Programme.*, p.11.

SchulteImWalde, S. (2006). "Experiments on the automatic induction of German semantic verb classes". In *Computational Linguistics* 32.2, pp. 159–94.

Sun, L., Korhonen, A., Poibeau, T., Messiant , C. (2010). "Investigating the cross-linguistic potential of VerbNet: style classification". In: *COLING 2010,* pp.1056-64.

Suzuki, Y. and Fukumoto, F. (2009). "Classifying Japanese Polysemous Verbs based on fuzzy C-means Clustering". In *Proceedings of the 2009 Workshop on Graph based Methods for Natural Language Processing* (TextGraphs-4). Suntec ,Association for Computational Linguistics: Singapore, pp. 32–40.

6. Language Resources References

Lexit (2012), Lenci, A., Lapesa, G., Bonansinga,G. (2012), "LexIt: A Computational Resource on Italian Argument Structure". In *Proceedings of the Eight International Conference on Language Resources andEvaluation (LREC'12)*, Istanbul, 23-25 May 2012, pp. 3712-18. Distributed by ELRA.

ValPal (2013), Hartmann, I., Haspelmath, M., Taylor, B. (eds.) (2013). *Valency Patterns Leipzig*. Max Planck Institute for Evolutionary Anthropology: Leipzig (Available online at: http://valpal.info).

7. APPENDIX

Sample of the most representative verbs for each class

1) VERBS OF KILLING

UCCIDERE V. : uccidere, ammazzare ("to kill")
 ASSASSINARE V.: assassinare ("to assassinate"), impiccare ("to hang"), fucilare ("to execute by firingsquad")

2) VERBS OF PUTTING

METTERE V: mettere ("to put"), caricare ("to load")
ACCOSTARE V.: accostare ("to put close"), allineare ("to align)
INSERIRE V.: inserire ("to insert"), introdurre ("to introduce")
ALZARE/ABBASSARE V.: alzare ("to raise"), abbassare("to lower")
VERSARE V.: versare ("to spill, pour"), spargere ("to spread")
ATTORCIGLIARE V.: arrotolare ("to roll, coil"), attorcigliare ("to twirl")
IMBUSTARE V.: avvolgere (" to envelop"), imbustare ("to put in an envelope"), incartare ("to wrap")

3) VERBS OF FILLING

RIEMPIRE V.: caricare ("to load"), riempire ("to fill")

IMBURRARE V.: asfaltare ("to asphalt"), imburrare ("to butter")

4) VERBS OF REMOVING

RIMUOVERE V.: levare ("to take away"), pulire ("to clear"), rimuovere ("to remove")
BANDIRE V.: bandire ("to banish"), cacciare ("to dismiss"), escludere ("to exclude")
RUBARE V.: rapire ("to kidnap"), rubare ("to steal"), sequestrare ("to sequester")
CURARE V.: curare ("to cure"), liberare ("to free")
DISOSSARE: disossare ("to debone"), scuoiare ("to skin")

5) VERBS OF CLEARING

VUOTARE V.: scaricare ("to unload"), svuotare ("to empty")

6) VERBS OF SENDING AND CARRYING

MANDARE V.: consegnare ("to deliver"), inviare, mandare ("to send")
PORTARE/PRENDERE V.: portare ("to bring"), prendere ("to take")
TRASPORTARE V.: spingere ("to push"), tirare ("to pull"), trasportare ("to carry")
GUIDARE V.: guidare ("to drive"), pilotare ("to pilot")
ATTIRARE V.: attirare("to attract"), respingere("to reject")

7) VERBS OF CHANGE OF POSSESSION

DARE V.: dare ("to give"), donare ("to donate")
 VENDERE V.: affittare ("to rent"), vendere ("to sell")
PROMETTERE V.: accordare ("to grant"), garantire ("to guarantee"), promettere ("to promise")
PROVVEDERE V. :
 PREMIARE/PUNIRE V.:premiare ("to award"), punire ("to punish")
 EQUIPAGGIARE V.: armare ("to arm"), dotare ("provide"), equipaggiare ("to equip")
ACQUISIRE V.:
 PROCURARE V.: accumulare ("to assemple"), conquistare ("to win over, to conquest"), ricavare ("to obtain")
  COMPRARE V.: affittare ("to rent from"), comprare ("to buy")
 OTTENERE V.: ereditare ("to inherit"), ottendere ("to obtain")
SCAMBIARE V.: cambiare ("to change"), scambiare ("to exchange")
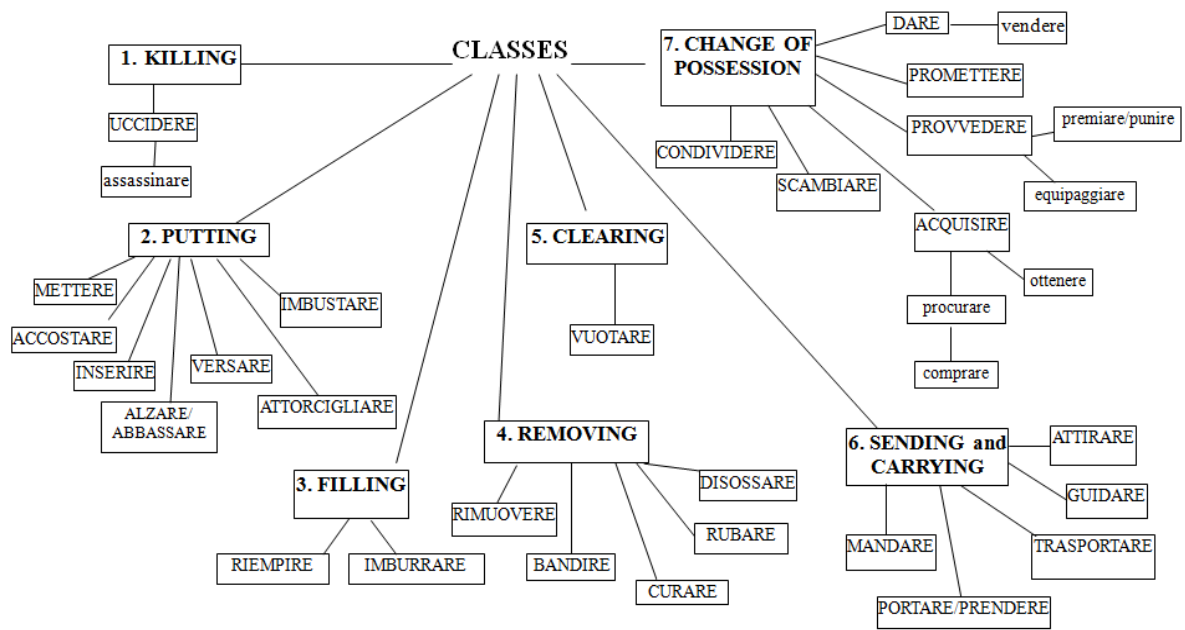CONDIVIDERE V.: condividere ("to share"), dividere ("to split")

Fig. 1: Synoptic schema of the analyzed classes