# Bootstrapping an Italian VerbNet: data-driven analysis of verb alternations

**Gianluca E. Lebani, Veronica Viola, Alessandro Lenci**

Computational Linguistics Laboratory, University of Pisa

via Santa Maria, 36, Pisa (Italy)

gianluca.lebani@for.unipi.it, alessandro.lenci@ling.unipi.it

### Abstract

The goal of this paper is to propose a classification of the syntactic alternations admitted by the most frequent Italian verbs. The data-driven two-steps procedure exploited and the structure of the identified classes of alternations are presented in depth and discussed. Even if this classification has been developed with a practical application in mind, namely the semi-automatic building of a VerbNet-like lexicon for Italian verbs, partly following the methodology proposed in the context of the VerbNet project, its availability may have a positive impact on several related research topics and Natural Language Processing tasks.

## 1. Introduction

In recent years, the study of the linguistic behavior of verbs at the syntax-semantics interface has gained a lot of interest in the Natural Language Processing (NLP) community. This topic, and in particular the development of automatic approaches to verb classification and characterization (for a review, see: (Korhonen, 2009; Schulte im Walde, 2009)), has greatly benefited from the availability of manually or semi-automatically built resources like WordNet (Fellbaum, 1998), FrameNet (Baker et al., 1998) and VerbNet (Kipper-Schuler, 2005). Such resources have also supported developments in those NLP tasks that benefit from verbal semantic knowledge, such as word sense disambiguation, machine translation, and information extraction (Korhonen, 2009). To a variable extent, however, a comparable range of lexical resources is lacking in languages other than English.

Notwithstanding the existence of Italian versions of both WordNet (Roventini et al., 2000; Pianta et al., 2002) and FrameNet (Tonelli et al., 2009; Johnson and Lenci, 2011), the lack of a verb-lexicon similar to VerbNet crucially undermines the future development of automatic classification methods for Italian verbs. In turn, this drawback can be traced back to the absence of a theoretical account of Italian verb alternations comparable to the one developed by Levin (1993) for English verbs.

The present work represents a first step towards the building of an Italian VerbNet. As the English lexicon, it will be based on the idea that *"verb behaviour can be used effectively to probe for linguistically relevant aspects of verb meaning"* (Levin, 1993, p. 1). Along these lines, we will root our classification on the notion of *diathesis alternation*. That is, we will classify verbs on the basis of the alternative syntactic ways in which they can express their arguments, as exemplified by the *dative* alternation in (1-2):

1. Gianluca gave <u>the book</u> <u>to Veronica</u>

2. Gianluca gave <u>Veronica</u> <u>the book</u>

With the notable exception of Ježek (2003), which however focused solely on transitive-intransitive alternations, there is no organic list of the syntactic alternations that Italian verbs can undergo. In order to overcome this limitation, we introduce an inventory of Italian argument alternations identified by means of a two-stage process. In a first step, for a sample of frequent Italian verbs, we manually extracted all the subcategorization frames (SCFs) reported in a monolingual Italian dictionary. Subsequently, we employed corpus-based methods to semi-automatically identify the most significant argument alternations shown by our sample.

This paper is structured as follows: the next section briefly reviews the relevant literature; in Section 3 we describe the techniques employed to identify and classify the alternations admitted by the most frequent Italian verbs; Section 4 is devoted to a description of our alternation classes, while a quick comparison against the classification proposed by Levin (1993) is outlined in Section 5.

## 2. Related Literature

The notion of diathesis (a.k.a. syntactic, argumental) alternation refers to the possibility for a verb to syntactically realize its arguments in more than one way. Even though the majority of research on this subject focused on English verbs, there is substantial evidence in favor of the idea that this phenomenon is consistent across languages (Guerssel et al., 1985; Ježek, 2003; Levin, 2013). Even if this notion is based on the idea that the same semantic argument can be realized in different syntactic positions, it does not entails that the meaning of the two alternating SCFs is identical, as shown by the *causative/inchoative* alternation in (3-4). Several authors, indeed, have stressed the idea that alternations can be seen as means to express some kind of semantic or pragmatic contrast (Beavers, 2006; Lenci, 2008).

3. Lucia broke <u>the window</u>

4. <u>The window</u> broke

Crucially, in the years several authors proposed to cluster together verbs in semantic classes on the basis of their regular alternating behavior under the assumption that this phenomenon can be better accounted for as semantically-driven, rather than as an idiosyncrasy (for a review, see (Levin and Rappaport Hovav, 2005)). The first large scale classification of this sort has been the one proposed by Levin (1993, henceforth LEVIN) that, by moving from the evidence reported in the linguistic literature, identified 79 argumental alternations involving nominal and prepositional phrases on the basis of which she classified 3024 English verbal lemmas (4186 verbal senses) into 49 broad semantic classes and 192 fine-grained classes.

The project VerbNet (Kipper-Schuler, 2005; Kipper et al., 2008, henceforth VN) extended this proposal by exploiting a semi-automatic procedure in order to increase the number and kinds of syntactic alternations, the lexical coverage and the number of identified semantic classes. In its most recent version (v. 3.2), VN covers more than 6,300 verbal senses, organized into 273 main classes and 214 subclasses[1] on the basis of their participation to a number of syntactic alternations that triples in size the original proposal by LEVIN, including alternations involving phrasal, adjectival, adverbial and predicative complements.

The kind of class-based information available in a a LEVIN/VN classification has proved to be useful both to further investigate verbal semantics, as well as for general NLP tasks, like language generation, machine learning and word sense disambiguation (Kipper et al., 2008). However, a resource of this kind is missing for languages other than English, mainly due to the lacking of inventories of syntactic alternations. A viable solution would be to derive the verb classes for the novel language from the English ones with a limited language-specific tuning (Sun et al., 2010). Such an approach has undeniable advantages, among which cost-effectiveness and a high inter-language consistency of the novel resource. However, it implicitly presupposes that the alternations on which the English classification is based are cross-linguistically constant, an assumption that holds only partially, as it is shown by the absence, in Italian, of a verb alternation similar to the English dative one, as shown by the contrast between (5) and (6):

5. *Gianluca ha dato il libro a Veronica*
   "Gianluca gave the book to Veronica"'

6. ***Gianluca ha dato Veronica il libro*
   "Gianluca gave Veronica the book"

Alternatively, a VN for a novel language could be based on a language-specific automatically built inventory of syntactic alternations. In the last 14 years few authors investigated the possibility to automatically detect which verbs may participate to which alternation. The exploratory investigations by McCarthy (2000) and Tsang and Stevenson (2004) were based on the notion of *slot overlap*, exploiting the intuition that syntactic alternations involving noun phrases and prepositional phrases could be detected by comparing the arguments in the slot positions of the alternating subcategorisation frames (SCFs). More recently, Parisien and Stevenson (2010; 2011) and Sun et al. (2013) proposed two Bayesian models able to identify verb alternations solely on the basis of the SCFs instantiated by a verb, abstracting away from the classes of arguments filling the slot positions. Finally, Baroni and Lenci (2010) showed how their vector space model, Distributional Memory, is capable of identifying transitivity alternations with a state-of-the-art accuracy. However, this methodology is still in an embryonic phase, and current systems still cannot be reliably exploited for the building of a large scale lexicon as the one we have in mind for Italian.

The only option left is to base a VN for a novel language on a manually identified language-specific set of recurrent syntactic alternations. The only classification of this kind available for the Italian language is the one developed by Ježek (2003), that identified 15 group of verbs on the basis of their possibility to occur in a combination of four (1 transitive, 3 intransitive) SCFs. By modeling solely on transitive-intransitive alternations, however, such a proposal appears scarcely usable for our practical goals. To overcome this limitation, then, we built a novel classification of the syntactic alternations admitted by the most frequent Italian verbs by exploiting the data-driven procedure described in the next section.

## 3. Carving Italian Diathesis Alternations

The development of a classification of argument alternations for Italian verbs has been carried out in a two-stage process. In the first stage, the SCFs for a sample of the most frequent Italian verbs were manually extracted from an Italian monolingual dictionary. In the second phase, we semi-automatically identified the most significant alternations in our annotated sample.

The manual extraction of SCFs was performed on the only Italian dictionary that marks the valency of each verbal sense, namely the *Il Sabatini Coletti* (Sabatini and Coletti, 2012, henceforth S&C). For instance, of the 9 reported senses of *imporre* ("to impose") associated with 5 distinct SCFs, 4 can occur with a transitive frame (`[subj-v-arg-prep.arg]`, `[subj-v-arg]`[2]), while the remaining 5 occur with pronominal frames (`[subj-v]`, `[subj-v-arg]`, `[subj-v-prep.arg]`). The choice of exploiting a monolingual dictionary over possible alternative lexicographic resources, such as the PAROLE lexicon (Ruimy et al., 1998), is due to the assumption than the proper locus of syntactic alternation is the verb sense, rather than the lemma (Roland and Jurafsky, 2002).

However, the formalism used by S&C neglects a crucial piece of information, namely the specification of the preposition introducing the prepositional phrases and the phrasal arguments. We overcome this shortcoming by resorting to LexIt (Lenci et al., 2012), an automatically built corpus-based lexical resources on Italian argument structure. In

---

[1]data from the Unified Verb Index: `http://verbs.colorado.edu/verb-index/index.php`

[2]SCFs represented in the S&C notation. The translated atomic slots labels are to be interpreted as follows: `subj` for "subject", `v` for "target verb", `arg` for "argument" and `prep.arg` for "argument introduced by a preposition".'

| SCF | example |
|---|---|
| subj#obj | [$_{subj}$ La situazione] impone$_v$ [$_{obj}$ una scelta] |
| subj#obj#comp-a | [$_{subj}$ La situazione] impone$_v$ [$_{obj}$ sacrifici] [$_{comp-a}$ al paese] |
| subj#inf-di | [$_{subj}$ La situazione] impone$_v$ [$_{inf-di}$ di fare dei sacrifici] |
| subj#si#0 | [$_{subj}$ Il modello] [$_{si}$ si] imporrà$_v$ |
| subj#si#comp-su | [$_{subj}$ L'azienda] [$_{si}$ si] impone$_v$ [$_{comp-su}$ sul mercato] |

Table 1: Subset of the structural configurations associated with the lemma "*imporre*". The constituents that occupy a given slot of the Subcategorization Frame (SCF) are enclosed by [$_{slot}$ square brackets]; target verbs are marked by the $_v$ tag.

LexIt, the range of SCFs in which a lemma may occur, are automatically identified from the reference corpora and their frequencies are collected. Being automatically built and being grounded on the notion of lemma, LexIt cannot serve as our reference resource, but, crucially, it provides a reliable description of the prepositions that most frequently introduce the prepositional argument(s) in a given lemma SCFs. We used the LexIt data to enrich the relevant SCFs available in S&C with the information about the prepositions introducing the prepositional phrases and the propositions and conjunctions introducing the phrasal arguments.This way, the S&C-derived SCFs for our example verb *imporre* can be further specified, and the set of SCFs associated to the different sense of this verb may be enlarged to include the LexIt-enriched[3] structural configurations similar to those reported in Table 1.

In order to have a sample of manageable size, we focused our analysis on a subset of the 1746 verbal lemmas that are marked as highly frequent in S&C, matching them with the corresponding verbs in the *La Repubblica* corpus (Baroni et al., 2004) and narrowing our sample down to the 1000 top frequent verbs. We then manually identified for each verb the corresponding frames registered in S&C, filtering out the technical, archaic and literary uses, and integrating the information about the prepositions available in LexIt.

We obtained a total of 4450 verb sense-SCF pairings, for which we also recorded the thematic roles by referring to the VerbNet role-set (VerbNet Annotation Guidelines[4], pp. 19-22), and the selectional preferences of the SCF syntactic slots using the LexIt inventory of categories, in turn taken from the 24 WordNet super-senses (Fellbaum, 1998).

In a second step, we moved on to identify: a) the argument alternations in our sample; b) the sets of verbs undergoing such alternations. We assumed an alternation to be a pair of SCFs that represent alternative realizations of a verb's arguments. Therefore, we looked for potentially alternating frames by identifying those SCFs pairs that tend to occur with the same verbs. We represented each SCF as a vector of binary values whose dimensions indicate whether a verb allows a frame or not, and discarded all SCF pairs whose correlation failed to reach the 0.2 threshold. Such a low threshold has been chosen to maximize recall over precision, because the next step employs a manual identification, which maximizes precision. The outcome of this procedure was a set of 174 potential argument alternations, for each of which we recorded the list of verbs allowing the candidate alternation.

Such a list of potential alternation cannot be taken as fully reliable, as part of the regularities found by means of simple correlation can be the by-product of various factors, the most influential being verb polysemy. For instance, the verb *sentire* may be realized within the SCFs: subj#fin-che and subj#inf-di. However, these two argument structures are paired with two different senses of the verb in question, the first of which corresponds to the English verb "to hear", while the second corresponds to the English verb "to feel", as shown by the contrast between (7) and (8). A minor consequence of this phenomenon is that an argument alternation valid for many verbs can also feature verbs for which it is not so.

7. *Gianluca ha sentito che Alessio vive a Utrecht*
"Gianluca heard that Alessio lives in Utrecht"

8. *Alessio ha sentito di dover lasciare la nostra band*
"Alessio felt the need to leave our band"

Regular patterns of verbal polysemy, moreover, may interest the whole set of verbs that our automatic procedure associated to a given potential argument alternation. In these cases, it is the frame alternation itself that needs to be marked as incorrect and, eventually, removed from our dataset. For instance, this is what happened with the alternation subj#comp-da/subj#inf-per registered for the verbs *ripartire* and *venire*: for both of these verbs, in fact, the two different frames simply refer to two different semantic patterns, the first one meaning "to leave a place again" and "to come from a place" respectively, and the second meaning "to head towards a destination" and "to go somewhere to achieve something" respectively.

Given the impossibility to handle these issues automatically, we decided to manually inspect our plausible frame pairs, verifying whether each verb associated with a given alternation was a proper case of argument alternation or not, thus filtering out those data that turned out to be inconsistent. In this phase, alternations involving only one verb were ruled out as well.

---

[3]Hereafter, SCFs will be labeled according to the LexIt notation (Lenci et al., 2012), i.e. by concatenating the labels referring to its atomic syntactic slots (e.g. subj for "subject", comp-a for "complement introduced by the preposition *a*") with the symbol "#". For example, the simple transitive SCF composed by a subject and an object is marked as subj#obj.

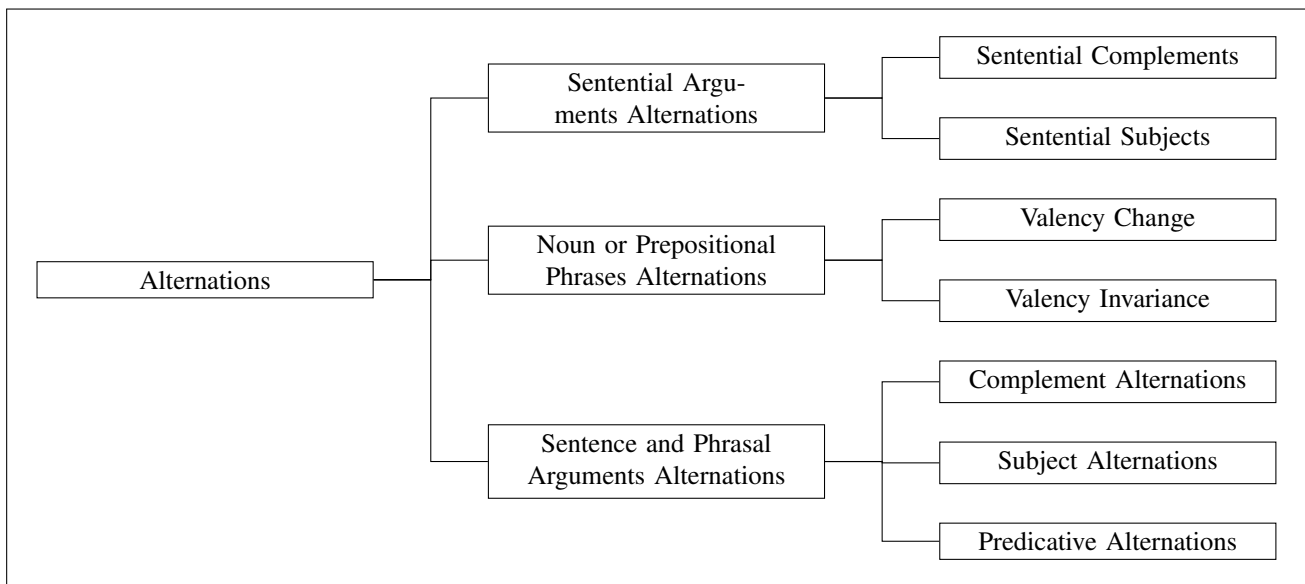[4]available at the URL: http://verbs.colorado.edu/verb-index/VerbNet_Guidelines.pdf

Figure 1: Taxonomy of Italian Syntactic Alternations

## 4. Towards and Italian VerbNet

By using this procedure, we were able not only to identify and classify 37 argument alternations valid for Italian verbs, but also to associate each of them to the subset of our target verbs undergoing this alternation, as illustrated in the Appendix. We decided to organise our alternations in the taxonomy shown in Figure 1, on the basis of the syntactic characteristics of the arguments involved.

### 4.1. Sentential Arguments Alternations

Alternations in this class present two possible syntactic realizations of the same argument, which in both cases is expressed as a dependent clause, as exemplified by the opposition between (9) and (10):

9. *Ti auguro che tutto vada bene*
   "I hope that everything goes well"

10. *Ti auguro di fare un buon viaggio*
    "I hope you have a good journey"

We identified 10 sentential arguments alternations, which, depending on their syntactic position, were in turn divided into two subclasses: 7 alternations involving sentences in a complement position, as exemplified above in (9-10), and 3 alternations involving sentences in a subject position, as illustrated in (11-12):

11. *Pare di non ottenere nessun risultato*
    "Seems$_{impers}$ to not achieve any results'"

12. *Pare che nessun risultato sia ottenibile*
    "Seems$_{impers}$ that no result is achievable"

### 4.2. Alternations involving NPs or PPs

Alternations of this kind involve noun phrases (NPs) or prepositional phrases (PPs), and they are, by far, the most studied cross-linguistically. We identified 9 alternations of this type and we divided them in the following two groups: alternations that cause a different syntactic realization of the same set of arguments, as shown in (13-14):

13. *Cataldo ha rimproverato Gianfranco per la sconfitta*
    "Cataldo blamed Gianfranco for the defeat"

14. *Cataldo ha rimproverato la sconfitta a Gianfranco*
    "Cataldo blamed the defeat on Gianfranco"

and alternations that bring about a change in the valence of the verb, as in (15-16):

15. *Lucia ha rotto la finestra*
    "Lucia broke the window"

16. *La finestra si è rotta*
    "The window broke"

### 4.3. Alternations involving a Phrasal and a Sentential Argument

This last class encompasses all those alternations in which an argument can be expressed either as a sentential argument (as in (17)), or as a phrase (as in (18)).

17. *Antonio ha garantito a Cataldo di occuparsi di Luigi*
    "Antonio promised Cataldo to take care of Luigi"

18. *Luigi ha garantito a Cataldo una partita maschia*
    "Luigi promised Cataldo a tough match"

This group was the most numerous one, consisting of a total of 18 valid alternations. We classified them depending on the syntactic position affected by the alternation, obtaining 11 alternations taking place in the complement position (as in (17-18) above), and 3 alternations taking place in the subject one, as illustrated in (19-20):

19. *A nessuno importano queste sciocchezze!*
    "To nobody matter these silly things!"

20. *Ai ragazzi importa che tu sia qui con noi*
    "To the guys matters$_{impers}$ that you are here with us"

Moreover, we also decided to keep separate those alternations that involve predicative complements (`cpred`), since, given their high level of similarity and the fact that they are all allowed by the verbs *credere*, *considerare* and *giudicare* ("to believe", "to consider", "to judge"), they can be considered a unique class on their own consisting of 4 alternations, exemplified by the sentences in (21-22):

21. *Lo considero il miglior giocatore del mondo*
    "I consider <u>him</u> the best player in the world"

22. *Considero sconveniente che tu rimanga qui*
    "I consider inappropriate <u>that you remain here</u>"

## 5. Comparison against LEVIN

The classification reported in these pages and the one proposed by LEVIN share the same overall goal, that is the identification of a set of syntactic regularities that can be used as a proxy to some aspects of the verbal semantics. Nevertheless, these two proposals differ both extensionally (i.e. in the number of verbs and alternations) and intensionally (i.e. in the nature of the identified oppositions and in the rationale of the overall inventory).

A first difference pertains to lexical coverage: the number of verbs considered in LEVIN roughly triples that of our verbs. Crucially, however, while LEVIN explicitly focuses on verbs taking NPs and PPs as complements, we chose our sample verbs solely on the basis of their corpus frequency. Such a departure from our inspirational model has been driven by the need to model also alternations involving non-phrasal complements, in contrast with what happens in LEVIN, and more similarly to what is modeled in the inventory of alternations exploited in VerbNet.

The large difference in the number of NPs/PPs alternations between the two classifications (9 vs. 79), moreover, is a direct consequence of the diverse kinds of evidence considered: while LEVIN moved from a comprehensive re-elaboration of the linguistic literature, we committed ourselves with a data-driven procedure. This allowed us to avoid two of the main issues of the LEVIN work, namely the partial semantic nature of its oppositions and the unclear role of polysemy in the whole framework.

Dang et al. (1998) and Baker and Ruppenhofer (2002) already noticed how the LEVIN semantic classes are partially semantically motivated. We feel that similar considerations can be applied to its inventory of syntactic alternations. As an example, three of the different kinds of *Unexpressed Object Alternations* in LEVIN (pp. 33-36) are characterized on the basis of the interpretation of the omitted object: as a bodypart ("I flossed my teeth"–"I flossed"), as a reflexive ("Jill dressed herself hurriedly"–"Jill dressed hurriedly") or unspecified ("Mike ate the cake"–"Mike ate"). Moreover, this alternations are kept in a separate class than other - superficially related - alternations such as the *Cognate Object Constructions*, characterized by the optionality of a zero-related object ("Sarah sang a song"–"Sarah sang").

Such an account faces two kinds of related problems. First of all, it implicitly assumes that the selectional preferences of a verb are part of the semantically relevant syntactic behavior of a verb, without explicitly establishing for them a clear role. As a consequence, it is not always clear if alternations distinctions such as the ones reported above are drawn on semantic or syntactic principles. In order to avoid such an issue, we opted for a more restrictive approach, according to which SCFs are opposed solely on the basis of the syntactic nature of their slots, independently of the semantic class of the arguments.

Another issue concerns the role played by lexical ambiguity, a problem that should be placed in the general discussion of the semantic effects of the syntactic alternations (Beavers, 2006; Lenci, 2008). According to S&C, indeed, the sentences in (23-24) instantiate two difference senses of the lemma *mangiare* ("to eat"), the former meaning "to ingest a solid substance", the latter meaning "to take a meal". In LEVIN, on the contrary, such cases of polysemy are somehow neglected, and the general strategy is to collapse, to a certain extent, the alternative senses of the ambiguous lemmas.

23. *Roberta mangia il cibo di Gianluca*
    "Roberta eats <u>Gianluca's food</u>"

24. *Roberta mangia*
    "Roberta eats"

In our work, we opted for a "sense preserving" strategy. We relied on the word sense distinctions in S&C and included in our inventory only those alternations involving the same sense of a given lemma. A major consequence of this choice has been the absence, in our inventory, of a set of alternations, like the object-drop in the example above. In the future, we plan to relax this constraint and to extend our analysis to "sense-shift" alternations too.

Overall, all the main differences outlined in this sections, as well as others not discussed here for space reasons (e.g. the absence of a passive alternation, the rejection of alternations admitted by just one verb) have to be ascribed primarily to the different nature of the two works. While LEVIN's goal was to conduct a preliminary large scale investigation of the behavior of English verbs at the syntax-semantics interface in order to provide some support to the hypothesis that some aspects of the semantics of the English verbs can be inferred by their linguistic behavior. On the other side, we committed ourselves to the creation of a machine-usable resource, thus addressing our efforts towards the creation of a resource less fine-grained, probably with less predictive power, but more consistent, more data-driven and based on the fewer possible theoretical assumptions.

## 6. Conclusions and Future Work

In this paper we introduced a classification of the syntactic alternations admitted by the most frequent Italian verbs. To our knowledge this is the first work that tries to organize and characterize Italian SCFs alternations in an organic and comprehensive way.

Even if this classification has been developed with a practical application in mind, namely the building of an Italian VerbNet-like lexicon, the availability of an inventory of Italian SCFs and of their alternations can constitute a valuable gold standard for tasks such as the automatic induction of SCFs and automatic identification of SFCs alternations (on these topics, see: Korhonen, 2009; Schulte im

Walde, 2009), and can support NLP tasks such as automatic verb classification, selectional preference acquisition, parsing, word sense disambiguation and machine translation. Our plan for future research is to use the manual classification presented here to bootstrap a process that exploits the information encoded in LexIt in order both to populate our alternation classes with novel verbs and to enrich the information associated to them with selectional preferences and semantic roles in a FrameNet-like fashion.

## 7. Acknowledgements

## 8. References

C. F. Baker and J. Ruppenhofer. 2002. FrameNets Frames vs. Levins Verb Classes. In *Proceedings of BLS-28*.

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of COLING-ACL '98*, pages 86–90.

M. Baroni and A. Lenci. 2010. Distributional Memory: A General Framework for Corpus-based Semantics. *Computational Linguistics*, 36(4):673–721.

M. Baroni, S. Bernardini, F. Comastri, L. Piccioni, A. Volpi, G. Aston, and M. Mazzoleni. 2004. Introducing the La Repubblica Corpus: A Large, Annotated, TEI(XML)-Compliant Corpus of Newspaper Italian. In *Proceedings of LREC 2000*, pages 1771–1774.

J. T. Beavers. 2006. *Argument/Oblique Alternations and the Structure of Lexical Meaning*. Ph.D. thesis, Stanford University.

H. T. Dang, K. Kipper, M. Palmer, and J. Rosenzweig. 1998. Investigating regular sense extensions based on intersective Levin classes. In *Proceedings of COLING-ACL '98*, pages 293–299.

C. Fellbaum. 1998. *WordNet: an Electronic Lexical Database*. The MIT Press.

M. Guerssel, K. Hale, M. Laughren, B. Levin, and J. White Eagle. 1985. A Cross-Linguistic Study of Transitivity Alternations. In *Papers from the Parasession on Causatives and Agentivity*, pages 48–63.

E. Ježek. 2003. *Classi di verbi tra semantica e sintassi*. ETS.

M. Johnson and A. Lenci. 2011. Verbs of visual perception in Italian FrameNet. *Constructions and Frames*, 3(1):9–45.

K. Kipper, A. Korhonen, N. Ryant, and M. Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation*, 42(1):21–40.

K. Kipper-Schuler. 2005. *Verbnet: A Broad-Coverage, Comprehensive Verb Lexicon*. Ph.D. thesis, University of Pennsylvania.

A. Korhonen. 2009. Automatic Lexical Classification - Balancing between Machine Learning and Linguistics. In *Proceedings of PACLIC 23*, pages 19–28.

A. Lenci, G. Lapesa, and G. Bonansinga. 2012. LexIt: A Computational Resource on Italian Argument Structure. In *Proceedings of LREC 2012*, pages 3712–3718.

A Lenci. 2008. Argument alternations in Italian verbs: a computational study. In *Atti del XLII Congresso della Società di Linguistica Italiana*.

B. Levin and M. Rappaport Hovav. 2005. *Argument Realization*. Cambridge University Press.

B. Levin. 1993. *English Verb Classes and Alternations*. The University of Chicago Press.

B. Levin. 2013. Verb classes within and across languages. In B. Malchukov and A. Comrie, editors, *Valency Classes: A Comparative Handbook*, pages 1–37. De Gruyter.

D. McCarthy. 2000. Using Semantic Preferences to Identify Verbal Participation in Role Switching Alternations. In *Proceedings of NAACL-2000*, pages 256–263.

C. Parisien and S. Stevenson. 2010. Learning verb alternations in a usage-based Bayesian model. In *Proceedings of CogSci 2010*, pages 2674–2679.

C. Parisien and S. Stevenson. 2011. Generalizing between form and meaning using learned verb classes. In *Proceedings of CogSci 2011*, pages 2024–2029.

E. Pianta, L. Bentivogli, and C. Girardi. 2002. MultiWordNet. Developing an aligned multilingual database. In *Proceedings of GWC 2002*, pages 293–302.

D. Roland and D. Jurafsky. 2002. Verb Sense and Verb Subcategorization Probabilities. In P. Merlo and S. Stevenson, editors, *The Lexical Basis of Sentence Processing: Formal, Computational, and Experimental Issues*, pages 325–346. John Benjamins.

A. Roventini, A. Alonge, N. Calzolari, B. Magnini, and F. Bertagna. 2000. ItalWordNet: a Large Semantic Database for Italian. In *Proceedings of LREC 2000*, pages 783–790.

N. Ruimy, O. Corazzari, E. Gola, A. Spanu, N. Calzolari, and A Zampolli. 1998. LE-PAROLE project: The Italian syntactic lexicon. In *Proceedings of EURALEX 1998*, pages 259–269.

F. Sabatini and V. Coletti. 2012. *Il Sabatini-Coletti: dizionario della lingua italiana*. Rizzoli-Larousse.

S. Schulte im Walde. 2009. The induction of verb frames and verb classes from corpora. In A. Lüdeling and M. Kytö, editors, *Corpus Linguistics. An International Handbook*, pages 952–972. Mouton de Gruyter.

L. Sun, A. Korhonen, T. Poibeau, and C. Messiant. 2010. Investigating the cross-linguistic potential of VerbNet-style classification. In *Proceedings of COLING 2010*, pages 1056–1064.

L. Sun, D. McCarthy, and A. Korhonen. 2013. Diathesis alternation approximation for verb clustering. In *Proceedings of ACL 2013*, pages 736–741.

S. Tonelli, D. Pighin, C. Giuliano, and E. Pianta. 2009. Semi-automatic Development of FrameNet for Italian. In *Proceedings of the FrameNet Workshop and Masterclass*.

V. Tsang and S. Stevenson. 2004. Using selectional profile distance to detect verb alternations. *Proceedings of CLS'04*, pages 30–37.

# Appendix. Chart of the Syntactic Alternations in Italian

## Sentential Arguments Alternations

| | Alternating SCFs | verbs |
|---|---|---|
| complements | `subj#fin-che#comp-a`<br>`subj#inf-di#comp-a` | augurare, garantire, ricordare, assicurare, promettere, gridare, confidare, annunciare, raccontare, chiedere, comandare, raccomandare, dire, dichiarare, permettere, scrivere, giurare, confessare, comunicare, riferire |
| | `subj#fin-che`<br>`subj#inf-di` | ignorare, disporre, scoprire, tacere, imporre, badare, escludere, vedere, riscoprire, protestare, stabilire, intuire, gridare, temere, confidare, sopportare, constatare, supporre, tollerare, annunciare, pensare, ammettere, deliberare, immaginare, dubitare, fingere, sapere, convenire, ipotizzare, nascondere, smentire, esigere, ritenere, sperare, pretendere, dire, aspettare, riconoscere, dichiarare, credere, dimenticare, sostenere, aggiungere, prescrivere, ottenere, decidere, negare, accettare, dimostrare, sognare, testimoniare, rivelare |
| | `subj#fin-che`<br>`subj#fin-come` | provare, sottolineare, notare, concepire, ricordare, vedere, giudicare, imparare, sapere, intendere, stabilire, prevedere |
| | `subj#fin-che#comp-a`<br>`subj#fin-come#comp-a` | spiegare, mostrare |
| | `subj#si#fin-che`<br>`subj#si#inf-di` | convincersi, augurarsi, assicurarsi, illudersi, ricordarsi, sorprendersi, immaginarsi, dispiacersi, attendersi, accorgersi, sognarsi |
| | `subj#fin-come#comp-a`<br>`subj#si#fin-come` | spiegare, domandare |
| | `subj#inf-di#comp-a`<br>`subj#si#inf-di` | proporre, permettere, imporre, augurare, assicurare, offrire, ricordare, impedire, rimproverare, risparmiare |
| subjects | `fin-che`$_{subj}$`#0`<br>`inf-di`$_{subj}$`#0` | sembrare, papere, accadere |
| | `fin-che`$_{subj}$`#0`<br>`inf-0`$_{subj}$`#0` | occorrere, bisognare |
| | `fin-che`$_{subj}$`#comp-a`<br>`inf-0`$_{subj}$`#comp-a` | dispiacere, convenire, risultare |

## Noun or Prepositional Phrases Alternations

| | Alternating SCFs | verbs |
|---|---|---|
| valency change | `subj#obj#comp-da`<br>`subj#si#comp-da` | rilanciare, buttare, escludere, difendere, gettare, dividere, staccare, spostare, lanciare, sciogliere, allontanare, sollevare, distrarre, trarre, ritirare, levare, sfilare, separare, liberare, ritrarre, riparare, salvare |
| | `subj#obj#comp-di`<br>`subj#si#comp-di` | convincere, privare, fornire, ricoprire, riempire, caricare, svuotare, dotare, investire, colmare, incaricare, circondare, coprire |
| | `subj#obj#comp-a`<br>`subj#si#comp-a` | disporre, affidare, donare, mostrare, alternare, opporre, adattare, abituare, mescolare, affiancare, costringere, agganciare, esporre, consacrare, adeguare, accordare, unire, iscrivere, preparare, sottrarre, paragonare, attaccare, raccomandare, presentare, votare, appassionare, associare, allineare, vendere, avvicinare, indirizzare, dichiarare, dare, predisporre, sommare, sottoporre, accostare, concedere, rivolgere, consegnare, interessare, dedicare |
| | `subj#obj#comp-in`<br>`subj#si#comp-in` | proiettare, integrare, immergere, situare, rinchiudere, specializzare, inserire, inquadrare, trasformare |
| | `subj#obj#comp-su`<br>`subj#si#comp-su` | fondare, proiettare, basare, concentrare |
| | `subj#obj`<br>`subj#si#0` | DIRECT REFLEXIVES: isolare, lavare, rinnovare, liberare, negare, allontanare, scoprire, schierare, esprimere, giustificare, assicurare, uccidere, ammazzare, allenare, licenziare, accettare, consolare, contraddire, tormentare, interrogare, valorizzare, ferire, escludere, spiegare, umiliare, nascondere<br>RECIPROCAL REFLEXIVES: conoscere, sospettare, rispettare, combattere, fronteggiare, controllare, attirare, ritrovare, rivedere, stimare, sfidare, abbracciare, rincorrere, inseguire, sfiorare, scegliere, dividere, stringere, picchiare, disturbare, odiare, respingere, frequentare, temere, incontrare, sposare, vedere, trovare, baciare, soccorrere<br>CAUSATIVE-INCHOATIVE ALTERNATION: chiudere, spaventare, confondere, emozionare, abbassare, restringere, staccare, ridurre, intrecciare, sbloccare, scatenare, rovinare, turbare, piegare, conservare, spezzare, spaccare, rompere |

Noun or Prepositional Phrases Alternations *(continued from previous page)*

| | Alternating SCFs | verbs |
|---|---|---|
| valency change | `subj#obj#comp-con`<br>`subj#si#comp-con` | conciliare, combinare, alternare, scambiare, confrontare, mescolare |
| valency change | `subj#obj#comp-*`<br>`subj#si#comp-*` | DIRECT REFLEXIVES: girare, disporre, accordare, orientare, buttare, muovere, trascinare, gettare, voltare, rilanciare, avviare, spingere, stabilire, spostare, lanciare, dirigere, piazzare, sistemare, ambientare, stendere, mettere, rivolgere, abbandonare, porre, informare, calare<br>CAUSATIVE-INCHOATIVE ALTERNATION: infilare, collocare, versare, aggiungere, aprire, insinuare, avvolgere, rovesciare, spargere, imprimere, stampare<br>INTENSIVE REFLEXIVES: fissare |
| invariance | `subj#0`<br>`subj#comp-*` | tornare, girare, correre, giocare, finire, accorrere, uscire, battere, emigrare, terminare, perdere, bastare, combattere, risalire, arrivare, durare, oscillare, riuscire, volare, nascere, giungere, ricadere, cadere, scivolare, avanzare, votare, venire, crescere, piombare, sorgere, comparire, reagire, saltare, slittare, lavorare, picchiare, salire, vagare, precipitare, rientrare |

## Sentential and Phrasal Arguments Alternations

| | Alternating SCFs | verbs |
|---|---|---|
| complements | `subj#obj#comp-a`<br>`subj#inf-di#comp-a` | rimproverare, imporre, risparmiare, garantire, ricordare, giurare, promettere, gridare, proporre, ordinare, confidare, proibire, annunciare, raccontare, chiedere, augurare, comandare, offrire, raccomandare, sussurrare, dire, predicare, vietare, dichiarare, permettere, suggerire, denunciare, scrivere, assicurare, confessare, concedere, comunicare, consigliare, riferire, impedire |
| complements | `subj#obj#comp-a`<br>`subj#fin-che#comp-a` | augurare, mostrare, garantire, ricordare, assicurare, promettere, gridare, confidare, ripetere, annunciare, raccontare, chiedere, comandare, raccomandare, dire, dichiarare, segnalare, spiegare, permettere, denunciare, proporre, scrivere, giurare, confessare, concedere, insegnare, comunicare, riferire |
| complements | `subj#obj#comp-a`<br>`subj#obj#inf-a` | disporre, autorizzare, invitare, stimolare, abituare, indurre, motivare, incoraggiare, costringere, condannare, esercitare, educare, ammettere, trattenere, spedire, sollecitare, delegare, forzare, ridurre, obbligare, mettere, destinare |
| complements | `subj#obj#fin-che`<br>`subj#obj#comp-di` | avvisare, convincere |
| complements | `subj#comp-a`<br>`subj#inf-a` | contribuire, arrivare, rinunciare, scappare, pervenire, provvedere, giocare, venire, badare, pensare, tenere, ritornare, mirare, aspirare, concorrere, tendere |
| complements | `subj#obj#comp-a`<br>`subj#si#inf-a` | disporre, convincere, costringere, obbligare, esercitare, indurre, abituare, ridurre, trattenere |
| complements | `subj#obj#comp-per`<br>`subj#si#inf-per` | preparare, sacrificare |
| complements | `subj#si#comp-per`<br>`subj#si#inf-per` | organizzare, preparare, sacrificare |
| complements | `subj#si#comp-di`<br>`subj#si#inf-di` | curare, convincere, accusare, assicurare, pentirsi, sorprendere, occupare, stupire, ricordare, accontentare, dimenticare, incaricare, accorgersi, vantare |
| complements | `subj#si#fin-che`<br>`subj#si#comp-di` | convincere, stupire, assicurare, ricordare, sorprendere, accorgersi |
| complements | `subj#si#comp-a`<br>`subj#si#inf-a` | preparare, rilanciare, costringere, determinare, obbligare, prestare, rassegnare, abbassare, rimettere, disporre, adattare, abituare, ridurre, indurre |
| subjects | `subj#comp-da`<br>`fin-che`$_{subj}$`#comp-da` | conseguire, risultare |
| subjects | `subj#comp-a`<br>`fin-che`$_{subj}$`#comp-a` | risultare, capitare, importare, dispiacere, sfuggire, convenire |
| subjects | `subj#comp-a`<br>`inf-0`$_{subj}$`#comp-a` | spettare, risultare, capitare, dispiacere, premere, convenire |
| predicative | `subj#obj#cpred`<br>`subj#inf-0#cpred` | credere, considerare, giudicare |
| predicative | `subj#obj#cpred`<br>`subj#fin-che#cpred` | credere, considerare, giudicare |
| predicative | `subj#inf-0#cpred`<br>`subj#si#cpred` | credere, considerare, giudicare |
| predicative | `subj#fin-che#cpred`<br>`subj#si#cpred` | credere, considerare, giudicare |