The Lexicon-Grammar of Italian Idioms

Simonetta Vietri
Department of Political,
Social and Communication
Sciences
University of Salerno, Italy
vietri@unisa.it

Abstract

This paper presents the Lexicon-Grammar classification of Italian idioms that has been constructed on formal principles and, as such, can be exploited in information extraction. Among MWEs, idioms are those fixed constructions which are hard to automatically detect, given their syntactic flexibility and lexical variation. The syntactic properties of idioms have been formally represented and coded in binary matrices according to the Lexicon-Grammar framework. The research takes into account idioms with ordinary verbs as well as support verb idiomatic constructions. The overall classification counts 7,000+ Italian idioms. In particular, two binary matrices of two classes of idioms will be presented. The class $C_1$ refers to the Verb + Object constructions, whereas the class $EPC$ refers to the prepositional constructions with the support verb essere. Pre-constructed lexical resources facilitate idioms retrieval both in the case of "hybrid" and "knowledge-based" approaches to Natural Language Processing.

1 Introduction

Idioms, and multi-word expressions in general, have always been "a pain in the neck", as Sag et al. (2001) state in the title of their paper. The formal representation and the construction of a computational linguistic model of idioms is not an easy task as shown by Gazdar et al. (1985), Pulman (1993), Abeillé (1995), Villavicencio et al. (2004), Muzny and Zettlemoyer (2013) to name a few of the many (computational) linguists who have carried out research on this topic.

It has always been pointed out that the main problem concerning the automatic analysis of idioms is the difficulty to disambiguate such constructions which are ambiguous by definition (Fothergill and Baldwin 2012, Li and Sporlender 2009, Fazly et al. 2009, McShane and Nirenburg 2014). However, given the flexibility of idioms, a more basic and still unsolved problem has to be taken into account: that is, the extraction and annotation of such constructions (Fellbaum 2011).

As Fazly et al. (2009, p. 61) point out "despite a great deal of research on the properties of idioms in the linguistics literature, there is not much agreement on which properties are characteristics of these expressions". The distinction drawn by Nunberg et al. (1994) between idiomatic phrases and idiomatically combining expressions has been adopted by most of the research on idioms. However, many problems still remain and they are due to two basic reasons. On one hand, idioms can be considered lexical units, given the fact that their "special meaning" is associated to a particular verb and one or more particular complements. On the other hand, idioms syntactically behave as non-idiomatic constructions. Passive is the syntactic construction more frequently analyzed by the linguistic research on idioms since it involves the occurrence of the fixed object to the left of the verb. However, idioms show a great deal of other syntactic constructions where the fixed object may not necessarily occur in postverbal position (see Vietri 2014, forthcoming).

It is for these peculiarities that idioms have also aroused the interest of the psycholinguistic researchers who have advanced several hypothesis on the processing of idioms (Swiney and Cutler,
The systematic description of French idiomatic and non-idiomatic constructions has been carried out by Gross (1982, 1988) and his colleagues (Leclère, 2002) on the basis of the formal principles of the Lexicon-Grammar methodology, as developed by Gross (1975, 1979). According to Gross, the basic syntactic unit is not the word but the simple or elementary sentence, and the Lexicon-Grammar of a language is organized into three main components: free sentences, frozen sentences (or idioms), support verbs sentences (Gross 1981, 1998). For each component, Gross and his colleagues built exhaustive classifications, systematically organized and represented by binary matrices (named Lexicon-Grammar tables), where each syntactic and/or distributional property is marked "+" or "-" if accepted or not by a certain lexical unit. In the Lexicon-Grammar methodology, idiomatic and non-idiomatic constructions are built according to the same formal principles. The difference between these two types of constructions mainly concerns the distribution: idioms show a higher level of restricted distribution than non-idioms. The French Lexicon-Grammars are available at http://infolingu.univ-mlv.fr/english/.

A classification of English idioms and phrasal verbs has been carried out according to the same formal principles and criteria, respectively, by Freckleton (1985) and Machonis (1985). A Lexicon-Grammar of European Portuguese idioms has been built by Baptista (2005a, 2005b).

The Lexicon-Grammar classification of Italian idioms has been implemented on the basis of Gross' methodology. It includes more than 30 Lexicon-Grammar classes of idioms with ordinary verbs (sec. 2) and support verbs (sec. 3), for a total of more than 7,000 lexical entries. The binary matrices are created in Excel format.

2 The Lexicon-Grammar of Italian Idioms with Ordinary Verbs

The Lexicon-Grammar of idioms using ordinary verbs includes 12 classes for a total of 3,990 entries (sec. 2.1, Table 1). Each class of idioms contains those constructions which share the same definitional structure. In the Lexicon-Grammar framework, the definitional structure is identified on the basis of the arguments required by the operators (see Harris 1982). In the case of idioms, the operator consists of the Verb and the Fixed element(s), while the argument may be the subject and/or a free complement. This section shows only the main differences between the idioms' classes C1 and CAN.

For example, idioms in (1) and (2) have two different definitional structures. On one hand, an idiom such as tagliare la corda in (1) is an operator that requires only one argument, i.e. the subject. On the other hand, an idiom such as rompere le balle in (2) is an operator that requires two arguments, the subject and the noun Amy within the prepositional complement. The prepositions a and di alternate and can be considered fixed:

1. Amy ha tagliato la corda
   Amy-has-cut-the-rope
   "to sneak off"

2. Joe ha rotto le balle (a + di) Amy
   Joe-has-broken-the-balls-(to + of)-Amy
   "to annoy sb."

Idioms such as (1) have been listed and analyzed in a class named C1, that counts about 1,200 entries. Furthermore, C indicates the "constrained" or "fixed" noun and I refers to its position in the sentence, in this case, the object position. These idioms have only one argument, that is the (non-fixed) noun in subject position. The definitional structure of the class C1 is N₀ V C₁ where N indicates the free noun, V the verb and C, as previously stated, the fixed element. The subscripts 0 and 1 indicate the position of the noun within the sentence in a linear order, in this case, the subject and the object position.

Idioms such as (2) have been listed and analyzed in the class named CAN, that counts 320 entries. The definitional structure of this class is N₀ V C₁ (a + di) N₂, since these idioms have two arguments,
i.e. the subject $N_0$ and the noun $N_2$. The alternation of the prepositions $a$ and $di$ is represented between brackets, and the "+" sign indicates "either/or".

Each class, formally represented by a table in the form of a binary matrix, contains a specific number of idiomatic entries associated with a specific number of distributional and syntactic properties. In particular, each row of the matrix corresponds to an idiom, and each column to a property (or a construction). If the idiom accepts that particular property, a "+" sign is placed at the intersection between the row and the column; otherwise a "-" sign occurs.

As a sample of this type of lexical resource, I will give an excerpt of the class $C_1$ in Figure 1. The central non-numbered columns indicate the "part of speech" assigned to each lexical element that constitutes the idiomatic construction. In an idiom like non alzare un dito (lit. not lift a finger), the negation non is obligatory. On the other hand, the si-pronominal form is obligatory in idioms like leccarsi i baffi (lit. lick-si the moustaches). The determiner can be Definite (Def), Indefinite (Ind), or null (Zero). As previously pointed out, $V$ refers to the verb and $C$ to the fixed noun.

The properties from [1] to [3] indicate the distribution of $N_0$, i.e. the subject. It can be expressed by [$\pm$ human] noun or a by a sentence [Ch F].

The distributional property [4] indicates if $C_1$ is expressed by a body-part noun, whereas the morphological property [5] indicates if $C_1$ can be in the plural form. Property [4] showed that 1,700+ idioms involve a body-part noun, at least the 24% of the overall classification. Property [5] is a useful piece of information because it refers to the possible variation of the fixed noun and, consequently, of the determiner.

| Neg | Pro | V | Def | Ind | Zero | $C_1$ |
|-----|-----|---|-----|-----|------|-------|
| +   | -   | - | allungare | il | -   | -    |
| +   | -   | - | non | alzare | - | un   | -    |
| +   | -   | - | alzare | la | -   | -    |
| +   | -   | - | chiudere | il | -   | -    |
| +   | -   | - | dipanare | la | una | -    |
| +   | -   | - | incrociami | le | -   | -    |
| +   | -   | - | mangiare | i | un  | -    |
| +   | -   | - | leccare | - | -   | baffi |
| +   | -   | - | mangiare | - | -   | -    |
| +   | -   | - | mostrare | - | -   | corda |
| +   | -   | - | perdere | - | +   | tempo |
| +   | -   | - | rizzare | gli | -   | orecchi |
| +   | -   | - | rompere | il | -   | ghiaccio |
| +   | -   | - | scoprire | l' | -   | acqua calda |
| +   | -   | - | scoprire | - | -   | carte |
| +   | -   | - | scoprire | l' | -   | America |
| +   | -   | - | tappare | il | un  | buco |
| +   | -   | - | vendere | - | -   | fumo |

Figure 1. The Class $C_1$
The syntactic properties are numbered from [6] to [10]. In particular, [6] and [7] refer, respectively, to the unaccusative (3b) and the adjectival passive (3c) constructs in which some idioms may occur, as in the following:

3a. *Liv ha dipanato la matassa*  "to solve a problem"
    Liv-has-unraveled-the-skein
3b. *La matassa si è dipanata*  The skein-si-is-unraveled
3c. *La matassa è dipanata*  The-skein-is-unraveled

Properties [8] and [9] indicate two more sentence structures in which idioms may occur. In particular, [8] refers to a sentence structure involving the verb *avere* ('to have') as in (4b):

4a. *Gli operai incrociano le braccia*  "to go on strike"
    The-workers-cross-the-arms
4b. *Gli operai hanno le braccia incrociate*  The-workers-have-the-arms-crossed

The syntactic property [9] indicates a particular structure where the verb is in the infinitive form and introduced by the preposition *da*, as in (5b):

5a. *Joe ingoiò un rospo*  "to swallow a bitter pill"
    Joe-swalled-a-toad
5b. *Joe ha un rospo da ingoiare*  Joe-has-a-toad-to-swallow

Notice that, in the constructions defined by properties [6]-[9], C1 does not occur in its canonical position but to the left of the verb.

Property [10] concerns the possibility of having a nominalization, as in (6b). Finally, the morpho-syntactic property [11] shows the formation of a VC compound, as in (7b). The VC compound is explicitly indicated in the corresponding column.

6a. *Joe ha alzato la testa*  "to rebel"
    Joe-has-raised-the-head
6b. *L'alzata di testa (di + che ha fatto) Joe*  The-raising-of-the-head (of + that-has-made)-Joe
7a. *Joe vende fumo*  "to be a snake oil salesman"
    Joe-sells-smoke
7b. *Joe è un venditumo*  Joe-is-a-sell.smoke

2.1 The Classes of Idioms with Ordinary Verbs

Table 1 contains all the classes of idioms with ordinary verbs. The first column indicates the name of the Lexicon Grammar class, while the second column refers to the definitional structure of the idioms belonging to the corresponding class. The third column contains an idiomatic example for each class. Finally, the fourth column refers to the number of idioms listed in each class. The last class of Table 1, i.e. PVCO, contains those idioms where the fixed verb is followed by a comparative clause introduced by *come*.

The figures in the fourth column are to be taken as an approximate quantity, since this is an ongoing research. Therefore, the classes are subject to updating. Although approximate, the figures are an important piece of information because they show the idioms' distribution throughout the syntactic patterns.

---

2 See also De Gioia (2001).
Table 1. Idioms with Ordinary Verbs

| Class  | Structure                      | Example                                      | Entries |
|--------|--------------------------------|----------------------------------------------|---------|
| C0     | $C_0 \, V \, \Omega$          | *il piatto piange*                          | 80      |
| C1     | $N_0 \, V \, C_1$             | *tirare le cuoia*                          | 1,200   |
| CAN    | $N_0 \, V \, C_1 \, (a + di) \, N_2$ | *rompere le scatole (a + di) N*              | 320     |
| CDN    | $N_0 \, V \, C_1 \, di \, N_2$ | *non vedere l’ora di N*                     | 90      |
| CPN    | $N_0 \, V \, C_1 \, Prep \, N_2$ | *attaccare bottone con N*                    | 550     |
| CPC    | $N_0 \, V \, C_1 \, Prep \, C_2$ | *prendere lucciole per lanterne*            | 450     |
| CPCPN  | $N_0 \, V \, C_1 \, Prep \, C_2 \, Prep \, N_3$ | *dire pane al pane a N*                     | 20      |
| NPC    | $N_0 \, V \, N_1 \, Prep \, C_2$ | *piantare N in asso*                        | 350     |
| PCNP   | $N_0 \, V \, Prep \, C_1 \, Prep \, N_2$ | *dare alla testa a N*                       | 100     |
| PC1    | $N_0 \, V \, Prep \, C_1$     | *parlare al muro*                           | 600     |
| PCPC   | $N_0 \, V \, Prep \, C_1 \, Prep \, C_2$ | *durare da Natale a Santo Stefano*          | 30      |
| PVCO   | $N_0 \, V \, come \, C_1$      | *fumare come un turco*                      | 200     |

3,990

3 The Lexicon-grammar of the French être Prep constructions has been built by Danlos (1988). The Portuguese constructions were analyzed by Ranchod (1983). A first classification of the Italian essere Prep constructions has been built by Vietri (1996). This early classification has been completely revised.
The distributional properties [1]-[4] have been previously illustrated (sec. 2, Figure 1). The properties from [5] to [8] indicate the possibility for the EPC constructions to occur with verbs other than essere. The verbs considered are stare\(^4\), in [5], restare and rimanere ('remain'), in [6], diventare ('become, get') in [7]. The property [8] indicates that a construction with the verb of motion andare ('to go') may be acceptable.

However, the acceptability of all these constructions is lexically dependant, as in the following examples:

| Example | EPC Construction | English Translation |
|---------|-----------------|---------------------|
| 10. Nelly (sta+ resta + *diventa + va) al settimo cielo | Nelly-(stays + remains + *becomes + goes)-at-the-seventh-sky |
| 11. Joe (*sta + resta + diventa + *va) in gamba | Joe-(*stays + remains + becomes + *goes)-in-leg |

EPC constructions can also enter complex sentence structures with causative verbs (see properties [9]-[10]) such as mandare ('send'), mettere ('to put'), ridurre ('make'), as in the following:

---

\(^4\) I will literally translate this verb as "to stay". However, there is no equivalent in English since this verb is to be found in Romance languages like Italian, Portuguese and Spanish.
12. Joe (mandò + mise + ridusse) Nelly al settimo cielo
Joe-(sent + put + reduced)-Nelly-at-the-seventh-sky
"to be in seventh heaven"

13. Joe (mandò + mise + ridusse) Nelly sul lastrico
Joe-(sent + put + reduced)-Nelly-on-the-pavement
"to be on the skids"

Finally, property [12] indicates that the link operator (see Gross 1981) avere ('to have') may produce an acceptable sentence, as in (14b):

14a. La situazione in Ukraina è sott'occhio
The-situation-in-Ukraine-is-under-eye
"to monitor N"

14b. Obama ha sott'occhio la situazione in Ukraina
Obama-has-under-eye-the-situation-in-Ukraine

3.1 The Classes of Idioms with Support Verbs

Table 2 lists only those classes of SV idioms containing at least 50 idiomatic entries⁵. As a general rule, the classes of idioms with the verb essere start with E, those ones with the verb avere start with A, and finally, those classes involving the verb fare start with F. The only exception is the class PECO which refers to the idioms of comparison where the verb essere is followed by a clause introduced by come⁶.

| LG class | Sentence structure | Example | N. |
|----------|--------------------|---------|----|
| EPC      | N₀ essere Prep C₁   | essere sulle spine | 530 |
| EPCModif | N₀ essere Prep Adj C₁, N₀ essere Prep C₁ Adj | essere di vecchio stampo essere in mani sicure | 130 |
| EPCPN    | N₀ essere Prep N₀ C, N₀ essere Prep C₁ | essere all'oscuo di N essere ai ferri corti con N | 140 |
| EPCPC    | N₀ essere Prep C₀ C₁ | essere nelle mani di Dio essere al passo con i tempi | 115 |
| EAPC     | N₀ essere Adj Prep C₀ | non essere dolce di sale | 100 |
| PECHO    | N₀ essere Adj come C₀ | essere sordo come una campana | 360 |
| AC       | N avere C₀ | avere polso, avere (buon) occhio | 80 |
| APCA     | N avere C₀ Adj | avere la memoriaorta | 400 |
| ACXC     | N₀ avere C₀ Prep C₀, i nervi a fior di pelle | avere i nervi a fior di pelle i nervi di N sono a fior di pelle | 180 |
| ACPN     | N₀ avere C₀ Prep N₀ | non avere la testa di N | 50 |
| ACPC     | N₀ avere C₀ Prep C₀ | avere il cervello tra le nuvole | 200 |
| FC       | N₀ fare C₀ | fare melina, fare lo gnorri | 300 |
| FCPN     | N₀ fare C₀ Prep N₀ | fare le buce a, fare man bassa di N | 300 |
| FCDC     | N₀ fare C₀ Prep C₀ | fare l'arte dei pazzi | 80 |
| FCPC     | N₀ fare C₀ Prep C₀ | fare un buco nell'acqua | 220 |
| FPC(PN)  | N₀ fare Prep C₀ (E + Prep N₀) | fare sul serio, farsi in quattro per N | 50 |
| **Total** |                         | **3,235** |    |

Table 2. Idioms with Support Verbs

---

⁵ See Vietri (2014, forthcoming) for the complete classification.
⁶ For the French classification of idiomatic comparisons see Gross (1984).
4 Annotating and Parsing Idioms

The Lexicon-Grammar classes of idioms can be exploited by the hybrid as well as the symbolic approach to Natural Language Processing. Some experimentation in this direction has already been carried out by Machonis (2011), who used NooJ to retrieve and disambiguate English phrasal verbs. NooJ is an NLP application developed by Silberztein (2003) that relies heavily on linguistic resources.

NooJ has been used to carry out experimentation on some of the Lexicon-Grammar classes of Italian idioms. The experimentation, still in progress, concerns the annotation and parsing of idioms. This application allows the construction of lexicons/dictionaries whose entries contain information such as the distributional and syntactic properties indicated in the Lexicon-Grammar classes. The Lexicon-Grammar classes of idioms can be converted in a NooJ dictionary of idioms. This dictionary, which contains thousands of entries, has to be linked to a grammar that describes the syntactic behaviour of idioms. By applying to a text such a dictionary/grammar pair, NooJ successfully annotates and parses idioms, also in case the constituents Verb + Fixed element(s) are discontinuous. An example of this is the sentence *John ha vuotato subito il sacco* (lit. John-has-immediately-emptied-the bag, "to spill the beans"), where the underlined adverb occurs between the verb and the fixed object.

However, the current NooJ version does not yet handle easily the syntactic flexibility and the lexical variation of idioms. 

5 Conclusion

The Lexicon-Grammar classes of idioms are a manually-built linguistic resource that provides information about variation and flexibility of idioms. These classes, being formally coded, constitute an invaluable linguistic resource that can be used for research in (psycho)linguistics, and computational linguistics. The overall classification, as illustrated in Tables 1 and 2, outlines the syntactic patterns of the idiomatic constructions. This is a piece of information that can be regarded as the syntactic map of Italian idioms. Furthermore, the lexico-syntactic information provided by the idioms' classes can also integrate the automatic Machine Translation evaluation methods.

The Lexicon-Grammar classes of idioms can be exploited by the hybrid as well as the symbolic approach to Natural Language Processing. Some experimentation in this direction has already been carried out by Machonis (2011) and by Vietri (2014, forthcoming). Both authors used the knowledge-based system NooJ. On the other hand, Baptista et al. (2014) used the Lexicon-Grammar classes of Portuguese idioms to test the hybrid system STRING.

Further experimentation will be conducted to evaluate the benefit of using the LG distributional and syntactic information in order to extract idioms from corpora. However, very huge corpora (consisting of documents in an informal language style) are needed, together with powerful tools able to perform complex searches on massive textual data.

References

Anne Abeillé. 1995. The Flexibility of French Idioms: a Representation with Lexicalized Tree Adjoining Grammar. In M. Everaert, E-J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: structural and psychological perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates: 15-41.

Jorge Baptista, Anabela Correia, and Graça Fernandes. 2005a. Léxico Gramática das Frases Fixas do Português Europeo», *Cadernos de Fraseoloxía Galega*, 7: 41-53.

Jorge Baptista, Anabela Correia, and Graça Fernandes. 2005b. Frozen Sentences of Portuguese: Formal Descriptions for NLP. Proceedings of the ACL workshop on Multiword Expressions: 72-79.

Jorge Baptista, Nuno Mamede, and Ilia Markov. 2014. Integrating a lexicon-grammar of verbal idioms in a Portuguese NLP system. COST Action IC1207 PARSEME meeting, 10-11 March 2014.

Cristina Cacciari, and Patrizia Tabossi. 1988. The comprehension of idioms. *Journal of Memory and Language*, 27: 668–683.

---

7 Cignoni and Coffey (1998) provides the corpus-based results of the lexical variations of idioms.

8 The complete Lexicon-Grammar classification of Italian idioms will be available at [unisa.it/docenti/simonettavietri/index](http://unisa.it/docenti/simonettavietri/index).

9 In this regard, see Giménez and Márquez (2010), Costa-jussà and Farrús (2013).
Federica Casadei. 1996. *Metafore ed espressioni idiomatiche. Uno studio semantico sull'italiano*. Roma: Bulzoni.

Laura Cignoni, and Stephen Coffey. 1998. A corpus-based study of Italian idiomatic phrases: from citation forms to 'real-life' occurrences. *Euralex 1998 Proceedings*: 291-300.

Marta Ruiz Costa-jussà, and Mireia Farrús. 2013. Towards human linguistic machine translation evaluation. *Literary and Linguistic Computing*, Online publication date: 6-Dec-2013.

Cooper Cutting, and Kathryn Bock. 1997. That’s the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, 25:57–71.

Laurence Danlos. 1988. Les phrases à verbe support être Prép. *Langages*, vol. 23, n. 90: 23-37.

Michele De Gioia. 2001. *Avverbi idiomatici dell’italiano. Analisi lessico-grammaticale*. Torino: L’Harmattan Italia.

Elisabete Ranchod. 1983. On the support verbs *ser* and *estar* in Portuguese. *Lingvisticae Investigationes*, VII(2): 317-353.

Afsaneh Fazly, Paul Cook, and Susanne Stevenson, S. (2009). Unsupervised Type and Token Identification of Idiomatic Expressions. *Computational Linguistics*, Vol. 35(1): 62-103.

Christiane Fellbaum. 2011. Idioms and Collocations. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning*, Vol. 1, Berlin/Boston: De Gruyter Mouton: 441-456.

Richard Fothergill, and Timothy Baldwin. 2012. Combining resources for MWE-token classification. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics*: 100-104.

Peter Freckleton. 1985. Sentence idioms in English. *Working Papers in Linguistics* 11, University of Melbourne: 153-168.

Gerald Gazdar, Ewan Klein, Geoffrey Pullum, and Ivan Sag. 1985. *Generalized Phrase Structure Grammar*, Oxford: Basil Blackwell.

Raymond Gibbs. 1995. Idiomaticity and Human Cognition. In M. Everaert, E-J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: structural and psychological perspectives*. Hillsdale, NJ: Lawrence Erlbaum Associates: 97-116.

Jesús Giménez, and Lluís Márquez. 2010. Linguistic measures for automatic machine translation evaluation. *Machine Translation*, 24:209-240.

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

Maurice Gross. 1979. On the failure of generative grammar. *Language*, 55(4): 859-885.

Maurice Gross. 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages*, 63, Paris: Larousse: 7-52.

Maurice Gross. 1982. Une classification des phrases "figées" du français. *Revue Québécoise de Linguistique* 11.2, Montréal: UQAM: 151-185.

Maurice Gross. 1984. Une famille d’adverbes figés: les constructions comparative en comme. *Revue Québécoise de Linguistique* 13.2:237-269

Maurice Gross. 1988. Les limites de la phrase figée. *Langages* 90, Paris: Larousse: 7-22.

Maurice Gross. 1998. La fonction sémantique des verbes supports. *Travaux de linguistique*, 37, Duculot: Louvain-la-Neuve: 25-46.

Daniela Guglielmo. 2013. Italian Verb-Adverbial Particle Constructions: Predicative Structures and Patterns of Variation. *Lingvisticæ Investigationes*, 36.2:229-243.

Zellig Harris. 1982. *A Grammar of English on Mathematical Principles*, New York: John Winsley and Sons.

Christian Leclère. 2002. Organization of the Lexicon-Grammar of French verbs. *Lingvisticæ Investigationes* XX(1): 29-48.

Linlin Li, and Caroline Sporleder. 2009. Classifier Combination for Contextual Idiom Detection Without Labelled Data. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP 2009) Singapore: 315-323.
Peter Machonis. 1985. Transformations of verb phrase idioms: passivization, particle movement, dative shift. American Speech 60(4): 291-308.

Peter Machonis. 2011. Sorting Nooj out to take MWE's into account. In K. Vučković, B. Bekavac, & M. Silberztein (Eds.), *Automatic Processing of Various Levels of Linguistic Phenomena*. Newcastle upon Tyne: Cambridge Scholars Publishing: 152-165.

Francesca Masini. 2005. Multi-word Expressions between Syntax and the Lexicon: the Case of Italian Verb-particle Constructions. SKY Journal of Linguistics 18, pp. 145-173.

Marjorie McShane, and Sergei Nirenberg. 2014. The Idiom-reference Connection. *STEP ’08 Proceedings of the 2008 Conference on Semantics in Text Processing*, ACL Stroudsburg, PA, USA: 165-177.

Grace Muzny, and Luke Zettlemoyer. 2013. Automatic Idiom Identification in Wiktionary. *Proceedings of the Conference on Empirical Methods in Natural language Processing* (EMNLP 2013), Seattle, Washington, USA: 1417-1421.

Geoffrey Nunberg, Ivan Sag, and Tom Wasow. 1994. Idioms. *Language*, Vol. 70, No. 3, 491-538.

Stephen Pulman. 1993. The recognition and Interpretation of idioms, Cacciari C. & Tabossi, P. (Eds.), *Idioms. Processing, Structure, and Interpretation*. New Jersey: Lawrence Erlbaum Associates: 249-270.

Ivan Sag, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2001. Multiword Expressions: A Pain in the Neck for NLP. *Computational Linguistics and Intelligent Text Processing*. Berlin Heidelberg: Springer: 1-15.

Max Silberztein. 2003-. Nooj Manual. Available for download at: [www.nooj4nlp.net](http://www.nooj4nlp.net).

Simone Sprenger, Willem Levelt, and Gerard Kempen. 2006. Lexical Access during the Production of Idiomatic Phrases. *Journal of Memory and Language*, 54: 161-184.

David Swinney, and Ann Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18: 523–534

Simonetta Vietri. 1984. *Lessico e sintassi delle espressioni idiomatiche. Una tipologia tassonomica in italiano e in inglese*. Napoli: Liguori.

Simonetta Vietri. 1996. The Syntax of the Italian verb ‘essere Prep’, *Lingvisticae Investigationes*, XX.2: 287-350.

Simonetta Vietri. 2014. *Idiomatic Constructions in Italian. A Lexicon-Grammar Approach*. Linguisticae Investigationes Supplementa, 31. Amsterdam & Philadelphia: John Benjamins (forthcoming).

Aline Villavicencio, Timothy Baldwin, T., and Benjamin Waldron. 2004. A Multilingual Database of idioms. *Proceedings of the Fourth International Conference on language Resources and Evaluation* (LREC 2004), Lisbon, Portugal: 1127-30.