

A formal scope on the relations between definitions and verbal predications

César Aguilar

Facultad de Lenguas y Letras
Universidad Autónoma de Querétaro
Centro Universitario, s/n C.P. 76010
Querétaro, México
Caguilar@iingen.unam.mx

Gerardo Sierra

Instituto de Ingeniería
Universidad Nacional Autónoma de México
Cubículo 3, Basamento, Torre de Ingeniería
C.U., C.P. 04510, México D.F.
GsierraM@iingen.unam.mx

Abstract

This paper outlines a formal description of grammatical relations between definitions and verbal predications found in Definitional Contexts in Spanish. It can be situated within the framework of Predication Theory, a model derived from Government & Binding Grammar. We use this model to describe: (i) the syntactic patterns that establish the relationship between definitions and predications; (ii) how useful these patterns are for the identification of definitions in technical corpora.

Keywords

Definition Extraction, Types of Definitions, Predication, Predicative Phrase.

1. Introduction

The (semi-)automatic recognition of terms and definitions in a corpus is an important task to research areas such as computational lexicography, terminology, language engineering and others. In the case of term recognition, several works report successful methodologies, computational tools and experiments that aim to identify and extract, in a no-supervised way, term candidates from large specialized corpora (e. g. Cabré, Estopà & Vivaldi 2001).

However, the automatic recognition of definitions presents a much higher degree of complexity, since definitions are linguistic structures used to formulate concepts (Sager, 1990). In contrast to terms, which are considered language units whose function is to refer specific entities in a scientific or technical knowledge domain, definitions condense information and establish several conceptual relations, with the purpose to delimitate the essential properties or attributes that characterize an entity in relation to others.

There are currently many authors that have proposed different methodologies for identifying candidates to

definitions, considering both linguistic and statistical points of views. Some relevant methodologies are:

- Definitional Sentences (fr. *énonces définitoires*): Auger (1997), Rebeyrolle (2000).
- Terms in Contexts: Pearson (1998).
- Knowledge-Rich Contexts: Meyer (2001).
- Mining Definitions on Texts: Malaisé, Zweigenbaum & Bachimont (2005).

In accordance with these methodologies, in this paper we present a methodology to identify different types of definitions in technical corpora, considering that these definitions are configured as grammatical patterns, in particular, as phrase structures. These patterns are linked to verbal predications with syntactic regularities.

For the syntactic analysis of these patterns, we use a formal model called Predication Theory (henceforth, PredT). This model is formulated within the framework of Government & Binding Grammar (Rothstein, 1983; Bowers 1993, 2001). So, the PredT allows us to describe, in a formal way, the grammatical relations that definitions establish with verbal predications. Taking this relationship into account, it is possible to identify good candidates to definitions considering their association with verbal predications, specifically when these definitions are introduced in scientific and technical texts.

2. Definitional Contexts

We situate this analysis within the framework of Definitional Contexts (or DCs) extraction. According to Sierra *et al.* (2008), a DC is a discursive structure that contains relevant information to define a term. A DC has at least two constituents: a term and a definition, and usually linguistic or metalinguistic forms, such as verbal phrases, typographical markers and/or pragmatic patterns. An example is:

1. In general, the **paraprofessional workers** are defined as those persons who are engaged in the provision of social care or social services, but who do not have professional training or qualifications.

According to this example, the term *Paraprofessional workers* is emphasised by the use of bold font; the verbal predication *are defined as* links the term *paraprofessional workers* to the actual definition *those persons who are engaged...* The term, the verbal predication and the definition are discursive units introduced by the pragmatic pattern *In general*. These are the three units that constitute the main syntactic sequence of a DC.

In this work we study this kind of DCs in Spanish, where the association between definitions and verbal predications is made explicit.

3. A formal description respect to predication

Taking into consideration that these sequences are composed of a verbal predication and definitions, several authors have found and reported such sequences for English (Pearson, 1998; Meyer 2001; Malaisé, Zweigenbaum & Bachimont, 2005) and French (Auger 1997, Rebeyrolle 2000). These authors have considered the use of these types of verbal predications as useful patterns for the (semi)automatic extraction of information associated to definitions.

However, none of these authors have analysed the nature of the relations between predications and definitions. In this paper, we focus on the description of their nature at the syntactic level, based on the PredT as a pertinent formal model for explaining the relations established between verbal predications and definitions.

3.1 Predication theory in GB grammar

Grosso modo, PredT is a model derived from Government & Binding Grammar, formulated by Chomsky (1981). PredT postulates that all predications indicate a semantic relationship between an entity and a particular property or characteristic feature. Syntactically, PredT explains all verbal predications as a type of phrase, structured around a relation *X-is-a-Subject-of/Y-is-a-predicate-of*. This relation is regulated by a syntactic rule named *rule of predicate linking*, proposed by Rothstein (1983). Examples of these relations are:

2. a. John is an intelligent professor.
- b. John considers his father as an intelligent professor.

Following Rothstein's explanation, Bowers (1993, 2001) develops a simple model to describe the syntactic configuration of these phrases, called Predicative Phrase (PrP). The PrP is mapped by a non-lexical head (that is, a functional head), and its grammatical behaviour is similar to that of phrases such as Inflexional Phrase (IP) or Complement Phrase (CP). A graphical tree representation of a PrP is:

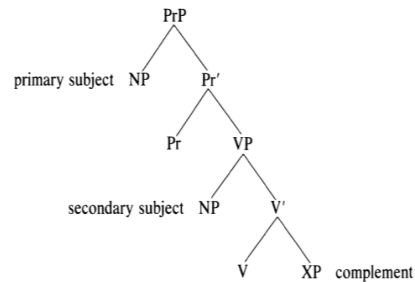


Figure 1: Tree representation for PrP, according to Bowers (1993: 596).

Figure 1 shows the basic elements that make up a PrP. Bowers recognise a functional head with the features $+/-$ predicative (Pr). This head maps two Subjects, a primary subject in the position of Specifier of PrP (represented by a Noun Phrase or NP); and a secondary subject, in the position of Specifier of Verbal Phrase or VP (often a NP). Finally, both subjects, the VP and the PrP are linked to one or several complements, which assume many phrasal representations (e.g. NP, IP, CP, and so on).

3.2 Primary and secondary predications

Based on this distinction between primary and secondary subjects, it is possible to recognise two types of predications:

- Simple or primary predication, consisting of a subject to the left of the verb (in position of Specifier of PrP), and a predicate to the right of the verb. An example in Spanish is:
3. [Una computadora [es [un tipo de máquina electrónica que sirve para hacer operaciones_{PrP}] VP] IP] (Eng. [A computer [is [a kind of electronic machine used to make operations_{PrP}] VP] IP]).
 - Double or secondary predication, which integrates a primary subject in a pre-verbal position, a secondary subject (situated as Specifier of VP), and the predicate. For example, again, in Spanish:
 4. [Turing [define una computadora [como un mecanismo electrónico que procesa conjuntos de datos_{PrP}] VP] IP] (Eng. [Turing [defines a computer [as a kind of electronic device that processes a set of data_{PrP}] VP] IP]).

In (4), the predicate *como un mecanismo electrónico...* (Engl. *as a kind of electronic device...*) affects the secondary subject *una computadora* (Engl. *a computer*), in accordance with the explanation provided by Bowers (1993). For our analysis, we consider both types of predications as regular patterns that syntactically codify sequences of terms, verbal predications and definitions.

4. Combinatory of patterns in DCs

Based on our formal description of PrP, it is possible to identify two types of patterns that structure particular sequences in DCs:

- In the case of primary predications, it codifies a sequence composed of a Term, a Verbal Predication and a Definition.
- In the case of secondary predication, it codifies a sequence composed of a specific Author, a Term, a Verbal Predication and a Definition.

4.1. Term + Verbal Predication + Definition

This sequence is a good example of a formulation of canonical definitional patterns, because the primary predication links directly a subject represented by a term, with a specific set of attributes codified in the PrP. These patterns are shown in the following sequences:

- a. [El contenedor refrigerado _{Term}] [es _{Verbal Predication}] [una forma especializada de transporte de perecederos _{Definition}] (Eng. [The refrigerated container _{Term}] [is _{Verbal Predication}] [a specialized form to transport perishable goods _{Definition}])
- b. [Un esquema XML _{Term}] [representa _{Verbal Predication}] [el significado y la estructura de la información recibida desde una aplicación _{Definition}] (Eng. [An XML schema _{Term}] [represents _{Verbal Predication}] [the meaning and structure of the information received from an application _{Definition}]).
- c. [Una jerarquía de dependencias _{Term}] [se refiere a _{Verbal Predication}] [todas las tablas que incluyen referencias mutuas _{Definition}] (Eng. [A hierarchy of units _{Term}] [refers to _{Verbal Predication}] [all tables that include references to each other _{Definition}]).

The sequence Term + Verbal Predication + Definition in cases 5 a-c is equivalent to the structure of primary predication. Therefore, the Term is situated in the position of Primary Subject, the Verbal Predication has the role of head of a VP, and the Definition is introduced through a PrP.

4.2. Author + Term + Verbal Predication + Definition

The second sequence we report here shows the sequence Author + Term + Verbal Predication + Definition. The characteristic feature of this pattern is that it explicitly points out the author (or authors) of the definition. This feature maps a semantic role, according to FrameNet (Baker, Fillmore and Lowe, 1998), concretely the author can be conceived as a Cognizer that associates certain Categories (the Definition) to a particular Item (that is, the Term). This is illustrated in the following examples:

- a. [Carlos Godino _{Author}] [define _{Verbal Predication}] [la arquitectura naval _{Term}] [como la ciencia que se enfoca en la construcción de los buques _{Definition}] (Eng. [Carlos Godino _{Author}] [defines _{Verbal Predication}] [naval architecture _{Term}] [as the science that focuses on the construction of ships _{Definition}])

- b. [El artículo _{Author}] [describe _{Verbal Predication}] [la evolución de ecología del paisaje _{Term}] [como una ciencia integrativa y transdisciplinaria _{Definition}] (Eng. [The article _{Author}] [describes _{Verbal Predication}] [the evolution of landscape ecology _{Term}] [as an integrative and interdisciplinary science _{Definition}]).
- c. [Ø Podemos _{Author}] [considerar _{Verbal Predication}] [las computadoras programables modernas _{Term}] [como la evolución de sistemas antiguos de cálculo o de ordenación _{Definition}] (Eng. [We _{Author}] [can consider _{Verbal Predication}] [the modern programmable computers _{Term}] [as the evolution of ancient systems of calculation and management _{Definition}]).

Hence, the pattern followed by this sequence clearly refers to the author of a definition, as shown in 6 a-c. However, a syntactic behaviour observed in this pattern is its recurrent configuration in non-personal forms, i.e. impersonal and passive forms, for example:

- a. [Se conoce como _{Verbal Predication}] [reenganche rápido _{Term}] [a la operación de cierre de un interruptor después de una falla _{Definition}] (Eng. [It is known as _{Verbal Predication}] [Quick Re-closing _{Term}] [to the operation of a switch after a fault _{Definition}]).
- b. [Los niveles relativos de los alcances de ola _{Term}] [fueron descritos como _{Verbal Predication}] [una función del parámetro de similitud de oleaje _{Definition}] (Eng. [The relative levels of the wave reach _{Term}] [were described as _{Verbal Predication}] [a function of the wave similarity parameter _{Definition}]).

In these examples, we observe the use of non-personal verbal patterns as in (7a), where the clitic *Se* is inserted (Eng. *It*) to make the sentence impersonal, or in (7b), which is in the passive form. So, when these sequences assume a non-personal pattern, they become equivalent to primary predications, where there is not an explicit mention of the author of a definition.

5. Types of definitions linked to predications

Another aspect that we found in the relation between predications and definitions is the influence of the predication on the selection of a particular type of definition. In fact, this influence is important because we can establish and formalise a possible grammar model that helps to identify different kinds of definitions, given a primary or secondary predication.

Following Sierra *et al.* (2008) and Aguilar (2009), we outline a typology with 4 types of definitions: analytical, synonymical, functional and extensional. These definitions are derived from Aristotle's model:

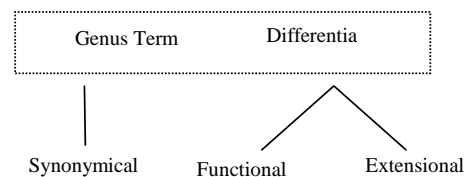


Figure 2. Typology of definitions based on Aristotle's model (Sierra *et al.* 2008: 81).

Figure 2 illustrates how these four types of definitions can be identified according to the presence/absence of Genus Term and/or Differentia in a good candidate of definition. So, when both the Genus Term and the Differentia are explicit, we have an analytical definition, which can be associated with two kinds of predication: primary predication or secondary predication.

When only the Genus Term is explicit, there is a certain relation of conceptual equivalence between the term and its definition. So, following Cruse (1986), we characterise this definition as synonymical.

In contrast, when only the Differentia is explicit, we have two options:

- The first one describes the use or the function of an entity represented by the term, that is, a functional definition.
- The second one enumerates all the components of a possible entity or a possible set, that is, an extensional definition.

In the following sections we briefly describe each type of definition.

5.1. Analytical definitions

This definition occurs associated with primary and secondary predications. In the case of primary predication, the analytical definition is integrated in a sequence Term + Verbal Predication + Definition. This definition does not explicitly state the author of a definition. For example:

8. [El apartarrayos_{Term}] [es_{Verbal Predication}] [un dispositivo_{Genus Term}] [que protege las instalaciones contra sobretensiones de origen atmosférico_{Differentia}] (Engl. [The lightning conductor_{Term}] [is_{Verbal Predication}] [a device_{Genus Term}] [that protects electrical systems against surges of atmospheric origin_{Differentia}]).

We propose a possible grammatical description model for this relation:

Definition	Genus Term	Differentia
Analytical (Simple Predication)	Noun Phrase = Noun + {Adjective Phrase/Prepositional Phrase}*	Complement Phrase = Relative Pronoun + Inflectional Phrase
		Prepositional Phrase = Preposition + Noun Phrase
		Adjective Phrase = Adjective + Noun Phrase

Table 1. Construction patterns derived from the relation between primary predication and analytical definition

In the case of secondary predications linked to analytical definitions, they follow the sequence Author + Term + Verbal Predication + Definition, where the Author is equivalent to the primary subject, the Term assumes the position of secondary subject, and the definition is introduced after the Verbal Predication. In this case, the adverbial particle *como* (Eng. *as/like*), or the preposition *por* (Eng. *for/by*) indicate the place of the definition:

Definition	Adverb/Preposition	Genus Term	Differentia
Analytical (Secondary Predication)	<i>Como Por</i>	Noun Phrase = Noun + {Adjective Phrase/Prepositional Phrase}*	Complement Phrase = Relative Pronoun + Inflectional Phrase
			Prepositional Phrase = Preposition + Noun Phrase
			Adjective Phrase = Adjective + Noun Phrase

Table 2. Construction patterns derived from the relation between secondary predication and analytical definition

5.2. Synonymous definitions

The synonymous definitions have a syntactic relation with primary predications, specifically with the Genus Term, but not with the differentia. An example is:

9. [La tensión de base_{Term}] [se le llama también_{Verbal Predication}] [tensión unidad_{Genus Term}]. (Engl. [The base tension_{Term}] [it is also called_{Verbal Predication}] [unit tension_{Genus Term}]).

In (9), we observe that the Term *la tensión de base* (Engl. *the base tension*) establishes a relation of cognitive equivalent with the Genus Term *tensión unidad* (Engl. *unit tension*). We formalise this relation in table 3:

Definition	Term	Genus Term
Synonymical (Primary Predication)	Noun Phrase = Noun + {Adjective Phrase/Prepositional Phrase}*	Noun Phrase = Noun + {Adjective Phrase/Prepositional Phrase}*

Table 3. Construction patterns derived from the relation between primary predication and synonymous definition

5.3. Functional definitions

The functional verbal pattern introduces a type of definition where the Genus Term is absent, but introduces a Differentia that describes the function or the use of a particular entity. The verbal pattern is also associated with a primary predication. The example is:

10. [La técnica de velocimetría de imágenes_{Term}] [permite_{Verbal Predication}] [medir la velocidad de un campo de flujo bi o tri dimensional_{Differentia}] (Engl. [The method of image velocimetry_{Term}] [allows_{Verbal Predication}] [to measure the speed of a flow field in two or three dimensions_{Differentia}]).

The formal description of this relation between predication and definition is:

Definition	Differentia
Functional (Primary Predication)	Infinitive Verb + Complement Phrase = Relative Pronoun + Inflexional Phrase + {Prepositional Phrase/Adjective Phrase/Adverbial Phrase/Complement Phrase}* *
	Infinitive Verb + Preposition + {Inflexional Phrase/Complement Phrase}* *
	Prepositional Phrase = Preposición + Noun Phrase + {Prepositional Phrase/Adjective Phrase/Adverbial Phrase/Complement Phrase}* *
	Noun Phrase = Noun + {Prepositional Phrase/Adjective Phrase/Adverbial Phrase/Complement Phrase}* *

Table 4. Construction patterns derived from the relation between primary predication and functional definition

5.4. Extensional definitions

Finally, extensional definitions provide a complete list of the parts, components or elements of an entity or set. In a similar way to functional definitions, extensional definitions are structured around a primary predication. An example is:

11. [La zona límite_{Term}] [incluye_{Verbal Predication}] [planicies costeras, marismas, áreas de inundación, playas, dunas y corales_{Differentia}] (Eng. [The border zone_{Term}] [includes_{Verbal Predication}] [coastal plains, salt marshes, flood areas, beaches, dunes and corals_{Differentia}]).

Our syntactic description of this pattern is:

Definition	Preposition	Differentia
Extensional (Primary Predication)	Con (Eng With) De (Eng. Of)	Noun Phrase = Noun + {Adjective Phrase/Prepositional Phrase}* *

Table 5. Construction patterns derived from the relation between primary predication and extensional definition

We can summarise all these patterns in table 6, considering some recurrent verbs in the position of head of PrP. These verbs are not exclusive, but their recurrence has been reported by Sierra *et al.* (2008), and Aguilar (2009):

Definition	Verbs	Associated Particles
Analytical (Primary Predication)	<i>referir</i> (to refer to) <i>representar</i> (to represent) <i>ser</i> (to be) <i>significar</i> (to signify/to mean)	<i>a</i> = to (preposition) (in the case of <i>referir</i> , it is a phrasal verb that inserts obligatory the preposition <i>a</i>)
Analytical (Secondary Predication)	<i>caracterizar</i> (to characterise) <i>comprender</i> (to include) <i>concebir</i> (to conceive) <i>conocer</i> (to know) <i>considerar</i> (to consider) <i>definir</i> (to define) <i>describir</i> (to describe) <i>entender</i> (to understand) <i>identificar</i> (to identify) <i>visualizar</i> (to visualise)	<i>como</i> = as/like (adverb) <i>por</i> = for/by (preposition)
Synonymy	<i>equivaler</i> (to be equivalent to) <i>llamar</i> (to call) <i>nombrar</i> (to name) <i>ser _ igual</i> (to be equal to) <i>ser _ similar</i> (to be similar to)	<i>también</i> = also (adverb) <i>a</i> = to (preposition) <i>igual a</i> = equal to (adverb phrase) <i>similar a</i> = similar to (adverb phrase)
Functional (Primary Predication)	<i>emplearse</i> (to employ + clicit "se") <i>encargar</i> (to be in charge of) <i>funcionar</i> (to function) <i>ocupar</i> (to occupy) <i>permitir</i> (to allow) <i>servir</i> (to serve) <i>usar</i> (to use) <i>utilizar</i> (to utilise / to use)	<i>de</i> = of (preposition) <i>para</i> = for (preposition)
Extensional (Primary Predication)	<i>componer</i> (to be composed of) <i>comprender</i> (to include) <i>consistir</i> (to consist of) <i>constar</i> (to consist of) <i>contar</i> (to have) <i>constituir</i> (to constitute) <i>contener</i> (to contain) <i>incluir</i> (to include) <i>integrar</i> (to integrate)	<i>de</i> = of (preposition) <i>por</i> = for/by (preposition) <i>con</i> = with (preposition)

Table 6. Verbs associated with definitions

6. Commentaries and conclusions

In this paper, we have outlined a formal description of the grammatical relations that can be established between definitions and verbal predications in DCs. We consider this is a pertinent way to analyse the syntactic behaviour of definitions in specialised texts, specifically when these definitions are linked to verbal predications.

We have described these verbal predications according to the PredT, a grammatical model useful to formalise patterns generated by the association of verbal predications to specific definitions. This description allowed us to distinguish:

- Two types of verbal predications: primary and secondary predications. Both predications entail particular types of definitions, depending on the verb that functions as the head of the predication.
- These predications play an important role in the selection and introduction of specific types of

definitions. In this paper, we have proposed a possible typology of definitions, based on the role played by predications. This typology considers four types of definitions: analytical, synonymical, functional and extensional.

- In addition, it is possible to observe that the relation established between the types of definitions with primary/secondary predications configure two sequences that structure two different kinds of DCs: (i) a sequence, Term + Verbal Predication + Definition, configured in primary predications which can be linked to analytical, synonymical, functional and extensional definitions and; (ii) another sequence, Author + Term + Verbal Predication + Definition, delineated by secondary predications which can be associated to secondary predications.

We think that the use of these patterns proposed in our analysis can sketch a useful grammatical model, applied to the task of (semi)automatic recognition and extraction of definitions in Spanish, from text corpora.

7. Acknowledgements

This paper was made possible by the financial support of the Consejo Nacional de Ciencia y Tecnología, CONACYT, and DGAPA-UNAM.

8. References

- [1] C. Aguilar. Análisis lingüístico de definiciones en contextos definitorios. Ph. D. Thesis, National Autonomous University of Mexico, Mexico City, 2009.
- [2] A. Auger. Repérage des énonces d'intérêt définitoire dans les bases de données textuelles. Thèse de Doctorat, Neuchâtel, Université de Neuchâtel, 1997.
- [3] C. Baker, Ch. Fillmore, & J. Lowe. The Berkeley FrameNet Project. In Proceedings of the COLING-ACL, Montreal, Canada, 1998.
- [4] J. Bowers. The Syntax of Predication. *Linguistic Inquiry*, 24(4): 591-636, 1993.
- [5] J. Bowers. Predication. In Baltin, M. & C. Collins (eds.). *The Handbook of Contemporary Syntactic Theory*. Oxford, Blackwell: 299-333, 2001.
- [6] T. Cabré, R. Estopà & J. Vivaldi. Automatic term detection. In Bourigault D., C. Jaquemin & M.C. L'Homme (eds.). *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia, John Benjamins: 53-87, 2001.
- [7] N. Chomsky. *Lectures on Government and Binding*. The Hague, Mouton de Gruyter, 1981.
- [8] D. Cruse. *Lexical Semantics*. Cambridge, Cambridge University Press, 1986.
- [9] V. Malaisé, P. Zweigenbaum & B. Bachimont. Mining defining contexts to help structuring differential ontologies. *Terminology* 11(1): 21-53, 2005.
- [10] I. Meyer. Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In Bourigault D., C. Jaquemin & M.C. L'Homme (eds.). *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia, John Benjamins: 279-302, 2001.
- [11] J. Pearson. *Terms in Contexts*. Amsterdam/Philadelphia, John Benjamins, 1998.
- [12] J. Rebeyrolle. *Forme et fonction de la définition en discours*, Thèse de Doctorat. Toulouse, Université Toulouse-Le Mirail, 2000.
- [13] S. Rothstein. *The Syntax Forms of Predication*, Ph. D. Thesis. Cambridge, Mass., MIT, 1983.
- [14] J.C. Sager. *Essays on Definitions*, Amsterdam/Philadelphia, John Benjamins., 1990.
- [15] G.Sierra, R. Alarcon, C. Aguilar & C. Bach. Definitional Verbal Patterns for Semantic Relation Extraction. In Auger A. y C. Barrière (eds.). *Pattern-based Approaches to Semantic Relation Extraction*. Special issue of *Terminology*, 14(1): 74-98, 2008.