

# Research Articles



**Photo:**  
**Nataly Villegas**

# SMART\_PV: A SOFTWARE APPLICATION FOR MANAGING ENGLISH PHRASAL VERBS

[SMART\_PV: UNA APLICACIÓN DE SOFTWARE PARA LA GESTIÓN DE VERBOS FRASALES EN INGLÉS]

## **Bell Manrique Losada**

holds a PhD in Engineering Systems and Informatics from Universidad Nacional de Colombia, Colombia. She currently works as full time and assistant professor at Universidad de Medellín, Colombia. Mailing address: Bloque 4-113, Universidad de Medellín, Campus Universitario Cra. 87 No. 30-65 Belén Los Alpes, Medellín, Colombia.

E-mail: bmanrique@udem.edu.co

## **Francisco Moreno Arboleda**

holds a PhD in Engineering Systems from Universidad Nacional de Colombia, Colombia. He currently works as associate professor at Universidad Nacional de Colombia, Colombia.

E-mail: fjmoreno@unal.edu.co

## **Guillermo Orrego Gil**

holds a Bachelor degree in Systems Engineering from Universidad Nacional de Colombia, Colombia. He currently works as teaching assistant at Universidad Nacional de Colombia, Colombia.

E-mail: georregog@unal.edu.co

## **ABSTRACT**

Phrasal verbs (PVs) are lexical units consisting of a verb and one or two particles. In this paper we present a characterization of English PVs. This characterization serves as the backbone for our web application called Smart\_PV. The purpose of Smart\_PV is twofold: i) to allow the input of PVs and ii) to detect PVs in texts. We designed web interfaces to register the PVs with their features and to detect PVs in the texts as follows: the user enters the text and starts the PVs detection process by splitting the text into words. Smart\_PV was validated inserting 80 PVs (including the 25 most common PVs in documents of the European Union) and detecting PVs in texts from different domains. Our results show the expediency of this kind of applications for teachers, students, translators, and common users, as a tool to support translation and text mining tasks. Although a database with more PVs and the analysis of more documents are required, our results demonstrate the feasibility and usefulness of our application.

**Keywords:** phrasal verbs, adverbial verbs, prepositional verbs, particle mobility, text mining

## **RESUMEN**

Los verbos frasales (VF) son unidades léxicas conformadas por un verbo y una o dos partículas. En este artículo se presenta una caracterización para los VF en inglés, la cual es usada como eje principal de una aplicación web para VF llamada Smart\_PV. El objetivo de Smart\_PV es doble: i) permitir el ingreso de VF, y ii) detectar VF en textos. Se diseñaron interfaces web para ingresar los VF con sus características y para detectarlos en textos así: el usuario ingresa un texto y comienza el proceso de detección de VF dividiendo el texto en palabras. Smart\_PV fue validada insertando 80 VF (incluyendo los 25 VF más comunes en documentos de la Unión Europea), y detectando VF en textos de diferentes dominios. Los resultados muestran la conveniencia de este tipo de aplicaciones para profesores, estudiantes, traductores y usuarios comunes como apoyo en tareas de traducción y minería de textos. Aunque es deseable una base de datos con más VF y el análisis de más documentos, los resultados obtenidos demuestran la viabilidad y utilidad de la aplicación.

**Palabras clave:** verbos frasales, verbos adverbiales, verbos preposicionales, movilidad de la partícula, minería de textos

## 1. INTRODUCTION

Phrasal verbs (PVs) are lexical units (Capelle, Shtyrov, & Pulvermuller, 2010) composed of a verb and one or two particles. This combination creates a grammatical construction with its own features. For a combination of a verb and particle(s) to be considered as a PV, the typical meaning of the verb should be modified by the particle(s). Although this paper focuses on the English PVs, such verbs also occur in other languages such as German and Italian.

These constructions are usually a language barrier for non-English speakers because of the challenges they present in reading and interpretation and because of their polysemic richness. We intend to cover and classify a significant portion of these constructions, and thereby facilitate both their identification and interpretation. In fact, several PVs classifications have already been proposed (Álvarez, 1990; Baldwin, 2005; Claridge, 2000; Gries, 2001). There are also several names for these verbs: phrasal verbs, verb-particle constructions, composite verbs, multi-word expressions, two/three part verbs, and two/three word verbs, among others (Colin, 2005; Hart, 1999; Hill & Bradford, 2000; Hoque, 2008; McArthur, 1992; Trebits, 2009). After the classification, we present the components of our PVs web application. The purpose of our application is twofold: i) creating a PVs database that allows the input of PVs and their features according to the classification presented, and ii) detecting PVs in texts using the PVs database.

Despite the large amount of software focused on both teaching and translating the English language, we did not find specialized tools about this subject. We believe our application can be useful for both teachers and students of the English language, and it may be of value in areas such as translation and text mining.

The paper is organized as follows: in Section 2 we present the PVs characterization; in Section 3 we describe our PVs web application; in Section 4 we

present experiments and results, and in Section 5 we describe future work and conclude the paper.

## 2. PVS CHARACTERIZATION

Next, we present a classification for PVs based on i) the role and the mobility of the particle, ii) the possibility that the verb allows passive transformation (PT) and iii) the relationship between the direct object (DO) and the prepositional object (PO). The classification is based mainly on the work of Lavin and Sánchez (1989).

### PVs Classification

**Group (A) - Adverbial verb:** verb + adverbial particle. The verb is accompanied only by a particle and there may be complementation. Example: The soil *gave off* radioactive carbon dioxide.

**Subgroup (A-1) - Adverbial verb - with DO - allows PT.** Example: I *brought up* my kids decently. (PT) My kids were *brought up* by me decently.

- Subgroup (A-1a) - Adverbial verb - DO before or after the particle - allows PT. Example: We *called* the picnic *off*, or we *called off* the picnic. (PT) The picnic was *called off*.
- Subgroup (A-1b) - Adverbial verb - DO before the particle - allows PT. These are adverbial verbs in which the DO comes preferably before the particle. These verbs have a static structure: "DO + particle". Note that the particle must be put at the end of the sentence, thus eliminating the possibility of being confused with a preposition. Example: The brandy will *bring* the girl *round*. (PT) The girl will be *brought round*.
- Subgroup (A-1c) - Adverbial verb - DO after the particle - allows PT. These are adverbial verbs in which the DO comes preferably after the particle. These verbs have a static structure: "particle + DO." Example: We *carried out* your orders. (PT) Your orders were *carried out*.

**Subgroup (A-2)** - Adverbial verb - without DO - does not allow PT. Here the absence of DO prevents the verb from admitting PT. Example: The car is *broken down*.

**Group (B) - Prepositional verb:** verb + prepositional particle. In this group, the verb is accompanied with a particle that plays a prepositional role which requires a complement. Example: You're *asking for* trouble!

**Subgroup (B-1)** - Prepositional verb - allows PT. Example: He *looked after* the kids for a long time. (PT) The kids were *looked after* for a long time.

- Subgroup **(B-1a)** - Prepositional verb - DO matches the PO - allows PT. Example: Her illness *accounts for* her absence. (PT) Her absence is *accounted for* by her illness.
- Subgroup **(B-1b)** - Prepositional verb - DO does not match the PO - allows PT. Example: They *took him for* a blind man. (PT) He was *taken for* a blind man.

**Subgroup (B-2)** - Prepositional verb - does not allow PT. Example: I promise to *abide by* anything you say.

**Group (C) - Adverbial prepositional verb:** These are PVs made up by two particles, the first one plays an adverbial role and the second one plays a prepositional one. Example: She *looked down on* my invitation.

**Subgroup (C-1)** - Adverbial prepositional verb - allows PT. Example: I am going to *put up with* your rudeness all afternoon. (PT) Your rudeness will be *put up with* all afternoon.

- Subgroup **(C-1a)** - Adverbial prepositional verb - DO matches the PO - allows PT. Example: We must *face up to* the truth. (PT) The truth must be *faced up to*.
- Subgroup **(C-1b)** - Adverbial prepositional verb - DO does not match the PO - allows PT. Example: They are *filling him in on* the subject. (PT) He is being *filled in on* the subject.

**Subgroup (C-2)** - Adverbial prepositional verb - does not allow PT. Example: It all *comes down to* nothing.

## Other PVs Features

**Equivalent Particles (D).** In some PVs the particle(s) may be exchanged for other(s) without changing the meaning of the resulting combination. The equivalent particles are separated by a slash. Example: Get about/around.

**Optional Particles (E).** Some PVs may be accompanied with an optional particle (especially a preposition), whose co-appearance in the construction depends on the existence of the prepositional complement, without changing the meaning of the sentence. The optional particles are enclosed in parentheses. Example: Be out (with).

**Reflexive Verb (F).** Some PVs allow reflexive pronouns. Example: Be beside oneself. Note that the inclusion of the reflexive pronoun implies that: i) the corresponding pronoun comes before the particle, as in *he picked himself up*, and ii) PT is not allowed.

**Style (G).** In some PVs is possible to identify their scope of use (formal, informal, slang, taboo). This feature does not apply to all PVs, since most of them are considered neutral and may be used in any environment. The following styles are considered:

**Formal (G-1).** PVs that are used more often in written language than in speech. Example: Call forth.

**Casual (G-2).** PVs that are used more often in speech than in written language. Example: Bump into.

**Slang (G-3).** PVs that are typical of everyday conversations. Many of these PVs come from certain social groups that use them to differentiate themselves (slang of students, soldiers, criminals, etc.). Example: Bum from/off.

**Taboo (G-4).** PVs that are considered obscene, of sexual connotation, and those with a brusque and inconvenient nuance. Example: Fuck off.

**Geographic Context (H).** There are PVs whose use is predominant in certain countries or regions. Examples: Line up (American English (**H-1**)), queue up (British English (**H-2**)), barrack for (Australian English (**H-3**)).

**Nominal Transformation (I).** Some PVs allow nominal transformation, thus generating nouns called *phrasal names* (PN). Example: Sell-out.

In Table 1 (see Appendix) we show a comparison among proposals for PVs characterisation. While all proposals consider characteristics such as the role and the mobility of the particle, only Lavin and Sánchez (1989) proposal along with our extension considers key issues such as passive transformation, different classifications for the same verb-particle combination, optional and equivalent particles, geographic context, and nominal transformation, among others.

### Different classifications for the same verb-particle combination

The same verb-particle combination (VPC), e.g. *go off* can generate several PVs, each with its respective classification and features. Consider the following example:

PV	Meaning	Classification (subgroup)
<b>go off</b>	Stop feeling, blow (bombs, fireworks)	<b>A-2</b>
<b>go off</b>	Lose interest for	<b>B-2</b>

In this example the *go off* VPC generates two PVs, the first one classified in Subgroup A-2 with the meaning “stop feeling” or “blow (bombs, fireworks)” and the second one classified in Subgroup B-2 with the meaning “lose interest for”. This example also describes the ability of some particles to play both an adverbial role as well as a prepositional one. Note also that a PV may have

multiple meanings. These aspects play a decisive role when designing the input interface of a PVs application, since a VPC can generate several PVs, each one with multiple meanings.

## 3. SMART\_PV

In this section we describe our web application called Smart\_PV that allows creating a PVs database and detecting possible PVs in texts.

### PVs Input

First, the verb and its particle(s) are entered into the system. Next, the system verifies its existence according to the flowchart on Figure 1.

Then, the user inserts features associated with PT and mobility (see Figure 2). Finally, the user inserts the additional features (see Figure 3).

The Web Interfaces for the PVs input are as follows:

### Preliminary Considerations.

We considered the following aspects for the design of our web interfaces:

- Reduce the possibility of input errors using lists and check buttons, and inferring some features, such as the PN form.
- Provide ease and clarity to the user using assisted inputs (e.g., drop-down selection fields), helpful tips on the inputs and error handling.
- Ensure that the user experiences a continuous visual perception: using some principles such as the one proposed in Scott and Neil (2009). Avoid transitions between pages as much as possible. Usually in web applications, the user performs actions that may lead him to new pages. These transitions generate breaks that prevent the user staying in the session. To reduce the number of transitions, the following techniques may be used:

*Input overlay:* A mini-page to enter data, overlaid on the current page. This type of input simplifies and allows the content of the main page to remain visible but disabled.

*Detail inlay:* A detail whose appearance depends on the actions of the user. This type of detail will display on the same page, thus preserving the context without disable the content of the main page. They are used mainly for data input.

*Light page:* The user should be offered with the minimum tools to facilitate the attainment of his goals.

*Always-Visible Tool:* These are tools that usually provide access to the most frequent or important user actions. For example, the help options contribute to the accurate input of PVs features; therefore, these options must always be visible. They should be used sparingly and should be placed strategically for easy access by the user.

*Hover-Reveal Tools:* These are tools that are displayed on demand, i.e., they remain hidden until an action that reveals them is performed. For example, a help message appears when the user places the mouse over an element (a *tooltip* [CartoWeb, 2010]).

## Description of web interfaces.

When entering a PV into the system, we must check if the corresponding VPC is already in the database; this validation is performed on the first interface. This in turn, along with the values inserted by the user, determines the content of the following interfaces:

**Interface 1.** It starts the insertion process and contains the following elements:

- *Links to pages.* Three links to pages that allow the input of particles, styles, and geographic contexts.
- *List.* A select list that shows the number of particles that makes up the PV. If the user selects “1”,

the role of the particle (adverbial or prepositional) is requested by using two check buttons (a *detail inlay*). Then the option to input the verb in a text field and select the particle from a list is offered (another *detail inlay*). If the number of particles is “2”, it is not necessary to ask for their role, the process goes directly to the input of the verb and the selection of particles by using two lists. At this point, the user presses a button to validate whether the PV exists in the database. If it does exist, the process continues at interface 2. If it does not exist, then the user must insert other PV’s features as follows: lists to select optional and equivalent particles, style, geographic context and buttons to indicate whether the PV is reflexive, if it allows PT, and if it is regular. If the verb is irregular, the participle and the past forms are requested by an *input overlay*. Note that if the verb is regular, these verbal forms can be deduced (Kendris, 2008); therefore, they are not requested.

- *Input fields.* Two text fields to input the meanings of the PV and the PN (if the PV allows it). As a PV can have several meanings, the user can enter them one at a time.

**Interface 2.** It presents to the user all the PV features that exist in the system with the same VPC and asks him/her whether he/she wishes to insert the PV in a different subgroup from the existing ones. If the answer is “NO”, the process ends, if the answer is “YES”, the process continues at interface 1, with the input of the other PV’s features.

**Interface 3.** The final interface depends on the parameters that were selected:

- *Form 1.* If the PV is composed of a particle with an *adverbial* role and allows PT, then a select list is presented where the mobility of the particle is asked for. If the answer is “YES”, it is asked if the DO must go before the particle or if it can go before or after it. If the answer is “NO”, then it means that the DO should go after the particle.
- *Form 2.* If the PV is composed of a particle with a *prepositional* role and allows PT, a select list

is presented where it is asked whether the DO matches the PO.

A button is provided at the end of the interface to indicate the end of the input process. The PVs are stored in a database, whose schema is shown in Figure 4. Type attribute indicates if the PV is A (adverbial), B (prepositional) or C (adverbial prepositional). If the PV has two particles, then it is of type C. PT attribute indicates if the PV allows passive transformation. Subtype attribute: if the PV is of type A, it indicates the position of the particle (before, after, or both); if the PV is of type B or C, it indicates if the DO matches with PO. Reflexive attribute indicates if the PV is reflexive.

#### PVs Detection Process

We developed an interface where the user can insert a text and a button to start the PVs detection process.

236

In order to detect PVs, the text is split into words. Using the PVs database, we find the following for each PV: its code, its verb forms (infinitive, past, past participle, gerund, and third person), and its first and second particle (if it applies). Once we have collected these data: i) we compare each word in the text with the verb forms of each verb; ii) if there is a match, we compare the following words in the text with the first particle of the PVC; iii) if there is a match and the number of particles of the PVC is one, then we have detected a PVC; and iv) if the number of particles of the PVC is two, we compare the following word in the text with the second particle, if they match, then we have detected a PVC, otherwise we return to step i) with the following word in the text.

Note that if any of the verb forms matches with a word in the text, we define two conditions to stop searching the first particle in the following words of the text: we find a punctuation mark or another verb (in any of its verb forms). Figure 5 shows the detection process.

## 4. EXPERIMENTS AND RESULTS

To validate our application, we initially populated the database with a sample of 40 PVs. Then we added another 20 verbs and then another 20 for a total of 80. Using these samples we analyzed 60 texts. The results are shown in Table 2. Among the indicators, we consider the following: *search efficiency* is a rate that is computed as: (total number of detected PVs/total number of PVs in the texts), *search deficiency* is a rate that is computed as: (total number of no-detected PVs/total number of PVs in the texts), and *search optimization* is a rate that is computed as  $(1 - (\text{total number of detected VPCs that were not PVs} / \text{total number of detected VPCs}))$ .

As expected, the search efficiency rate increased as the PVs sample database increased. A similar behavior happened to the search deficiency rate, indeed the number of detected VPCs that were not PV decreased as the PVs sample database increased because more of these VPCs were likely to be found in the PV database.

In addition, we analyzed the frequency of PVs in the texts. In Table 3 we present the more frequent PVs that our application detected. Note that 10 of our most frequent PVs are also included in the 25 more frequent PVs in documents of the European Union (Trebits, 2009).

## 5. CONCLUSIONS AND FUTURE WORK

We presented a classification for PVs and a specialized web application called Smart\_PV. Smart\_PV allows input and detection of PVs. The application was validated with samples of 40, 60, and 80 PVs, which included the 25 most common PVs in documents of the European Union according to Trebits (2009) and with the detection of PVs in texts from different domains. Although a database populated with more PVs and the analysis of more documents are required, our preliminary results show the feasibility and usefulness of our application.

A key aspect of our application is that we take into account the polysemic richness of the PVs, i.e., a VPC can correspond to several PVs. However; we believe that more work is needed in order to reduce ambiguity, for instance, because a PV may have several meanings, we would like to indicate which of these meanings is the most appropriate in the context of the text. We also considered several techniques in the design of the interfaces to facilitate the user's input and detection tasks.

As future work we consider doing the following:

a) adding a module of games specialized in PVs; b) developing an application that automatically extracts and populates the database from a specialized dictionary of PVs; c) adding multimedia components that complement and illustrate the use of each PV; d) supporting the PVs ambiguity, because the meaning of a PV can not be deduced separately from its components (verb and particle(s)). In Indranil, Ananthakrishnan and

Sasikumar (2004) an example-based technique for disambiguating English PVs is proposed. Another option is to consider ontologies (Staab & Studer, 2009). The idea is to develop ontologies for the verbs and for the particles, to discover possible relationships between these elements. Also, we use the Case Based Reasoning (CBR) (Sankar, David, & Kalyan, 2007) methodology in order to develop a PVs suggestion system. Our goal is that the application suggests PVCs to the user in those cases where exist PV in the text, but do not exist in our database. An expert user can then analyze whether it really is a PV and provide feedback on the application. The goal is to identify similar cases of accepted and rejected suggestions so that the suggestions of possible PVs become more accurate. In addition, these suggestions may be enriched with the derivation of potential PVs from the relationships between the elements of the ontologies (synonymy, antonymy, paronymy, equivalency, among others).

## REFERENCES

- Álvarez, G. (1990). ¡Partículas, partículas, partículas!: aproximación pedagógica a los verbos adverbiales y los verbos preposicionales ingleses. *Cauce Revista de Filología y su didáctica*, 5(13), 239-249.
- Baldwin, T. (2005). Deep lexical acquisition of verb-particle constructions. *Computer Speech and Language*, 19(1), 398-414.
- Capelle, B., Shtyrov, Y., & Pulvermuller, F. (2010). Heating up or cooling up the brain? MEG evidence that phrasal verbs are lexical units. *Brain and Language*, 115(3), 189-201.
- CartoWeb (2010). *CartoWeb documentation*. Retrieved from <http://www.cartoweb.org/doc/cw3.3/xhtml/user.tooltips.html>.
- Claridge, C. (2000). *Multi-word verbs in early Modern English. A corpus-based study. (Language and Computers 32)*. Atlanta, GA: Rodopi Bv Editions.
- Colin, B. (2005). Learning about the meaning of verb-particle constructions from corpora. *Computer Speech and Language*, 19(1), 467-478.
- Gries, T. (2001). A multifactorial analysis of syntactic variation: particle movement revisited. *Journal of Quantitative Linguistics*, 8(1), 33-50.
- Hart, C. (1999). *The ultimate phrasal verb book*. Hauppauge, NY: Barron's Educational Series. Retrieved from: [http://books.google.com.co/books?id=1XixeA7xHrcC&printsec=frontcover&hl=es&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](http://books.google.com.co/books?id=1XixeA7xHrcC&printsec=frontcover&hl=es&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false)
- Hill, S., & Bradford, W. (2000). *Bilingual grammar of English-Spanish syntax: a manual with exercises and key*. Maryland, MD: University Press of America.
- Hoque, M. (2008). An empirical framework for translating of phrasal verbs of English sentence into Bangla. *CUET Journal*, 2(8), 43-49.
- Indranil, S., Ananthakrishnan, R., & Sasikumar, M. (2004). *Example-based technique for disambiguating phrasal verbs in English to Hindi translation*. Mumbai, India: CDAC Mumbai.
- Kendris, T. (2008). *Complete English grammar review for Spanish speakers*. New York, NY: Barron's Educational Series.

- Lavin, E., & Sánchez, F. (1989). *Diccionario de verbos ingleses con partícula*. Madrid, Spain: French & European Pubs.
- McArthur, T. (1992). *The Oxford companion to the English language*. New York, NY: Oxford University Press.
- Sankar, K., David, W., & Kalyan, M. (2007). *Case-based reasoning in knowledge discovery and data mining*. Washington, DC: John Wiley & Sons.
- Scott, B., & Neil, T. (2009). *Designing web interfaces: Principles and patterns for rich interactions*. New York, NY: O'Reilly Media.
- Staab, S., & Studer, R. (2009). *Handbook on ontologies. International handbooks on information systems*. Berlin, Germany: Springer.
- Trebits, A. (2009). The most frequent phrasal verbs in English language EU documents – A corpus-based analysis and its implications. *SYSTEM*, 30(37), 470-548.

**How to reference this article:** Manrique, B., Moreno, F., & Orrego, G. (2012). Smart\_PV: A software application for managing English phrasal verbs. *Íkala, revista de lenguaje y cultura*, 17(3), 231-243.

## APPENDIX

**Table 1:** Comparison of proposals for PVs characterization

Author	Lavin + Our contribution	Álvarez (1990)	Baldwin (2005)	Claridge (2000)
<b>Name</b>	Phrasal verbs	Phrasal verbs	Verb-particle constructions	Multi-word verbs
<b>Role of the particle</b>	It defines the group of the verb, (adverbial or prepositional)	It focuses on particles that can play the adverbial and prepositional role at the same time	It defines the type of VPC, depending if the preposition is transitive	It considers prepositional and no prepositional particles
<b>Mobility of the particle</b>	It is applicable to the adverbial verbs and depends if the VPC allows mobility	It is used to differentiate between adverbial and prepositional verbs	It establishes two configurations: “joined” where the verb and particle are adjacent and the DO follow the particle and “split” where the DO occurs between the verb and the particle	It considers the type of the particle that can be transitive or intransitive
<b>Passive transformation</b>	A PV allows PT when the DO is included	It is not considered	It is not considered	It is not considered
<b>Relationship between DO and PO</b>	Sometimes the DO and the PO are the same	It is not considered	It is not considered	It is not considered
<b>Thematic transformation</b>	It is not considered	It distinguishes the subject and focus on the information contained in the sentence, where the subject is information already known and the focus is on the new information	It is not considered	It is not considered
<b>Use of other semantics categories that complement to the verb</b>	It is not considered	It is not considered	It is not considered	The author proposes classifications for the PVs, depending if they are made up by a verb and an adjective, a nominal element or other verb
<b>Other features</b>	Equivalent and optional particles, reflexivity, style, geographic context and nominal transformation	They are not considered	They are not considered	They are not considered
<b>Different classifications for the same VPC</b>	A same VPC, e.g. “go off” can generate several PV	It is not considered	It is not considered	It is not considered

Table 2: Detection process results

Indicator	Value		
<b>PVs database size sample</b>	<b>40</b>	<b>60</b>	<b>80</b>
Total number of PVs with one particle in the texts	30	30	30
Total number of PVs with two particles in the texts	10	10	10
Total number of detected PVs with one particle in the texts	21	23	24
Total number of detected PVs with two particles in the texts	5	5	5
Total number of VPCs with one particle in the texts	26	26	26
Total number of VPCs with two particles in the texts	10	10	10
Total number of detected VPCs with one particle in the texts	19	20	20
Total number of detected VPCs with two particles in the texts	5	5	5
Total number of detected VPCs with one particle in the texts that were not PVs	5	4	2
Total number of detected VPCs with two particles in the texts that were not PVs	1	1	2
Search efficiency	65%	70%	72%
Search deficiency	35%	30%	28%
Search optimization	83%	87%	90%

Table 3: More frequent PVs

Ranking	PV	Number of occurrences	% with regard to the total of detected PV
1	Set up	50	13%
2	Make up	45	12%
3	Find out	38	10%
4	Get on	36	9%
5	Take up	31	8%
6	Look for	26	7%
7	End up	25	6%
8	Work on	20	5%
9	Base on	13	3%
10	Back up	12	3%
11	Break in	8	2%
12	Call on	7	2%
13	Get around	5	1%
14	Carry out	5	1%
15	Hold on	4	1%

Figure 1: Flowchart to validate PVs existence

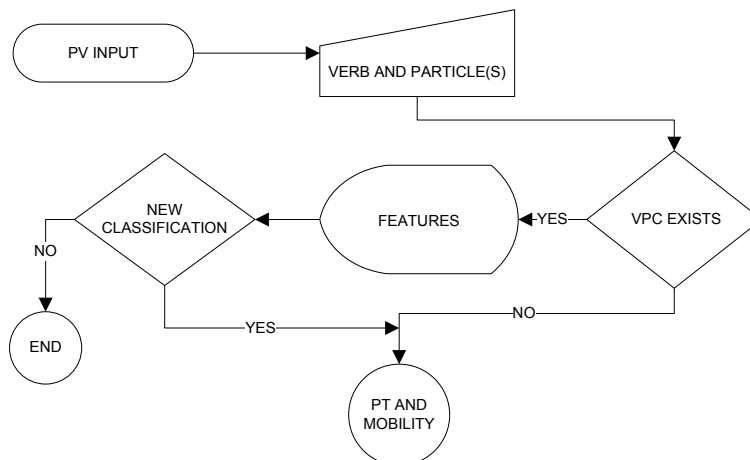


Figure 2: Flowchart to input PT and mobility of the particle

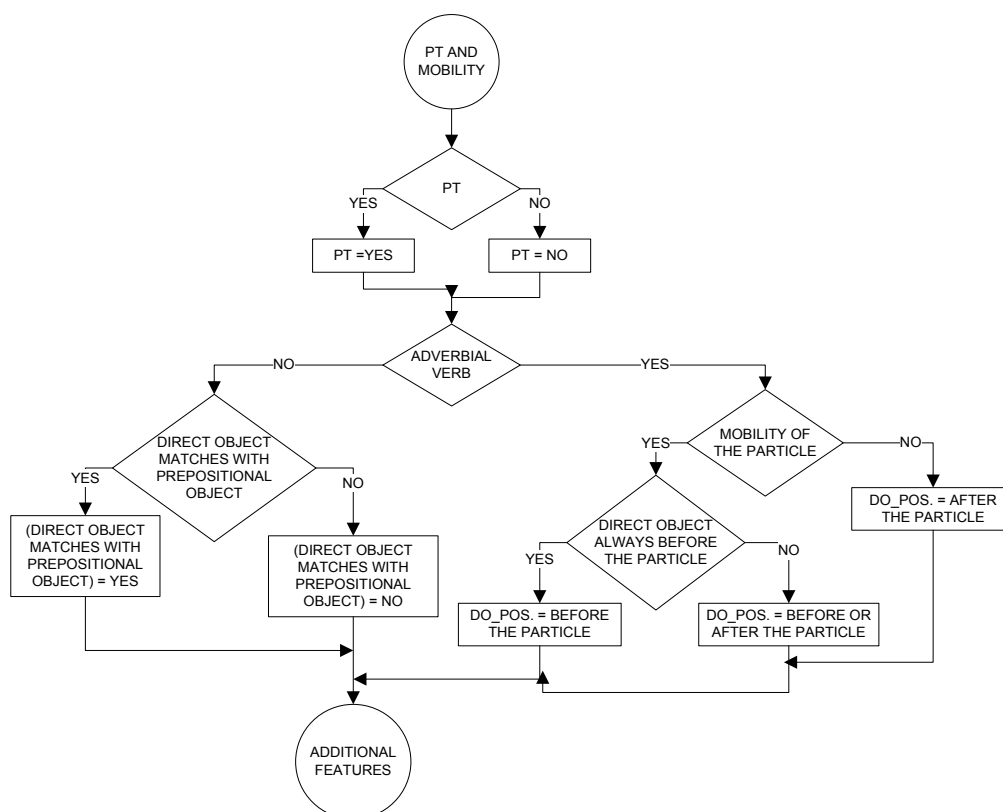
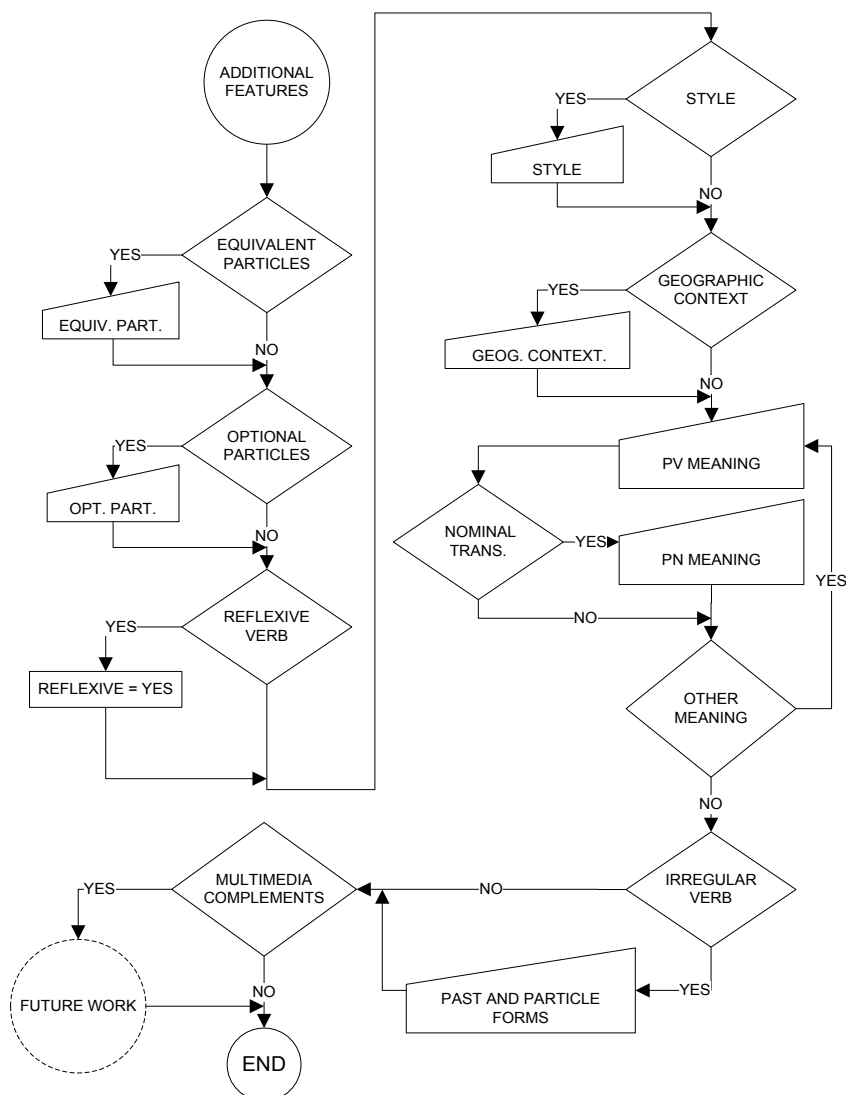
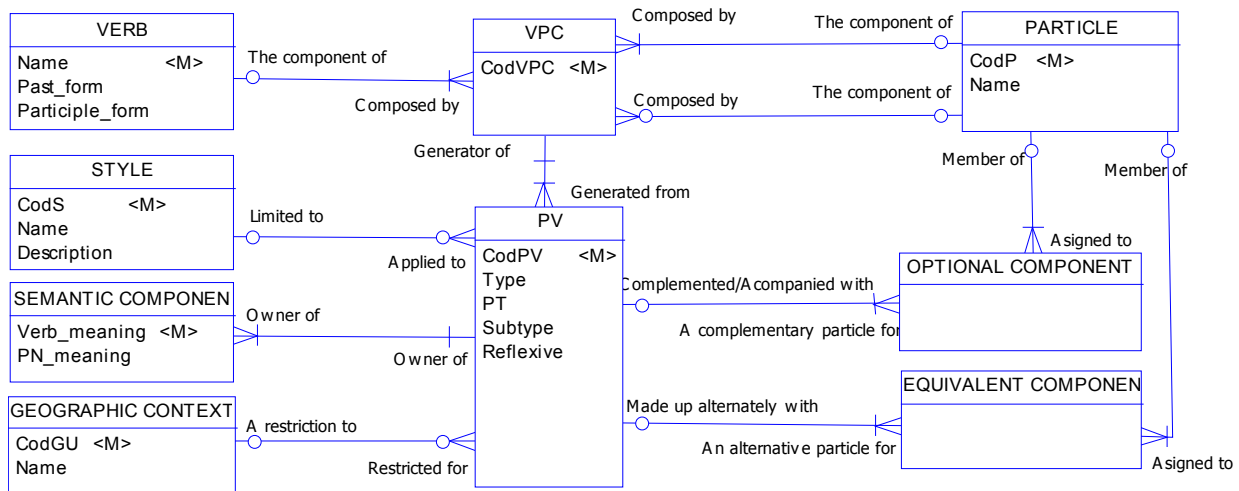


Figure 3: Flowchart to input other PVs features



**Figure 4:** Database schema for the PVs application



**Figure 5:** Detection Process

