MWEs: Support/light Verb Constructions vs Fixed Expressions in Modern Greek and French
Angeliki Fotopoulou, Voula Giouli

To cite this version:
Angeliki Fotopoulou, Voula Giouli. MWEs: Support/light Verb Constructions vs Fixed Expressions in Modern Greek and French. Workshop on Multiword units in machine translation and translation technology, Gloria Corpas Pastor; Johanna Monti; Violeta Seretan; Ruslan Mitkov, Jul 2015, Malaga, Spain. pp.68-73. hal-03194262

HAL Id: hal-03194262
https://hal.science/hal-03194262
Submitted on 9 Apr 2021

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

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
MWEs: Support/light Verb Constructions vs Fixed Expressions in Modern Greek and French

Angeliki Fotopoulou
Institute for Language and Speech Processing,
R.C. “Athena” www.ilsp.gr
afotop@ilsp.athena-innovation.gr

Voula Giouli
Institute for Language and Speech Processing,
R.C. “Athena” www.ilsp.gr
voula@ilsp.athena-innovation.gr

Abstract
The paper reports on a study aimed at defining the limits between fixed expressions and Support Verb Constructions. To this end, a set of formal criteria that are applicable for the efficient classification of verbal MWEs were checked consistently against data in Greek and French. Ultimately, the delineation of the two types of constructions and their intermediate class is crucial not only for linguistic and lexicographic purposes, but also for Natural Language Processing tasks.

1 Introduction
Generally, verbal multi-word expressions (MWEs) fall into two general classes, namely idiomatic expressions and Support Verb Constructions (SVCs). Idioms or frozen/fixed expressions are defined as having a non-compositional meaning that cannot be deduced from the meaning of their parts (Bobrow & Bell, 1973; Chomsky, 1980; Fraser, 1970; Swinney & Cutler, 1979; M. Gross, 1982, 1988; Van der Linden, 1992). On the contrary, SVCs (also referred to in the literature as Light Verb Constructions, LVCs) consist of a support verb and a predicative noun. In between those two categories, however, a number of MWEs are proved to exhibit properties that are characteristic of both classes.

In this paper, we will try to discern the limits between fixed expressions and SVCs focusing on MWEs which fall in-between, comprising, thus, a grey-zone. The comparative study of MWEs in two languages, namely, Greek (EL) and French (FR), has a two-fold purpose: (a) to test the validity of the criteria set on the basis of cross-lingual similarities; and (b) to draw conclusions that might be useful for a range of NLP applications, either in single- or multi-lingual settings.

The paper is outlined as follows: section 2 briefly outlines the observations that motivated the present study; initial definitions and the criteria for disambiguation are presented in section 3. EL and FR data and the tests applied on them are presented in section 4; conclusions are discussed in section 5, whereas final conclusion and future research are outlined in section 6.

2 Motivation and Scope
The present study was triggered by two observations. On the one hand, the distinction between SVCs and fixed expressions is not always easy or straightforward and the limits between the two are often fuzzy. One could even maintain that there is a visible scalar passage between the two types of structures. In other words, a number of expressions seem to bear properties normally inherent to SVCs despite of their primarily being classified as fixed expressions and vice-versa. This is better illustrated in sentence (1) below taken from M. Gross (1981):

(1) il y a de l’eau dans le gaz
   lit. it has of the water in the gaz
   things are not running smoothly

The constituents of the expression in (1) cannot be modified if the meaning of the
sentence is to be preserved; nevertheless, they allow for a paraphrase with the operator verb *mettre* (=put), that is a structure typical of SVCs:

(2) la venue de Max a mis de l’eau dans le gaz  
*lit.* the arrival of Max has put of the water in the gaz  
Max’s arrival has caused irritation  
On the contrary, a paraphrase with the verb *être* (=to be) is ungrammatical:  
(3) *de l’eau est dans le gaz*  
*lit.*of the water is in the gaz

One step further, evidence from FR and EL data shows that the phenomenon is not language-specific. So the current study might prove to be useful for the analysis and representation of a phenomenon which is quite important for Greek - and not at all negligible for French and other languages alike. Based on these observations above, further criteria for the distinction between the two classes need to be elaborated, complementing, thus, the existing ones. These criteria are formal, depicting, thus, syntactic properties of MWEs.

3 **Fixed Expressions and SVCs**

Although there is nothing exceptional in their syntactic behavior or in their lexical content, fixed expressions are considered to be exceptions to the normal rules of the language. In an attempt to identify fixed expressions, linguistic tests are employed. The primary one is intuitive yet sufficiently operational: the meaning of fixed expressions is *non-compositional*, i.e., it cannot be computed from the meaning of its parts. The second test checks whether the constituents of the expression can be modified. At least two elements of a fixed expression do not allow for modification (*non-modifiability*); usually, one of the two is the verb. These criteria constitute a mechanism for distinguishing between a *fixed* expression and a *free* one.

SVCs, on the other hand, are defined as expressions comprising a support verb (*Vsup*) (*écho*/*avoir* (=have), *prendre*/*perdre* (=take), *γίνω*/*παίρνω* (=miss) in EL and FR respectively) or the operator verb *δίνω*/*donner* (=give) and a predicative noun (*Npred*) in object, subject or *PP* complement position. SVCs are defined as “semi-phrasal expressions formed by two lexical items *l*₁ and *l*₂ in which *l*₂ is taken in an arbitrary way to express a given sense and/or syntactic role in function of *l*₁’s choice” (Alonso Ramos, 2004). In this respect, their semantics is more or less transparent and their meaning is *semi-compositional*.

A systematic treatment of SVCs can be found in Gross (1981) and Danlos (1986), Vives (1993) in French. According to these studies, although highly idiosyncratic, SVCs exhibit regularities. For example, a nominal predicate forming a SVC with the *Vsup* *avoir* (=have) can also form aspectual variants when combined with other support verbs, i.e., *prendre* (=take), *perdre* (=loose), etc. or operator and causative verbs. An example of this type of expressions is shown in Table 1 below.

<table>
<thead>
<tr>
<th>EL</th>
<th>FR</th>
<th>EN</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>écho κουράγιο</em> to-have</td>
<td>avoir de courage</td>
<td>(=have courage)</td>
</tr>
<tr>
<td><em>παίρνω κουράγιο</em> to-take</td>
<td>prendre du courage</td>
<td>(=take courage)</td>
</tr>
<tr>
<td><em>γίνω κουράγιο</em> to-loose</td>
<td>perdre son courage</td>
<td>(=loose my courage)</td>
</tr>
<tr>
<td><em>δίνω κουράγιο</em> to-give</td>
<td>donner du courage à</td>
<td>(=give courage to)</td>
</tr>
</tbody>
</table>

Table 1. Examples of SVCs

Finally, the SVC is semantically equivalent to a full verb and can, thus, be paraphrased as an elementary sentence; in this context, the *Vsup* licenses *nominalisation*, a term which, in this context, refers to the relation between two sentences, one of which is a verbal construction (as shown in (4) below) and the second one is a SVC (depicted in (4a)):

(4) Marie respecte son père  
Marie respects her father  
(4a) Marie *a du respect* pour son père  
(Marie has respect for her father)

The relation that holds between the two elementary sentences (4) and (4a) is of the form:  
**N0 V N1 = N0 Vsup Det Npred Prep N2**

The implication is two-fold: (a) in terms of semantics, the *Vsup* *does* contribute to the meaning of the expression, maintaining, thus, some sort of lexical meaning; and (b) even though there is a great deal of idiosyncrasy in the formation of SVCs, regularities with respect to the alternations are attested.

However, as we will show below (examples 8a- 8c) grey zones arise when these regularities –

---

Proceedings of the Workshop on Multi-word Units in Machine Translation and Translation Technology, Malaga, Spain, 1-2 July, 2015  
Corpas Pastor, G., Monti, J., Seretan, V., Mitkov, R., eds.
alternations are attested in expressions initially classified as idiomatic and vice versa. In this context, the intuitive criterion of compositionality seems to be insufficient for the robust classification of MWEs; to this end, we test a set of formal criteria that complement the semantic criterion.

3.1 Criteria for disambiguation

In the following, we will present the structural properties of SVCs; these may be further employed as formal criteria for disambiguation when classification seems to be problematic. More precisely:

(i) An elementary sentence of the form $N0 \ V \ N1$ or $N0 \ V \ Prep \ N1$ may be classified as SVC if it can be paraphrased as a Vsup sentence, that is, one that comprises one of the most common Vsup, namely $\dot{\varepsilon}z\omega/avoir (=have)$ and $\varepsilon\eta\mu\alpha/\text{être}$ and a predicative noun (Npred):

$$N0 \ V \ N1 = N0 (\dot{\varepsilon}z\omega/avoir+\varepsilon\eta\mu\alpha/\text{être}) \ Npred$$

Our hypothesis can then be formulated as follows: a MWE (which is otherwise difficult to be classified due to its exhibiting also properties of fixed expressions) falls into the class SVC, given that it can be associated with a simple Vsup construction and a Prep besides the predicative Noun (N). If a typical SVC can replace the original sentence, then the expression in question is considered to be an aspectual or lexical variant of a SVC. Otherwise, if the Vsup sentence is ungrammatical, the construction is considered to be a proper fixed expression instead.

(ii) An elementary sentence of the form $N0 \ V \ N$ (or $N0 \ V \ Prep \ N$) is classified as SVC if an equivalence relation is proved to hold between this sentence and a structure of the form Npred de $N0$ for French and Npred $N0_{Gen}$ for Greek; the latter structures are the result of a nominalization transformation. According to this requirement, therefore, if the so derived nominal group is acceptable, then the candidate expression is classified as SVC otherwise, it is a fixed expression.

4 The Data

The afore-mentioned tests were applied on EL and FR datasets developed within the Lexicon-Grammar framework (Gross, 1982). (Fotopoulou, 1993). The dataset comprises c. 1020 MWEs in EL c. 3700 expressions in FR. Iterative checks were performed over the EL and FR data to accurately perform MWE classification. In this regard, a number of structures were observed as displaying the characteristics of both MWE classes. Disambiguation, of grey expressions was guided by the application of the formal tests proposed in section (3.1) above.

In the remaining, we will present the cases of disambiguating MWEs based on the criteria proposed, i.e., capitalizing on the properties of two types of MWEs: (a) those with the Vsup $\dot{\varepsilon}z\omega/avoir (=have)$ and (b) verbal MWEs with $\varepsilon\eta\mu\alpha/\text{être} (=to be)$ and their variants.

4.1 Classification: $\dot{\varepsilon}z\omega/avoir (=to have)$ test

Standard (lexical) tests employed for the identification of SVCs show that a number of verbs function as support verbs. This property, however, can be blocked when these verbs are used with specific nouns. For example, the verb $\chi\alpha\rho\omega (=to enjoy)$ is used in SVCs of the form:

(5) Ο Νίκος χαίρει σεβασμό
theSING.Nom NikosSING.Nom enjoys3.SG
respectSING.Gen

Nikos is respected.

However, the expression $\chi\alpha\rho\omega \ \acute{\alpha}κ\acute{\alpha}ς \ \upsilon\gamma\epsilon\iota\alpha\varsigma (=to be healthy) was identified as a fixed expression on the basis of the criteria defined above. The test employed was of the form Vsup $C = (\dot{\varepsilon}z\omega/avoir) \ C$. As a matter of fact, the noun $\upsilon\gamma\epsilon\iota\alpha\varsigma (=health) along with its modifier $\acute{\alpha}κ\acute{\alpha}ρ\alpha (=extreme)$ form together a unique combination that allows for no modification; the formation of a Vsup construction with the verb $\dot{\varepsilon}z\omega (=have)$ results in an unacceptable construction. The Vsup paraphrase depicted in (5b) is unacceptable:

(5a) Ο Νίκος χαίρει ύπαρξης υγείας
o Nikos cheri akras igias
theSING.Nom NikosSING.Nom healthSA3.SG

Nikos has extreme health

(5b) Ο Νίκος δεχεί σεβασμό
theSING.Nom NikosSING.Nom enjoys3.SG

Nikos is respected.

The application of the first test, i.e. the substitution of the verb by $\dot{\varepsilon}z\omega (=have)$ has proven the sentence to be fixed. Along the same lines, the operation implied by the second criterion above is applied, i.e., elimination of the
\(Vsup\) and formation of a nominal group; the resulting structure is also unacceptable:

\[(5c) \quad \text{ο Νίκος χαίρει εξαιρετικής υγείας} \]
\(\text{o Nikos cheri eksiaretikis ugias}\)
the\_SG\_Nom Nikos\_SG\_Nom
Nikos’ extreme health

However, when the modifier \(άκρα\) (=extreme) is replaced with another synonym, e.g. \(εξαιρετική\) (=excellent), paraphrasing the sentence with the use of \(έχω\) (=have) is permitted:

\[(5d) \quad \text{ο Νίκος έχει εξαιρετική υγεία} \]
\(\text{o Nikos exei eksiaretikia ugia}\)
the\_SG\_Nom Nikos\_SG\_Nom
Nikos enjoys excellent health

\[(5e) \quad \text{ο Νίκος έχει εξαιρετική γνώση} \]
\(\text{o Nikos exei eksiaretikia gnose}\)
the\_SG\_Nom Nikos\_SG\_Nom
Nikos has excellent health

Now, when the modifier is replaced, the formation of the nominal group is allowed:

\[(5f) \quad \text{η εξαιρετική υγεία του Νίκου} \]
\(\text{he egaeoretika ugia tou Nikou}\)
the\_SG\_Nom Nikos\_SG\_Nom
Nikos’ good health

We therefore conclude that on grounds of formal criteria, the expression in sentence (5a) is a fixed one. On the contrary, based on the criteria proposed, the expression \(τρέφω\) \(ελπίδες\) (=nourish hopes) was proved to be a pseudo-fixed expression since properties of SVCs were identified:

\[(6) \quad \text{Ο Γιάννης τρέφει ελπίδες \ldots} \]
\(\text{O Gianis trefi elpioed}\)
lit. the John feeds\_3.SG hopes for
John hopes…

\[(6a) \quad \text{Ο Γιάννης έχει ελπίδες \ldots} \]
\(\text{O Gianis exei elpioed}\)
lit. the John has\_3.SG hopes
John hopes…

\[(6b) \quad \text{οι ελπίδες του Γιάννη} \]
\(\text{oi elpioed tou Gian}\)
lit. the hopes of-the\_SG\_Gen John\_SG\_Gen
John’s hopes

Evidence from French follows:

\[(7) \quad \text{Max voulait une admiration à} \]
\(\text{Max voule une admiration a}\)
lit. Max pays admiration for
Max admires

\[(7a) \quad \text{Max a de l’admiratation pour} \]
\(\text{Max a de l’admiratation pour}\)
lit. Max has admiration for
Max admires

\[(7b) \quad \text{l’admiratation de Max} \]
\(\text{l’admiratation de Max}\)
lit. the admiration of Max
Max’s admiration

Nevertheless, the expression \(παίρνω\) \(σάρκα\) \(και\) \(οστά\) (=become true) which includes the \(Vsup\) \(παίρνω\) (=take), doesn’t meet the formal criteria. As a matter of fact, neither the formation of the nominal group is acceptable, nor can the verb \(παίρνω\) (=take) be substituted by \(έχω\) (=have):

\[(8) \quad \text{to \(έργο\) παίρνει \(σάρκα\) και \(οστά\)} \]
\(\text{to erjo perni sarka ke osta}\)
the\_SG\_Nom project\_SG\_Nom takes\_3.SG flesh and bones
the project starts

\[(8a) \quad \text{σάρκα και \(οστά\) του \(έργου\)} \]
\(\text{sarka ke osta tou ergou}\)
flesh and bones of-the\_SG\_Gen project\_SG\_Gen
flesh bones of the project

\[(8b) \quad \text{το \(έργο\) \(έχει \(σάρκα\) και \(οστά\)} \]
\(\text{tou ergo exei sarka ke osta}\)
to ergo exei sarka ke osta
the project has flesh and bones

However, a semantically similar expression that includes the causative operator \(δίνω\) (=give) is also possible:

\[(8c) \quad \text{κυβέρνηση \(δίνει\) \(σάρκα\) και \(οστά\)} \]
\(\text{kivernisi di sarka ke osta}\)
lit. the government gives\_3.SG flesh and
bones to the project

The same is attested in French expressions with lexical variations. The application of the two tests on the French expression \(\text{donner corps}\) results in a grammatical sentence when paraphrased with a simple \(Vsup\) (9a) and an ungrammatical nominal group (9b):

\[(9) \quad \text{Je donne corps à un projet} \]
\(\text{lit. I give body to a project}\)
I give flesh to a project

\[(9a) \quad \text{le projet prend corps} \]
\(\text{lit. the project takes body}\)
the project comes into being

\[(9b) \quad \text{le corps du projet} \]
\(\text{lit. the body of the project}\)

4.2 Classification: \(\text{être} (=\text{to be})\) test

A number of expressions in our dataset seem to be paraphrases of structures displaying the form \(\text{être} (=\text{to be})\) \(\text{Prep C1}\). Their aspectual variants contain either a verb that is prototypically defined as denoting movement (\(Vmt\)) or movement causative verb (\(Vcmt\)). To better account for properties these expressions have, a reference to the properties of \(Vmts\) and \(Vcmts\) is in order. These are depicted in examples (10)-(12) below.
The relation that holds between simple structures with Vmt predicates of the type (10), (10a) and Vcmt structures (11), (11a) involves a semantic alternation. Sentences with a Vmt or a Vcmt have a dynamic aspect, whereas sentences of the form είμαι (=be) Prep have a static one. The dynamic aspect of the former can also be ascribed, in case of (11), (11a), to the causative operator that introduces the subject of the sentence. The same phenomenon is also attested in MWEs of the form be Prep X and their variants comprising verbs that are prototypically Vmt or Vcmt predicates, as illustrated in examples (13)-(18a) below:

N0 Vcmt N1 Prep N2
(11) η Μαρία έστειλε τον Νίκο στην εξοχή
lit. the Maria sent Nikos to the countryside
Maria sent Nikos to the countryside
(11a) Maria a envoyé Nikos à la campagne
Maria sent Nikos to the countryside

N1 είμαι (=to be) Prep C2
(12) ο Νίκος είναι στην εξοχή
lit. the Nikos is at the countryside
Nikos is at the countryside
(12a) Nikos est à la campagne
Nikos is at the countryside

The expressions in (19) and (20) are MWEs bound with the alternation implied by their Vmt and Vcmt predicates, even though the be Prep X structure has a compositional meaning:

Discussion
Summing up the presentation of the different cases, we observe two main classes of MWEs: (i) fixed expressions, which allow for regular syntactic alternations, and which do not correspond to a simple SVC; and (ii) fixed expressions that follow the be Prep X paradigm,
in that they have aspectual variants or causative operations, yet they lack the basic be Prep X form. In our view, an efficient computational treatment of MWEs within NLP applications requires a detailed analysis of each expression and the definition of their so-called fixed zone (zone fixe) - a term coined to denote a fixed set of words or tokens within an MWE that would be optionally amenable to morphological variation. In this sense, the support verb within a support verb construction (SVC) is not considered to be part of the SVC fixed zone (Laporte, 1998). However, variation within the SVC can also be lexical/aspectual. It is, therefore, crucial to redefine and delineate the notion of fixed zone and to define more elaborate linguistic features that apply to MWEs regardless their initial classification. This will have an impact on their treatment in NLP. For example, expressions that do not correspond to the SVC paradigm, but, nevertheless, allow for certain operations should be encoded as appropriate.

6 Conclusion

This study is aimed at delineating the boundaries between fixed expressions and SVCs constructions (or collocations) in two languages, namely French and Greek. A set of formal tests suitable for the classification of verbal MWEs have been checked. The focus is primarily on expressions that seem to exhibit features inherent to both fixed expressions and SVCs. Future work involves the application of these tests to ultimately all MWEs in the Lexicon-Grammar tables and beyond, aiming at the elaboration of an accurate and consistent classification thereof. The latter is particularly important in view of processing MWEs in a number of applications including alignment and paraphrasing, translation, etc.

References

Laporte, É., Nakov, P., Ramish, C., and Villavisencio, A., (eds) 2010. Proceedings of the COLING Workshop on MWEs: from Theory to Applications (MWE 2010), Beijing, China. ACL.

1 The term elementary sentence is defined as the linguistic unit of meaning in accordance with Z. Harris’ Transformational Grammar and the Lexicon-Grammar framework (Gross 1975)

2 The notation employed here is the one adopted by Lexicon-Grammar (Gross 1975).