UNIVERSITAT ROVIRA I VIRGILI

A Formal Characterization of Fuzzy Degrees of Grammaticality for Natural Language

Adrià Torrens Urrutia

Adrià Torrens Urrutia

A Formal Characterization of Fuzzy Degrees of Grammaticality for Natural Language

DOCTORAL THESIS

Supervised by Dra. María Dolores Jiménez López
Departament de Filologies Romàniques

UNIVERSITAT ROVIRA i VIRGILI

Tarragona
February 2019

.

UNIVERSITAT
ROVIRA I VIRGILI

DEPARTAMENT DE FILOLOGIES ROMÀNIQUES

Av. Catalunya, 35 (Campus Catalunya)
43002 Tarragona
Tel. +34 977 559 555
Fax +34 977 558 386
E-mail: secdfr@urv.cat
http://pizarro.fll.urv.es

I STATE that the present study, entitled "A Formal Characterization of Fuzzy Degrees of Grammaticality for Natural Language", presented by Adrià Torrens Urrutia for the award of the degree of International Doctor, has been carried out under my supervision at the *Departament de Filologies Romàniques* of *Universitat Rovira i Virgili*, and that it fulfills all the requirements for the award of International Doctor.

Tarragona, 26 February 2019

Doctoral Thesis Supervisor

Dra. M. Dolores Jiménez López

Acknowledgements

For this PhD dissertation, I would firstly like to acknowledge María Dolores Jiménez López. I couldn't have had a better partner for this journey. She has challenged me continuously during our discussions. She has always been very supportive, and I will remember her kindness forever. She has been the best professional and mentor to me. I dedicate this PhD to her.

I also want to acknowledge Prof. Philippe Blache and to Prof. Vilem Novák for being supportive, kind and dedicated.

In my first and second stay, I had the honor of collaborating with Prof. Blache in his lab *Laboratoire Parole et Langage* in Aix-en-Provence. Our discussions have been vital for this PhD I really appreciate all the time that he spent with me and his involvement. This PhD would never be this way without him. He was very sharp when we were discussing different linguistic regards. He was very kind, and I am thankful for that. I hope we can collaborate again in the future on more applications with linguistic constraints.

In my third stay, I had the honor to collaborate with Prof. Novák. Regardless of my inexperience in any kind of fuzzy logic or mathematical formalism, he accepted me at his research institute *Institute for Research and Applications of Fuzzy Modeling* in Ostrava. He was very patient and utterly involved in giving me lectures about fuzzy logic. We discussed intensely how to apply fuzzy logic into linguistics. This PhD would never be as it is without his help. He was very kind, and I am thankful for that. I hope we can collaborate again in the future to model more linguistic aspects with fuzzy logic.

I also want to thank both the PhD program in Humanistic Studies from *Universitat Rovira i Virgili* for their economic support and to the *Departament de Filologies Romàniques* from the same university for their funding and for allowing the use of their facilities during my PhD.

I want to thank my colleagues, with whom I have shared thoughts, problems, and happiness. Colleagues from my university, *Universitat Rovira i Virgili*, from my stay in Aix-en-Provence, *Laboratoire Parole et Langage*, and from my stay in Ostrava, *Institute for Research and Applications of Fuzzy Modeling*. I want to give special thanks to Maria Bargalló, Natalia Català, M. José Rodríguez Campillo, Xue Li, Rebeca Ramírez, Toni Brosa, Grégoire Montcheuil, James Sneed German, Elisa Sneed German, Emmanuele Chersoni, Adam Wilson, Jorane Salsifis, Elora Ribiéro, and Nhung Cao.

Regarding my personal life, I want to thank my parents, who always believed in me and who have been very supportive in the decision of embarking on a PhD degree. I would like to thank my closest friends with whom I have shared victories, preoccupations and a lot of craziness concerning this PhD Jordi N., Joan B., Jordi S., Cristian M., Joel P.

I also want to thank Rachel, who has been both supporting and loving.

Abstract

This dissertation presents a new system which formally characterizes the fuzzy degrees of grammaticality in natural language. It proposes the basis of a fuzzy grammar, its features, fuzzy and evaluative formulas to calculate the values of grammaticality, and it provides a proof-of-concept of a Fuzzy Property Grammar for the syntax of Spanish.

The notion of grammaticality is essential to the process of defining a natural language grammar. Linguists need to distinguish between what is in the grammar (grammatical) and what is not (non-grammatical). The categorical view of grammaticality is widely defended in theoretical linguistics. However, deviations from the norm are inherent to the spontaneous use of language. Hence, either a natural language grammar or linguistic analysis tools need to account for different levels of grammaticality. In linguistics, degrees of grammaticality are well accepted by certain authors (Aarts, 2004a; Bolinger, 1961; Keller, 2000; Lakoff, 1973; Ross, 1972). These authors mainly represent the degrees of grammaticality under the notion of grammaticality judgments and linguistic gradience (Aarts et al., 2004). The most well-known theories that take into account linguistic gradience are Optimality Theory and its variations. Keller (1998, 2000, 2006) and his Linear Optimality Theory showed how grammaticality and acceptability judgments are a gradient object. However, these theories together mostly evaluate the degree of acceptability of an input, since they represent the speaker's judgments by a formal approach such as Optimality Theory. Linguistics still lacks a grammar framework which can deal with degrees of grammaticality regarding linguistic competence.

This dissertation both defends and states that degrees of grammaticality can be both a matter in terms of linguistic competence and performance. This dissertation focuses on reviewing and providing a critical analysis concerning the concept of grammaticality in linguistics, and on proposing a fuzzy grammar for representing degrees of grammaticality in natural language.

We recognize a grammar as a system which both produces and recognizes inputs. We also claim that a grammar is a compilation of linguistic constraints which defines the linguistic competence of a speaker. Any input that violates the constraints of a grammar will be associated with a degree of grammaticality of that specific grammar. Therefore, degrees

of grammaticality are understood as a vague object that can take into account degrees of membership regarding a linguistic input.

In order to explain this, a new system is proposed that combines Property Grammars from Blache (2016) and the Fuzzy Natural Logic tools from Novák (2015). This new approach allows us 1) to provide a fuzzy grammar for dealing with degrees of grammaticality; 2) to create a framework for defining borderline constraints and categories regarding a grammar; and 3) to define grammaticality in terms of linguistic competence (not performance).

The extraction of the fuzzy properties (our fuzzy linguistic constraints) is based on: new constraints behavior proposed in this dissertation, the features of a fuzzy grammar, the combination of theoretical reasons and frequencies using the Spanish Universal Dependencies Treebank and the MarsaGram tool.

Many future benefits may come from this new approach. The first advantage is theoretical. Our proposal can define a grammar taking into account both a formal method, which represents objects in terms of degrees; and a grammar with constraints, which can define any linguistic input. Thus, this combination is useful for representing the concept of the degrees of grammaticality in a grammar with a gradient approach. The second advantage is related to language technologies and computational applications for users. This approach might improve human-machine interfaces since the machine would be able to process any input in terms of degrees. It would classify any linguistic input in a scale of degrees of grammaticality. In consequence, this could have an impact on the development of more flexible computational tools that facilitate our interaction with machines.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

1.1 Objectives

This work presents a new approach which combines different interdisciplinary methods to explain fuzziness in a natural language grammar. Since this topic is vast and complex, this work has focused its efforts on dealing with only one fuzzy (or vague) linguistic phenomenon: the *degrees of grammaticality*. Consequently, in the present dissertation fuzziness in natural language refers to one of the most well-known gradient phenomena in linguistics. The hardest part here is explaining the degrees of grammaticality regarding a natural language grammar (an objective and formal perspective) without involving degrees of acceptability (a subjective and psycho-linguistic perspective) in the measurement of grammaticality. To achieve this objective and to deal with this fuzziness, this new approach has drawn on different disciplines and points of view such as discrete linguistics, gradience linguistics, grammars with constraints, fuzzy logic reasoning, fuzzy grammars, vagueness in natural language, etc. Consequently, this dissertation deals with different sensitive topics in science, particularly in linguistics.

Given that our primary aim is to study grammaticality from a fuzzy (and not discrete) point of view, we start with three research questions:

1. Is it possible to measure the degrees of grammaticality of a given linguistic input?

2. Which are the best formal tools to calculate the different levels of grammaticality?

3. Can we provide a fuzzy approach to the concept of grammaticality (not acceptability) that takes into account linguistic competence (but not performance)?

The hypothesis that emerges from these research questions is the following:

- A formal model which combines fuzzy logic and a grammar with constraints can represent fuzzy degrees of grammaticality regarding linguistic competence.

In order to test this hypothesis, we will develop a formal model to deal with the degrees of grammaticality.

Therefore, the main aim of this work is to introduce a new linguistic model which can both represent and calculate degrees of grammaticality by considering the notion of grammaticality as a fuzzy phenomenon of natural language.

In order to achieve our main aim, we establish the following research objectives:

1. To review and provide a critical analysis of the concept of grammaticality in linguistics.

2. To review and provide a critical analysis of the models of gradience regarding their concept of grammaticality.

3. To set forth the basis for a definition of a fuzzy grammar to deal with degrees of grammaticality and its fuzzy features.

4. To extract/determine the linguistic properties that will define the linguistic inputs for taking into account degrees of grammaticality.

5. To provide a proof-of-concept of a fuzzy grammar with properties for Spanish syntax which can represent the fuzzy degrees of grammaticality.

Additionally, this work also aims to make the following contributions to the topic:

- To deal with the theoretical fragmentation between competence and performance in linguistics. A new point of view has been proposed regarding the classic description of Chomsky's work.

- To provide the theoretical bases that give sense to creating a grammar in terms of degrees.

- To explain the common points that linguistic gradience and fuzzy logic share.

- To propose a new linguistic model to deal with degrees of grammaticality based on real data from real speakers.

This proposal aims to increase collaboration between the disciplines of fuzzy logic and linguistics. We believe that many language technologies, such as translators, electronic proof-readers, or human-machine interfaces, would greatly benefit from this partnership.

We believe that our work offers a satisfactory (and alternative) solution to account for degrees of grammaticality. This has been possible thanks to the application of notions from fuzzy logic and fuzzy sets theory. However, our work is mainly linguistic and in our theory the constraints are the basis of our formal grammar.

Nevertheless, this work does not seek to build a new grammar system which deals with the **whole** fuzziness of a natural language. Neither is it seeking to propose a grammar which defines all of the syntactic constraints in the Spanish language. This dissertation does not aim to offer a manual of the Spanish language with its syntactic constraints. Many constraints have been extracted, but it is not the aim of this work to extract all of them. The constraints are a tool to explain the degrees of grammaticality of a natural language; which is the primary objective of this research.

The pages of this dissertation do not contain a full-comprehensive system which deals with the final value of grammaticality in terms of degree. In order to provide the full value of grammaticality from an input, it shall be necessary to study all linguistic modules exhaustively. This task cannot be done in a single PhD thesis.

It should be clarified that this is not a dissertation about grammars with constraints. Here, grammars with constraints are just a tool. The core aim of this work is not to determine which grammar with constraints is the best and why; it is to formally represent a fuzzy phenomenon such as grammaticality in natural language. This study has considered different grammars with constraints before finally selecting the one considered most suitable for our primary objective. To the best of our knowledge, this dissertation proposes the first formal approach in linguistics of a fuzzy grammar as a model to explain degrees of grammaticality in natural language. Until now, much literature has been written concerning linguistic gradience and fuzzy grammars. However, as far as we know, nobody has ever proposed a model of a grammar which would be able to deal with degrees of grammaticality regarding linguistic competence whilst not including acceptability judgments. Consequently, this work presents the first step towards a multi-modal fuzzy grammar which, in the future, could deal with much more vague phenomena of natural language.

1.2 Justification

The nature of natural language is evident when people use it in natural conversation. When communicating with each other, we often hesitate over what we are going to say and abandon the discursive thread as well as repeating words and phrases. Hence, grammatical niceties are not often respected, as is explained in Hayes et al. (1981):

> When people use natural language in natural conversation, they often do not respect grammatical niceties. Instead of speaking sequences of grammatically well-formed and complete sentences, people often leave out or repeat words or phrases. Break off what they are saying and rephrase or replace it, speak in fragments, or use otherwise incorrect grammar.
>
> (Hayes et al., 1981)

Furthermore, Baldwin et al. (2013) gathers these binding characteristics under the banner of *noisy text*. Remarkably, humans can still decode the received input even though grammatical niceties are not often respected. Lesmo and Torasso (1984) stressed this phenomenon and proposed that the parsers should be able to decode ill-formed sentences just as the humans do:

> The first problem that must be faced is the following: how can the parser be prevented from rejecting sentences that are syntactically ill-formed, but could be interpreted correctly if they are passed to the other components of the system?
>
> (Lesmo and Torasso, 1984:26-28)

As stated by Eisenstein (2013), "media has brought computational linguistics in ever-closer contact with bad language: text that defies our expectations about vocabulary, spelling, and syntax." Therefore, the challenge is to provide tools which can deal with such "non-grammatical" utterances.

Following these claims, grammaticality should be considered in its broad sense from the perfect utterances to the ill-formed ones. This regard may help the linguist to explain more accurately both how natural language works and how to create an interface which can decode all kinds of natural language inputs.

Nevertheless, the main trends in linguistics have considered grammaticality as a discrete or categorical object: an utterance is abruptly either full well-formed or ill-formed (that is to say that input is either grammatical or ungrammatical).

Discrete grammars focus solely on well-formed utterances. However, speakers very often produce non-canonical inputs regarding natural language. Thus, a problem arises regarding the discrete conception of language as it cannot describe all natural language productions. Given that humans are able to decode grammatical deviations in natural language processing, a formal grammar which aims to represent natural language must also do the same.

Additionally, Lavie (1996) points out that parsers have generally been set up according to a dichotomous view of grammaticality and therefore crash when receiving input that even slightly violates grammatical rules:

> By the definition of their recognition process, these algorithms are designed to detect ungrammatical input at the earliest possible opportunity, and to reject any input that is found to be ungrammatical in even the slightest way [...]. Such parsers are thus unsuitable for parsing spontaneous speech, where completely grammatical input is the exception more than the rule.
>
> (Lavie, 1996)

These algorithms are inspired by discrete grammars which consider that linguistic competence is perfect. Furthermore, they are not interested in and cannot deal with spontaneous speech, hence their rejection of an "ill-formed" input. The grammar on which the algorithm is based makes no attempt to parse that kind of input. Discrete grammars have progressed by using the categorical notion from studies of linguistic competence. They have developed the vast majority of the linguistic concepts that we use nowadays by following these criteria. However, there are several references which point out the limits of discrete grammars because of their approach, specifically when it comes to explaining the nature of natural language: Bolinger (1961), Ross (1972), Lakoff (1973), Manning (2003), Aarts et al. (2004), Aarts (2004a), Keller (2000, 2006), Fanselow et al. (2006), Prost (2008), Bresnan and Nikitina (2009), Goldberg (2009), Blache (2000, 2016).

In this manner, a grammar must accept gradient phenomena and degrees of grammaticality according to the reality of natural language. In this regard, gradience is a well-known linguistic term that designates these concepts. Aarts (2004b) defines gradience as a term to designate the spectrum of continuous phenomena in language, from categories at the level of grammar to sounds at the level of phonetics. However, most of the gradual phenomenon studied in the processing of natural language is given by gradient acceptability; examples of this are: Aarts (2004a,b); Aarts et al. (2004), Chomsky (1965, 1975), Bolinger (1961), Ross (1972), Prince and Smolensky (1993), Keller (2000, 2006), Manning (2003), Legendre et al. (1990a).

Nevertheless, we distinguish two types of research in linguistic gradience:

- Research which has increased understanding of the concept of gradience. This approach has often been philosophical rather than formal or pragmatic.
- Formal applications aimed at capturing linguistic gradience. These approaches take into account gradience for mostly representing degrees of acceptability rather than

grammaticality. However, they did not propose a model which can capture solely grammaticality as a fuzzy-gradient object.

Some of the formal systems on linguistic gradience which have endeavored to represent the degrees of acceptability are: Harmonic Grammars (Legendre et al., 1990a); Standard Optimality Theory (OT) (Prince and Smolensky, 1993); Linear Optimality Theory (LOT) (Keller, 2006); and Probability Theory (Manning, 2003). Nevertheless, most of these approaches take into account grammaticality by considering degrees of "ungrammaticality". Using the concept of "ungrammaticality" cannot provide a satisfactory explanation for grammaticality because:

- Degrees of ungrammaticality still imply a discrete approach for grammaticality: one linguistic expression is grammatical if it is optimal and if it satisfies all the constraints; whereas one linguistic expression is ungrammatical in terms of degree depending on its violated constraints. This therefore has to mean that grammaticality is discrete while ungrammaticality is gradient.
- Most of these approaches illustrate grammaticality as acceptability, using the notion of *grammaticality judgments*. This is the idea argued in this dissertation. We discuss how these theories actually represent acceptability rather than grammaticality. The main reason behind this claim is that a judgment always implies a subject such as a speaker. Therefore, a speaker is never able to judge grammaticality without being naturally conditioned by extra-linguistic elements. Consequently, in those approaches, a grammaticality judgment is always an acceptability judgment.

In the end, there is still no framework to represent degrees of grammaticality. Consequently, there is no grammar that can formally represent the degrees of grammaticality regarding linguistic competence alone — in other words, that represents grammaticality strictly from a grammar perspective without involving the speaker and extra-linguistic facts.

1.3 Methodology

In order to test our hypothesis and to reach our aims, we have applied the methodology sketched in this section.

The first step concerns a deep acknowledgment of the scientific literature regarding both the notion of degrees of grammaticality and systems which can deal with linguistic gradience and degrees of grammaticality.

We realized that one of the first obstacles for representing degrees of grammaticality was the distinction between competence and performance which determines the definitions of grammaticality and acceptability. The first solution over this problem was to provide a critical analysis regarding these concepts as well as proposing new definitions and considerations which open up the way for calculating degrees of grammaticality as a fuzzy object.

Gradience and fuzziness seek to show how the relationship between two categorical objects is a question of degree rather than discrete; that is, each word belongs to "a class in which the transition from membership to non-membership is gradual rather than abrupt" (Zadeh, 1965). This fuzzy reasoning can be applied to linguistic gradience to determine the grammaticality of an input. Thus, rather than classifying an utterance as non-grammatical if it features some grammatical deviations (discrete reasoning), it can be classified as more or less grammatical according to the constraints that are violated or satisfied (gradient or fuzzy logic).

In the literature, Aarts et al. (2004) and Lakoff (1973) suggest how fuzzy logic can be a legit tool for representing degrees of grammaticality. Moreover, Blache and Prost (2008) and Blache (2016) propose a formal model for describing any kind of input regardless of its "ill-formedness". Therefore, his proposal is a model which can take into account linguistic information concerning all degrees of grammaticality. Thus, we conclude that a model is needed which combines the fuzzy perspective with the formal model of the Property Grammars (PGs) in Blache (2016) in order to describe the degree of grammaticality of any input.

A Property Grammar for Degrees of Grammaticality. Blache and Prost (2008) proposed a framework with algorithms based on a Property Grammar (PG) for taking into account degrees of grammaticality judgments. They point out the importance of cumulative effect for both satisfied and violated properties (*constraint counterbalance*), rather than just taking into account violated properties, which is in contrast with the Optimality Theory approaches. The concepts that PGs apply to model degrees of grammaticality judgments are: *cumulativity*, *constraint violation*, *constraint weighting*, *violation position*, *constraint density*, and *propagation*. These notions will be explained in section 3.4.

PGs propose several algorithms and formulas in the literature to calculate weights for properties, degrees of acceptability, ungrammaticality, degrees of grammaticality and degrees of complexity. The formula that we propose to account for grammaticality is inspired by the notions developed by PGs as well as by some of these formulas.

In general terms, Property Grammars are ideal to account for the gradual phenomena of language such as grammaticality. They are highly descriptive. Likewise, their constraints are totally autonomous which enables them to be independently weighed. They also collect a lot of linguistic information regardless of whether the input is grammatical or ungrammatical. This differentiates them from other grammars, which successfully describe canonical inputs but offer very little information about the elements that trigger violations. Therefore, PGs are an excellent tool for building syntactic models that tolerate different degrees of grammaticality.

Using fuzzy logic for a fuzzy grammar. Grammaticality is acknowledged as a vague concept. Since fuzzy logic is the right tool to capture vague objects, we state that fuzzy logic can represent a grammar which can deal with degrees of grammaticality.

We show in our work why grammaticality is a vague notion rather than an uncertain one. We made a distinction between different phenomena, some of the most important differentiation are the following: *vagueness* vs. *uncertainty*; *gradience* vs. *indeterminacy*; *fuzziness in linguistics*; *fuzzy grammars* vs. *gradient grammars*.

We are showing the application of Fuzzy Natural Logic (FNL) from Novák (2015) in order to define a grammar which can capture the vague notion of grammaticality. In section 6.5, we show our proposal for evaluating grammaticality in a fuzzy grammar.

Finally, we claim that FNL is a highly suitable tool for a Property Grammars for dealing with degrees of grammaticality.

The application of fuzzy logic to the property grammar has been supervised by members of the *Institute for Research and Applications of Fuzzy Modeling* (IRAFM) Center of Excellence IT4 Innovations of Ostrava (Czech Republic).

There are many different approaches to formalizing fuzzy logic as well as different theories. For our work, we have chosen Fuzzy Natural Logic by Novák (Novák, 2005; Novák, 2015); a high-order fuzzy logic; fuzzy type theory (FTT) with Łukasiewicz algebra. This theory is highly suitable for modeling natural language and other linguistic concepts. This theory is genuinely linguistically motivated and highly influenced by Lakoff (1970) and Montague (1970).

Extracting a Property Grammar from Universal Dependency Treebank. A Spanish Property Grammar and its extraction will be presented. This Property Grammar is determined by the basis of Blache (2016) together with the fuzzy logic from Novák (2015) and original evaluative bases from this work. This method provides a new perspective since no grammar has ever been able to define their bases from a gradient grammaticality perspective before. This new framework might be called Fuzzy Property Grammars, and it can represent the degrees of grammaticality that are found in the different constructions, regarding linguistic competence.

The syntactic properties have been extracted automatically by applying the MarsaGram tool by Blache et al. (2016) to the Universal Dependency Spanish TreebankCorpus. This corpus is obtained from the Universal Google dataset (version 2.0). It consists of 16,006 tree structures and 430,764 tokens. It is built from newspaper articles, blogs, and consumer reviews.

We have applied this new interdisciplinary approach to the description of Spanish syntax. Property Grammars have been used in our work to define the different constructions and linguistic elements of Spanish. Our property grammar has been modified in order to bring up descriptions with fuzzy logic. In this way, we have defined a fuzzy grammar that can represent the different gradual phenomena and variability that take place in Spanish.

Property Grammars and Fuzzy Natural Logic, the work with a corpus and the critical review of the literature on grammaticality and gradience have made it possible to propose a model based on a formal grammar with constraints combined with fuzzy logic that can represent certain gradual-fuzzy phenomena of language, such as the degrees of grammaticality that are found in natural language, regarding linguistic competence. This method provides a new perspective since a grammar has never been able to define their bases from a fuzzy gradient perspective before.

1.4 Structure of the Dissertation

This thesis is divided into three parts:

- **Part I** presents the background, review and critical analysis over the notions of *grammaticality*, *gradience* and *fuzziness*. It also introduces the main features of the Property Grammars framework.

  - **Chapter 2** consists of an original review and a critical analysis of the notions of *competence, performance, grammaticality* and *acceptability*. Different points of view regarding these conceptions are reviewed and clear distinctions between what is grammaticality and acceptability on one hand, and competence and performance on the other are shown. Finally, this section offer some conclusions regarding these concepts in our work.
  - **Chapter 3** spells out the background assumptions on linguistic gradience. It provides a history of gradience in linguistics. It describes the frameworks which dealt with gradience up to the 21st century. It provides an original review and a critical analysis over the methods which model gradient data, the methods which apply linguistic judgments for representing degrees of grammaticality, and the methods based on probabilities. We briefly mention the methods for dealing with

semantic gradience. Finally, we offer a review and a critical analysis regarding the notions of grammaticality, acceptability and linguistic weights concerning the frameworks on linguistic gradience.

- **Chapter 4** is a presentation on the bases of fuzzy logic and the features that it shares with linguistic gradience. We provide an original review exposing the traits shared by linguistic gradience and fuzzy logic. We introduce concepts from fuzzy logic such as *indeterminacy, vagueness, uncertainty, randomness*, and *ambiguity*. We introduce the fuzzy set theory and its truth values. We state that these definitions are extremely useful for linguistics. We describe an application of fuzzy set theory and truth values for linguistics and grammaticality. We provide a review and a critical analysis of the notions of linguistic vagueness and linguistic uncertainty, and the necessity of distinguishing between fuzziness from gradience in linguistics.
- **Chapter 5** describes Property Grammars. We introduce them by explaining their properties, their application over lexical entries, their descriptive potential with features, the way they work, the notion of construction in Property Grammars, and their parsing approach. We briefly compare the Property Grammars theory with derivational tools, and we reveal why is a non-hierarchical theory.

• **Part II** presents our proposal for the definition of a fuzzy grammar and the extraction of the properties for a Spanish property grammar as a proof of concept of a fuzzy grammar.

- **Chapter 6** presents the *formal* core of our work. It reveals the bases of our Fuzzy Property Grammar. We justify the necessity of a fuzzy property grammar for representing degrees of grammaticality. We introduce Fuzzy Natural Logic as a tool for building a fuzzy grammar. We present a Fuzzy Grammar formula which takes into account degrees of grammaticality regarding linguistic competence. We show the mechanics of a fuzzy grammar with fuzzy logic. We define the structure of a *word* with fuzzy logic. The theoretical bases of a Fuzzy Property Grammars are provided taking into account: the different constraint behavior in a fuzzy grammar, the notion of *xCategory*, and a proposal with tools for dealing with linguistic variability for representing degrees of grammaticality. A proposal adding some computational value is made. We offer an abstract example of the application of a Fuzzy Property Grammar.
- **Chapter 7** presents the *linguistic* core of our work. It presents our work regarding Universal Dependency Treebank and the use of MarsaGram. The use of

frequencies and the extraction of properties for a fuzzy grammar are described. We discuss the previous considerations concerning the extraction of a Spanish Property Grammar. We highlight the importance of a proof-of-concept of subject construction for the Spanish language. We provide fuzzy property grammar descriptions for the different elements which are present in the subject construction in Spanish such as the verb, the noun, the modifier, the proper noun, the pronoun, and the determiner. We describe how each property grammar has been extracted, explain why its properties are as they are, and give examples in which they are applied. We certify as a proof of concept how the fuzzy property grammar approach can deal with language variability and fuzzy degrees of grammaticality.

- **Part III** closes the dissertation by briefly summarizing its content, presenting conclusions in **chapter 8** and giving some directions for future research in **chapter 9**.

Part I

Through Grammaticality towards a Fuzzy Grammar with Constraints

Chapter 2

Grammaticality in Natural Language and Linguistics

Any attempt to define what natural language grammar is must include the notion of grammaticality. The first step for linguists is to differentiate between what is grammar-related (grammatical) and what is not (non-grammatical). However, the difficulties in establishing the formal and objective boundaries regarding whether an utterance is grammar-related (or not) are well-known. To help define these fuzzy boundaries, scientific studies have proposed various ways of dealing with this notion of grammaticality, but linguistics has never been able to reach common agreement. Two main perspectives in this topic have been the subject of controversy throughout the history of linguistics:

- To consider grammaticality as a discrete object.
- To consider grammaticality as a gradual-fuzzy object.

The origin of this disagreement can be found in the theoretical basis of competence and performance. This distinction leads and constrains linguistics when it comes to describing natural language. Let us introduce an old (but still current) statement from Jespersen in his work *The Philosophy of Grammar*. Jespersen (1924) asserts that natural language is spontaneous, immediate and ambiguous, and often produced with grammar violations:

> Apart from the fixed formula a sentence does not spring into a speaker's mind all at once, but it is framed gradually as he goes on speaking. [...] Anyone who listen carefully to ordinary conversation will come across abundant evidence of the way in which sentences are built up gradually who will often in the course of the same sentence or period modify his original plan of presenting his ideas, hesitate, break off, and shunt on to a different track.
>
> (Jespersen, 1924:26-28)

This assertion satisfies both *discrete linguistics* and *gradient linguistics*. According to discrete linguistics, it is not necessary to explain natural language in a broad sense since linguistic competence is regarded as perfect. From this perspective, all grammatical deviations stem from performance when an output "is framed gradually as he [the speaker] goes on speaking". As discrete linguistics was not (usually) interested in performance, it used to disregard these properties. However, gradient linguistics considers that linguistics should deal with the vague nature of language. However, in the end, both points of view (discrete and gradient) used to agree that the gradient approach belongs to performance under the slogan "Grammar is grammar and usage is usage". This slogan is found in the title of the 2003 presidential address to the Linguistics Society of America (as cited in Bresnan et al. (2007)). Competence therefore remains perfect, so there are no degrees in grammaticality but in acceptability, such as in Keller (2000), or Fanselow et al. (2006).

This PhD thesis goes one step further. *It shows how the gradient phenomenon is a matter of linguistic competence*, which is why we are talking about gradient grammaticality (not acceptability). This perspective is closer to fuzzy grammars, such as in Lakoff (1973), than to gradient grammars, such as in Keller (2000) and Sorace and Keller (2005).

Taking into account the various theoretical reasons in linguistic literature, this chapter offers a holistic statement. It concerns how the concept of grammaticality needs to be explained as a fuzzy-gradual concept rather than a discrete or abrupt one. A fuzzy grammar therefore needs to be produced in which the fuzzy boundaries of grammaticality in natural language grammars can be explained.

2.1 Grammaticality as a Discrete Concept in Linguistics

Linguistics has been greatly motivated by understanding how children acquire and develop linguistic competence. The general aim of linguists used to be to abstract the linguistic notions that assemble a natural language in every single modality. To attain this objective, they both define and elaborate a grammar. A grammar may therefore be considered the tool developed by linguists to provide a framework for either generating or defining (or both) a natural language (as humans do).

However, linguists found nebulous boundaries when attempting to extract both linguistic notions and grammar from natural language, probably because language is involved in many variables, including cognition-processing and social variables. With this evidence, linguists are fully interested in delimiting what language is and what it is not. Therefore, what is language is within linguistics and what is not language is outside linguistics. The interdisciplinary areas in linguistics maybe considered a middle field. However, even in these interdisciplinary linguistic areas, linguists used to establish clear boundaries about what is related to the language domain and what is not.

This preoccupation with framing the precise limits of language is found in certain post-Bloomfieldian authors, who considered it was necessary to employ the discrete approach over the boundaries of the language regarding the language domains and their treatment. For example, Joos (1958) and Hockett (1955) hold that discreteness is the only way to define language. Hockett (1955:17) states that "if we find continuous-scale contrasts in the vicinity of what we are sure is language, we exclude them from language". Similarly, Joos (1958) accepts that there are different variables around language, some of which are in continuous gradation. However, for Joss all of these variables are definitely out in the design of the language:

> True, the sounds (and thus all the forms) occurring in the use of the language are variable by continuous gradation, and so are not only temperatures but all things and phenomena spoken of. But in the design of the language we never find a trace of those facts!
>
> (Joos, 1958:351)

Although these are some of the most well-known claims when it comes to doing linguistics through discrete concepts, discrete grammaticality has been closely related to Chomsky's work, and therefore to Generative Grammar. This theoretical approach is framed under several dichotomous and categorical notions such as the discrete notion of grammaticality. This notion is a consequence of a whole theoretical background in Chomsky's work that is utterly motivated by some of his most famous concepts, including I-language, X-bar theory, the innate linguistic abilities of children and, especially, the difference between *competence* and *performance*. The theoretical partition of competence and performance is the most important reason for defending a discrete concept of grammaticality in generative grammar. This partition is one of the most critical partitions in the history of linguistics; it conditions a long working tradition in linguistics by means of discrete notions. In the next subsection, we explain the theoretical concepts of Chomsky in order to clarify why generative grammar finally decided to work with categorical notions.

We should note, however, that in his early work Chomsky was undecided regarding if the notion of grammaticality should be gradient or not. Chomsky (1975:129) (from his 1955 PhD thesis) recognizes that: "a partition of utterances into just two classes, grammatical and non-grammatical, will not be sufficient to permit the construction of adequate grammars". Chomsky (1957:31) claims that a language's real description needs to understand all language as well as the unclear boundaries. He also states that he disagrees with Hockett (1955). For Chomsky, Hockett's approaches are schematic since he tries to fit language phenomena into a discrete conception; grammatical/non-grammatical, possible/non-possible. Chomsky's most influential work on this topic can be found in *Aspects* (Chomsky, 1965). He also discussed the notion of 'degrees of grammaticalness' by differentiating between acceptability and grammaticalness (Chomsky, 1965:11,77). However, in the end, he clearly stands for a discrete approach for grammaticality and linguistic work.

2.2 Competence and Performance

> We make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of the language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence.
>
> (Chomsky, 1965:4); see also Chomsky (1975:14-18)

This distinction between competence and performance has been fundamental for both generative linguistics and other grammars with a discrete approach.

This distinction, which attempted to determine what is worth researching in linguistics and what is not, conditioned linguistic research for years. In Chomsky's words, competence or Language-I is what matters in linguistics since performance or Language-E "cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline" (Chomsky, 1965:4). To do that, the idealization of language is necessary by means of perfect speaker-hearers and a homogeneous language community to provide the generative grammar that aims to generate perfect linguistic outputs. The goal of the generative linguistic theory is therefore clear: to describe the knowledge of language independently from the role this knowledge plays in its production.

Linguistic data from performance was considered "false" (Chomsky, 1965:4). Doing linguistics by underpinning the linguistic theories on empirical data was therefore considered "a joke" or "second class". The conclusion was to acknowledge linguistic competence as a perfect, purely abstract and theoretical object.

2.2.1 Chomsky's Competence-Performance

I believe that Chomsky's work was influenced by Vygostky's*Thinking and speech* (1934), which was published in Vygotsky (1964). In *Thinking and speech* Vygotsky tackled the vague boundary concerning what is thinking and what is speech in order to clarify these obscure limits.

> Verbal thinking is partitioned into its elements; it is partitioned into the elements of thought and word and these are then represented as entities that are foreign to one another [...]. The conclusion of this study was that inner speech facilitates the consolidation of the material and creates an impression of what must be understood. *When inner speech was included in the processes involved in understanding, it helped the subject to sense, grasp, and isolate the important from the unimportant in the movement of thought. It was also found that inner speech plays a role as a facilitating factor in the transition from thought to overt speech [...] In structural terms, there are no significant differences between whispered and normal speech and, more importantly, that whispered speech manifests none of the characteristics of inner speech.*
>
> (Vygotsky, 1964:4,69)

Vygotsky's work stands out (especially in linguistics) for distinguishing between thinking and speech, for demarcating what language is and what it is not. Vygotsky proposes a new definition of what is inner speech, which in his regard is "predicative", and a medium level which arranges the language before it is explicitly performed (Vygotsky, 1964:202). He differentiates categorically between *inner speech* and *overt speech* and considers overt speech as either written or oral. He treats inner and overt speech as "polar opposites" and believes that their "syntaxes are different". He states that "The grammar of thought" can be found in the mind (Vygotsky, 1964:164).

In linguistics these concepts seemed confusing, so linguists have gone one step further. Chomsky's work strove to explain the mental phase in the human mind that generates language from thought to speech. However, in Chomsky (2002), he claims that inner speech is speech itself: "Inner speech is most of speech". In this sense, Chomsky was influenced by Vygotsky's concepts of "inner speech" and "overt speech", however, he provided new perspectives for them.

Chomsky claims that the processes of language generation in our mind and the processes of performing speech are not the same. In this sense, language is generated by a system of rules that embody the linguistic competence of every speaker. For Chomsky, ill-formed utterances occur in the final stages of generating language due to linguistic performance. In his opinion, there is no reason to consider more than one grammar and more than one linguistic competence since any variables in an output are a consequence of linguistic performance. Chomsky established the features of linguistic competence/performance in his work *Aspects of Syntax Theory* (Chomsky, 1965):

> We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use of language in concrete situations). Only under the idealization set forth in the preceding paragraph is performance a direct reflection of competence. In actual fact, it obviously could not directly reflect competence.
>
> (Chomsky, 1965:4)

From this distinction, Chomsky renamed this discrete classification, claiming that competence is embodied in the I-language, a concept previously understood as grammar. In this context, I-language plays the leading role in the linguistic process: "I-language, [...] the system of knowledge of language attained and internally represented in the mind/brain" (Chomsky, 1986:24). We may, therefore, consider that the distinction we find in Vygotsky regarding *Thinking-Inner Speech-Speech* and *Inner Speech-External Speech* evolved in Chomsky to I-language (grammar/competence) and E-language (language/performance). This comparison is illustrated in Table 2.1.

Vygotsky	Chomsky
Thinking Inner Speech	I-language Grammar Linguistic Competence
Speech External speech	E-language Language Performance

Table 2.1 Contrasting concepts between Vygotsky and Chomsky.

In summary, for Chomsky, competence, and therefore I-language, is a linguistic mind process that is different from any kind of speech. Chomsky's distinction between I-language (grammar/competence) and E-language (language/performance) was also highly influenced by Saussure's *language* (competence) and *parole* (speech). However, Chomsky mentions that the I-language concept is borrowed from Otto Jespersen:

> Otto Jespersen, who held that there is some "notion of structure" in the mind of the speaker "which is definite enough to guide him in framing sentences of his own," in particular, "free expressions" that may be new to the speaker and to others. Let us refer to this "notion of structure" as an "internalized language" (I-language).
>
> (Chomsky, 1986:21-22)

Chomsky states that linguistic science is committed to explaining I-language and its rules but not to explaining the phenomena in E-language since it "cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline" (Chomsky, 1965:4). In this regard, linguistics must explain only a natural language by means of its rules and how those rules generate language.

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.
>
> (Chomsky, 1965:3)

These rules must be extracted from a "homogeneous speech community" to idealize language and explain linguistic competence. In this context, doing linguistics with an ideal speaker-hearer is necessary because the grammatical deviations occur during the surface production of an utterance (performance) and not during its deep structure generation (competence). This reasoning directly constrained the concept of grammaticality and led to quotes such as "we believe, as do most linguists, that native speakers do not make mistakes" (Andersson and Trudgill, 1990:111).

Grammaticality is named in *Aspects* as grammaticalness and "belongs to the study of the competence" (Chomsky, 1965:11). Since grammar, competence and the speaker-hearer are always perfect, there is no place for unacceptable grammatical sentences. Hence, for Chomsky, an utterance is always grammatical since grammaticality is a matter of competence and its rules generate perfect utterances (without deviations). Grammatical violations therefore do not exist in competence or in the I-language. *Only degrees of acceptability exist in the performance*.

Since there is no imperfection in competence or I-language, there is no reason to take degrees of ungrammaticality into account. Chomsky pursues the idea of understanding grammatical universals as rules in the I-language in order to fully comprehend what linguistic competence is.

The theoretical assumptions for defining a grammar are based on a categorical perspective since they are not interested in generating ill-formed utterances. They deny any structure which is not well-formed or perfect. Generative grammar is willing to *generate* outputs and thus generate utterances without deviations: "a sentence is 'grammatical' if it is generated by the grammar, 'ungrammatical' if it is not" (Newmeyer, 1983), i.e. from the deepest structure (I-language/competence) to the most superficial structure (E-language/performance).

Chomsky's view is based on implying that children do not assume all the infinite forms of language but that they are competent with a few rules that form the basis and models of the linguistic structures. He distinguishes between two kinds of structures: *deep structures* and *surface structures*. Deep grammatical structures of the language represent formations which are *transformed* to surface grammatical structures in all their variety. Deep structures belong to competence. They widely reproduce the structure of surface structures. They establish hierarchy and general semantic notions. The same deep structure can be performed roughly in different surface structures. For example, after watching a dog eating a bone, we could say *"the dog was eating the bone"*, *"my dog ate a bone"*, *"the dog ate the bone"*, etc. These sentences have distinct surface forms that derive from a common deep one. The deep structure in these sentences therefore holds the basic main rules, while the surface structure shows some variety which refines the utterance. In this regard, Luria (1980) points out how Chomsky's X-bar theory clearly represents the process that was mentioned in Vygotsky from *Thinking-Inner Speech-Speech*. In this sense, taking the above sentence as an example *"the dog ate the bone"*, at first the thought is blurred and the inner speech sorts out that thought by ordering the language around the verb *ate*:

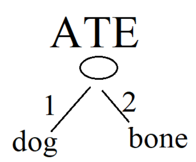


Fig. 2.1 Representation of Luria's regard about X-bar theory in human mind.

The main difference between Vygostky's view and Chomsky's is that, for Vygotsky, inner speech is 'predicative' while Chomsky's deep structures are abstract categories. For Chomsky, the step from the deep structure to the surface structure is conditioned by means of a filter mechanism that rejects non-acceptable structures for a specific language. This is why in a transformational grammar the two structures (deep and surface) do not coincide. During the transition from deep (perfect) to surface and then to external/performed structure, many things can happen and it is in the surface structure where grammatical deviations and ill-formed structures occur. Generative Grammar aimed to look for the universal rules which allow the transition from deep structure to surface structure. If we follow this reasoning, we have another perception of the X-bar theory and generative grammars: they aim to explain the transition from a deep structure (in relation to a thought or a motive) to performed speech by generating language as a perfect output.

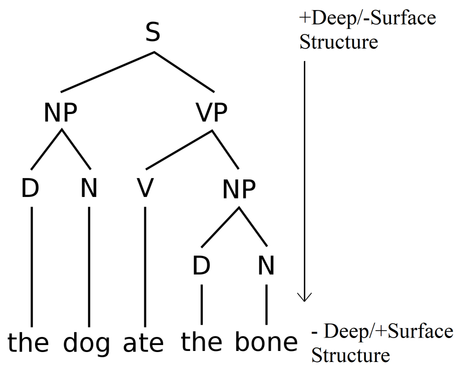


Fig. 2.2 Generative Grammar treebank.

Luria (1980:30), who agrees with Chomsky, summarizes the process of a generative grammar from linguistic motivation to external speech in the following way:

1) Motivation: fundamental sense of the phrase.

2) Semantic representation in inner speech.

3) Deep syntactic structure.

4) Surface syntactic structure.

5) External speech.

From a generative grammar, items 1 to 3 are, therefore, a matter of linguistic competence, while items 4 and 5 are a matter of performance. Even though this can be treated in linguistics, it has often been considered non-interesting since all is blurred, confusing and obscure:

> This is the point I discussed in *Rules and Representations* [1989] in a somewhat different context. It was a mainly debate with philosopher who feel that the notion of language is somehow clear and the notion of internally represented grammar is somehow suspect. It seems to me that it is actually the other way around. The notion of internally represented grammar is quite clear [...]. On the other hand, the concept of language is very obscure, and it is not obvious that there is any intelligible or significant notion.
>
> (Chomsky et al., 1982:108)

In summary, many claims and statements are made in this way, leaving external speech aside, while calls are made for methodological reasons to defend the discrete approach as the only way to explore and understand grammar and linguistic competence.

The methodological reasons for a discrete approach to explain a natural language lie in the theoretical considerations we mentioned earlier. Generative linguists have dismissed the notion of fuzziness from their framework, claiming that that their discrete approach to understanding linguistic competence (grammatical universals) is underpinned by methodological reasons, such as Bever (1975):

> To give up the notion that a grammar defies a set of well-formed utterances is to give up a great deal. This is not to say that it is impossible in principle that grammars are squishy [squishy means gradient/fuzzy]. Rather the possibility of studying precise properties of grammar and exact rules becomes much more difficult... Thus, if we can maintain the concept of discrete grammaticality, we will be in a better position to purse an understanding of grammatical universals.
>
> (Bever, 1975:601)

The need for the idealization of language has constantly been stressed by the generative perspective. Chomsky stressed the idealization of language as “the only reasonable way to approach a grasp of reality” (Chomsky, 1995:7) (Chomsky, 1998:115). This conception has constrained the way linguistics is done (almost as a dogma) since it considers that no other acceptable way to build a grammar exists. However, this reasoning cannot explain all that happens within natural language production and processing. Some generative approaches to gradience in grammar have been made by Keller (2006) and are summarized in Fanselow et al. (2006). However, these mainly apply the gradient approach to specific language fields such as performance level or degrees of acceptability. In any case, they propose a system or formulation of a grammar that can formally explain degrees of grammaticality. Because of these discrete theoretical notions, they fail when it comes to explaining linguistic inputs with grammatical deviations. To fully understand natural language processing, especially in terms of degrees, it is therefore necessary to review the competence/performance and grammaticality/acceptability concepts. The extended generative view according to these concepts does not fit into a gradient/fuzzy grammar with degrees of grammaticality that can explain linguistic inputs according to a grammar and not in relation to a speaker's judgment.

2.2.2 Competence and Grammar

From the generative point of view, *grammar* has two different uses:

- Firstly, a grammar describes the linguistic knowledge and competence of a native speaker in relation to a specific language, i.e. native speakers know the grammars of English or French or Spanish, and so on. Speakers are competent in their native language. Here, the aim of study is to identify the mental capacities that generate sound/meaning pairs for an unbounded number of linguistic structures.

- Secondly, a grammar describes the most basic deep structures for all languages. This is known as *Universal Grammar*. In this sense, two languages may look different from the outside (surface structure) but deep down they are the same (deep structures). For Chomsky, two languages which are different on the surface (performed) are the same in their deep structures (competence).

Consequently, a native speaker who acquires a second language in adventitious circumstances already has a compilation of grammar universals that facilitate the acquisition of the second one. This conception of grammar is the most abstract and theoretical. For Chomsky, this is the essence of linguistic competence. Here, the object of study is the mental capacity to derive any input of a grammar from the principles of a universal grammar.

In summary, the rules in the deep structures are perfect, which is why competence is still immutable. Theoretically, this reasoning claims that English and Spanish are the same in their deep structures but different on their surface and therefore in their performance. From this point of view, doing linguistics without taking into account "performance" was sort of beneficial since all the noise from the "performance" had to be removed in order to better extract the deep structure rules. Any linguistic rule which is not merely structurally derivative is performance. In this way, the rules of semantics, pragmatics and prosody, and so on, were therefore considered rules of performance. A violation of these rules is therefore (in this perspective) a matter of *acceptability* rather than *grammaticality*.

2.2.3 Performance and Grammar

Lakoff (1973) divides the features of Chomsky's performance into three types of linguistic performance and discusses them along with the notion of competence. This is important because all the mechanisms considered to be performance would be components which resolve the degrees of acceptability. While the mechanisms considered to be competence would be components which resolve the degrees of grammaticality:

- **Performance 1** is what people actually do when they speak, excluding the linguistic mental apparatus. Lakoff points out that this idea of performance is extracted from *Aspects* (Chomsky, 1965:4).
- **Performance 2** is the perceptual mechanisms that are common to all human beings. These mechanisms underlie what people actually do. These"mechanisms" are considered to be extra-linguistic but they can still condition the language. An example of this is found with self-embedded sentences which are, from Chomsky's perspective, grammatically correct, because they can be generated by a grammar, but unacceptable from the performance point of view since the "perceptual mechanism" constrains such embeddings.
- **Performance 3** is the abstract linguistic rules which arrange the surface structures of a particular grammar. Here, Lakoff points out that Chomsky is using the term "performance" in a completely new way without specifying what the difference is between these abstract rules and those of competence. In fact, Chomsky's idea was clear: transformational grammars were born from the conception that every language shares a universal mechanism which transforms the deep structure into surface structures. The deep structure rules are related to competence while the surface structure rules are related to performance. However, even if we understand that, differentiating between competence and performance rules is quite difficult because both are abstract and both structure the language of a grammar. If this lead is to be followed, linguists must accept that almost all the work done is related to performance.

Following the reasoning of *Performance 3*, any grammar or system that aims to describe surface structure (i.e. utterance data from a corpus) is considered *linguistics of the performance* (in the worst possible sense). Theories or grammars that describe language based on a corpus that is already "performed" must explain at which point from the competence and performance distinction they are located. Some such theories are: Optimality Theory (Prince and Smolensky, 1993), Lexical Functional Grammar (Kaplan and Bresnan, 1982), Construction Grammars (Goldberg, 2009), Probability Theory (Manning, 2003), Property Grammars

(Blache, 2000, 2016), and General Phrase Structure Grammar (Gazdar et al., 1985). These theories work with *constraints* rather than *rules*. This is because the generative theories have taken possession of the word *rule*. A grammar that does not work with generative rules (which are linked to competence) must therefore say that it uses constraints, not rules. This is silly, however, since both constraints and rules are related to properties of the language. Both try to explain how language works. In essence, the rules of the X-bar theory are a constraint itself since you cannot generate a specifier in an X'. The same is true of the other way round: a distributional constraint that says that a determiner must precede a noun is essentially a rule.

Lakoff's summary of Chomsky's performance reveals *performance 3* as one of the main points of argument in linguistics. Because *performance 3* is the study of the abstract linguistic rules, these rules seem little different from competence (if competence is taken to be the knowledge of a grammar).

For many linguists, the term 'performance' has been used in so many discussions in so many ways, leading to misunderstanding and confusion. Lakoff (1973) suggests discarding that term and talking about *actual speech*, *perceptual mechanisms*, *abstract linguistic rules*. In my opinion, the term 'performance' is too widely extended in linguistics and this will not change. It is necessary to find agreement within the community and discuss this theoretical topic until the problem is fixed forever. One of the reasons for all this disagreement is that in the end every linguist rules the criteria of his/her grammar. In a sense, a general agreement can be found by considering:

- *Grammaticality* as a notion of *competence*, where competence is the linguistic knowledge that speakers have of the grammar of their language.
- *Acceptability* as a notion of *performance*, where linguistic and extra-linguistic factors such as memory and processing, and so on are mixed.

This differentiation between competence and performance is shared among linguists. If we accept this differentiation, *performance 3* is never a matter of performance but a matter of competence, bearing in mind that there is a value of grammaticality for each linguistic rule regardless of whether these rules are from deep or surface structures from the transformational grammar perspective.

From the minimalism perspective, deep structures are ruled by syntax since it is necessary to set up a structure to have an output. Semantics and other modules are considered modules of what is called, in Lakoff (1973), *performance 3*. Chomsky's sentences *Colorless/Furiously* are excellent for showing the independence between syntax and semantics. However, this independence is relative to such specific conditions and authors such as Sadock (1991), Montague (1970), Bolinger (1961) and Jacobson (2004) are committed to showing the connection between the semantic-syntactic interface.

Chomsky's sentence may be grammatically correct if grammaticality is only a matter of syntax. If we understand linguistic competence as the linguistic knowledge of a language, grammaticality can also be an issue in semantics since the semantic module also has rules. *"Colorless green ideas sleep furiously"* would be syntactically grammatical but semantically non-grammatical. Consequently, Chomsky's old quote –"Grammar is best formulated as a self-contained study independent of semantics. In particular, the notion of grammaticalness cannot be identified with meaningfulness" (Chomsky, 1957:106)– is out of the picture here since *performance 3* is no longer performance but competence, and the semantic-syntactic interface determines specific linguistic rules for different expressions. In the Spanish example *"fuego rojo"* (*red fire*) and *"rojo fuego"* (*fire red*) have different meanings since the alteration of the standard structure *"fuego rojo"* triggers new functions, new semantic rules, new thematic roles and a new meaning. In this sense, we cannot decide whether the alterations in the syntax have an impact on semantics or whether it is semantics that constrains the syntactic structure. Obviously, there is no reason to decide which one does what to which since both work as a pair. What is important here, however, is to point out that, even if theoretically both structures share the same deep structure, the linguistic knowledge triggered in the hearer to understand each of these expressions is different. A grammar must therefore take into account this phenomenon regarding linguistic competence.

Again, within these approaches the separation between linguistic modules is based on methodology. Language works as a whole system of rules that determine competence and grammaticality, which is why a multi-modal approach to what is competence and performance would lead to a more realistic description of the natural languages. What we cannot assure is that that the sum of the parts will actually represent the complexity of the whole system.

Performance 1 and *2* also have certain problems regarding whether they can be considered in terms of performance alone. If we analyse them more deeply in a theoretical sense, they cannot be radically split from the field of competence.

In *performance 2*, Lakoff (1973) explain how a 'perceptual mechanism' can constrain the language. In some cases, these constraints maybe purely extra-linguistic. However, if they can constrain language to modify the system of rules in a speaker's grammar, these 'perceptual mechanisms' are no longer merely extra-linguistic but both linguistic and extra-linguistic. For example, the 'perceptual mechanism" of a person with Alzheimer's is damaged. This is clearly an extra-linguistic fact. However, when this has an unavoidable effect on the speaker's grammar, the speaker is not only affected by this damage to the 'perceptual mechanism' but also by the oversight of certain constraints, or rules, that define their linguistic knowledge in relation to a grammar. A damaged 'perceptual mechanism' may have an effect such as producing structures with deviations and evidence of less syntactical resources such as the inability to generate relative clauses, semantic deviations and so on. We therefore cannot claim that this type of *performance 2* is unrelated to the field of competence. Methodologically, in abstraction, competence, performance 3, and performance 2 can be split up. However, if the objective is to describe what is happening to the speaker's competence, the 'perceptual mechanism' in the speaker's grammar must be explained as a component of competence too since a person with Alzheimer is obviously not as competent as a common speaker in the grammar of a language.

In fact, *performance 1* is the only one that could be considered almost entirely extra-linguistic. This performance considers aspects such as the disposition of the speaker-hearer (whether they are ready to listen), pace, vocalization, and so on. Again, however, we cannot completely exclude the field of competence here either. For example, a speaker who vocalizes badly is violating the phonological rules or constraints of the grammar of the hearer. In this sense, if we wished to formally represent the consequences of a bad input in the linguistic competence of the hearer, we would need to be able to represent how a bad performance triggers grammatical violations in the grammar of the hearer since these grammatical violations (or constraints' violations) make the task of understanding the message such as an impossible one.

2.2.4 Competence vs. Performance

Lakoff (1973) revealed the fuzziness along the boundaries between performance and competence. At some point it appears difficult to continue to consider certain types of performance (i.e. *performance 3*) as matters of performance alone especially when linguistic constraints that define the linguistic knowledge of the grammar of a language are found.

Clearly, the core of all this discussion is theoretical. However, this discussion is mainly related to notions about what is competence and what is performance rather than to degrees of grammaticality or acceptability. Defining what is competence and performance determines the notions of grammaticality and acceptability.

Following the reasoning in Chomsky's distinction and the concepts of competence and performance, it seems that the speaker always generates well-formed structures. However, it is in performance where the problems arise.

Many linguists who talk about gradience focus on talking about degrees of acceptability in the field of competence. Even those who try to explain degrees of grammaticality, such as Keller (1997, 1998, 2000, 2002, 2006), cannot avoid following Chomsky's distinction from *Syntactic Structures* (Chomsky, 1957) and *Aspects* (Chomsky, 1965).

Could we say that people who have Alzheimer's are still competent in a language when they perform "non-acceptable" utterances because those utterances are constrained due to a cognitive problem that belongs to performance? Could we say that a person who speaks systematically without determiners is still competent in English but not in the performance of English? In terms of rules, if a speaker can produce language from only 100 rules of grammar while another speaker can produce language from 150 rules, this must be because the second speaker is more competent than the first. How can we say that both have perfect knowledge of a language (perfect competence) when the numbers are so different? Is it possible to claim that the systematical lack of use of 50 rules is due to a performance issue caused by distractions and extra-linguistic matters? If the extra-linguistic issue affects our brain, and linguistic competence is placed in the brain, how can we say that someone is still competent if that person can barely produce language? This is where the nonsense in the distinction between competence and performance arises.

Methodologically, the distinction can help us to work in linguistics. However, when we attempt to grasp a real picture of natural language processing there is no reason to make this distinction. Without the boundaries of the idealization of language, a person who is not competent in a language is not going to perform that language better than what their competence allows them to. If we accept that we have inborn linguistic skills, we should accept that, as humans, we have logic capacities for learning a language. However, these logic capacities do not reflect competence in relation to a grammar as a set of rules but only the capacity both to acquire and learn a grammar.

The theories that divide competence and performance are due to methodological purposes or the fact that they do not consider grammar in a multi-modal way. The competence and performance distinction has been merely syntactic, dismissing other linguistic modules that play a role in language. A full grammar of natural language must include pragmatics and prosody as linguistic modules full of rules. In linguistics, the rules we may find in these modules have usually been considered a "second class" field as well (negatively) as "performance linguistics". However, following on from the previous examples, I do not think we could say that a speaker with *Asperger* who is unable to understand *irony* or *metaphors* is as competent as someone who can deal with that linguistic capacity in pragmatics. In this sense, a person who can deal with such linguistic abilities is more competent than someone who has no idea how to understand irony or metaphor, or formulate a question following the tonal rules in the prosody of a specific language. The problem with the rules in both *pragmatics* and *prosody* modules arises because these rules are much more difficult to extract or define than those used in syntax, morphology, phonology or semantics.

Any distinction between competence and performance rules that places some modules of grammar in performance is nonsense, since every rule is needed to be successful in communication and to produce grammatically natural language outputs. Both pragmatics and prosody rules are rules, which means that they are part of the grammar and therefore have an important role in providing a value of grammaticality. In other words, they belong to competence. The fact that these rules are difficult to explain does not mean that we should ignore them. Following this multi-modal view of what language is, it seems impossible to split competence and performance when talking about natural language processing. However, it is convenient to split them from a methodological point of view in order to define every aspect of what language is. Once linguistic knowledge is complete, we will gather all linguistic knowledge together and talk holistically about what language is by taking into account its fuzzy boundaries.

In summary, when we take into account a grammar and we are able to evaluate its rules, constraints or whatever theoretical tool that defines the knowledge of a natural language with regard to that grammar, we are evaluating competence and not acceptability judgments since the grammar evaluates the inputs by itself rather than through the judgment of a speaker. A speaker with a strong linguistic competence has a strong capacity to resolve different grammars through the received inputs. Speakers are universally competent when they are able to compare and extract common features from different languages while generating different grammars that enable them to understand different languages, dialects or linguistic variations. On the other hand, a speaker with poor linguistic competence will be clumsier when it comes to dealing with linguistic variability. If a speaker is in a linguistic context in which he/she does not know the grammar, they will be forced to generate new grammars. Lack of competence in carrying out this process will generate frustration since they will be incapable of generating and understanding the received linguistic inputs. Obviously, this lack of competence will have a consequence for the speaker's performance.

However, following on from this point of view, even with a really low level of linguistic competence, performance may still have an independent value if we consider features such as pace, attitude, disposition in performance, and others that are extra-linguistic and not definable by linguistic constraints. For example, a foreigner may be asking for help in a country in which he/she is barely competent with the local language. A local person is very competent in the local language but not competent in the foreigner's language. However, the local may not be disposed to collaborate when they try to communicate with each other in the local language. Poor attitude will lead to poor performance, which will result in a breakdown in communication.

The distinction between competence and performance may exist in a theoretical field. However, a full definition of the linguistic competence of a natural language will always depend on the rules that define that language. Competence is perfect when it is provided by those rules which are known or able to be used by the speaker. On the other hand, competence is not perfect when those rules are unknown or unable to be used for some reason. The idealization of a perfect speaker-hearer is an idealization itself and the generative enterprise needs it; it is not misconceived at all, as is shown in subsection 2.2.1. However, if linguistics wishes to explain every phenomenon in the language, it must aim to understand language without differentiating between competence and performance in a way that constrains the field of competence itself. Linguistic competence must be a broad concept that includes every feature that is able to explain natural language processing. This includes explaining why we understand utterances with grammatical violations.

Competence is for producing outputs as well as for understanding inputs. If as humans we can process violated inputs, with violated rules, this must be because our linguistic competence is able to decode those violations in such way that we are able to understand them. This decoding process is different from one person to another and some speakers are more skilled at them than others, which must be because their linguistic competence for processing language is higher than that of others. Natural language speech is everything but homogeneous. Linguistics does not have the privilege to reject phenomena that are actual language by tagging those phenomena as "performance", especially when the phenomena are actually linguistic competences such as the degrees of grammaticality.

2.3 Grammaticality and Acceptability

Chomsky establishes that *grammaticality* belongs to *competence*, while *acceptability* belongs to *performance* (Chomsky, 1965:11).

Note that Chomsky accepted the existence of degrees of grammaticalness as well as degrees of acceptability (Chomsky, 1965:11,77). Degrees of grammaticality judge the rules that generate language or describe a grammar. Chomsky claimed that the degree of grammaticalness has an impact on the degree of acceptability:

> Like acceptability, grammaticalness is a matter of degree [...] but the scales of grammaticalness and acceptability do not coincide. Grammaticalness is one of the factors that interact to determine acceptability.
>
> (Chomsky, 1965:11)

However, these degrees are different since the degree of acceptability takes into account both linguistic and extra-linguistic factors. It determines whether a sentence is acceptable to the judgment of the speaker, which is affected "by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attentions and interest, and errors (random or characteristic) in applying his knowledge of the language" (Chomsky, 1965:3).

Along these lines:

- **Grammaticality** is a notion that belongs to linguistic competence: a sentence is grammatical if it is generated by the grammar of the speaker, in accordance with the speaker's linguistic knowledge.

- **Acceptability** is a notion that belongs to linguistic performance: a sentence is acceptable in the judgment of the speaker, according to the behavior that the speaker exhibits, his/her performance data, etc.

In the end, the degrees of "grammaticalness" will be dismissed. Competence is perfect, so there is no reason to take into account degrees of grammaticality. Either the degrees of grammaticalness do not exist or they are useless. Competence is related to grammar, and its rules remain immutable. Any variation from a deep structure to a surface one is a matter of performance.

On the other hand, the degrees of acceptability are widely accepted. Acceptability is imperfect and distorted by psycho-linguistic and extra-linguistic effects. Chomsky acknowledges that degrees of acceptability are real and can be evaluated by subjective psycholinguistic judgments:

> Obviously acceptability will be a matter of degree, along various dimensions. One could go on to propose various tests to specify the notion more precisely [...] for the present purposes, it is unnecessary to delimit it more carefully.
>
> (Chomsky, 1965:10)

2.3.1 Degrees of Grammaticality and Acceptability Judgments

In *Aspects*, Chomsky asserted:

> Like acceptability, grammaticalness, is no doubt, a matter of degrees [...], but the scales of grammaticalness and acceptability do not coincide. Grammaticalness is only one of the many factors that interact to determine acceptability.
>
> (Chomsky, 1965:11)

Chomsky (1961:236) attempts to formally solve the description of degree of "grammaticalness" by assigning to each string of words "a structural description that indicates degree of grammaticalness, the degree of its deviation from grammatical regularities, and the manner of its deviation" (Chomsky, 1965). This is basically adding information about the deviation identified in each string.

Askedal et al. (2010) summarizes how Chomsky had a clear way of measuring the degrees of grammaticalness:

> The degree of grammaticalness is a measure of the remoteness of an utterance from the generated set of perfectly well-formed sentences, and the common representing category sequence will indicate in what respects the utterance in question is deviant.
>
> Askedal et al. (2010)

This is establishing a degree from the gold standard to the bottom of the deviation feature. However, it is not clear whether this degree of non-grammaticalness can be less than zero (infinite) or a closed scale on which after a certain amount of violations an utterance is always zero.

As we noted earlier, however, because of the abstract concept of competence and grammaticality, *grammaticalness* is finally considered in Chomsky's work and in generative grammar as an abstract perfect object, and therefore without degrees. The reasons for an unacceptable grammatical sentence therefore have to do "not with grammar, but rather with memory limitations, intonation and stylistic factors, "iconic" elements of discourse (for example, a tendency to place logical subject and object early rather than late [...]) and so on" (Chomsky, 1965:11).An unacceptable grammatical sentence is therefore a matter of performance. However, if both are considered elements of degree (since they are different), acceptability judgments are not sufficient to determine the grammaticality of the sentence, so the degree of grammaticalness cannot be extracted from them (Chomsky, 1965:11).

However, an important study in Keller (2000:200), Keller (2006) and in Sorace and Keller (2005) explores how competence factors interact to determine the degree of acceptability. Keller aimed to "fill this gap providing a systematic experimental investigation of how competence factors interact to determine the degree of acceptability of a linguistic structure" (Keller 2000:20). He pursued competence explanations in the absence of systematic performance explanations (Keller 2000:44). Extracting the degree of grammaticality from the degree of acceptability from linguistic data by means of rules (constraints): "our investigation is to extend the empirical scope and the theoretical reach of models of linguistic competence" (Keller, 2000). He claimed that with his methodology it is possible to isolate the degree of grammaticality from the degree of acceptability provided by a speaker in a psycho-linguistic test underpinned with a specific methodology. Keller, probably inspired by the ideas of Chomsky in measuring grammaticalness, extracts the gold standard or the *optimal* from the acceptability judgments. Then, depending on how many *re-rankings* a constraint (rule) needs, an input will either be more grammatical or more non-grammatical. Actually, he is providing a **degree of non-grammaticality** since this degree tells us how bad an input is, not how good it is. This degree is always going to be a negative number:

> The degree of grammaticality of a structure depends on the number of re-rankings necessary to make it optimal: the more re-rankings a structure requires, the more ungrammatical it becomes.
>
> (Sorace and Keller, 2005:1517)

Keller asserts that the linguistic rules (or constraints) are the objects that define the knowledge of the grammar. Since the rules of a grammar are theoretical objects, the distinction between competence rules and acceptability is also theoretical:

> We assume that some aspects of gradient data are due to factors that pertain to grammatical competence, while other aspects are due to performance effects. The decision which aspects to subsume under competence and which ones to treat as performance is ultimately a theoretical one; it cannot be settled on purely empirical grounds.
>
> (Keller, 2000:28)

Following on from this, the rules of a sentence can be isolated by means of violations, and the degree of grammaticality can be extracted from a native speaker.

Although Keller's work provides an interesting and necessary approach to the matter of degrees in linguistics, his theoretical bases are a little confusing. He provides degrees of grammaticality following Chomsky's distinction but, at the same time he aims to explain competence in a multi-modal sense by means of degrees, which, in my opinion, is a contradiction:

> It does not make any sense to speak of grammaticality judgments given Chomsky's definitions because people are incapable of judging grammaticality –it is not accessible to their intuitions [...]. The present thesis follows Chomsky's definitions and treats the terms acceptable and grammatical as distinct [...]. We will work on the assumption that gradience is best analyzed in terms of linguistic competence.
>
> (Keller, 2000:28-29)

Keller is interested in maintaining the distinction between competence and performance in order to justify the theoretical extraction of grammaticality from acceptability. If he accepts that both are the same, he must also accept that either the degree of acceptability is equal to the degree of grammaticality or it is impossible to extract grammaticality from acceptability because all the extra-linguistic factors that occur in an acceptability judgment cannot be split from the grammaticality value. However, he proposes a method in which he claims that by taking into account linguistic constraints in a specific linguistic context he can split the grammatical value from the acceptability value.

The main problem with working with acceptability judgments to explain grammaticality concerns linguistic multi-modular interference. Speakers parse sentences from a holistic point of view, so all the linguistic domains are involved when a speaker provides a "grammaticality judgement". Although Keller's methodology is good, he admits that some constraints (soft constraints) depend on multi-modal compensation between the domains of the grammar of a natural language:

> The hypothesis underlying the work reported in this paper is that hard constraints are purely structural (i.e. syntactic) in nature, while soft constraints are at the interface between syntax and other domains (i.e., semantics or pragmatics).
>
> (Sorace and Keller, 2005:1521)

Some problems arise here:

- Firstly, following Chomsky's distinction, treating the terms of grammaticality and acceptability as different is a contradiction since, in Chomsky's distinction, pragmatics are a component of performance, while Keller considers them a component of competence that plays an important role in the final value of grammaticality. For Keller, pragmatics is a domain that can soften the value of grammaticality and for this reason he must place it in competence. The concepts of competence and performance are therefore not the same for Chomsky as for Keller. Chomsky's distinctions leave no place for considering that interference between syntax and pragmatics is a matter of grammaticality. Given Chomsky's definition of grammaticality and competence and taking into account the deep structures defining linguistic competence, grammaticality is almost merely structural. Pragmatics are not a component of linguistic competence according to Chomsky's distinction since they take place in surface structures at the level of performance. The definitions of competence, performance, grammaticality and acceptability, and the distinction between competence and performance should therefore be reviewed by Keller rather than fully agreed. Additionally, the theoretical point here is that once we accept that pragmatics takes place as a component of linguistic competence, there is no reason to keep Chomsky's distinction and conception of competence and performance. One may think: where, then, is the limit between competence and performance? Once every domain of the language is considered a component of competence, what is performance? Unfortunately, these theoretical questions are not solved in Keller's impressive attempt to explain the degrees of grammaticality and acceptability.

- Secondly, we do not know the weight for each rules (constraints) in the grammar. The syntactic constraints have not been fully isolated to determine each weight for each constraint independently of other multi-modal "interference". The structural constraints are considered hard because they cannot be balanced with any other constraint from other modules. Without a holistic evaluation of the whole constraints of a grammar, we cannot be sure that hard violations come exclusively from the syntactic module and not from various violations of other constraints from other modules. The structure of a sentence might therefore be considered to be strongly deviated from a grammar when the violated syntactic constraints trigger (or is a reflection of) a chain reaction of other constraints that have been violated in other modules in the same input. The soft weights are therefore soft because other modules balance them since no other rule (apart from the syntactic one) has been violated in an input.

From Keller's point of view, therefore, we do not know the weight of the syntactic rules from any other module's constraints. The total weight of grammaticality is mixed through a multi-modal point of view. The only possible conclusion is that it is improbable (nowadays) to extract, from the acceptability judgments, single-modular grammaticality such as the degrees of grammaticality regarding specific constraints in the syntax of natural language. We therefore conclude that the only way to extract the weights of grammaticality is from a mathematical-theoretical point of view rather than extracting them from acceptability judgments from speakers. Linguistic competence is considered an abstract-theoretical notion.

For this reason, grammaticality weights must be extracted following an abstract-theoretical method that is closer to mathematical and formal approaches than to psycho-linguistic ones.

In summary, dealing with degrees of grammaticality without tackling Chomsky's distinction of competence and performance seems impossible. Keller had to agree to distinguish between the concepts of grammaticality and acceptability in order to provide degrees of grammaticality by extracting grammaticality from acceptability judgments. If he had provided anything else, such as "grammaticality judgments", he could never have claimed that he was working on linguistic competence. However, he could not avoid the importance of pragmatics in the phenomenon of grammaticality. This creates a problem for the classic distinction between competence and performance since pragmatics are clearly placed in performance, and yet this domain clearly has constraints that condition and represent the linguistic knowledge of a grammar, such as pragmatic universals from Kasper (1992), maxims from Grice (1975), constraints in relation to the speech acts of Searle (1969), and so on. The classic distinction between competence and performance cannot hold pragmatics in the competence level nor can they hold other domains that are typically tagged in the performance level. I will discuss this theoretical puzzle in the following subsection.

2.3.2 Classifications for Grammaticality and Acceptability

To clarify a certain confusion, Chomsky described several situations in which grammaticality and acceptability differ (Chomsky, 1965:10-15; 148). Some authors, such as Sternefeld (1998) and Lakoff (1973),classified these situations.

Sternefeld (1998) derives the features of acceptability and grammaticality to the following classification:

- **Grammaticality without acceptability**. Some sentences are considered by linguists as grammatical but are still rejected by native speakers on performance grounds. Examples of this are the garden path sentences or center embeddings such as: *"Yesterday I spoke with the father of the mother of the cousin of the sister of the nephew of your friend's grandfather"*. Clearly, such sentences are very hard to process and are therefore deemed unacceptable. However, they are grammatical since they are structurally correct and the grammar can generate them.

  One of the most famous cases of this point is from Chomsky's *Syntactic Structures* (Chomsky, 1957) with *"Colorless green ideas sleep furiously"*. Chomsky claimed that such sentences are grammatically correct but semantically nonsensical: "the notion of *grammaticalness* cannot be identified with *meaningful* or *significant* in any semantic sense" (Chomsky, 1957:15). He used this sentence to assert that grammaticality and competence are purely structural and a matter of syntax. By contrast, he pointed out that *"Furiously sleep ideas green colorless"* is non-grammatical because while the first sentence is well-structured, the second one is upside down.

Chomsky's proposal aimed to provide a counter-example to statistical approaches (Markov) by showing that two remote sentences from English do not necessarily have the same value of grammaticalness. He therefore claims that "grammatical in English cannot be identified in any way with the notion 'high order of statistical approximation to English"' (Chomsky, 1957:16-17). However, these example sentences treat grammaticality as a purely structural feature. If we consider grammaticality as a wholly grammatical value, we must accept that neither is fully grammatical but that the first sentence satisfies more rules than the second one. Since the first sentence satisfies more syntactic rules than the second one, presumably it is still more probable than the second one because native speakers tend not to produce fully non-grammatical sentences because they wish to be understood.

Moreover, this example contradicts Chomsky's future definitions of competence, performance, grammaticality and acceptability since he rejected "sense" (semantics) as a feature for "grammaticalness" but used performance traits such as intonation and acceptability judgments to explain grammaticality between these two sentences:

> These sentences a speaker of English will read (1) with a normal sentence intonation, but he will read (2) with a falling intonation on reach word; in fact, with just the intonation pattern given to any sequence of unrelated words. He treats each word in (2) as a separate phrase. Similarly, he will be able to recall (1) much more easily than (2), to learn it much more quickly, etc.
>
> (Chomsky, 1957:16)

To clarify the confusion between grammaticality and acceptability, Chomsky resolved this point, as we mentioned earlier, by considering competence as a purely structural perfect object and dismissing grammaticality as a matter of degrees. Additionally, in language it is possible to find grammaticality with meaningless structures if a grammar considers semantics outside a grammar, i.e. in the performance level: "Grammar is best formulated as a self-contained study independent of semantics. In particular, the notion of grammaticalness cannot be identified with meaningfulness" (Chomsky, 1957:106).

- **Acceptability without grammaticality**. This situation is very common. Because of the idiosyncrasy of natural language, we often produce sentences that are not fully correct, though we accept them. However, some linguists have pointed out that a sentence can be highly deviated and yet still acceptable. Examples can be found in Gibson and Thomas (1999), which provides the results of an experiment with nested relative clause structures. They consider the appearance of the verb as a very important constraint in English. With this criterion, they point out how, given a three-nested relative clause structure, a sentence with only two verb phrases is sometimes more acceptable than one with the grammatically required three:

  – (++Grammatical) *"Jack met the patient who was admitted by the nurse who the clinic had hired."*
  – (+Acceptable, -Grammatical) *"The patient who the nurse who the clinic had hired."*
  – (-Acceptable, +Grammatical) *"The patient who the nurse who the clinic had hired admitted met Jack."*

- **Overdetermined and undetermined cases**. This point comprises constructions where one grammatical rule (or more) is applicable in its multi-modality (Sadock, 1998). The subject-verb agreement rule in English, which concerns to both semantic and grammatical number, would be an example of a multi-modal rule. Because of the multi-modal complexity of the rule, these situations appear to manifest variable judgments.

Sternefeld (1998) approach is simply an amplification of the first approach from Chomsky, and we could find more linguists who follow or clarify them.

2.4 Grammaticality and Acceptability in this Work

Grammaticality is the value that represents how satisfied an input is according to the linguistic knowledge (constraints) that defines a natural language grammar. The degree of knowledge of a grammar is related to a speaker's degree of linguistic competence. This value is fuzzy since it takes into account numerous criteria that make it vague. The vagueness of the fuzzy value of grammaticality will be determined by a value in terms of degrees giving a certain amount of satisfied or violated criteria based on a grammar with constraints.

This definition may satisfy many linguistic approaches in which a distinction between *grammaticality* and *acceptability* exists. While also making this distinction, we take into account that every object that could be explained by rules or linguistic constraints is a component of competence. In other words, once a linguistic object is a rule/constraint, it is a component of competence.

Distinguishing between *grammaticality* and *acceptability* can be theoretically accepted in our work following next clarifications:

- Agreed is Chomsky's idea that considers the fact that the degree of grammaticality is a component of the degree of acceptability but that both degrees are not the same.
- The degree of grammaticality is the result of the satisfaction and violation of the linguistic rules that characterize the linguistic knowledge, while the degree of acceptability is essentially a subjective evaluation.
- The degree of grammaticality evaluates linguistic competence only, while the degree of acceptability subjectively calculates linguistic competence plus all the other variables involved in the performance of the language. This subjective evaluation can be influenced by attitude, predisposition and various psycho-linguistic variables. This subjectivity is inherent in the judgment of an input. A speaker provides an estimation based on his/her subjective perception of the weights of each rule. In this sense, two speakers who are judging an input can provide two very different results. On the other hand, an input's degree of grammaticality is always the same for any given grammar since the rules are weighted theoretically using formal, mathematical methods.
- The degree of grammaticality is a theoretical value representing the empirical value of grammaticality implied within acceptability judgments. However, because human cognition parses the language altogether, it appears impossible with empirical methods to provide a value of grammaticality for each rule, domain or input. Formal, mathematical methods are a better option for accurately explaining the phenomenon of grammaticality. They enable us to propose theoretical weights and values for each rule or constraint in every module of the language. In this respect, applying mathematical and formal methods provides a theoretical solution to a theoretical problem. The empirical and theoretical values could be contrasted once this solution is applied to every module of the grammar. A positive result would qualify this method as empirically valid.

In Table 2.2, we summarize the distinction between degrees of grammaticality and acceptability.

While the notion of *correctness* is related to the notion of satisfying a prescriptive grammar, since there is an incorrect option or an error, grammaticality belongs to descriptive linguistics. Grammaticality refers to the degree of satisfaction with the rules of an utterance

Linguistic Value	Evaluation of	Object of Evaluation	Nature of the Object
Degrees of Grammaticality	Competence	Rules/Constraints (Knowledge of a Language)	Theoretical
Degrees of Acceptability	Performance	Grammatical objects -Grammaticality Extra-linguistic objects -Attitude -Psycholinguistic variables	Theoretical & Empirical

Table 2.2 Diagram of features of degrees of grammaticality and acceptability.

by a specific grammar of a natural language. Sometimes the notion of non-grammatical object is mentioned. However, a real degree of grammaticality will always be provided positively in the strict sense for grammaticality and not in the negative sense of ungrammaticality or non-grammaticality. We are interested in determining to *which degree an input satisfies a grammar rather than evaluating how much an input violates a grammar.* The motivation is therefore clearly descriptive since it takes into account the degree. This is very different from the prescriptive motivation, which involves saying what is correct and what is not.

Traditionally, most authors in linguistics have used obscure marks such as “*”, “?”,“??”,“???”, to define grammaticality. These marks are typically used in acceptability judgments. However, they have often been used to mark grammaticality, in which case they have become quite superfluous. The use of these marks is inconsistent and based on subjective assessments. There are no objective reasons to use them to determine why an input is “??” instead of “?” or even “*”. In this thesis, we provide new values of grammaticality in terms of degrees; ranging them in a continuum from 1 to 0 and using them for representing grammaticality as competence) rather than a component of acceptability judgments (performance).

Chapter 3

Gradience in Linguistics

Gradience is an essential concept in order to provide a framework to deal with the degrees of grammaticality. Gradience offers indispensable theoretical concepts that have laid the foundations of our work. However, there are different approaches to linguistic gradience, and not all of them are focused on explaining the degrees of grammaticality phenomenon. In this chapter, we first introduce the concept of gradience, which is the theoretical core of our framework. Secondly, a brief history of gradience in linguistics is shown, as well as the most recent theories from gradience which make an effort to deal with the degrees of grammaticality. However, we will show how the degrees of grammaticality are still a topic that gradience has not fully solved with a framework. Finally, a discussion of how gradient data is modeled in order to represent gradience in either a grammar or a theory is provided.

3.1 What is Gradience?

First approaches of what is gradience can be found in ancient philosophy. Aristotle in *The Metaphysics* defended that two things that are not equally wrong cannot be equally even since one is more true than the other (Aristotle, 1984):

> Again, however much all things may be so and not so, still there is a more and a less in the nature of things; for we should not say that two and three are equally even, nor is he who thinks four things are five equally wrong with him who thinks they are a thousand. If then they are not equally wrong, obviously one is less wrong and therefore more right. If then that which has more of any quality is nearer to it, there must be some truth to which the more true is nearer. And even if there is not, still there is already something more certain and true,

> and we shall have got rid of the unqualified doctrine which would prevent us from determining anything in our thought.
>
> (Aristotle, 1984) (cited in Aarts (2004b))

Gradience has appeared throughout the history of linguistics. Probably the first author who introduced the term *gradience* in linguistics was Bolinger (1961) with *Generality, gradience, and the all-or-none*. He worked on a detailed compilation of phenomena in natural language which needed of a gradient approach, rather than a discrete one, in different linguistic domains such as syntax, semantics, and phonology.

In Aarts (2004b:1), gradience is defined as "a cover term to designate the spectrum of continuous phenomena in language, from categories at the level of the grammar to sounds at the level of phonetics". Along with this, gradience is a term used in linguistics to design phenomena which, because of their apparent blurred nature, fit better in a non-discrete explanation such as an explanation in terms of degrees.

Therefore, the basic idea of gradience is sorting out elements that are not equally even but vague in terms of degrees.

Aarts (2004b) defines two basic types of gradience in grammar:

1. **Linear gradience**: We have two different categories clearly defined: α and β. Gradience here is found in the third group of elements fitting to the middle field between these two categories. The idea is to sort out the linguistic objects at the linear scale in three degrees. Thus, in a scale of [0,1], α is 1, β is 0, and the middle-blurred area is 0.5.

α- - - - - - - ///////////- - - - - - - β

Fig. 3.1 The representation of linear gradience (Aarts, 2004b).

2. **Set-theoretic gradience**: We have two different categories clearly defined (α and β). Here gradience is found when an intersection between the two categories occurs: $\alpha \cap \beta$. The elements which share features from both categories are pointed out in the intersection.

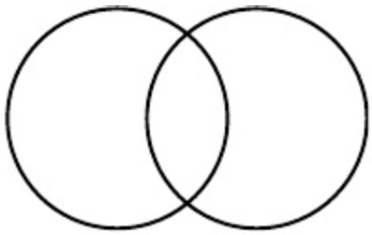


Fig. 3.2 The representation of set-theoretic gradience (Aarts, 2004b).

Aarts (2004a) in *Modelling linguistic gradience* distinguishes between two specific forms of lexical gradience: Subsective Gradience (SG) and Intersective Gradience (IG):

1. **Subsective gradience** is defined as "intra-categorical in nature, and allows for members of a class to display the properties of that class to varying degrees" (Aarts, 2004a:1). This type of gradience takes into account a prototype as the highest degree on a lexical scale. An element x would be closer to the prototype of a category α than an element y if the element x satisfies more prototype requirements than the other. As an example, in English, *"house"* would be a more prototypical noun than *"home"* with respect to determiners and quantifiers; *"house"* is also less subject to idiomatic use.
2. **Intersective gradience** is conceived as a phenomenon between two categories in which two (or more) categories can adopt a non-prototypical role (or fit, such as either syntactic or semantic fit). In this case, intersective gradience represents the relationship between two categories α, β (and more if so) concerning a set of elements γ. This set of elements share properties from both categories, but those elements cannot fit completely in any of α and β category. Aarts (2004a) argues that "α and β themselves are strictly bounded, and do not overlap [...]. The intersection is between γ and the full set of α-like properties, and between γ and the full set of β-like properties". As an example, when a noun is used as an adjective and vice-versa such as *"race horse"*, *"tennis shoes"*, and *"the rich is powerful"*, *"the blind is losing hope"*. We also find morpho-syntactic examples in some verbs with *-ing* as a set of elements which share nominal and verbal properties.

However, all of these cases of gradience are based on considering that a candidate does not entirely fit a prototype category or a canonical; thus this canonical has been violated in some degree, and it needs to be both represented and placed. Thus, the main objective in linguistic gradience is to rank and explain an element in terms of degrees either with a linear or a set-theoretic representation; either if it is a subsective or an intersective situation.

At first sight, these types of gradience show up some problems. On the one hand, *linear gradience* is too discrete, because it just shows three degrees. Hence, it looks merely like a discrete notion with a third degree. On the other hand, *set-theoretic gradience* is too vague. It just explains the shared features without pointing out any degree.

Some detractors of gradience use to show up that these approaches are not as gradient as they claim. Muñoz Cuadros (2006) points out that it is contradictory to define an element as vague but yet to classify it discretely, even if this discrete way has three degrees. For these objectors, there is no need to talk about gradience since the authors who use gradient approaches keep using mostly two degrees with a third one in which they fit all they cannot define. Hence, in the end, it is one way or the other. This kind of critics to gradience are frequent. They are motivated due to a misunderstanding of what is gradience. The discrete view is gradient itself, but it is a very basic kind of gradience. The main difference between the discrete and the gradient approach is the following one: the discrete approach deny gradience while gradience accepts the discrete approach and yet try to improve it in order to establish a rank between that 1-0 notion. Where the discrete approaches just see *black* (α) and *white* (β), gradience accept *black, white*, and all the *possible grays* between those α and β, just like in Figure 3.3. It is necessary to accept that it is already worth to classify an object with only three or four degrees if just by adding those degrees we are nearer to explain the world in a better way.

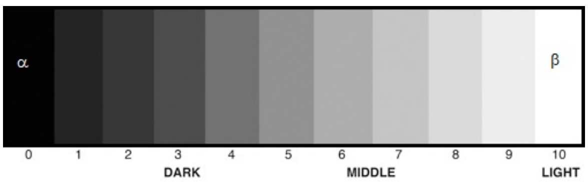


Fig. 3.3 *Black* (α) and *white* (β) are discrete objects but in a gradient relationship.

Actually, in our regard, set-theoretic gradience and linear gradience work together. While set-theoretic gradience describes the shared features of α and β in γ, linear gradience provides the degree of membership of γ according to the number of features shared from α and β. This mixture is going to be shown in the features of our fuzzy grammar.

3.2 History of Gradience in Linguistics

We can find an excellent paper by Aarts (2004b), *Conceptions of gradience in the history of linguistics*, in which the author provides us a vast and detailed history of linguistic gradience from the ancient philosophers to the 20th century. This section presents a summary of this paper in order to provide to the reader a general knowledge about the history of linguistic gradience and why it is still a matter of interest. A better understanding of the history of gradience will show up the convenience of proposing a new theoretical grammar framework which represents the degrees of grammaticality in natural language.

3.2.1 Gradience in the Early 20th Century

Grammarians used to be fully discrete until the 20th century. Otto Jespersen, in the early part of the 20th century, has been considered as an Aristotelian descriptive grammarian. However, in my opinion, he has described as well some bits of the nature of the language as vague (rather than discrete) in terms of linguistic variability, especially in his work (already mentioned) *The Philosophy of Grammar* from 1924. He offered many examples of linguistic variability and the consequent vagueness in language. Such as the behavior of the linguistic categories (or part-of-speech) in the closed/opened sentences (Jespersen, 1924:12). He considered that some sentences were totally "closed" such as *"Long live the King"*, in which the substitution of one of these words without having an odd sentence is not possible. While some sentences such as *"My house is pretty"* allow more variation without triggering a rare sentence since: *"The house number 10 in Jefferson street is beautiful"* and *"John's house looks amazing"* can be the same sentence from a deep point of view, all the houses are the same, and all are expressing (in a general regard) the same sense. However, exchanging some of these parts for other categories or words would trigger an odd one such as *"My house is a dog"*.

Consequently, Jespersen (1924) did not mention a gradient or vague relation between categories, but he exposed how the language exhibits a certain degree of variability in terms of open and closed expressions. Somehow, from my point of view, he was considering the utterances as a pair between structure and meaning which determines how discrete or vague an utterance could be in terms of linguistic variability. He claimed that because of this nature from the structures in the language, the speaker could not be just a slave of the habits, he needs to change them in order to fit with the new communication necessities which can trigger both new linguistic habits and grammatical forms (Jespersen, 1924:23).

Curme (1935), another descriptive grammarian, described a similar phenomenon in which he represented the dual nature existing between categories such as, in this case, between the adjective and the noun. He distinguished three degrees without using, explicitly, the word gradience:

1. The *noun degree*, being the noun formally used as an adjective and perceived as a noun (*"stone bridge"*).

2. The *medium noun-adjective degree*, that is when a noun is used with adjective elements, these are used as a bonded element to the adjective but perceived as a noun (*"John's book"*).

3. *The adjective degree*, those are plain adjectives of indisputable use (*"happy woman"*).

3.2.2 Sapir, Bloomfield, and post-Bloomfieldans

Even though some authors in linguistics were making some gradient approaches as an exception, in the 20th century, Sapir and Bloomfield agreed in the fact that linguistics should face the indeterminacy in category boundaries within a continuum approach.

The post-Bloomfieldans such as Hockett (1955) or Joos (1958) reacted to these claims by assuming that the language must be defined by discreteness. Some claims were very strong such as: "If we find continuous-scale contrasts in the vicinity of what we are sure is language, we exclude them from language" (Hockett, 1955:17).

The discussion if either *discreteness* or *gradience* should represent natural language had a philosophical and digressing core rather than functional, in terms of proposals, for a long time. This topic was exhorted in some congresses such as in the Eighth International Congress of Linguists where Wells (1958) defended the necessity of being friendly with the gradient approach. Crystal (1967) proposed a new way to define membership of part-of-speech through a multi-modal approach in terms of degrees, and to introduce some "bridge classes". Quirk (1965) defended gradience from those who stated hardly for discreteness in language. He kept working on gradient descriptions for linguistics, and defining gradience in Quirk et al. (1985:90) as "a scale which relates two categories of description (for example two word classes) in terms of degree of similarity and contrast". Therefore, from this author's perspective, gradience is in every part of the grammar, and not strictly related to the categories or part-of-speech relations and membership.

However, the big reaction against the post-Bloomfieldans came from Dwight Bolinger (1961) in *Generality, gradience, and the all-or-none*. This one was a brave attempt in which the author exhibits some gradient examples, defending gradience from the point of view of the acceptability judgments. However, this important work was focused on two main tasks.

- Firstly, on highlight the necessity of applying gradience in linguistics by doing pedagogy and giving philosophical and pragmatical arguments against some contexts in which the discrete notions were non-sense.

- Secondly, on providing linguistic examples in which the author attributed, instinctively, degrees of acceptability between sentences and then he checked his assumptions by means of test applications.

His work was received with skepticism in the linguistic community. Some authors hardly criticize him such as Stockwell (1963). However, Bolinger's work is, without doubt, pioneering, but he was a man of his time. Some of his examples were anecdotal, and his proposals needed a better methodology.

3.2.3 Gradience and the Generative Grammar

Chomsky was following the lead of considering language such as a gradient object when it comes to talking about degrees of grammaticalness and acceptability (Chomsky, 1957, 1961).

Chomsky claimed that "a partition of utterances into just two classes, grammatical and non-grammatical, will be not sufficient to permit the construction of adequate grammars" (Chomsky, 1965). However, in the end, Chomsky's work, transformational grammars, and many of the generative linguists have been rejecting the gradient and fuzzy or vague notions in grammar approaches. Either by means of methodological arguments or because of a lack of interest in non-discrete approaches.

This attitude towards gradience comes from the theoretical distinction of competence and performance which rules generative frameworks. Because of this distinction, they used to avoid empirical data since they conceived grammar as a perfect mathematical framework in which everything is either-or-that in the deep structures, placing the vagueness, randomness, and empirical data as things that might occur in the performance. This trend has been prevalent in linguistics, and it has been discussed in the chapter 2 of this PhD thesis.

Newmeyer (1998:696) has staunchly defended the necessity of the discreteness in linguistics, claiming that gradient approaches are weak since they come from facts placed in the linguistic performance. However, more recent works, in the generative perspective, do not close the door to the possibilities that a gradient approach can offer to lay down the basis for building linguistic frameworks such as in Pinker (1999), Borsley and Kornfilt (2000), Van Riemsdijk (1998, 2001), and Corver and van Riemsdijk (2001).

3.2.4 Zadeh and Fuzzy Logic Feeding Linguistic Gradience

During the 1960s, Lofti A. Zadeh, and those who were following his approach dealt differently with gradient phenomena in comparison with the generative theories and other discrete trends.

Zadeh (1965, 1972, 1975) mathematical description of gradience phenomena is well known. He describes the variable semantic values of words, or fuzzy phenomena, in terms of degrees.

Zadeh's contributions were mostly mathematical and yet he dealt with specific linguistic phenomena in terms of degrees. However, he did not develop a formal linguistic framework to describe fuzziness in natural language grammar.

Both linguistic gradience and fuzzy logic aim to represent the relationship between two categorical objects as a scale of degrees rather than discrete. In other words, "a class in which the transition from membership to non-membership is gradual rather than abrupt" (Zadeh, 1965).

This fuzzy reasoning can be applied to linguistic gradience in order to provide different kinds of values of an input such as category membership, grammaticality, lexical and semantic value, and so on. Thus, rather than classifying linguistic items as non-grammatical when those show some grammatical deviation (discrete reasoning), these can be classified as a more or less grammatical according to the features that are violated or satisfied in them.

Zadeh's work was pioneering, and a beacon for those who were attempting to work with gradient linguistics by means of a serious theoretical base. Many authors from the Prague circle of linguists were inspired by this new model called fuzzy logic; some of the most notable were: Daneš (1966), Vachek (1966), and Neustupnỳ (1966).

Without a doubt, Zadeh's work had a significant repercussion in many fields and many authors, as well as in linguistics. Mainly, regarding linguistic gradience, it is especially necessary to highlight his effect on the work of Ross (1972) and Lakoff (1973).

3.2.5 Ross: An Effort to Work with Gradience in Linguistics

John Robert (Haj) Ross was the linguist who most argued for category conflation during the 1960s and the 1970s. Ross approach was probably very much influenced by Zadeh (1965, 1972) fuzzy logic and his way to deal with vague objects.

His work is classified as generative semantics standing as a reaction between the descriptive tradition of Bolinger, and the generative theories, in the 1970s. Generative semantics were using data, such as Bolinger (1961), to deal with their formal theories. They tried to fit the gradient perspective among with the formal theories.

Ross created the notion of *squishy* categories to deal with gradience in category tagging. By *squishy* categories, Ross meant those categories which can exchange categorical fits or roles between one to each other such as in 'Adjectives as nouns' (Ross, 1969a).

Aarts (2004b) points out that Ross dealt with two types of gradience, *gradience within a category* (Ross, 1973a) and *gradience between categories* (Ross, 1972, 1973b), without actually making any distinction between them. This classification is close to the one proposed in Aarts (2004a), concerning *subsective gradience* and *intersective gradience*. Other works by Ross in this topic are Ross (1969b) and Ross (1974).

Ross' work was more accurate than Bolinger in representing gradience. Harris (1993:220) points out that "squishiness was not a hit" in the history of linguistics, because Ross used to list valuable observations and mentioning phenomena rather than proposing a grammar system for gradience which includes his notion. He had some subsequent detractors such as Pullum (1976) and Bever and Carroll (1981), but also some defenders such as Schutze (1996).

3.2.6 Lakoff and the First Fuzzy Grammar to Deal with Gradience

Lakoff has been the first linguist that aimed to developed a fuzzy grammar in order to represent linguistic gradience, with particular attention towards the degrees of grammaticality.

In 1970, he published the paper *Linguistics and Natural Logic*, in which he established the bases of what a fuzzy grammar should be, as well as significant claims defending gradience and its vagueness: "The violation [of a rule of grammar] only guarantees that the sentence will be ungrammatical relative to a given reading. A sentence will be fully ungrammatical only if it is ungrammatical relative to all readings" (Lakoff, 1970:154-155).

In 1973, Lakoff published the first paper in which a linguist mixes fuzzy logic formalism with fuzzy figures and linguistic examples (Lakoff, 1973). In this work, Lakoff represents in a more precise way how natural language sentences are not either true or false, or nonsensical, but they are true or false to a certain extent, or in certain respects.

In April 1973, Lakoff released his work in which the phrase *fuzzy grammar* was heading a linguistic paper for the first time: *Fuzzy Grammar and the Performance/Competence Terminology Game*. Here, he provided the first theoretical base framework that claimed for a fuzzy grammar, represented language with his formalism, and argued strongly against the conception of the competence and performance established by Chomsky.

Lakoff admitted that his work was highly influenced by Zadeh's and Ross' works results, with a special mention to this one, since he put so much effort in his undertaking with gradience in the absence of a theory of a fuzzy grammar.

Lakoff's proposal of fuzzy grammar reaches its peak in 1987. He finally approaches his model towards to the prototype theory in *Cognitive models and prototype theory* (Lakoff, 1987a) and in *Women, fire and dangerous things: What categories reveal about the mind* (Lakoff, 1987b). Even though that prototype theory and fuzzy logic are not the same, both tools can complement each other, since *prototype theory* is a cognitive theory representing the mind and *fuzzy logic* is a mathematical theory representing logic descriptions.

3.2.7 Prototype Theory in Linguistic Gradience

Prototype theory had a significant impact on cognitive linguistics, as well as for explaining linguistic gradience from its point of view.

This theory became very strong in linguistics thanks to the work of the psychologist Rosch (1973a,b, 1975, 1978).

The main idea is to represent how different objects which have a resemblance to the prototypical object are not categorically excluded from that object, but related. In this sense, humans are able to make general assumptions even if they are not entirely true because of the prototypes. Hence, when we can claim that *"all birds can fly"*, even though it is not entirely true, it is profoundly true, and the prototype is the one who let us understand each other because of the representation of that prototype in our mind.

This idea has been applied in linguistics in order to deal with those categories which are a problem for the discrete classifications. Linguistics works with prototype theory in order to provide flexible categorization regarding mind representation of that category. Lakoff (1987a), Taylor (1995), Labov (1973), Bybee and Moder (1983), and Moure (1994) are some of the linguists which worked fuzzy category boundaries with prototype theory.

This trend received strong criticism, usually because many detractors misunderstood that this trend does not fully deny the Aristotelian classifications. As well as, others point out that the new gradient classifications with prototype theory do not provide new solid classifications which are better or much different than the discrete ones. Some of the most rigid critics come from Bouchard (1995), Newmeyer (1998) and Muñoz Cuadros (2006).

3.3 Constraints: A Tool for Linguistic Gradience Theories

Gradience received much attention during the 20th century. However, during this period, we found works nearer to the compilation of phenomena rather than framework proposals which can deal with the phenomena of gradience for a natural language grammar. Most of the works on gradience are focused on word-classification, or phonology, rather than on syntactical phenomena.

Some linguistic trends might accept or even implement the notion of grammatical vagueness in their work, such as the descriptive grammarians who are following Matthews (1981) and Quirk et al. (1985). Some authors such as Huddleston (1984) and Huddleston and Pullum (2002) feel comfortable by admitting vague-gradient approaches in the grammar partially, but they do not recognize the full presence of gradience in grammar. Therefore, for example, Huddleston defends that some cases are all-or-none.

These half-way statements such as in Huddleston (1984) are common in the history of gradience. Grammaticality and syntax seem to be too valuable for being explained in terms of degrees since Chomsky's classifications of competence-performance and grammaticality-acceptability seemed strongly accepted by almost the whole community. However, in the 90s and the 21st century, using linguistic constraints as a theoretical tool to model natural language processing starts to gain prominence. Constraints appear as an alternative to the generative rules. Their basic form is very similar to an *if-then* rule, and it represents properties that an object must satisfy.

Frameworks with constraints are used mainly for two proposes:

1. For evaluating an input determining if that input is acceptable or not.

2. For characterizing an input through a grammar with constraints.

Hence, the constraint became an essential tool for building theoretical frameworks for both evaluating natural language inputs and describing natural language inputs by a grammar.

Some of the most famous frameworks using constraints for dealing with gradience are: Harmonic Grammars (Legendre et al., 1990a), Optimality Theory (Prince and Smolensky, 1993) and its variations such as Linear Optimality Theory from Keller (2000) and Keller (2006), Lexical Functional Grammar in Bresnan and Nikitina (2009), Head-Phrase Structure Grammar in Malouf (1999, 2000) and Hudson (2003), Model-Theoretic Syntax recently worked by Pullum and Scholz (2001), Womb Grammars by Dahl and Miralles (2012) and Property Grammars from Blache (2000, 2016), Prost (2008) and Duchier et al. (2011).

In this period, probability theory appears as a tool which can help linguistics to deal with difficult linguistic phenomena as well as with gradience by using the notion of probability. Manning (2003) provides new perspectives with his probabilistic syntax in order to work with corpus frequencies and probabilities. These stochastic models are going to be helpful to the linguists to deal empirically with specific phenomena such as gradient phenomena and categorization problems. These models open new perspectives and solutions supported by data towards some problems which cannot be fully solved by merely theoretical-discrete perspectives neither by categorical frameworks. These models also work with constraints as their fundamental tool.

This section is dedicated to briefly explain the theories and grammars which use constraints for dealing with gradience. The reader can have with these sections a better idea of what is the situation of gradience and degrees of grammaticality, especially in the 21st century which is when most of these theories has been updated proposing a system to deal with gradience. Even though, in the end, most of them have been conditioned by the competence-performance distinction, and, therefore, assuming gradience only in the acceptability, and not as a phenomenon taking into account grammaticality.

3.3.1 Models with Constraints

The concept of *constraint* comes from computer science. In linguistics, it has been used mostly in syntax and phonology. However, it can be applied to any component of the grammar of a language. In general, the concept of constraint is recognized as very useful for both in the description and in the processing of natural language (Blache et al., 2014). Therefore, many linguistic theories use constraints in their models. The notion of constraint begins to take an important place in linguistics, especially as a tool able to unify different grammars.

The theories that use this notion understand that natural languages have constraints that define the relationship between two elements. These constraints indicate the properties that an object must satisfy. That is, an input is accepted or rejected depending on whether it satisfies or violates the constraints of a language. It is, in short, to stipulate properties that either discard or eliminate structures that do not belong to the language.

Traffic constraints are a clear parallelism with the notion of a linguistic constraint. They stipulate what the driver should do; however, the driver decides either to respect them or to violate them in some degree. In a zone in which we cannot speed up over 40 km/h, we can drive at 30 km/h, or 60 km/h, or 100 km/h and so on. It is up to us to decide at what speed we want to drive. However, the consequences might be more serious; the higher the violation is. Hence, satisfying or violating a constraint has a relation of degree between the driving speed and the constrained-rule. As speakers, we all use a language, the constraints of a language helps us to use it as well as understand what is being said, something similar to acknowledging when we can speed up or not while driving. Depending on what constraints we are violating, our "crash" or success during communication is going to be estimated in terms of degrees concerning the violated constraint and its amount.

There are two types of constraints:

1. **General or universal constraints** that are valid for any language.
2. **Specific constraints** that are applicable to a specific language

Under the generic term *constraint grammars* we find models such as the following: Functional Unification Grammar (Kay and Fillmore, 1999), Lexical Functional Grammar (LFG) (Kaplan and Bresnan, 1982), Categorial Grammar (CG) (Buszkowski et al., 1988), Head-Driven Phrase- Structure Grammar (HPSG) (Pollard and Sag, 1994), Tree Adjoining Grammar (TAG) (Joshi et al., 1975), Optimality Theory (OT) (Prince and Smolensky, 1993), etc.

In the bibliography, a difference is established between three types of formalisms:

1. Models based on the **evaluation of constraints**.
2. Models based on **describing inputs with constraints.**
3. Models which combine **generative mechanisms with constraints.**

The first type is focused on evaluating linguistic inputs, but they need of a grammar with constraints in order to have constraints to evaluate. This first block is comprehended by the Optimality Theory and its variations (Boersma and Hayes, 2001; Keller, 2006; Prince and Smolensky, 1993). This limitation is overcome by the second and the third block of models.

In the second type, the models are based exclusively on constraints to define a grammar which can represent any input to be processed. Therefore, they are not interested in the generation of sentences. Some of these models are Property Grammars (Blache, 2000, 2004, 2016), Womb Grammars (Dahl and Miralles, 2012), or Model-Theoretic Syntax (Duchier et al., 2011; Prost and Lafourcade, 2011; Pullum and Scholz, 2001).

The third type of systems describe without problems the syntactic structures in canonical constructions but offer little information when an input presents some sort of violation. There are theories such as HPSG (Pollard and Sag, 1994), LFG (Kaplan and Bresnan, 1982), TAG (Joshi et al., 1975) and Construction Grammar (Goldberg, 1995, 2003, 2009) among others, that tend to focus on the use of constraints to define the argument requirements between constituents. In fact, the purpose of these models is to generate structures and combine generative mechanisms with constraints. However, both models which define and generate need of a model from the first type if they want to evaluate inputs.

3.3.2 Evaluating with Constraints in Gradience: Optimality Theory

Harmonic Grammars (Legendre et al., 1990a,b, 1991); (Smolensky et al., 1992); (Smolensky et al., 1993); (Smolensky and Legendre, 2006) are the predecessor of Optimality Theory. This theory assumes that every constraint has a numerical weight. The addition of all weighted constraints in a linguistic input would reveal its harmony according to the grammar. The optimal candidate is highlighted with a symbol ($\rightarrow$ in this case). This theory is still discrete since it assumes just one optimal. Table 3.1 shows an example of this formalism.

Differently, from Optimality Theory, Harmonic Grammars take into account both satisfied and violated constraints rather than just the violated ones. Prince and Smolensky (1993:200) claims that “Optimality Theory [...] represents a very specialized kind of Harmonic Grammar, with exponential weighting of the constraints”. Harmonic Grammars are still being improved in recent work such as in McCarthy and Pater (2016).

Weight	4	3	2	1	*H*
Structure	C_1	C_2	C_3	C_4	
a. $Candidate_1$	-1				-4
b. $Candidate_2$		-1			-3
c. $Candidate_3$			-1		-2
d. $\rightarrow Candidate_4$				-1	-1

Table 3.1 Harmonic Grammar, hypothetical example.

The Optimality Theory (OT) by Prince and Smolensky (1993) is probably the most well-known constraint theory nowadays. OT changed the paradigm in the 90s by offering a new framework available for phonology and syntax, primarily as an alternative to Sound Pattern of English (SPE) of Chomsky and Halle (1968) when it comes to evaluating outputs. SPE was considered over-complex and over-generative, while OT offered simplicity and clarity. OT was a hit representing how we understand an input. However, OT does not generate the inputs, also neither the constraints. OT is focused on the evaluation of the candidates in a set. That is why OT, in general, is not considered a grammar but a theory. OT needs from a source, either from a grammar or from a linguist, which provides the rules or constraints which are meant to be evaluated.

Structure	C_1	C_2	C_3	C_4
a. $Candidate_1$	*!			
b. $Candidate_2$		*!		
c. $Candidate_3$			*!	
d. $\rightarrow Candidate_4$				*

Table 3.2 Standard OT, hypothetical example.

Table 3.2 shows how standard OT is rather discrete than gradient. The theory is merely interested in finding the most optimal candidate in a specific set while dismissing the others as non-optimal. The constraints are ranked from the most important one to the less important one. The constraint ranking cannot be changed (C_1 will always be ranked above C_2: $C_1 \gg C_2$) and the constraints are not represented by any weight more than its ranking. Following this reasoning, standard OT impedes that an input which violates the first constraint could be better than the rest of them. The optimal candidate is highlighted by a symbol ($\rightarrow$, in this case).

Boersma and Hayes (2001) offered Probabilistic Optimality Theory (POT) as a variation of OT in which every constraint has a value of acceptability concerning its frequency in a corpus. This approach provides the possibility of presenting constrain re-ranking. Every constraint is represented with a numerical strictness value in a scale of constraint strictness.

In this case, C_1 and C_2 are ranked as $C_1 \gg C_2$, because C_1 has a higher strictness value than C_2 because of the frequencies in a corpus. However, this constraint ranking might change if when adding a small amount of random noise in the strictness values of the constraints (i.e., between C_1 and C_2), the added noise exceeds the distance between the constraints on the strictness scale, re-ranking them as $C_2 \gg C_2$ instead. In this case, POT offers a probabilistic constraint evaluation assuming that, if a re-ranking happens, those candidates that satisfy C_2 are probably more optimal than the others. In POT, grammaticality has attempted to be represented by looking for the optimal candidate, such as in OT. Therefore, again, one candidate is grammatical, that is the optimal one, while the rest are ill-formed. This is illustrated in Table 3.3 and Figure 3.4.

/input/	C_3	C_1	C_2	Freq./Accept.
S_1		*		3
S_2		*	*	2
S_3	*			1

Table 3.3 Probabilistic Optimality Theory, hypothetical example (Keller, 2006:14).

/S, O, V/	VERB	NOM	PRO	Acceptability
O[pro,acc] S[nom] V		*		.2412
O[acc] S[pro,nom] V		*	*	−.0887
V S[pro,nom] O[acc]	*			−.1861

Fig. 3.4 POT table with scores for word order in German (Keller, 2002).

As a response to OT and POT features, Keller (2000, 2006) proposed a variation from standard OT called Linear Optimality Theory (LOT). LOT is a mixed between Harmonic Grammars and Optimality Theory; as seen in Figure 3.5. Keller reveals how *cumulativity* and *ganging up effect* matters when it comes to deciding the most optimal candidate of a series. Therefore, LOT takes into account both the weight of the constraints plus how many constraints have been violated in an input. Thus, an input who violates several constraints can be less harmonic than an input which violates the most important constraint.

LOT defends that it can model gradient acceptability better than OT and POT since it takes into account the ganging up effect, and harmony distinguishing between soft and hard constraints. Actually, Boersma (2004) recognizes that POT has problems to represent cases of harmonic bounding and cumulativity. Boersma proposes a new model called POT' to satisfy those phenomena. However, Keller (2006) denotes some problems still, according to POT' when it comes to model ganging up effects. LOT is able to represent two candidates

$w(C)$	C_1 4	C_2 3	C_3 1	$H(S)$
S_1		*	*	−4
S_2		*	**	−5
S_3			*	−1
S_4	*			−4

Fig. 3.5 Linear Optimality Theory, hypothetical example, displaying cumulativity and weights (Keller, 2006).

as either equally grammatical or acceptable than the other, when in POT' most of these candidates are always represented as one being over the other.

LOT computes grammaticality by comparing the harmonic value of two candidate structures in a candidate set. Therefore, in LOT the absolute notion of well-formedness is called *harmony* while the relative value of ill-formedness is called *grammaticality*.

> Harmony is an absolute notion that describes overall well-formedness of an structure. Grammaticality, on the other hand, describes the realtive ill-formedness of a structure compared with another structure. While it is possible to compare the harmony of two structures across candidates sets, the notion of grammaticality is only well defined for two structures that belong to the same candidate set (i.e., share the same input) [...] we can define the optimal structure in a candidate set as the one with the highest relative grammaticality.
>
> (Keller, 2006:5)

Even though OT has many variations and alternative versions, OT is probably one of the most consistent framework for modeling linguistic gradience; especially when it comes to model acceptability.

Representing grammaticality with OT has been more controversial since OT pictures either one grammatical (optimal) candidate or a grammatical degree of ill-formedness. In this sense, a positive degree of grammaticality is missing. Because of grammaticality means well-formedness, a real degree of grammaticality must provide positive values of grammaticality rather than a scale of negative values, depicting which candidates are better than others but still both well-formed and grammatical, rather than ill-formed.

OT and its use of linguistic constraints was an essential premise for linguists to propose grammatical frameworks with constraints. These grammars with constraints brought to OT what it needed, a grammar build on grammatical constraints that could evaluate linguistic candidates or inputs. Therefore, the phenomena of gradience were mostly proved theoretically in this period until nowadays by combining those frameworks which define a natural language with constraints plus applying OT as a framework to evaluate linguistic inputs with constraints.

Generative trends have accepted OT as one of their theories for explaining linguistic gradience. In *Gradience in Grammar: Generative Perspectives* edited by Fanselow et al. (2006), we can find how these trends widely accept that OT just evaluates candidates for a set of an input on the performance level. Acknowledging that OT can formally represent gradience for acceptability judgments, but OT does not evaluate degrees of grammaticality. Keller (2006) work, where a degree of (un)grammaticality is provided, is included in this publication edited by Fanselow et al. (2006). However, his paper is defined as "LOT is designed to model gradient acceptability judgment data" (Fanselow et al., 2006:14). Even though the attempt of Keller of representing grammaticality, such opinion is based on the fact that neither OT and any of its framework variations are convincing enough for modeling grammaticality regarding Chomsky's competence-performance distinction. Moreover, in my opinion, those systems themselves do not provide a grammar in the sense of having a set of a type of constraints or rules which can actually parse inputs rather than just evaluate them. The degrees of grammaticality must be found in the grammar itself in relation with an input; they must be represented concerning a whole grammar system, rather than just taking into account an input with some general constraints such as agreement.

3.3.3 Grammars with Constraints in Gradience

Grammars with constraints are an alternative to the generative grammars with phrase-structure rules. A constraint is more susceptible to be evaluated than a phrase-structure rule. This trait is why grammars with constraints are very attractive for working with gradience. Besides, any set of constraints can be implemented in OT or any of its variations. As an example, a constraint such as $V \Rightarrow N$ (a verb (V) requires a noun (N)) or $DET \Rightarrow N$ (a determiner (DET) requires a noun (N)) can define several linguistic phenomena independently. While, by using phrase-structure rules, we would need to take into consideration a derivation for evaluation a single feature. This problem contrasts very clearly with the model based on restrictions of the Tree Adjoining Grammar (TAG) from Joshi et al. (1975). TAG can evaluate trees independently employing constraint requirements. The application of TAG illustrates very clearly the flexibility of working with restrictions (Figure 3.6), even while working with trees, in comparison to the derivational rules.

Grammars with constraints are not related to gradience in the same way as OT. Those grammars do not pursue explaining an input in terms of degrees, neither acceptability nor

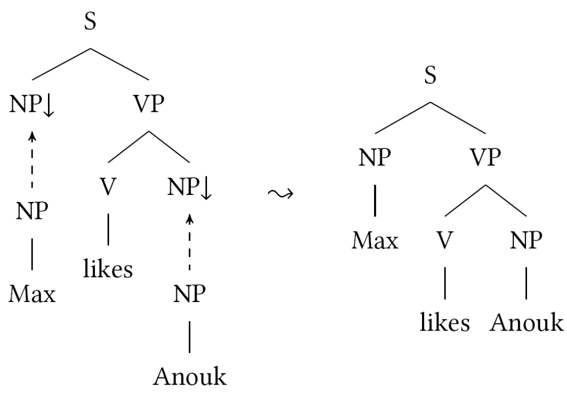


Fig. 3.6 TAG trees displaying their constraint requirement (Müller, 2016).

grammaticality. They offer frameworks to explain natural language and to generate inputs. Examples of this type of grammars are: Lexical Functional Grammar (LFG) in Bresnan and Nikitina (2009), Head-Phrase Structure Grammar (HPSG) in Malouf (1999, 2000) and Hudson (2003), Construction Grammar in Goldberg (1995, 2003, 2009) and Kay (2002), and Extensible Dependency Grammar in Debusmann et al. (2005).

In grammars with constraints, the grammar must generate just well-formed inputs, since there is no place for rules which generate deviated inputs. However, some of these grammars tried to go further by being able to define linguistic inputs with violations, instead of just ignore them. In this sense, some grammars with constraints offer the alternative to parsing inputs with deviations and saying why those inputs are ill-formed.

Some of these grammars take into account gradience when it comes to describing those categories which are very difficult to classify in a discrete way since they present features from different prototypical categories. This trait is shown clearly in LFG in which different categories can have different heads which may display non-prototypical functions. Some authors in HSPG, such as Malouf (1999), acknowledge constructions that have lexical items with more than one category at the same time, such as the gerunds in English being considered at the same time as nouns and verbs.

Therefore, the radical difference between these grammars and OT is that:

- Grammars with constraints provide a grammar framework for actually representing natural language with its interrelations between part-of-speech and features, generation, parsing, and tagging (sometimes);
- while OT represents an evaluation of candidates once a set of constraints is given, being unable to capture fuzzy features such as vague boundaries between categories or constructions.

The following grammars or frameworks with constraints have been specialized in the parsing process rather than in the generation: Model-Theoretic Syntax (MTS) recently worked by Pullum and Scholz (2001), Womb Grammars by Dahl and Miralles (2012) and Property Grammars by Blache (2000, 2016), Prost (2008), and Duchier et al. (2011). These grammars display better systems to deal with inputs which incur to be violated.

Pullum and Scholz (2001) present a system in which they mix the Generative-Enumerative Syntactic (GES) and Model-Theoretic Syntax (MTS). They consider that an utterance is grammatical only if it can be generated by derivation by rules from the grammar. In their framework, they differentiate GES for generating from MTS for parsing an input. A grammatical utterance is called an *expression*. However, MTS represents grammaticality by taking into account syntactic and semantic features. In MTS, an utterance is fully well-formed a called an *expression* when both semantic and syntactic features are satisfied. The utterances are called *quasi-expressions* when they violate some constraints allowed to be violated by the framework. Considering *quasi-expressions* as those expressions which are not fully well-formed, but they still show some form of syntactic structure and properties. They define the *quasi-expressions* in terms of ungrammaticality.

Among all the grammars with constraints, Blache's Property Grammars (Blache, 2000, 2004, 2016) arises as the grammar that better can capture gradience. Property Grammars have presented an independent way to evaluate and predict acceptability without using Optimality Theory, demonstrating that its predictions are not as far as from the judgments from native speakers (Blache and Prost, 2008; Prost, 2008). Property Grammars display many constraints in one single grammar, which make these grammars very rich and powerful when it comes to defining inputs. Because they take into account many constraints form an utterance, they can provide accurate values in terms of degree. Even though the theory is mainly focused in syntax, it can be applied in a multi-modal way such as in Chersoni et al. (2016) with semantics; and in Blache et al. (2008); Blache and Prévot (2010) with phonology, morphology, syntax, pragmatics, and prosody.

Another valuable positive aspect of this theory is that it does not make any distinction between *expressions*, *quasi-expressions*, and so on. Every utterance is a construction in Property Grammars without considering how ill-formed or deviated an input is. The grammar can parse any kind of input and providing a list of its satisfied and deviated inputs. This characteristic makes Property Grammars the perfect theory to deal with both vague or fuzzy constructions as well as to provide a degree of grammaticality taking into account the both satisfied and violated inputs according to a property grammar.

In this PhD, we dedicate chapter 5 to this theory since it is the theory that we have used to represent the fuzzy degrees of grammaticality in natural language. This theory has attempted to describe degrees of grammaticality, but it has not been used to represent grammaticality. In this work, we added some new tools to a property grammar as well as combining the grammar with some fuzzy logic formalism in order to make it entirely suitable to describe fuzzy degrees of grammaticality.

3.4 Modeling Gradient Data: Features and Properties

Many different theories and grammars brought some light to linguistic gradience but, as a fact, all they share that they are modeling gradient data. In this section, the bases of the main features that belong to gradient data are presented.

Gradient data is extracted mainly from two sources: 1) corpus data; 2) and/or linguistic judgments employing psycholinguistic tests.

Newmeyer (2003) claims against gradience in grammar by arguing that all the linguistic data come from performance sources: "the evidence of probabilities being associated with grammatical elements seems pretty weak. The numbers are overwhelmingly epiphenomenal". However, it seems a bit odd that this kind of claims do not consider weak-performance-data those list of examples that rank sentences with '?', '??', and '*' to argue which is the most well-formed and why. These kind of discrete approaches are yielding into a gradient method with performance data in order to extract conclusions. Moreover, it is usually data analyzed by instinct and self-knowledge rather than being supported by any objective number such as frequencies, or by any psycholinguistic test application.

The "performance" data is the only empirical data that linguistics can use for their research. It seems impossible to work in linguistics without considering any linguistic data and yet hoping for a realistic framework to represent linguistic phenomena.

Serious gradient approaches have concluded different features regarding gradience. The most important of them are the following ones:

(1) Using a framework with constraints. It is necessary to choose a grammar or a model which takes into account constraints. Linguistic works related to gradience are based on constraints.

(2) Context effects. One of the first strategies to deal with gradience is to check how marked is a sentence. The concept of *markedness* arises to represent the importance of context for a word. A sentence α is more marked than a sentence β if α is acceptable in less contexts than β. Therefore, β is less marked since it can occur in more contexts. These traits point out that a specific sentence α which is not fully acceptable in various contexts such as C^1, C^2, C^3, and so on, can be fully or mostly acceptable in a specific context such as C^5. Müller (1999) claimed that *markedness* can be determined either by the judgments of the speakers or by extracting the number of possible context types for a sentence. Keller (2000) points out that "a constraint is context-dependent if the degree of unacceptability triggered by its violation varies from context to context".

(3) Constraint ranking. It takes into account how some constraint violations are more significant than others. Therefore, those more significant constraints are going to be ranked higher than the others. Constraint ranking is especially essential for representing degrees of acceptability since it seems clear that speakers find some violations more notable than others.

(4) Cumulativity. This effect is present in those structures that violate multiple constraints in contrast to those structures that violate a single constraint which is highly ranked. As an example, we have two structures α and β and three ranked constraints $C_1 \gg C_2 \gg C_3$. α just violates C_1 while β violates $C_2 \gg C_3$. Therefore, β is more unacceptable than α.

(4) Constraint counterbalance. This notion is found in Blache and Prost (2008:7) as an alternative use of cumulativity. Constraint counterbalance claims that "cumulativity must take into account both violated and satisfied constraints; in contrast with standard cumulativity which takes into account **only** the violated ones."

(7) Ganging up effect. This effect shows up when a constraint has been violated multiple times in a structure. Acknowledging this effect allows us to consider that a constraint, which might be ranked below another one, can trigger more unacceptability if it has been violated more repeatedly than another which is ranked higher and violated just a single time. As an example, we have two structures α and β and three ranked constraints $C_1 \gg C_2 \gg C_3$. α just violates C_1, however β violates C_3 five times. Therefore, β is more unacceptable than α because of the ganging up effect.

(8) Soft and hard constraints: Goldsmith (1993) already revealed the notion of "soft universals" as those "which we could informally characterize as phonological states of affairs that are cross-linguistically preferred or avoided but are not universally, categorically true or false" (Guy, 1997). However, more recently, soft and hard constraints are proposed as a paired concept by Keller (2000, 2006). The conception of *markedness deeply influences this way of dealing with constraints* and for applying constraint ranking. Both constraints share features such as universal effects of being ranked, being cumulative and performing a ganging up effect. However, they also have features that distinguish them in Keller's approach:

- Hard constraints trigger strong unacceptability when violated; while soft constraints trigger mild violations.
- Hard constraints are independent of their context while soft ones are context dependent.
- Hard constraints are mandatory during the acquisition process of a language as both for a native or as a second language acquisition; while soft constraints display optional traits when they are being acquired.

Sorace and Keller (2005) conclude that constraints are hard and soft equally across languages since they made a cross-linguistic comparison between German, French, Dutch, and German. Find in Table 3.4 a summary of these properties:

<table>
<tr><td>Properties of hard and soft constraints</td><td>Hard Constraints</td><td>Soft Constraints</td></tr>
<tr><td>Universal effects</td><td colspan="2">Ranking effects
Cumulativity effects
Ganging up effects</td></tr>
<tr><td>Type-Specific effects</td><td>Strong unnaceptability
No context effects
No developmental optionality</td><td>Mild unnaceptability
Context effects
Developmental optionality</td></tr>
<tr><td>Crosslinguistic effects</td><td>Constraints are hard across crosslinguistic effects</td><td>Constraints are soft across crosslinguistic effects</td></tr>
</table>

Table 3.4 Properties of hard and soft constraints from Sorace and Keller (2005).

(9) Violation position. This notion is also from Blache and Prost (2008:7) and points out how the value of a violation of a constraint might differ from one syntactic structure from another. In their work, following this consideration, they decide to weight equally all the same type of constraints (i.e., all constraints regarding word order, known as *linearity*, weighted 20). However, they assume that the same constraint of a grammar would be heavier, or more *dense*, in one construction than another. Following the same example as before, the constraint of *linearity* would be more serious in an adjective construction defined with 7 constraints than in a nominal construction defined with 14 constraints.

(10) Weights and rules. Frequently, linguists who work in gradience weigh constraints according to their ranking, context effect, and how hard and soft they are. The weights of constraints are deeply dependent on the perceived, extracted or intuited impact on native speaker's acceptability. Constraint weighing is present in several works such as in Blache and Prost (2008); Blache et al. (2006); Bresnan and Nikitina (2003); Gibson and Thomas (1999); Keller (2000); Keller (2006); Sorace and Keller (2005); and Van Rullen et al. (2006). Usually, the degree of grammaticality and acceptability of a linguistic input is computed as the sum of the weights of the **violations** triggered by an utterance.

However, modeling gradient grammaticality with weighted constraints is a choice not exempt from controversy. Generally, describing degrees of grammaticality are based both on studies about acceptability judgments and on stochastic approaches. However, the results of psycholinguistic studies regarding gradience have been the main reason to weight linguistic constraints on formal approaches. These studies understand that speakers acknowledge some violations as much more severe than others. While stochastic approaches, models based on corpus with frequencies, use to justify the most critical constraints as the most frequent ones. Two problems concern the discussion concerning these two criteria regarding degrees of grammaticality:

- Firstly, the fact of justifying a theoretical concept such as grammaticality through an empirical perception of a speaker. The linguist is **interpreting** the speaker's constraint evaluation during an exercise of acceptability judgment. The truth is that the linguist does not know if a speaker has that particular constraint in his mind.
- Secondly, a frequency based approach is genuinely more impartial for modeling a theoretical framework with weights. Nevertheless, the necessity of weighting constraints is motivated for proposing formal models focused on predicting either acceptability or judgments rather than providing degrees of grammaticality.

In contrast with these approaches, other linguists who defend gradient grammaticality, such as Aarts (2007), claim that the weight of all constraints in a formal model ought to be the same.

3.5 Modeling Grammaticality with Linguistic Judgments

Certainly, linguistic judgments have been one of the most used criteria to justify linguistic arguments. Even in discrete approaches, such as in the generative trends, it is quite common to see a large display of examples using symbols like "?, ??, ?*, *, **" and so on, to justify a grammar rule by contrast among a battery of examples. The use of these marks are unsystematic, inconsistent, and mostly based on the intuition of the researcher as discussed in Bard et al. (1996); Schutze (1996); Cowart (1997); Keller (2000); and Wasow (2007). The main aim of these discrete grammars is to explain why the sentence which does not have any symbol is better than the sentences with a question mark(s) or asterisk(s). The odd part is that these categorical trends are ranking in somehow those sentences without agreeing with gradient theories. In conclusion, justifying a full-discrete approach utilizing a gradient relationship, and yet denying gradience exposes a contradiction.

Somehow, generative grammar has based their theories in gradient data, utilizing acceptability judgments (of the linguist) by justifying why the well-formed phrase is well-formed and not ill-formed like other displayed utterances in the set of examples. Even though these methods were based on intuitive judgments, they brought decades of extensive linguistic work (Wasow, 2007).

Probably, the main difference between discrete grammars and gradient theories is that the first ones are hiding their interest in any ranking while the second ones are actually trying to rank linguistic notions such as acceptability or grammaticality.

As we have seen through the history of gradience, gradient approaches used psycholinguistic methods based on speaker's judgments as one of the main sources to justify gradience and its degrees. Schutze (1996) defends the use of judgment data with various arguments. A couple of his most interesting arguments are:

- Linguists can use judgment data to investigate data that strangely occur in natural speech.
- Judgments allow to test and check certain structural properties of a language under a controlled context.

Despite the fact of these inspiring works, this section shows up some criticism over using judgments to explain the degrees of grammaticality.

3.5.1 Degrees of Grammaticality and Grammaticality Judgments

The features of grammaticality judgments are underpinned by the observations made on acceptability judgments data. The research made on linguistic judgments has used tests with Magnitude Estimation (Bard et al., 1996; Cowart, 1997; Keller, 2000, 2006; Sorace and Keller, 2005). This technique asks subjects to evaluate an object with their personal judgments by using a physical stimulus for evaluation, such as brightness, volume, and so on, instead of the classical numerical evaluation scales of 1-10, 1-7, 1-5, among others.

In the linguistic literature, the notion of grammaticality judgments concerns a speaker who has been asked for evaluating grammaticality of an utterance in discrete terms (Bader and Häussler, 2010:278). In contrast, magnitude estimation methods use the notion of acceptability judgments to approach gradient grammaticality as a part of the acceptability judgment. On the other hand, other authors, such as Schutze (1996), point out that acceptability and grammaticality judgments have been mixed up in the same way for pragmatical reasons. Differently, to avoid controversy by using the terms of grammaticality or acceptability, Bader and Häussler (2010:278) uses the term of perceived well-formedness as an alternative, involving in it the process of judging grammaticality and acceptability all-together.

Accordingly, grammaticality judgments arise as a concept because some research interprets that speakers can judge the grammaticality of a structure in a specific context on a psycholinguistic procedure (i.e., some of the contributions on Fanselow et al. (2006); Lau et al. (2014)).

Furthermore, some theoretical frameworks, such as Model-Theoretic Syntax (MTS) in Pullum and Scholz (2001), Prost (2008), Duchier et al. (2011), endeavors to formally represent grammaticality judgments.

Pullum and Scholz (2001) claim that:

> Speakers produce utterances that even they would agree are grammatically imperfect - not by some external authority's standard but by their own. However, experienced users of a language are also aware that some ungrammatical utterances are much closer to being grammatical than others.
>
> Pullum and Scholz (2001:26)

This claim is an extended affirmation from those who stand for grammaticality judgments. However, the assertion is half-true, and we must highlight two points:

1. firstly, the term grammaticality judgment is misled;
2. and, secondly, accepting gradient grammaticality and gradient grammaticality judgments is not the same.

Regarding the first point, we should agree on the fact that the notion of grammaticality judgments is misled since giving a grammaticality judgment is not the same as giving a pure linguistic evaluation. A grammaticality judgment involves both processing the linguistic properties from grammar's competence plus the extra-linguistic features from the performance mechanisms. This mixture is indivisible, which is the reason why linguistics uses to name speakers' judgments such as *acceptability judgments*, gathering in the acceptability the linguistic part plus the extra-linguistic features. We must agree on the fact that grammaticality is a theoretical concept. Speakers process language all-together, they are unable to acknowledge where is the line that divides their linguistic knowledge from the extra-linguistic one. Therefore, speakers cannot emit a grammatical value without being influenced by their prejudices, attitudes towards the conversational partner, physical context influence, real-world plausibility and so on. A speaker can evaluate grammaticality as part of the whole acceptability judgment. However, he is not able to explain the judgment neither to evaluate it rationally and so extracting just grammaticality from it. In other words, linguists with their linguistic training are already struggling to evaluate grammaticality in their "examples" with phrases. Consequently, it is a belief that a native speaker without any linguistic training would be specially blessed by judging grammaticality aside from acceptability without any problems. This consideration is agreed by those linguists already mentioned who have been working with acceptability judgments.

Concerning the second point, recognizing gradient grammaticality means agreeing on the fact that an utterance actually has degrees of grammaticality even if some constraints have been violated in that input. This conception contrasts with considering that an input is grammatical solely when it is entirely well-formed, but an input is ungrammatical in a degree to covering from the slightest violation to the most severe ones. Optimality Theory and Model-Theoretical Syntax agrees on that. In both positions, generally, grammaticality is discrete, and ungrammaticality is gradient (further explanations will be provided in section 3.8). This reasoning exposes why the term grammaticality judgment is chosen on purpose since there is a will of not acknowledging grammaticality as a gradient concept from the theoretical point of view; but from the empirical point of view in terms of speaker's judgments (it would be more precise to call it *ungrammaticality judgment*).

3.5.2 Degrees of Grammaticality and Gold Standard on a Judgment

The gold standard which shall rank the grammaticality (judgment) of the utterances is not always justified. Usually, one might hesitate about how do we know for sure that the phrase which is meant to be grammatical is actually the gold standard. If there are no explicit criteria behind a linguistic gold standard constraint, such as a corpus methodology, one might think that the linguist trusts his knowledge (and his instincts) in order to provide a gold standard. The linguist also sometimes assumes a language with a gold standard without taking into account the dialects. This fact might lead the results to a misunderstanding.

If, as linguists, we assume that every dialect of a language is defined as a different system from each other, we must differentiate those dialects in order to take into account its linguistic constraints. In this sense, the constraints of Geordie, Cockney, Scourse, *Hinglish*, and so on, are going to be defined with a different grammar than a grammar based on the Cambridge and Oxford English variety. The same happens in Spanish; we cannot define the dialects of Spanish with a single grammar of Spanish, such as using a grammar of standard Spanish (Castilian), for representing other varieties as Andalusian, Mexican, Cuban, Colombian, Chilean and so on, which are going to need their particular constraints for being fully defined.

This reasoning might also fit from a generative perspective. All these dialects may share the same deep structure, however, defining a language must take into account deep and surface structure in order to represent the full system which a language is. Once it is assumed that every linguistic rule or constraint belongs to the competence, we must specify what constraint is belonging to which dialect. Therefore, we would obtain a grammar with constraints for each dialect.

3.5.3 Justifying Constraints and its Weights with Judgments

Another problem with the judgments regards their lack of efficiency for extracting linguistic constraints and its weights. If we would want to extract all the constraints of a language through psycholinguistic tests, we would need to prepare tests for each constraint in order to justify the constraints and extract their weight. Besides, we would never be sure if the weight of a constraint is merely a single-module or a cross-modal weight. Maybe a constraint C_1 is hard without considering that when C_1 is violated is, in fact, violating more constraints in other domains and, therefore, the *hard* weight effect is a consequence of a multi-modal ganging up effect. *Number agreement* would be a good example since it might perform a violation in morphology, syntax, semantics, and pragmatics.

The multi-modal perspective is also a problem for the weights either extracted or based on linguistic judgments. Nothing fully guarantees that a subject is actually evaluating the syntactic constraint that we are attempting to see evaluated. Even if it is assumed that the psycholinguistic tests are able to control the majority of the extra-linguistic variables, in the end, grammaticality or acceptability judgments come from the perception of the speaker, and they are influenced by their education, linguistic skills, sociological background and so on. A subject cannot produce single module evaluations in terms of weights since the language is parsed in a multi-modal sense by the participant. Splitting the linguistic modules is a reality in theoretical linguistics; it is not an empirical fact in the psycholinguistic reality of the subject. Thus, representing linguistic weights by domain has to be done from a theoretical perspective, without taking into account psycholinguistic data. We cannot rely on them to cover up a theoretical value such as grammaticality since these judgments have too many extra-linguistic variables. We cannot split both linguistic and extra-linguistic features hoping for the speaker to provide precisely the grammaticality data what we want to.

3.6 Modeling Grammaticality with Probabilities

In the past, linguistics dismissed the use of probabilities and frequencies as a tool for research. Chomsky (1969:57) pointed out that the notion of probability is not related to grammaticality, nor to any linguistic fact: "It must be recognized that the notion 'probability of a sentence' is entirely an useless one, under any known interpretation of this term". This claim was exemplified by some of the most famous Chomsky's old arguments with: *"Colorless green ideas sleep furiously"*, and *"Furiously sleep ideas green colorless"*. Although both examples are assumed by Chomsky to have a 0 probability, the first is considered grammatical, while the second is ungrammatical. In this sense, a probabilistic approach should be unable to distinguish grammaticality between them since both phrases have a probability close to zero.

However, probability models and frequency-based theories have updated these claims. Pereira (2000) demonstrates how the computation of probabilities and occurrences based on corpus bi-grams coincide with Chomsky identifying the above first sentence as 200.000 times more probable than the second one. Therefore, it is wrongly assumed that both have 0 probabilities, being the grammatical option more probable than the second one. Bresnan et al. (2007) claims that:

> Theoretical linguists have traditionally relied on linguistic intuitions such as grammaticality judgments for their data. However, the massive growth of computer-readable text and recordings, the availability of cheaper, more powerful computers and software, and the development of new probabilistic models for language have now made the spontaneous use of language in natural settings a rich and easily accessible alternative source of data.
>
> (Bresnan et al., 2007:2)

Frequencies of occurrence have been studied as one of the main phenomena to explain linguistic variability. Greenbaum (1980:102) points out how frequencies of occurrence reveal the knowledge that a native speaker has about a language. Greenbaum compares pairs of expressions with different frequencies due to its syntactic variability: *"Tom gave a dime to the boy"*; *"Tom gave the boy to tell"*. The result of his research concludes that there is a direct relationship between less frequent syntactic variation and acceptability.

Tremblay (2012) claims that the frequency of appearance is a pattern directly related to the difficulty of perception. Those constraints which are the gold standard are going to be frequent and easy to process. While those which are less frequent, and inconsistent, are going to be violations which are more difficult to process. In short, a constraint is empirically justified by its frequency; while frequency is justifying the difficulty of processing. Therefore, frequencies are a perfect element to both describe language theoretically and empirically. Even though our model is accounting for degrees of grammaticality regardless of a cognitive approach, it is stimulating to know that a theoretical model based on frequencies might be empirically valid for computing the grammaticality value within the total value of acceptability processed by speakers.

Manning (2003) explains how frequencies of occurrence revel as phenomena regarding natural language. He claims that much of the phenomena in natural language has a probabilistic basis; since "human cognition has a probabilistic nature: we continually have a reason for incomplete and uncertain information about the world, and probabilities give us a well-founded tool for doing this" (Manning, 2003:2). However, not all language phenomena, such as grammaticality, can be explained with probabilities. Manning admits that sentences with little probability in a model can be grammatical; therefore, probability and grammaticality are not directly proportional. Nevertheless, Manning emphasizes that the connection between probabilities and grammaticality is possible while both are still treated as separate concepts. This connection must be accompanied by the study of the forms, meanings and context in which those possibilities appear: "A profitable way to connect grammaticality and probability is perhaps to begin with the joint distribution *P*(*form*, *meaning*, *context*)" (Manning, 2003:15).

In this sense, probabilistic methods offer an alternative to work with judgments' data. Native speakers are not asked to make considerations in relation to frequencies in language. Frequencies occur naturally outside the speaker's consciousness. Therefore, frequencies of occurrence are more objective data than acceptability (or "grammaticality") judgments concerning the speaker's participation. Judgment data might be considered more obscure since we actually do not know what the speaker is evaluating in a judgment process:

> However, while humans clearly have some ability to consider sentences in an imagined favorable context when judging (syntactic) 'grammaticality', sociolinguistic and psycholinguistic research has shown that judgments of grammaticality are strongly codetermined by context and that people don't automatically find the best context [...]. Nevertheless, Keller (2000) argues for continuing to judge grammaticality via acceptability judgments (essentially still intuitive judgments, though from a more controlled experimental setting), and modeling

> soft constraints by looking at relative acceptability. The issues here deserve further exploration, and there may be valuable alternative approaches, but, as exemplified at the beginning of this section, *I tend to feel that there are good reasons for getting at synchronic human grammars of gradient phenomena from large data sets of actual language use rather than from human judgments. In part this is due to the unclarity of what syntactic acceptability judgments are actually measuring, as just discussed.*
>
> (Manning, 2003:16)

Bresnan (2007); Bresnan et al. (2007); Bresnan and Nikitina (2003, 2009) claim that "linguistic intuitions of ungrammaticality are a poor guide to the space of grammatical possibility". They display the success of the models based on frequencies; demonstrating that those are reliable for both extracting linguistic constraints and predicting linguistic behaviors such as the problem of dative alternation in English.

Another example of the success of methods with frequencies can be found in machine learning. Dunn (2017) introduces a grammar induction algorithm based on the notion of construction (Goldberg, 2009) that is able to learn a (small) grammar of a language by using merely statistical properties, without any innate structures. In this sense, Dunn's work reveals how probabilities are related to the grammar competence as well; and not only a matter of usage:

> However, these results are sufficient to provide empirical evidence against the poverty-of-the-stimulus line of reasoning for Universal Grammar. This source of evidence, further, is unique in providing large-scale corpus-based evidence for a question which in the past has been approached with small-scale intuition-based evidence. [...] What this means is that grammatical representation can be learned from observed frequencies. While there are always technical improvement to be made, the current algorithm shows that the learning of grammatical structures in this way is possible and in this sense provides converging evidence with many other empirical sources that have been collected.
>
> (Dunn, 2017:28)

3.7 Modeling Gradience in Semantics

Gradience, as well as the degrees of grammaticality, is a multi-modal phenomenon. However, as we have seen before in the history of gradience, studies of gradience have focused mostly on classifying both part-of-speech and syntactic constraints. After all, syntactic information has a limit in which syntax meets with semantics. This vague boundary between syntax and semantics interface has been pointed out quite frequently in linguistics. Jacobson (2004, 2007) and her *direct compositionality* framework is an example of this:

> The hypothesis of Direct Compositionality - familiar (in one form) from, the work of Montague (1970) and more recently from considerable work within Categorial Grammar -is that the syntax and semantics work "in tandem". The syntax is a system of rules (call them "principles" if one prefers) which prove the well-formedness if linguistic expressions while the semantics works simultaneously to provide a model-theoretic interpretation for each expression as it is proved well-formed in the syntax theoretical, empirical and psycholinguistic tools, and probabilistic methods with corpus methodology.
>
> (Jacobson, 2004:1)

Nonetheless, there are linguistic issues which are purely semantic. Some combinations of words seem to be syntactically correct such as the syntactical relation between a noun as a subject with its verb, and a verb with its direct object. However, even though the syntactic requirements of these words are satisfied, their semantic relation seems odd. In other words, their semantics do not fit each other while their syntax does. These words which semantically do not fit each other have a gradient relation since in some cases their relationship does not seem radically inappropriate. Such semantic gradient distinction is illustrated between these examples (extracted from Chersoni et al. (2018)):

(1) *"The hamster explored a backpack."*

(2) *"The hamster lift a backpack."*

(3) *"The hamster entertained a backpack."*

Distributional semantic models with selectional preferences (Baroni and Lenci, 2010; Erk et al., 2010) are a good option in order to model the differences between such examples. Chersoni et al. (2018); Santus et al. (2017) present an interesting purpose for modeling such inputs. They use the notion of *thematic fit* on such models. The distributional models of thematic fit have been evaluated by comparing the plausibility scores produced by the models with human-elicited judgments (Baroni and Lenci, 2010; Erk et al., 2010; Greenberg et al., 2015; Santus et al., 2017), showing significant correlations. The notion of thematic fit has been supported by recent psycholinguistic research which states that humans activate a rich array of knowledge during sentence processing determined by their expectations (Altmann and Kamide, 1999; Ferretti et al., 2001; Hare et al., 2009; Matsuki et al., 2011; McRae et al., 2005, 1998). The concept of thematic fit is related to the classical notion of selectional preferences but taking into account gradient compatibility between elements; whereas the latter conceives such compatibility as boolean constraint evaluated on discrete semantic features (Lebani and Lenci, 2018). In this sense, a thematic fit is close to the gradient notion of a prototype. The prototypical elements are extracted employing frequencies of occurrence through a corpus methodology with vectors. The distance in the vectors between the elements concerning the thematic fit establishes the degrees of satisfaction of the selectional restrictions.

According to the last example in this section, Chersoni et al. (2018) distinguish between three degrees: *prototypical thematic fit* (1); *complement of coercion* (2); and *restriction violation* (3). The first example would be a plausive example; the second one would be an implausible condition with no violation of selectional restrictions; the third one shows a sentence with a selectional restriction violation.

Somehow, this model accounts for semantic "grammaticality", since the framework does not consider coercion as a violation. In order to demonstrate the efficiency of the system, the model attempts to identify just the sentence with violated selectional restrictions.

Therefore, it is essential to take into account models like this one which can support the syntactic computation of degrees of grammaticality. Sometimes, while we are modeling syntax, linguists can establish constraints for specific categories, such as adjectives, nouns or prepositions. However, not all the elements which fit into a category behave in the same way concerning its syntactic features, such as the subject-verb-object relation, the noun-adjective modifier relation, and so on.

Our work to account for degrees of grammaticality presents some limitations regarding semantics; since we approach the problem from the syntactic point of view. Therefore, our constraints cannot deal with semantic problems. In the future, it would be exciting to combine our work with frameworks in distributional semantics in order to compute grammaticality taking into account both semantics and syntax.

3.8 Degrees of Grammaticality vs. Degrees of Ungrammaticality

Our proposal attempts to formally represent how degrees of grammaticality have a place in the grammar of a natural language. Therefore, the model we propose aims to approach the concept of grammaticality from a theoretical point of view, without involving extra-linguistic features, in the hope that it will tip the debate on grammaticality and linguistic competence in favor of linguistic gradience.

Gradience studies strive to represent both degrees of grammaticality and acceptability. In order to extract these degrees, linguists have used several tools.

Theoretical tools for gradience are mainly divided into two groups:

1. Theoretical tools designed to characterize fuzzy language phenomena;

2. and theoretical tools for evaluating inputs in terms of degrees through formal models.

We have seen that the most studied fuzzy phenomenon over time has been the categorization of the words in a natural language. Interest in this phenomenon has been aroused by the discrete classification of parts-of-speech of a language. The behavior of syntactic categories needs to be represented in such a way as to satisfy the different possible linguistic contexts, such as a noun being used as an adjective (i.e., *"a stone-bridge"* and *"a bridge made of stone"*). Linguistics need formal tools to deal with these fuzzy phenomena since the degree of grammaticality and acceptability is often conditioned by the labeling of word classes and their consequent syntactic function.

Likewise, it is clear that grammars and theories that use constraints are very effective for evaluating grammaticality and acceptability in terms of degrees, probably because of the properties displayed by constraints as a tool.

However, firstly, many of these models and studies in the literature do not account for degrees of grammaticality but for *degrees of judgments*. This is due to the fact that many of these models mix the theoretical concept of grammaticality with the empirical concept of acceptability. These models conclude on avoiding gradient grammaticality, but admitting gradient judgments, adopting some concepts such as *grammaticality judgment*; which is a theoretical odd-one, as we have seen in subsection 3.5.1. Secondly, another frequent issue in the literature is that degrees of ungrammaticality are regarded as synonymous with degrees of grammaticality, which is not exactly true.

In this section, considering all the above, we discuss why the models that deal with degrees of *ungrammaticality* do not fully account for degrees of *grammaticality*.

Different approaches and authors conceive the degrees of grammaticality as degrees of ungrammaticality. Here we comment on some examples.

Approaches which combine degrees of grammaticality and degrees of ungrammaticality.

Ross (1987, 2000) is one example. He claims that linguists must allow the notion of membership to be explained in terms of degrees. He uses the idea of a prototype to propose a degree of 'viability.' He assumes that acceptability will not be dropped at the sightless deviation and that its degree of 'viability' will decrease as a result of a cumulative effect of grammatical deviations. This 'viability' provides two degrees of the syntactic well-being of a sentence: on a scale of 0 to 100, if the linguistic input is under the prototype of 50, the linguistic input has a *degree of ungrammaticality*, while if the linguistic input is over the prototype of 51, it has a *degree of grammaticality.*

Approaches in which grammaticality is discrete and ungrammaticality is gradient. These approaches use Optimality Theory (OT) and Model-Theoretic Syntax (MTS) to represent degrees of grammaticality. Standard Optimality Theory, together with most of its variations, is a framework that looks for the *optimal* candidate in a group of candidates. Every candidate which is not optimal is *suboptimal* (ungrammatical); therefore, it has a degree of ungrammaticality. Even those OT frameworks with constraint re-ranking have just one *optimal* candidate. Additionally, OT does not aim for constraints to capture the vague boundary between categories, or attempt to be a descriptive grammar such as Dependency Grammars (DG), Construction Grammars (CxG), Lexical-Functional Grammar (LFG), Head-Phrase Structure Grammar (HPSG), or Property Grammars (PGs) among others. However, OT does usually need other descriptive grammars (or approaches) to identify constraints to be evaluated. In the end, some OT frameworks actually attempt to explain gradient degrees by acceptability rather than grammaticality (Keller, 2000, 2006).

For MTS, an input is an *expression* if it satisfies its constraints, but a *quasi-expression* if it violates some of them. This model assumes that a *quasi-expression* has a degree of ungrammaticality, since only *expressions* can be grammatical. Unlike OT frameworks, these models are descriptive. They can account for vague structures using the notion of prototype, such as in Prost (2008:64): "Therefore, we prefer —and adopt— a model-theoretic notion of grammaticality with graded ungrammaticality, where classes of utterances are constructions in the sense of CxG. Constructions should have sharp boundaries, with prototype members". However, MTS uses a system to either predict judgments or to model degrees of acceptability, such as in Prost (2008:217): "In this work, we have addressed part of the problem concerned with replacing the traditional binary notion of grammaticality with intermediate degrees of acceptability in natural language processing".

3.9 Grammaticality, Weights and Corpus Methodology in this Work

3.9.1 Why do we Claim Grammaticality rather than Ungrammaticality?

All the above conceptions have a fundamental problem: they put the concepts of degrees of grammaticality and degrees of ungrammaticality on the same level.

Degrees of ungrammaticality should not be understood as a synonym of degrees of grammaticality. Speakers cannot understand the unintelligible even though they can predict sense from an unintelligible utterance.

When speakers understand deviant input, they, in fact, understand the grammatical parts and reconstruct the violated ones. Thus, a speaker understands a degree of grammaticality (a positive value, rather than a negative one). Let us consider a linguistic input which would be profoundly violated, i.e., α= 0.001. Even though input α can be described as being 0.999 "ungrammatical", it is still 0.001 grammatical, so the speaker will surely understand that grammaticality, the well-formed part.

In this respect, theoretically, while a linguistic object still satisfies some constraints from a grammar, those satisfied constraints are grammatical, so the linguistic input is still grammatical in relation to that grammar. So, it is false to classify an input just by its ungrammaticality because it still has grammatically satisfied constraints.

On the other hand, under no circumstances can the degree of grammaticality fall below 0 from the point of view of a grammar. A degree of grammaticality 0 is equivalent to expressing that an input is *fully unintelligible* for a grammar. In short, there is not a single constraint in the input that belongs to the grammar. Therefore, a grammar cannot distinguish between a grammaticality membership of 0 and -31 because both values state that the constraints of an input do not belong to the grammar at all. The input would be too ill-formed.

Similarly, the competence of a speaker is not dealing with fuzziness anymore when an input has a degree of grammaticality of 0 or below 0. The competence of a speaker cannot find any constraint from his grammar in such input. However, a speaker can use linguistic predictions based on probabilities to deal with linguistic uncertainty and deduce an unintelligible utterance.

Thereby, an input can be more or less true in terms of the number of satisfied and violated linguistic elements in relation to a grammar. In order to determine whether an element is satisfied or violated, a grammar has to represent the vague elements which are neither fully satisfied nor fully violated. These vague elements can be identified by a method in which we attempt to represent linguistic variability. In a corpus, linguistic elements that are not performing their canonical role can be identified, even though they may have a different function, such as a noun being used as an adjective. These kinds of elements are not uncertain but fuzzy or vague. They are systematically and naturally repeated as a result of the competence of a speaker. They are not random or predictable. They are part of the grammar.

A linguistic element is used fuzzily when it has a degree of truth in relation to canonical use. Thus, some elements are more true than others in a given linguistic situation for a grammar. In this regard, it is more true to say that an adjective is used for purposes of modification than a noun, even though a noun can also be used as a modifier. This vague phenomenon of linguistic variability is confirmed by a systemic and repeated frequency of appearance. Following this reasoning, all these kinds of elements have a degree of grammaticality, a positive value, since they are part of the grammar.

For all the above, we propose a model that accounts for *degrees of grammaticality*, and which theoretically represents different degrees of truth concerning the notion of grammaticality. We dismiss the criteria of non-grammaticality, which is mostly influenced by the criterion of optimality and the acceptance of gradient grammaticality not in terms of competence but in terms of acceptability (performance).

In summary, there is no theoretical reason to represent the grammaticality of linguistic inputs in negative terms such as degrees of ungrammaticality. They can and should only be classified by considering the degree of grammaticality in terms of the degree of membership between a linguistic input and a grammar.

In line with this, the theories that represent grammaticality as a negative number are probably mainly representing either acceptability or predictability.

Acceptability is a concept too vague to be rated as 0 since it is difficult to say what full unacceptability is from an objective point of view. For this reason, expressing acceptability as a negative number, which could be infinite, is a good option since it is more probable that an input is indefinitely non-acceptable rather than not acceptable at all. As an example, Rickford and Wasow (1995) made a study in which the participants had to rate 20 sentences on a scale of acceptability. They found that extreme responses to acceptability and unacceptability are rare. None of the subjects rated any sentence with a single zero.

It should be pointed out that studies that assume degrees of grammaticality to be degrees of ungrammaticality mostly focus on predicting, describing, and modeling either acceptability or grammaticality judgments. These studies do not provide a theoretical solution to a theoretical matter such as the formal representation of degrees of grammaticality in natural language without taking into account acceptability.

As well as their own corpus which contains many probabilities, speakers might use logical mechanisms to deal with a less grammatically uncertain sentence. Speakers can try to reconstruct or re-build an uncertain input by combining probabilities and their linguistic competence. Bresnan (2011:71) points out that "language users have powerful predictive capacity". So, degrees of ungrammaticality might make sense if ungrammaticality is a homonym of uncertainty. Frameworks based on Optimality Theory would be a good way of modeling uncertainty. They could implement computational tools to make a *suboptimal* input *optimal*, and provide mechanisms of sentence reconstruction in terms of degrees of probability.

Consequently, we have to make a distinction between what is *uncertain* and what is *vague* or *fuzzy*. A fuzzy mechanism would help us to deal with grammaticality since vagueness has (positive) degrees of truth regarding the membership of an object to a group. On the other hand, probabilistic tools would help us to deal with negative degrees of ungrammaticality or acceptability. These negative degrees can be understood as the value of uncertainty of a linguistic input. The uncertainty of an input can be expressed in terms of degrees of probability or prediction.

In our system, a negative number or a zero means that a linguistic input cannot be given a degree of truth. However, it can be given a degree of probability or uncertainty. This decision relies on the basic idea of the distinction between vagueness and uncertainty. In this study, we explore fuzzy degrees of grammaticality since a truth value can be assigned to an input which has grammatical violations.

3.9.2 Corpus Methodology with Occurrences

Someone might find a contradiction on using frequency of occurrence as the main criteria to justify canonical constraints of a grammar in favor of degrees of grammaticality. However, there is a fundamental distinction between using such frequencies with empirical value in contrast with the empirical data from judgments.

Working with frequencies is a task carried through the rational criteria of a professional such a linguist; based on theoretical criteria whereas judgments' data are always relying on subjects which are unable to make the distinction between linguistic and extra-linguistic elements (even if it happens under a highly controlled experimental context).

From the above comparison, we can establish theoretical explanations concerning with which one is the gold standard and under which criteria, regarding Manning (2003) consideration of: P (*form*, *meaning*, *context*). However, since frequencies and grammaticality are not the same, the raw frequency of appearance might not be enough for weighting a single-domain-constraint for grammaticality; and some considerations might be added in order to propose a theoretical weight.

Representing grammaticality in our work requires a grammar with constraints which are going to be either satisfied or violated in a degree. Extracting those constraints from a specific corpus means that we are going to obtain a grammar for that specific corpus (our constraints and the corpus extraction are presented in chapter 7). The fact that we are evaluating grammaticality with a grammar of a dialect rather than with a grammar of a language might be a possible criticism towards our methodology. We are fine with that since our proposal justifies the use of the concept of a dialect rather than a language (cf. section 6.4). Actually, grammars based on constraints extracted by frequencies are most probably representing one dialect of a language. Hence, a corpus based on Spanish from Spain is building a grammar more Castilian-like than Mexican-like.

Following the mentioned tradition, we have chosen to work with methods relying on frequency rather than with judgments. We have used *MarsaGram* as a tool for this work. *MarsaGram* provides the possibility of extracting all constraints by one method. This tool was built to extract a property grammar, which is the grammar framework that we are going to use [*)].

3.9.3 Degrees of Grammaticality and Weights

In this section, we briefly contribute to the debate on the constraint weights for grammatical gradience. The notion of weight is closely related to both the degrees of ungrammaticality discussed in subsection 3.9.1 and the degrees of frequency discussed in subsection 3.9.2. The theories which work with degrees of ungrammaticality build weights in relation to the number of violations. That is, an input can be calculated as -0.25, -2.75 and so on (Boersma and Hayes, 2001; Keller, 2006; Prost, 2008). The theories which weight a constraint by frequency can mislead the weight if they take into account solely frequencies disregarding theoretical reasons for representing constraints with weights.

Taking into account the discussions in subsection 3.9.1 and in subsection 3.9.2, the model proposed in this dissertation will not take into account weights in its constraints for the following reasons:

We propose a theoretical model because grammaticality is a theoretical concept. Therefore, the notion of grammaticality does not coincide with the cognitive value of an acceptability judgment, which is where we find the weights. A fuzzy model of grammaticality should include the relationship between inputs and a grammar, establish

[*)] Check chapter 7 for further details regarding this tool and our extraction method.

the relationship between degrees of satisfaction and violation of the constraints of a grammar, and provide a grammatical value as feedback. However, this feedback is, in our model, the partial theoretical value (of grammaticality) that complements the final value of acceptability in terms of linguistic performance. In short, our model provides the value of only one of the elements of the formula: *Acceptability Value = Grammaticality Value + Value of the Mechanisms of Performance*. This value of grammaticality is not necessarily weighted in terms of the acceptability value since we do not know how a theoretically weighted system would perform taking into account multi-modality.

Our model takes into account the multi-modal perspective of a grammar, but we have only worked with the syntax module. The behavior of restrictions should be checked theoretically to see how their weights perform in all the modules of the grammar. From a multi-modal perspective, the restrictions of a module (i.e., semantics) could be interfering with the restriction from another module (i.e., pragmatics). If this happens, we will be providing biased values of grammaticality.

Our model concerns vagueness, and does not account for uncertainty in acceptability judgments. Weighted constraints are much more likely to predict acceptability judgments in a computational model, such as in Keller (2000) or in Prost and Lafourcade (2011). However, with the right tools, our model could also do this in the future.

We assume that violations of two constraints such as C_1 and C_2 may be equally serious. However, an acceptability judgment test might perform differently. The speaker might be conditioned by the high frequency of C_1 being violated in contrast to the low frequency of C_2 being violated.

We also consider that a single constraint such as C_1 may display different values of violation in acceptability depending on whether it has been violated or simply omitted. One example is the omission or violation of the restriction of *linearity* (word order) in Spanish. Spanish requires a determiner (DET) to precede a noun (N), as an example: *"El chico juega a fútbol"* (*The kid plays football*). However, structures such as *"chico juega a fútbol"* (*kid plays football*) and ***"chico el** juega a fútbol"* seem to have different grammaticality values. The first one seems much less serious than the second one. Perhaps, because it is commonplace to hear non-native speakers speak without determiners but not to change the order of the determiner and the noun, the perception of severity is perceived as minor. If this is true, constraints would again be weighted in terms of acceptability, not grammaticality. The weight would be associated with greater tolerance of the violation due to the frequency of occurrence. Therefore, the speaker's level of tolerance of a violation is an element that should be modeled in terms of both linguistic and extra-linguistic aspects.

Paradoxes of probabilities for grammaticality. Frequency cannot always model neither grammaticality or a grammar. As an example, a Spanish corpus with frequencies such as Universal Dependencies strangely provides examples of personal pronouns being used as subjects (almost not at all). If we built grammaticality being guided just by frequencies, we would be forced to say that, in Spanish, the use of the personal pronoun as a subject is non-grammatical. However, no native Spanish speaker would consider such a thing. Sometimes, working with a corpus taking into account frequencies presents such paradoxes. These issues illustrate why frequencies are a tool for the linguist who combines these data supported with theoretical reasons. Such paradoxes are one of the main reasons why we cannot model grammaticality *only* under the spectrum of frequencies. Grammaticality needs theoretical reasons plus objective data such as frequencies in order to be modeled with weights.

Chapter 4

Fuzzy Logic and Gradience

4.1 Why Together?

Fuzziness and *gradience* are related to each other. Gradience is a linguistic term which gathers all-together different gradient notions such as vagueness and uncertainty. Therefore, it is necessary to talk about fuzziness for a full understanding of the linguistic notion of gradience. Even though some linguists like Bolinger (1961) or Ross (1972) attempted to approach gradience from linguistics, it is in mathematics where we can find formal approaches of gradience concerning many quotidian incidents, such as gradient sense of temperature, colors, or linguistic expressions such as the gradient relation between *tall-short*, *big-small*. Nevertheless, in mathematics, the phenomenon of revealing degrees of truth between objects in a set is called fuzzy phenomenon, and the fuzzy logic is the right tool to represent these vague-fuzzy relations formally.

In this work, we treat grammaticality as a fuzzy-vague object rather than discrete, uncertain or predictable. This reason is why we use fuzzy logic as one of the tools to represent grammaticality.

In this chapter, an introduction to Fuzzy Logic is provided. Linguistics is going to be nourished from the better understanding of the term gradience through the vast field of the fuzzy logic. This research shows that fuzziness and gradience are not exactly the same; even though both terms are usually mixed up in the linguistic literature. Additionally, one of the main differences is that *fuzziness* is a mathematical term while *gradience* is a linguistic one. Nevertheless, in subsection 4.6.1 a differentiation of both terms is proposed. New definitions are displayed to procure a clear idea about the different theoretical terms that a fuzzy grammar requires.

4.2 Fuzziness as a Conception in Linguistics

Fuzziness is a conception concerning with those objects which are difficult to be classified categorically at first sight. Linguists, such as Almela Pérez (2003:62), claim that studying linguistic objects with a non-discrete approach seem legit since such objects in natural languages are prone to poly-signification, and they are essentially context-dependent.

Aarts (2004b:365-366) clarifies that the term vagueness in linguistics is present on many different proper linguistic terms such as *core, periphery, generality*, and *gradience*.

Other linguists, like Neustupnỳ (1966:39), have defended the necessity to deal with the fuzzy phenomena in language claiming that "the main task today's linguistics is to determine the full extent of vagueness, to analyze and explain it and to make possible the combination of its thorough consideration with the stream of the world linguistic tradition". He makes different distinctions of vagueness for different elements:

- **Discourse vagueness** referring for vagueness in real-world objects.
- **Systemic vagueness** as the vagueness which is present in a system.
- **Approximation vagueness** for those linguistic elements which are similar to each other (close to value 1 between each other).
- **Annihilation vagueness** for those linguistic elements which are almost entirely different (close to value 0).

Neustupnỳ (1966) applied these concepts in phonetics and phonology in an attempt to show to the linguistic community how concepts and notions from the fuzzy logic theories might help to better represent specific linguistic phenomena in contrast with some of the existent classifications in the linguistic theory.

In linguistics, part-of-speech categorization has been considered one of the most acknowledged fuzzy objects; since, for example, it is very easy to find words which typically perform as a part of speech such as a noun, but, in some specific contexts, they seem to perform as an adjective rather than nouns. In this sense, a discrete classification for this type of words seems to be unsatisfactory since those words can perform in both ways; and, therefore, they are displaying fuzzy boundaries. Ross (1972) has been one of the linguists who most notably worked the part-of-speech in natural languages regarding fuzziness.

Grammaticality is another of the most famous fuzzy objects in linguistics; and the principal matter in this work. As it has been exposed in chapter 2 and chapter 3, linguistics have continuously discussed if grammaticality is actually a vague object which can be defined by degrees. However, this topic has been generally concluded by accepting grammaticality as discrete for being part of linguistic competence while acceptability has degrees because it relies on speakers' perception. In this sense, some linguists might accept gradient grammaticality as the partial value of an acceptability judgment which relies on the grammatical features of an input. However, accepting fuzzy grammaticality merely regarding linguistic competence has been an impossible task for theoretical linguistics until now, due to the lack of a theoretical system for representing fuzziness.

Lakoff (1973) was the first linguist to work on the concept of a *fuzzy grammar* as a solution to deal with fuzzy grammaticality. He was highly influenced by Zadeh (1965) and Ross (1969a,b, 1972). He applied fuzzy logic as a tool for theoretical linguistics. His paper *Linguistics and Natural Logic* from 1970 presents fuzzy figures and logic mechanisms in order to deal with grammaticality and some other fuzzy traits of natural language.

In front of such remarks, several linguists specify out the necessity of a formal model with the following features: *permissive, flexible*, and able to conceive *fuzzy reasoning*.

> The challenge is to develop a grammatical framework that is permissive enough to account for gradient data without idealizing it, but restrictive enough to allow us to formulate precise, testable linguistic analyses.
>
> (Sorace and Keller, 2005:3)

Therefore, such model must be capable of representing linguistic objects such as part of speech and grammaticality in its gradient and fuzzy nature; getting-together both the discrete and the gradient perspectives and using each other when it is necessary.

Almela Pérez (2003) states that linguistics would be better nourished by dealing with the fuzzy and continuous phenomena with the right tools such as the prototype theory and fuzzy logic.

Prototype theory (Labov, 1973; Moure, 1994; Taylor, 1995) has been quite used in linguistics as shown in subsection 3.2.7. Lakoff, who used fuzzy logic in his work, also explored prototype theory (Lakoff, 1987a,b) to deal with fuzzy linguistic objects. Additionally, Moure (1994) applied prototype theory in Spanish regarding part of speech and some syntactic and semantic phenomena.

However, in contrast with the quite explored applications of prototype theory in linguistics, there is a generalized lack of adapting fuzzy logic tools for linguistic research either for Spanish or for other languages.

The application of fuzzy logic as a tool would help to solve theoretical problems by creating new perspectives; such as building new grammar frameworks attending for linguistic fuzziness. Fuzzy logic has appeared in this work as the most powerful tool to demonstrate formally the fuzzy concept of degrees of grammaticality concerning a natural language grammar.

It is necessary to point out that discrete approaches are still useful for linguistics. Linguists who stand by fuzzy approaches clarify that. Almela Pérez (2003:64) suggests that there is no need to banish binary oppositions where they give a good explanation of linguistic facts. Nor should we extend the non-discrete method beyond its possibilities. Hayes (1997) (in Keller (2000:19)) also claims that "we don't have to trash existing theories of what constraints are like just to get gradient well-formedness".

However, the above statements do not contradict the fact of approaching linguistics to fuzziness and gradient phenomena. Fuzzy tools do not compete against the discrete ones. Discrete approaches are gradual itself, but they only consider two degrees, strictly 1 and 0. Therefore, a fuzzy grammar can hold both resolutions for different objects: infinite intermediate points between 1 and 0, and binary oppositions as a particular solution of some intermediate point between two poles.

In the following section, a brief introduction to what is fuzzy logic is provided. Even though this work is mostly focused in linguistics rather than in mathematics, we think it is necessary to provide some bases about this tool in order to show its possibilities as a tool for linguistics as well as for a better understanding of our work. The application of fuzzy logic formulas for our grammar is very simple, but, at the same time, it is hugely effective in order to represent the degrees of grammaticality as a matter of the linguistic competence; disregarding of both the degrees of acceptability and the speakers' judgments. In fact, we would be unable to represent the degrees of grammaticality formally without the application of these fuzzy logic formulas.

Concerning our new approach, we do not want to provide a system for uncertainty; which is a system which represents how a speaker (or a grammar) predicts linguistic information taking into account an utterance. We want to provide a system which represents how a grammar deals with vagueness; that is a system which represents how a grammar acknowledges an input which has a membership degree of belonging in that grammar. In other words, a linguistic input is true in a certain degree for a specific grammar, without the necessity of predicting anything. In the future, this system could be extended to the speakers' cognitive representation of dealing with fuzziness in natural language, illustrating how the logical inference that a speaker does about a specific linguistic input is true in a degree rather than just probable. Therefore, a complete fuzzy grammar could be compared with cognitive approaches; testing how a fuzzy grammar processes linguistic inputs in contrast to how speakers process them. However, our current goal is not cognitive but theoretical, especially since grammaticality is a theoretical object.

4.3 Fuzzy Logic and Natural Language

Fuzzy logic's bases appeared in the paper *Fuzzy Sets* of Lofti Zadeh (1965). His system changed the paradigm in mathematics proposing fuzzy logic as a solution for those objects and context which were a trouble for Boolean logic.

Sorites paradox or the bald man paradox of Eubulides are some of those scenarios which mean a problem for classical logic. Sorites paradox sets out the mathematical problem of acknowledging at what point a grain of sand become a heap of sand. The bald man (*falakros*) paradox is pretty similar; it sets out that a man with a head full of hair is obviously not bald. On the other hand, a man with almost no hair is obviously bald. However, a problem arises when we want to establish a boundary to decide in which moment a man has lost enough hair to be considered bald. The question of this paradox is: how many hairs a man has to have, at least, to do not be considered bald?

Classical logic cannot solve these paradoxes since it needs to decide abruptly at what point the transition occurs either from a non-heap to a heap; or from a head full of hair to baldness. Fuzzy logic solves this paradox since there is not an abrupt transition from one membership class to the other. In fuzzy logic, a heap or baldness are truth in a degree. These paradoxes are mathematically solved in Hájek and Novák (2003) who claim that "sorites is not paradoxical at all".

Zadeh (1965) claimed that "a fuzzy set is a class of objects with a continuum of grades of membership. Such set is characterized by a membership (characteristic) function which assigns to each object a grade of membership ranging between zero and one."

Zadeh published *A Fuzzy-Set-Theoretic Interpretation of Linguistic Hedges* in 1972 (Zadeh, 1972). In this work, he displayed more definitions regarding what is a fuzzy set and fuzziness. These definitions established the bases of fuzzy logic:

> Roughly speaking, a fuzzy set is a class with unsharp boundaries, that is, a class in which the transition from membership to non-membership is gradual rather than abrupt. [...] The pervasiveness of fuzziness in human thought processes suggest that much of the logic behind human reasoning is not the traditional two-valued or even multi-valued logic, but a logic with fuzzy truths, fuzzy connectives and fuzzy rules of inference. Indeed, it may be argued that it is the ability of the human brain to manipulate fuzzy concepts that distinguishes human intelligence from machine intelligence.
>
> (Zadeh, 1972:4)

In Novák et al. (1999:9) fuzzy logic is defined as: "many-valued logic with special properties aiming at modeling of the vagueness phenomenon and some parts of the meaning of natural language via graded approach".

Fuzzy set theory resembles human reasoning in its use of vague information to generate decisions. People succeed on processing information by using language that is imprecise rather than precise such as *"very pretty"*, *"quite good"*, *"almost poisonous"*, *"London is close to Paris"*. Thanks to fuzzy logic, we could have a better definition of how, for example, a door, is not entirely closed (0) or open (1); it could be half-open (0.5), quite open (0.8) or almost closed (0.1), and so on. In the same way, we would like to capture how an utterance can be reasoned as *very grammatical* (0.8), *quite grammatical* (0.5), *almost grammatical* (0.9), and so on.

Therefore, the main aim of fuzzy logic is to capture vagueness; as well as the vagueness featured in natural language. If we want an artificial system, such as a formal grammar, to imitate human thinking (and linguistic gradience), we have to apply the right tools for that such as fuzzy logic. It is worth to point out that Novák (2008c:192) acknowledges that "it is important to stress that fuzzy logic *does not* explain vagueness but provides a reasonable mathematical model that enables us to deal with vagueness".

Novák et al. (1999) and Novák (2015) distinguish two types of fuzzy logic and two basic classes of methods with fuzzy set theory as a base.

Regarding the *types* of fuzzy logic they distinguish between:

1. *Fuzzy logic in narrow sense* (FLn) as "a special many-valued logic which aims at providing formal background for the graded approach to vagueness" (Novák et al., 1999:9).

2. *Fuzzy logic in a broader sense* (FLb) as "an extension of FLn and it aims at developing a mathematical model of natural human reasoning, in which principal role is played by the natural language" (Novák et al., 1999:9).

Regarding the *classes of methods*, Novák (2015) distinguishes between:

1. methods with linguistic motivation;

2. and methods with non-linguistic motivation.

Obviously, this work is utterly interested in the first class of methods which are also divided in two types by Zadeh (1996, 2004):

1. *computing with words*;
2. and *precisiated natural language*.

The first type of linguistic method is focused on classifying what in fuzzy logic is called *linguistic expression* which are mostly quantifiers and evaluative adjectives such as *very small*, *very big*, *quite medium*, *average*, and so on. The main objective uses to be substituting values of numbers on scales in questionnaires by these expressions; and allowing machines to deal with these linguistic expressions instead of being able to understand just numbers. In this sense, just dealing with these expressions is not actually using neither capturing natural language.

Precisiated natural language is the method which is more linguistically oriented in fuzzy logic and is focused on developing:

> A reasonable working formalization of the semantics of natural language without pretensions to capture it in detail and fineness. The goal is to provide an acceptable and applicable technical solution. The concept of PNL is based on two main premised: (a) much of the world's knowledge is perception based, (b) perception based information is intrinsically fuzzy. [...] as has been convincingly argued by many authors, vagueness is an unavoidable feature of natural language semantics. We argue that the idea of fuzzy sets an fuzzy logic provides a reasonable model of vagueness.
>
> (Novák, 2015:142)

Novák (2015) framework is motivated by Lakoff (1970). His approach for dealing with semantics in natural language is called *Fuzzy Natural Logic*. In this framework, he defines with high order fuzzy logic several linguistic objects such as nouns, adjectives, evaluative linguistic expressions, linguistic quantifiers and determiners, noun phrases and others. However, this framework is deeply oriented in semantics, pragmatics, and it gives attention to some morphological objects.

Thus, somehow, we can say that both formal linguistics and fuzzy logic have an objective in common: the formalization of natural language. In this sense, our work is innovative not only in the field of linguistics, but in the area of fuzzy logic too, since frameworks in fuzzy logic are focused in semantics while our work is pursuing to represent syntactic objects such as constraints between word classes and constructions, and degrees of grammaticality.

The new framework introduced in this dissertation has applied the bases of Novák's *Fuzzy Natural Logic* in order to define certain linguistic objects in terms of degrees such as a *fuzzy grammar*, the structure of the concept of *word*, the representation of degrees of grammaticality computed with words, and the mathematical representation of *the degrees of grammaticality as part of the linguistic competence*.

Before we introduce some fuzzy logic bases and its degrees of truth, the following subsections provide the definitions of several concepts which use to be confused one to each other. Most of these concepts are considered all-together in linguistics under the notion of gradience. Fuzzy logic can help linguistics by procuring definitions which distinguish all those concepts that use to be classified as gradience. These concepts are: *vagueness, indeterminacy, ambiguity, randomness* and *uncertainty*.

4.3.1 Indeterminacy and Vagueness

Vagueness plays the main role when it comes to talk about concepts in fuzzy logic. Sometimes *vagueness* is misunderstood as *uncertainty*. These both concepts are gathered together under *indeterminacy*. In Novák et al. (1999), *indeterminacy* is considered as the definition in Merriam Webster Dictionary:

> Indeterminacy: the quality or state of being indeterminate.
> Indeterminate: 1. a) not definitely or precisely determinate or fixed. VAGUE. b) not know in advance c) not leading to a definite end or result.
>
> (Merriam Webster Dictionary: https://www.merriam-webster.com/)

While 1.a) belongs to the definition of vagueness, 1.b) belongs to uncertainty. 1.c) could be interpreted as the non-discrete nature of *indeterminacy*.

Therefore, *indeterminacy* is a concept which gathers both vagueness and uncertainty at the same time. This way of treating this concept could help linguistics to identify these phenomena which are vague and uncertain at the same time.

However, it is worth to mention that sometimes in fuzzy logic literature, some authors use the notion of *indeterminacy* understood just as *uncertainty*.

4.3.2 Vagueness as a Main Concept of Fuzziness

Vagueness is the main phenomenon which is attempted to be represented in fuzzy logic. Vagueness is found when we want to group in X certain objects (x, y, z...) which have a certain property φ. In Novák et al. (1999:3) the following formalization is provided:

$$X = \{x \mid x \ \ has\, the\ property \ \ \varphi\} \tag{4.1}$$

Two types of objects can be found when dealing with vagueness:

1. **Borderline objects**
2. **Typical** or **prototypical objects**.

The first type of objects show up problems for being grouped in X since it is unclear if they have the property φ. The second type refers to those objects which clearly presents the property φ . In this sense, fuzzy logic deals with vagueness taking into account the typical objects and comparing them to other candidates, checking if the property delineated in the typical one is still remaining on the borderline case in somehow.

One of the most characteristic features of vagueness is its *continuity*. Continuity immediately refers to the sorites' paradox and Zadeh's definitions in which the transition of memberships occur gradually rather than abruptly. In this sense, there is not an exact point in which a non-heap become a heap, or a hair man become bald; it is a gradual transition in which every new state is closer or farther concerning to the typical object.

In Novák and Dvorák (2011:7), vagueness of natural phenomena is defined as: "phenomena recordable by human mind (usually using natural language) is a consequence of the indiscernibility phenomenon. Characterization of indiscernibility using degrees leads us to the concept of fuzzy equality (fuzzy equivalence)".

4.3.3 Uncertainty and Vagueness

Uncertainty arises due to the lack of knowledge over the occurrence of some object. The more input we gather, the more capacity we have to predict probabilities over a context or an object. The truth value of an uncertain object is not known. It can display potential outcomes, solution or choices. Frequently, uncertainty slightly disappears as long as we have tools to predict the possible outcomes. The mathematical tool for dealing with uncertainty is probability theory. Probabilities can predict in which degree a linguistic phenomenon might occur.

Linguistics has been genuinely interested in uncertainty and into probability theory because of the capacity of a speaker to predict sentences and meaning. Once a speaker starts to receive information about an uncertain linguistic input, a speaker can start to deal with uncertainty to try to grasp some information. That is when an interlocutor is speaking with someone with low competence in a language, and the interlocutor cannot do more than guess what the other interlocutor is talking about. The more knowledge he has, the more he can figure out and starting to re-build the received sentences in order to grasp some sense.

As stated by Bresnan (2011):

> Most remarkably, language users have powerful predictive capacities, which can be measured using statistical models of spontaneous language use. [...] New models based on the mathematics of uncertainty have become more widely available to linguists. These provide much more powerful ways to cope with the complexity and variability of language, and these, too, reveal hidden structure and even richer mental capacities.
>
> (Bresnan, 2011:71-72)

If we contrast uncertainty with vagueness, at first sight, we can say that uncertainty is a matter of possibilities and predictions, while vagueness is a matter of truth taking into account non-sharped boundaries and transitions from one membership to another.

The concepts of *actuality* and *potentially* might help us with this distinction:

> In classical set theory, every set is understood to be actualized, i.e., we imagine all its elements to be already existing and at disposal to us in one moment. This concerns both finite as well as infinite sets. Though we can always see only part of the infinite set, our reasoning about any set stems from the assumption that it is at our disposal as a whole. On the other hand, most events around us are only potential, i.e. they may, but need not, to occur or happen. Thence, the difference between actuality and potentiality corresponds to the difference between vagueness and uncertainty.
>
> (Novák et al., 1999:5)

In this sense, an *actualized* group of objects refers to a group X defined with non-sharp boundaries. While a *potential* group of objects refers to speculating the existence of a whole X; however just a part of it exists. If we consider an object x with the property φ, we know it belongs to the group X; hence, it is part of the actualized grouping of objects. However, if x might never exist we do not know of its belonging to X; therefore we cannot recognize its truth.

According to these statements, we can recognize that vagueness just concerns the object itself while uncertainty relies on external conditions rather than on the object's features. Vagueness remains on time while uncertainty is temporary.

I would like to provide a popular example which pictures very well the difference between uncertainty and vagueness. Imagine a man who is being in captivity and who is extremely thirsty. He is provided with two bottles of water. One bottle is labeled as *0.1 membership in the class of poisonous*. The other bottle is labeled as *10% of probabilities of being poison*. If he chooses the first bottle for some sips, he will probably survive. Being the bottle 0.1 in the class of poisonous will not affect the prisoner enough to kill him. However, if the prisoner would choose the second bottle, he might die in a case that bottle is part of that 10% of probabilities of being full poison; which is being 1 membership in the class of poisonous. The features of the first bottle remain on time, while once the second bottle is tested, we would know if it belongs to that 10% or not.

Following this reasoning, a grammar gathers a compilation of rules or constraints; those constraints have a degree of truth between one each other and its part-of-speech when we take into account grammaticality as a vague object. However, a linguistic input might be uncertain for a speaker in case he is not fully competent in a language and, therefore, he has a lack of knowledge of the constraints of the grammar. A grammar might find a linguistic input uncertain when an input belongs actually to other grammar; i.e., parsing with a Spanish grammar a German input. The grammar cannot understand truth from it; however, with a stochastic tool, it could guess features.

Therefore, a grammar tool for parsing the linguistic input and its vagueness (how many constraints are satisfied or violated) in order to determine a degree of grammaticality differs from other tools such as predicting structures, predicting constraints or re-building uncertain sentences through probabilities.

4.3.4 Randomness and Ambiguity

Randomness is a particular form of uncertainty. Probability theory is the right tool to predict this phenomenon. Randomness is characterized in terms of probability as our expectation to certain phenomena which occur in the world. This concept is essential for atomic physics and quantum mechanics which recognize it as a deep feature of nature.

Ambiguity refers to the fact of having an object which can be interpreted equally in various ways. It frequently occurs in natural language expressions when one single linguistic object displays more than one possible meaning such as *"I saw a man on a hill with a telescope"*. We know that this sentence has at the same time all the following meanings, therefore is ambiguous; even though a speaker might find one more probable than the other:

(1) *"I was using a telescope with which I saw a man on a hill."*

(2) *"The man on a hill has a telescope and I am watching him."*

(3) *"I am sawing a man on a hill and the hill as a telescope on it."*

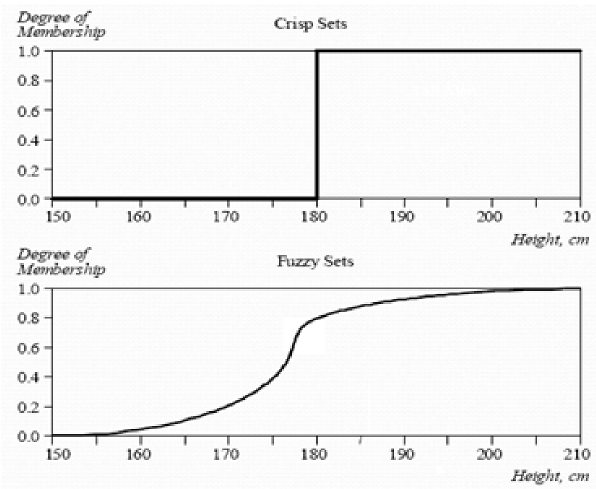
(4) *"I am on a hill and I saw a man using/carrying a telescope."*

4.4 Fuzzy Set Theory and Truth Values

Fuzzy logic is based on fuzzy sets. As humans, we manage vague concepts everyday such as *young*, *old*, *tall*, *short*, *close*, *far* and so on. In linguistics, we also deal with such vague concepts in terms of meaning as well as with abstract theoretical concepts i.e. *grammaticality*, *complexity*, *acceptability*, *coercion*, etc. In fuzzy logic, a fuzzy set is expressed with a function:

$$A\colon U \longrightarrow L \tag{4.2}$$

Each element x (208cm; 179cm; 167cm...) is an object in our large set called Universe U (a group of people), $x \in U$. A fuzzy set A (i.e. "being tall") assigns every object x from an Universe U with a value L between 0 and 1 [0,1]; we can represent this in the following expression: $A(x) \in L$. The value of this last expression is the membership degree of x in the fuzzy set A.



Name	Height, cm	Degree of Membership	
		Crisp	*Fuzzy*
Chris	208	1	1.00
Mark	205	1	1.00
John	198	1	0.98
Tom	181	1	0.82
David	179	0	0.78
Mike	172	0	0.24
Bob	167	0	0.15
Steven	158	0	0.06
Bill	155	0	0.01
Peter	152	0	0.00

Fig. 4.1 Crisp graph in contrast with a fuzzy graph. Both represented in a table (https://slideplayer.com/slide/12816051/).

Regarding Figure 4.1, a fuzzy set is able to express the membership of an apparent ambiguous concept such as "being tall or short":

- In the scale L, an object (x) has a value of 1 when it has the property of being "tall" (φ); like *Chris* and *Mark*.

- An object has a value of 0 when it doesn't have the property φ at all; like *Peter*.
- An object x has a value of a degree between 1 and 0 when it partially has the property φ. The closer is the object to the maximum criteria, the taller it is; like *John*, *Tom*, and *David*. The farther is the object from the maximum criteria, the less tall it is, therefore, the shorter; like *Mike*, *Bob*, *Steven* and *Peter*

As we can see in the chart, classical logic fits perfectly with a fuzzy expression; therefore, it is not true that fuzzy logic is *against* classical logic; fuzzy logic is a more precise and extended expression of the discrete degrees; in fact, stretching this reasoning, classical logic is a type of fuzzy logic. In this sense, fuzzy logic models those objects which are not entirely well represented with just two values neither with just probabilistic approaches.

In fuzzy logic, it is possible to model *degrees of truth* of every object taking into account the context and other parameters. For example, the degree of truth for every object regarding "tall" is not the same in Sweden than in Philippines. In Sweden, a person of "170cm" might be "not very tall" while in Philippines it might be "definitely tall". However, the values in fuzzy logic are universal, since we all know that "very tall" means a high value regardless of the criteria behind it in order to achieve that value.

Additionally, there are no negatives degrees of truth in fuzzy logic. Objects which are negative numbers can be modeled, i.e., modeling "freezing" taking into account objects such as 0ºC, -5ºC, -10ºC. The degrees of truth L will consider those negative values in a range of 0-1 for the fuzzy set of "freezing". The degrees of truth can model an infinite amount of objects always between 0-1 transforming any value into that range, i.e., 0ºC corresponds to the hypothetical value of 0.2; -5ºC corresponds to 0.5, and -10ºC corresponds to 0.76.

There are different methods to assign degrees of membership for an object; some of them are: individual criteria (objective or subjective), methods based on experimental procedures, methods based on distributional criteria, and so on. It is frequent to see interdisciplinary methodologies working together in order to assign a degree of membership for a group of objects.

4.4.1 Fuzzy *IF-THEN* Rules

These rules constitute the tool in order to represent *approximate reasoning* for vague linguistic variables such as *very*, *almost*, *less*, *young*, *old*, etc. These rules are also called *linguistic description*. They consist of at least two evaluative linguistic predictions united by an implication. Fuzzy IF-THEN rules are generally represented as:

$$IF\ X\ is\ \mathscr{A}\ THEN\ Y\ is\ \mathscr{B} \tag{4.3}$$

Equation 4.3 is the basic rule applied in many electronic devices such as in air conditioners. We could easily find a rule like: IF *temperature is high* THEN *the amount of cool air* is *big*. We leave here a more complex example in Figure 4.2 providing a more complex reasoning.

Condition:	IF the obstacle is *near* AND the car speed is *big* THEN break *very much* IF the obstacle is *far* AND the car speed is *rather small* THEN break *a little*
Observation:	the obstacle is *quite near* AND the car speed is *big*
Conclusion:	break *quite much*

Fig. 4.2 Approximate reasoning scheme example from Novák et al. (1999).

Many complex ways of reasoning can be found in fuzzy logic; however, we are not going to explain them here. Our work applies fuzzy sets to linguistics rather than IF-THEN rules.

4.5 Applying Fuzzy Sets to Grammaticality

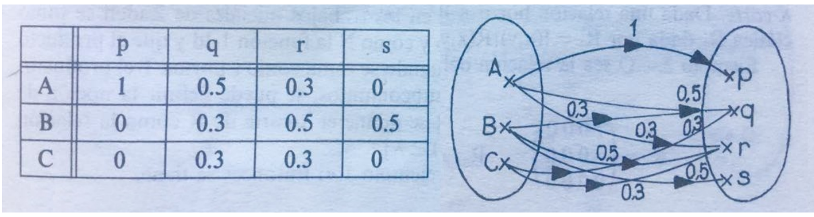
Some parallelism between fuzzy sets and grammaticality are presented in this section. The first attempts of how to apply fuzzy sets to grammaticality are made here, and they are original work of this PhD.

4.5.1 Grammaticality as a Fuzzy Set

The basic idea in our work is taking into account a fuzzy set for representing grammaticality as a vague object. The first idea was considering a fuzzy set a grammar G which assigns a degree of truth L to each constraint in a linguistic input I; according to their importance and considering if they have been satisfied or violated.

$$G\colon I \longrightarrow L \tag{4.4}$$

The idea is that the grammar does not consider every rule as a 1 or a 0 regarding a linguistic input. The grammar takes into account special rules which are triggered when some canonical rules are violated; capturing vague linguistic phenomena.



	p	q	r	s
A	1	0.5	0.3	0
B	0	0.3	0.5	0.5
C	0	0.3	0.3	0

Fig. 4.3 Fuzzy representation from Trillas and Ríos (1992) with both tables and graph.

Taking into account Figure 4.3, we can substitute *p*, *q*, *r*, *s* for constraints in the grammar, and *A*, *B*, *C* for linguistic inputs such as utterances, sentences, constructions, even words. Every linguistic element is matched regarding a rule from a grammar in terms of degree.

	C_p	C_q	C_r	C_s
CandidateSentenceA	1	0.5	0.3	0
CandidateSentenceB	0	0.3	0.5	0.5
CandidateSentenceC	0	0.3	0.3	0

Table 4.1 Capturing the values from Figure 4.3 with an Optimality Theory representation.

It seems clear that Figure 4.3 and Table 4.1 are very similar. We could say that Optimality Theory is close to a fuzzy expression. However, as we have shown before, Standard OT would select just the most optimal candidate, as well as other theories with OT base. On the other hand, all the candidates in Table 4.1 are optimal in terms of degree. Taking into account their fuzzy values, we could rank the candidates in Table 4.1 as $CandidateSentenceA \gg CandidateSentenceB \gg CandidateSentenceC$. This picture is what we want to capture in our work. All the candidates remain with a positive value, being grammatical to a certain degree regarding a grammar.

In contrast with the degrees of ungrammaticality and degrees of acceptability with a negative value, we would need to establish a clear rank in order to transform those degrees in fuzzy degrees. We mentioned the impossibility of a final value of grammaticality with a negative number in section 3.8 and subsection 3.9.1. As we see here, fuzzy logic coincides with the fact that it is not possible to characterize a fuzzy set with a degree lower than 0. Therefore, we cannot express grammaticality as a fuzzy object with a value such as -15. Even though such thing would happen, the degree of the negative number should be either 0 or being adapted to a new scale of values between 0-1.

4.5.2 Grammaticality with IF-THEN rules

In this work, we do not look for a formal application of the fuzzy IF-THEN rules. In contrast, we have used fuzzy sets in order to represent formally a fuzzy grammar and its degrees of grammaticality; as well as other linguistic objects.

In our opinion, IF-THEN rules are going to be more convenient once the fuzzy grammar has evolved towards a more robust multi-modal framework. Linguistics might find fuzzy IF-THEN rules as a useful tool in order to compute with words some theoretical concepts.

As an example, we show some basic reasoning with IF-THEN rules taking into account grammaticality. For this reasoning, we treated the value of grammaticality in terms of the number of violations as in Figure 4.4.

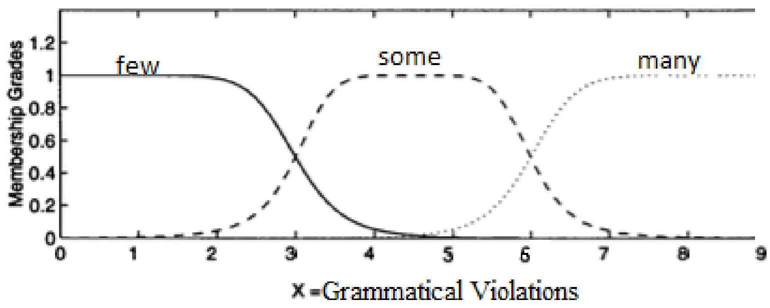


Fig. 4.4 Fuzzy graph of hypothetical values regarding number of violations.

Therefore, we consider (1-4) few violations, (2-7) some violations, (5-9 and more) many violations. Also, we take into account approximate fuzzy reasoning such as:

- IF *an input* is violated *few times* THEN *the value of grammaticality* is *high*.
- IF *an input* is violated *some times* THEN *the value of grammaticality* is *medium*.
- IF *an input* is violated *many times* THEN *the value of grammaticality* is *low*.

4.6 Fuzziness vs. Gradience in Linguistics

In the linguistic literature, all the gradual phenomena are underpinned under the term of *gradience*. As we have seen, fuzzy logic provides a good amount of theoretical definitions in order to deal with all those gradient phenomena which are not the same such as: *randomness, possibility, probability, predictability, vagueness* and so on. We suggest that linguistics would be nourished from the distinction of these concepts.

Distinguishing fuzziness from gradience and other phenomena will help linguistics for proposing appropriate theoretical frameworks in order to deal with specific language phenomena. In this section, we clarify some linguistic concepts by taking into account the already mentioned concepts from fuzzy logic. The following definitions are part of the original work that we present in this PhD.

4.6.1 Fuzziness in Linguistics

We find fuzziness in linguistics regarding those objects which their membership cannot be classified satisfactorily within a discrete approach of all-or-none.

Firstly, fuzzy objects in linguistics are understood as vague objects in fuzzy logic. Some of them are clearly vague such as the degrees of grammaticality. The degree of grammaticality would determine the belonging of a linguistic input towards a grammar. Therefore, a linguistic object is fuzzy, and the degrees are the method to define those objects (Almela Pérez, 2003).

Secondly, in linguistics, a fuzzy object can be understood as those linguistic objects which can fit on roles for which they canonically were not meant to be. In syntax, we find part of speech such as adjectives in noun fits (*"the young and the old were cheering up together"*), nouns in adjective fits (*"I have never seen a horse race"*), and so on. In semantics, we find the phenomenon of coercion allowing some words with specific meaning performing its semantics without triggering a full violation, like in *"The hamster explored a backpack"* (satisfied); *"The hamster lift a backpack"* (coercion); *"The hamster entertained a backpack"* (violation). Other semantic phenomena such as metonymy and metaphors could be understood as vague objects.

There are two problems concerning these fuzzy objects:

1. The first is found on its barely gradient nature stating for an intermediate degree. We cannot classify these cases categorically as a black-or-white matter and neither as a clear continuum. And yet, they seem clearly are vague by means of theoretical reasons.

2. The second problem constitutes of the extreme difficulty of classifying them in a continuum (degree) regardless of frequencies and distributional corpus methodology.

4.6.2 Gradience, Indeterminacy, and Uncertainty

Gradience is the linguistic term which takes into account all gradual phenomena in natural language. In this sense, gradience accounts at the same time for both vague and fuzzy phenomena such as degrees of grammaticality; as well as for uncertain phenomena such as degrees of probability and predictability of linguistic elements.

In this sense, gradience would be understood as what *indeterminacy* is in fuzzy logic: a phenomenon that we meet with both vague and uncertain aspects at the same time.

Degrees of acceptability are going to be considered as degrees of indeterminacy since acceptability both combine degrees of truth plus degrees of possibility, predictability and other uncertain objects.

Uncertainty refers to those linguistic objects which need of probabilistic tools. Probabilities will help us to deal with utterances with a meager degree of grammaticality as well as with ambiguous sentences; since for the speakers, despite the fact of ambiguities, some possible meanings are acknowledged as more probable than others. Tools for uncertainty will help to recreate the ability of a speaker in order to re-build or to reconstruct unintelligible sentences. Linguistics has already widely explored the probabilistic traits in the language.

4.6.3 Gradient Systems and Fuzzy Systems

Gradient systems (grammar) are formal frameworks in which gradient phenomena such as degrees of probability, acceptability, grammaticality or predictability are defined. It might use different formal approaches to do that; however it does not *necessarily* explain fuzzy relations between different linguistic objects. A gradient system does not *necessarily* need to take into account constraints. Stochastic grammars could be a type of a gradient system (or grammar).

Fuzzy systems (grammar) are formal frameworks which define any kind of linguistic information in any context (as humans do). A fuzzy grammar is able to describe linguistic objects taking into account its fuzzy membership and vague features when it is needed. *This framework is set through a flexible constraint system which describes a natural language grammar*. These constraints are known as properties, and they work as logical operators that represent grammatical knowledge. They are flexible because they can be violated or satisfied to different degrees. A fuzzy grammar might include phenomena described by a gradient system.

In contrast with a gradient system, a fuzzy grammar is a deeply oriented-focused system on defining both vague memberships and objects. Not all gradient systems are a fuzzy grammar, but a fuzzy grammar can be a type of gradient system. Nevertheless, a gradient system such as probabilistic grammars can be the perfect complement for a fuzzy grammar.

4.6.4 Processing Linguistic Gradience and Linguistic Fuzziness

Processing linguistic gradience refers to our capacity to process multiple kinds of gradient objects in the language. This capacity can be either related to fuzziness or uncertainty.

Processing linguistic fuzziness refers to our capacity to understand vague inputs as well as classifying them in sets. These inputs can be valued by a scale of degrees, such as degrees of grammaticality, degrees of meaning, degrees of complexity, and so on. In fuzzy logic, this might be understood as truth values.

4.7 What Fuzzy Theory is Better for Linguistics?

As we have shown, we can find both uncertain and vague phenomena in the same natural language utterance.

Tools for dealing with probabilities will reproduce the ability of a speaker in order to predict structure in a continuum plus to assume different possible meanings over a sentence. The first refers to the speakers' ability who is able to finish sentences before the other interlocutor actually finishes them. The second concerns to the fact of understanding your interlocutor over the ambiguity that characterizes natural language.

Tools for dealing with vagueness will provide a solution to define those elements which are vague such as grammaticality. A sentence is grammatically truth in a degree depending on how many constraints are satisfied or violated taking into account a grammar. This concept of grammaticality is not a matter of probabilities, since there is not an acknowledgment of how grammatically probable is a sentence. We are defining in what degree a sentence is true concerning a grammar.

In this sense, grammaticality seems obviously a vague concept. However, defending vague relations between part-of-speech seems more controversial.

In my regard, tools for vagueness will also imitate the ability of a speaker in order to acknowledge linguistic variability regarding a grammar. This ability refers to the ability of the speaker's capacity to acknowledge some linguistic constraints which seems contradictory, but they are not. A clear example might be the precedence of the adjective concerning a noun in Spanish. All Spanish native speakers know that the noun typically precedes the adjective in Spanish such as in *"un coche bonito"* (*a car pretty*). However, they also know that, with those elements, they can use the adjective before the noun to emphasize the meaning of *pretty*: *"un bonito coche"*. This variability is not a matter of probability; this represents the vague distributional relationship between the noun and the adjective which must figure in a grammar of Spanish language. Nothing has to be predicted here; any native speaker acknowledges this vague relation. Moreover, at the same time, it is not just a matter of occurrences, the second structure is linguistically more marked than the first one since some specific elements do not allow the special precedence of the adjective over the noun such as in *"un chico bonito"* (*a boy pretty*) and *"un bonito chico"* (*a pretty boy*). Consequently, this is not a matter of probability; this is a matter of how certain constraints which allow or not the vague relations within linguistic categories.

Therefore, if we take into account all adjectives in Spanish and we analyze its flexibility within a nominal structure, we could speak of a general degree of truth regarding its precedence or not. Nevertheless, the most controversial part in this reasoning is the fact that these linguistic variability does not seem as gradual as grammaticality also neither as other fuzzy objects. Concerning the noun-adjective case, if we avoid its general perspective and we focus on the particular cases, it seems a more discrete phenomenon rather than gradual since it just seems that some specific adjectives allow linguistic variability while others do not.

In conclusion, a grammar which has to deal with language, just like humans do, needs of tools for dealing with uncertainty and vagueness. However, uncertainty in language has been more explored than vagueness. In this work, we offer a new perspective acknowledging grammaticality as a vague object which can be defined in terms of degrees. The degree of grammaticality has a positive value since grammaticality is a vague object which determines the extent to which an input belongs to a grammar in terms of degrees of truth.

Linguistics has already considered fuzzy logic as a tool in works such as Lakoff (1970) and Ross (1972). Some mathematicians already have applied fuzzy logic in order to formalize features from natural language.

Nevertheless, this work is innovative because of its interdisciplinary methodology: it uses formal tools such as fuzzy logic and grammar with constraints together with corpus analysis methodology in order to extract a property grammar of Spanish Syntax.

However, what fuzzy theory is better for linguistics? There are many different approaches in order to formalize fuzzy logic as well as different theories: Fuzzy Logic in a broad sense (FLb), Fuzzy Logic in a narrow sense (FLn), Fuzzy Type Theory, and so on.

For our work, we have chosen Fuzzy Natural Logic by Novák (2005); Novák (2015); a high-order fuzzy logic; fuzzy type theory (FTT) with Łukasiewicz algebra. This theory is highly suitable for modeling natural language and other linguistic concepts. This theory is genuinely linguistic motivated and highly influenced by Lakoff (1970) and Montague Grammars (Montague, 1970).

Novák (2016) points out the main expected contributions by this theory:

- *Development of methods for construction of models of systems and processes on the basis of expert knowledge expressed in genuine natural language.*

- *Development of algorithms making computer to "understand" natural language and behave accordingly.*

- *Help to understand the principles of human thinking.*

A more technical explanation of this theory is provided in subsection 6.2.1.

Chapter 5

Property Grammars: A Fully-Constrained Theory to Deal with Degrees of Grammaticality

Property Grammars (PGs) were introduced by Blache (2000) who defined them as a formalism based exclusively on the use of constraints. The framework has been updated several times; most importantly by Blache (2004); Blache and Balfourier (2001); Blache and Prost (2008). The state-of-the-art in Property Grammars is reported in Blache (2016), which provides an extensive explanation of the theory, clarifies the tools available and describes their potential for linguists who would like to implement them in a natural language grammar. Our proposal is largely based on the newest model proposed in 2016.

In this chapter, we introduce Property Grammars, their main features and their advantages over other frameworks. In general terms, Property Grammars define linguistic inputs under several more constraints than other theories which mostly focus on just one or two. These constraints work as logical operators and are known as properties. They define the grammatical relations between the parts-of-speech of an input. However, Property Grammars stand out for their capacity to represent any kind of input regardless of its form or grammatical violation. A property grammar displays all the linguistic information of an utterance with constraints that are both satisfied and violated. All these characteristics make them ideal for representing natural language inputs, which are frequently not fully grammatical because of their natural idiosyncrasy: spontaneity, immediateness and *noisy text*. For these reasons, this theory is the one that best explains grammaticality as a fuzzy-gradient object.

Nevertheless, these constraints on parts-of-speech are displayed differently in every language. The constraints require corpus methodologies and theoretical reasons to be extracted so that a grammar of a natural language can be built. Property Grammars have been applied mostly in French and English but they have yet to be used to describe Spanish. Therefore, our work is the first to implement PGs to represent Spanish, the second most spoken language. Hence, we feel that Spanish Property Grammars can make an important contribution to improving the analysis of all kinds of spontaneous language in Spanish.

Throughout the history of linguistics, grammars have used the concept of constraint as a key notion in their theoretical proposals. Some of these grammars are Categorial Grammar (CG) (Ajdukiewicz, 1935) and Dependency Grammars (DG) (Tesnière, 1959). However, most grammars and constraint models were proposed in the $70^{\prime s}$ and $80^{\prime s}$ as alternative theories to the transformational models of Chomsky (1965): Tree Adjoining Grammar (TAG) (Joshi et al., 1975), Lexical Functional Grammar (LFG) (Kaplan and Bresnan, 1982), Generalized Phrase Structure Grammar (GPSG) (Gazdar et al., 1985), Optimality Theory (OT) (Prince and Smolensky, 1993), Head-Driven Phrase-Structure Grammar (HPSG) (Pollard and Sag, 1994), among others.

All these theories share the use of constraints as a tool to represent the properties that linguistic input must satisfy. Constraints have been used specially in syntax and phonology. There are two types of constraints: *universal* (valid for any language); and *specific* (valid for a specific language). Therefore, the general idea is to stipulate properties that discount or eliminate structures that do not belong to a language. Linguistic constraints are also named *properties* in Property Grammars. The properties are represented as *logical operators with universal traits*. However, these properties *might interact differently with parts-of-speech from one specific language or another*.

Although all these constraint theories use constraints to make detailed descriptions of the syntactic aspects of a language, they end up neglecting other constraints which are also relevant. For this reason, Property Grammars were created to collect the most important constraints in the different grammars with constraints, and bring them together in a single and unique homogeneous model. The idea is to represent natural language processing with just one formal grammar. Property Grammars approach grammar as a set of statements which make it possible to describe any kind of input, rather than as a device for generating language (Blache, 2004, 2016).

Property Grammars are unlike all previous theories since they do not attempt to generate structures. Their main aim is to describe any type of input regardless of whether it is well formed or not (that is to say, regardless of its level of grammaticality). This is possible thanks to the notion of construction presented by the Construction Grammars (Goldberg, 1995; Kay and Fillmore, 1999), which understand a linguistic construction as a pair of function and structure such as in the example of *"What is [this scratch] doing [on the table]?"*. These approaches acknowledge this utterance as a whole which cannot be fully explained by the collection of its parts. Therefore, its structure and meaning must be explained together. In this regard, a grammar must be prepared to acknowledge structures which cannot be simply explained by phrases and well-formedness. We can even find constructions without heads such as in *"Monday, ironing; Tuesday, shopping; Wednesday, cleaning"*. Property Grammars borrow the concept of construction from construction grammars, use properties to describe them, and conclude that "In a grammar: construction = set of properties" (Blache, 2016:209).

5.1 Properties for a Highly Descriptive Grammar

Property Grammars aim to describe the local language phenomena for representing natural language processing. PGs assume that linguistic information must be decoded immediately and independently from the form of the final construction, an idea inherited from both the TAG and the HPSG approaches. Because of these notions, PGs describe any input regardless of their final degree of grammaticality.

PGs consider that there are two ways of looking on grammar: as a set of descriptions that explain the knowledge that a speaker has about the language; or as a set of operations, or rules, which are capable of generating language. Therefore, PGs focus on the first way, since their interest is to propose a formal model that can explain the knowledge that a speaker has of a language regardless of the final grammatical value of a generated input (Blache and Prost, 2008). By default, this will help to represent the gradual phenomenon of language more effectively.

5.1.1 The Properties

The entire Property Grammars model uses constraints we call *properties*. These properties express the linguistic relationships between the lexical categories of a language. A property is defined in the literature as a constraint that represents relationships between words without requiring a superior structure or other elements. In short, "a property is a relation between two words, nothing more" (Blache, 2016:188). Consequently, properties can be described independently at the same level and define syntactic relations between categories. Therefore, they are able to describe a grammar as well as the knowledge of a speaker regarding a language: "the grammar is a set of constructions, each described by a set of properties" (Blache, 2016:210). The constraints, or properties, responsible for defining the different types of linguistic information are presented here.

Linearity ($\prec$)

Precedence order between two elements. A precedes B. Therefore, a violation is triggered when B precedes A. A typical example of this property can be found with the precedence relation between the determiner (DET) and the noun (NOUN) in English: *"The kid"*: $DET \prec NOUN$. Blache (2016:190) points out that this property is inspired by GPSG. This relation is independent from hierarchical, requirement or dependency features.

Co-occurrence ($\Rightarrow$)

Co-occurrence between two elements: A requires B. A violation is triggered when A occurs, but B does not. A typical example of this property in English: *"The girl plays football"*: $NOUN \Rightarrow DET$. A violation would be: *"girl plays football"*. Besides, co-occurrence demands at the same time that B requires A. This property is non-hierarchic and non-headed. Therefore, the co-occurrence property must figure in both categories since, if A is missing, the property will not be triggered because the tagging process of a grammar cannot find A (the $NOUN$ category), and yet a violation is happening. In *"The plays football"* there is no A (a $NOUN$ in this case), so the requirement $A \Rightarrow B$ and its violation cannot be triggered. In conclusion, because of the bilateral idiosyncrasy of this property, co-occurrence must be explicit for both co-occurring property descriptions: 1) in NOUN: $NOUN \Rightarrow DET$; 2) in DET: $DET \Rightarrow NOUN$.

Exclusion ($\otimes$)

Exclusion between two elements: A and B never appear in co-occurrence in the specified construction. That is, only A or only B occurs. A violation is triggered if both A and B occur. An example of this property in English is the exclusion between the pronoun (PRON) and the noun (NOUN): *"He boy does yoga"*: $PRON \otimes NOUN$. Unlike co-occurrence, this property does not necessarily figure in both property descriptions. In Spanish, for example, the determiner can occur with any nominal element (N) (either a noun, a proper noun or a pronoun) for reasons of linguistic variability: *"El yo está presente en toda la obra"* (*The* [philosophical] *I is present in the whole work*); *"El Manchester juega muy bien"* (*The Manchester* [United] *plays very well*). However, both proper noun (PROPN) and PRON exclude the determiner since they canonically appear without a DET throughout the Spanish language. Therefore, thanks to Property Grammars and this exclusion property, we can represent vague cases of linguistic variation by defining the following: $PROPN \otimes DET$ and $PRON \otimes DET$, while $DET \prec N$.

Uniqueness (*Uniq*)

Neither a category nor a group of categories (constituents) can appear more than once in a given construction: In a construction X, $Uniq = \{a, b, c, d\}$. A violation is triggered if one of these constituents is repeated in a construction. A classic example in English is the non-repetition of the determiner and the relative pronoun concerning the nominal construction: *"The **the** kid that **who** used to be my friend"*: Nominal construction: $Uniq = \{Det, Rel\}$.

Dependency ($\leadsto_x$)

An element A has a dependency$_i$ on an element B. A violation is triggered if the specified dependency does not occur. A classic example in English is the relation between an adjective (ADJ), with the dependency of a modifier, and a noun: *"Colombia is a big country"*: $Adj \leadsto_{mod} NOUN$. One such violation might be: *"Colombia is a big badly"*. This property can be perceived as the syntactic property which connects syntactic features with semantic ones.

The dependency properties have to satisfy some formal characteristics illustrated in Blache (2016:197):

- Antisymmetric: if $A \leadsto_x B$, then $B \not\leadsto_x A$
- Antireflexive: if $A \leadsto B$ then $A \neq B$
- Antitransitive: if $A \leadsto_x B$ and if $B \leadsto_x C$ then $A \not\leadsto_x C$.

It is worth mentioning that this dependency property does not have exactly the same notion of dependency as that of Dependency Grammars (Tesnière, 1959). In PGs, dependencies strictly define the relation between two categories rather than express a hierarchical relation between a head along with its dependents, which maybe more than one.

In PGs, a dependency head is reduced to its mildest expression; just one element in relation to another. An extended explanation on this can be found in section 5.3. Besides, Blache (2016:198) mentions that "the main difference in PG is that the dependency graph is not necessarily connected and does not necessarily have a unique root".

Figure 5.1 display all the dependencies included in PG.

Besides, we show in Figure 5.2 a graph representing the dependency properties in PGs from Blache (2016:197).

dep	generic relation, indicating dependency between a constructed component and its governing component
mod	modification relation (typically an adjunct)
spec	specification relation (typically *Det-N*)
comp	the most general relation between a head and an object (including the subject)
subj	dependency relation describing the subject
obj	dependency relation describing the direct object
iobj	dependency relation describing the indirect object
xcomp	other types of complementation (for example between *N* and *Prep*)
aux	relation between the auxiliary and the verb
conj	conjunction relation

Fig. 5.1 List of dependencies in PG from Blache (2016:195).

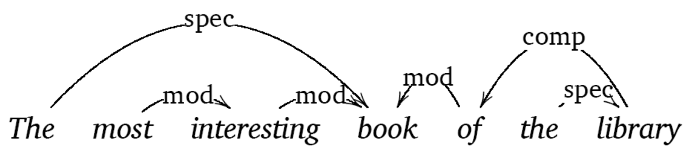


Fig. 5.2 PG dependency graph from Blache (2016:197).

Older properties

PGs definitely use several constraints to describe the syntactic relations between local language phenomena. However, they had different properties in older versions of PGs which are not included in the 2016 version. Blache (2004) explains that the linguist is free to use all of the properties, just some of them, or to add new ones to produce a proper set of descriptions which represent the knowledge a speaker has about the features of the language. In order to provide an overview of PG, some of its older properties are briefly explained:

- **Constituency** (*Const*). This property was used to identify categories and produce maximal set of categories. This property was the first to be applied in the first parsing phase of PG, called *characterization* (Blache, 2000, 2004; Blache and Prost, 2008). In the literature on PG, constituency is sometimes dismissed as a property so that it can be considered as a part of the process of parsing properties.
- **Obligation** (*Oblig*). This property determines which elements are heads (Balfourier et al., 2002; Blache, 2000, 2006; Blache and Azulay, 2002; Blache et al., 2003; Blache and Prost, 2008; Guénot and Blache, 2005).
- **Agreement** ($\leftrightarrow$). This property represents that two constituents must have the same agreement (Blache, 2004; Guénot and Blache, 2005). Nowadays, PG can express agreement by means of the dependency property, specifying in the features that the agreement is as well dependent: $DET_{[agr]_i} \rightsquigarrow_{spec} NOUN$.

5.1.2 Properties and Lexical Entries

PGs aim to represent most of the information by means of properties. Consequently, there is significantly less information for a lexical entry than with other approaches such as LFG and HPSG. The triggered properties of a structure depend on its categories. Therefore, the lexical information in PGs is mostly based on setting up the categories for each word. Besides, the lexical information might also contain information such as agreement, morpho-syntactic features, sub-categorization lists and grammar functions if they are need.

Lexical entries can also be used to refine linguistic information from a multi-modal point of view, so that it includes features from other domains such as pragmatics, semantics, phonology, and so on.

On the other hand, there is not just one system for defining a set of properties for a given category. PGs use different tools to define a PG: description by items; formalisms from HPSG or CxG, among others. In subsection 6.5.5, we use our own system for defining properties.

5.1.3 Properties Supported by Features

The properties can be supported with features that specify when those properties are going to be applied to a category. The typical feature to be represented is a function: $X_{[subj]}$ understood as X as a subject.

For example, a property for an English grammar such as $N \prec V$ might be inaccurate since the noun can both precede and be preceded by a verb. Thanks to the features we can specify functions and other values for a category in order to provide proper linguistic information. Therefore, PGs can specify that a noun *as a subject* precedes a verb: $N_{[subj]} \prec V$. Likewise, they can indicate that a verb precedes a noun when this noun has a function as a *direct object*: $V \prec \mathrm{N}_{[dobj]}$. Besides, a property can define a relation of two words and their features with an uncertain word category and a certain word category: i.e. $V_{[ditrans]} \prec \mathrm{X}_{[iobj]}$ (either a nominal or prepositional construction).

Features can also represent other linguistic information, such as agreement: $DET_{[agr]_i} \rightsquigarrow_{spec} NOUN$: the determiner, with an agreement feature$_i$, is dependent on the NOUN and its agreement feature $_i$, which has to be the same ($_i$).

Features reinforce properties as a tool that can describe linguistic information independently of a context, and more precisely represent grammatical knowledge by taking into account linguistic variation.

5.1.4 Can Properties be Triggered by Other Elements?

Properties are defined as *a constraint that represents the relation between two words*. This creates a problem: *what is a word?* I have been unable to find in the literature on PGs a definition of what a word is for this theory. I assume that PGs are not a framework that constrains the linguist's task; hence, it is the linguist who has to decide the most convenient definition of *word* so that the properties can be applied. Otherwise, properties are triggered by categories rather than words. This can be helpful in English because sometimes two words can easily be considered as one category, i.e *"dining-table"* is one word with one category rather than two words with one category or two words with two categories. However, in Spanish, this is more difficult because words are composed without *dash* (-), setting out a problem to decide if a pair are actually two words or one word composed by two elements: because of their vague characteristics: *"padre nuestro"* (*our fathers' prayer*); *"guardia civil"* (military-policeman); *"pavo real"* (*peacock*), and others. We can even find extreme cases with examples of full constructions that can be construed as one word i.e. *"corre ve y dile"* (literally, *run-go-and-tell-him/her*; means *gossiper*); which sometimes appears also as *"correveidile"* or *"corre-ve-y-dile"*. Linguistic periphrasis poses a problem for the notion of word, too. Some closed phrases behave as a single category, known in Spanish grammar as *locuciones*. For example, verb phrases such as *"morirse de risa"* (*laugh out loud*) and prepositional phrases such as *"a falta de"* (*without*), among others. These examples show that it can be difficult to decide if they are syntactically one word, two words, one category, two categories, or more. Therefore, it is up to the linguist to decide whether this kind of example is a single lexical entry in the lexicon, which triggers one category, or, in contrast, are two (or more) separate categories that trigger individual properties separately from one other.

Bearing in mind that the notion of word is vague, I think that it might be useful in some cases to expand the notion of property to *a constraint that represents the relation between two linguistic objects*. So, a property would also define relations between constructions, between constructions and categories, categories and features, features and meanings and so on.

5.1.5 Properties and Derivational Tools

The idiosyncrasy of the properties shown when linguistic relations are defined and inputs described has a practical advantage over the generative model of transformational grammars (Chomsky, 1995). The generative model needs to complete all the operations to generate a syntactic tree. The transformational and minimalist models focus on describing the hierarchical relationship between constituents because of their derivational process. If we want to check a linguistic relation, we must start from the lowest part of the tree, and travel through the nodes until we finally have a complete picture. PGs, however, can define and identify a linguistic relationship immediately, with no need for either a derivation process or a hierarchical setting. This works very well in such languages as English, in which the hierarchical structure of constituents in a sentence is respected much more frequently than in Romance languages. The reason behind this is that English does not have such clearly defined morphology as Spanish. The more morphological information and case-marking a language has, the freer is its distribution in an utterance. This does not pose a problem for a grammar when it deals with inputs that respect the expected hierarchy. However, it poses a problem for managing structures which display elements with an apparently free syntactic distribution, even though they are perceived as grammatical. Therefore, in a language such as Spanish, the hierarchy of constituents might not be enough to explain syntactic relations in spontaneous language. Spanish would benefit more from a framework such as PG that defines more syntactic properties to deal with syntactic idiosyncrasy. In addition, the grammars that base syntactic description only on constituents, whether they use phrase-structure rules or constraints, do not manage to define dependency relations. However, this problem is usually solved by combining *constituency-based grammars* with *dependency grammars* (Müller, 2016). In this way, Property Grammars have another advantage: they can describe constituents and dependencies at the same time. Their extensive and flexible list of properties is subject to the mechanisms and needs of a linguistic system and, therefore, facilitates the description of language knowledge.

5.2 Notion of Construction in Property Grammars

PGs adopt the notion of construction from Construction Grammars (CxG). The concept of a construction grammar was proposed by Fillmore (1988) as a reaction to the phrase-structure approaches and in favour of language variability.

Constructions are understood as *pairs of form and function*. As mentioned above in chapter 5, this can be seen in Kay and Fillmore (1999) with their example: *"What is this scratch doing on the table?"*. The meaning of the verb *to do* is completely conditioned by the structure. In this respect, the verb is linguistically marked by the structure since we cannot find the verb with this exact meaning in another structure.

Goldberg (2009) states that Construction Grammars "emphasize that speakers' knowledge of language consist of systemic collections of form-function pairings that are learned on the basis of the language they hear around them".

Notions from Construction Grammars together with PGs can represent linguistic variability and the fuzzy properties of a construction very well. In this regard, these approaches show how some words can function in some constructions with a non-canonical fit (for example, a noun functioning as an adjective): "Kaschak and Glenberg (2000) have demonstrated experimentally that subjects rely on constructional meaning when they encounter nouns used as verbs in novel ways" (Goldberg, 2009). Therefore, constructions dictate the function of elements according to the form in which they appear. The constructional approach uses linguistic frequencies to objectively justify the belonging of the linguistic constraints towards a grammar. However, the frequency of a construction is not always enough for constructions to be accepted as a part of the grammar; theoretical reasons are also required with such concepts as *markedness* or *context effects*. Some of the constructions in language are displayed in Figure 5.3.

In my opinion, the conception of *construction* is not opposed to *phrase*. Both notions complement each other since *a phrase is a type of construction*. Nevertheless, they differ from each other because of the hierarchical notion of *head*. For *phrases*, the notion of head is mandatory, while for constructions it is not. PGs (and our work) are based on the notion of construction since it is the most flexible.

Word	e.g., *tentacle, gangster, the*
Word (partially filled)	e.g., *post-N, V-ing*
Complex word	e.g., *textbook, drive-in*
Idiom (filled)	e.g., *like a bat out of hell*
Idiom (partially filled)	e.g., *believe* ⟨one's⟩ *ears/eyes*
Covariational	The Xer the Yer
Conditional	(e.g., *The more you watch the less you know*)
Ditransitive	Subj V Obj1 Obj2
	(e.g., *She gave him a kiss; He fixed her some fish tacos.*)
Passive	Subj aux VPpp (PP_{by})
	(e.g., *The cell phone tower was struck by lightening.*)

Fig. 5.3 Examples of constructions from Goldberg (2009:94).

According to PGs, a construction is defined as the result of the convergence of many properties, which allows the pairing of form and function (Guénot and Blache, 2005). PGs can identify a construction by noticing its properties. Therefore, *a construction is a set of categories that are related by a set of properties* (Blache, 2016:209). This definition allows us to maintain the notion of construction even in constructions with violated properties, i.e.: *"house red lot a like I"*. Even though the low degree of grammaticality of this construction, we can see that different properties are related to each other: we see that the noun *"house"* precedes the adjective *"red"*; the pronoun *"I"* is in agreement with the verb *"like"*, and so on. We can recognize these relationships because the properties are directly identifiable, and they are activated automatically once the categories begin to establish relations with each other and higher level of information is not required. Likewise, PGs identify the constituents of a construction of a language when a set of properties is recurrently repeated given a set of categories. This may explain why Spanish speakers recursively identify a subject construction with a nominal construction and the properties of a specific nominal construction with the following characteristics (∧ states for AND, while ∨ states for OR):

Construction	Properties
Subject Construction	$\text{Nominal Construction}_{[subj]} \Rightarrow \text{Verb}_{[ditrans]} \vee {}_{[trans]} \vee {}_{[intrans]}$
Nominal Construction	$\text{PRON} \otimes \text{PROPN} \wedge \text{NOUN}$ $\text{PROPN} \otimes \text{PROPN} \wedge \text{NOUN}$ $\text{NOUN} \otimes \text{PROPN} \wedge \text{PRON}$
$\text{Pron}_{[subject]}$	$\text{Pron} \otimes \text{ADP} \wedge \text{ADJ} \wedge \text{ADV} \wedge \text{DET} \wedge \text{PRON}$ $\text{Pron}_{[agr_i]} \rightsquigarrow_{subj} \text{Verb}_{[agr_i]}$

Table 5.1 Constructions and its properties.

PGs can characterize different constructions with a PG graph. In our grammar, we do not provide any property graph. We have focused on building the grammar such as shown in Table 5.1. However, an example of a property graph is given in order to illustrate the formalism in Figure 5.4.

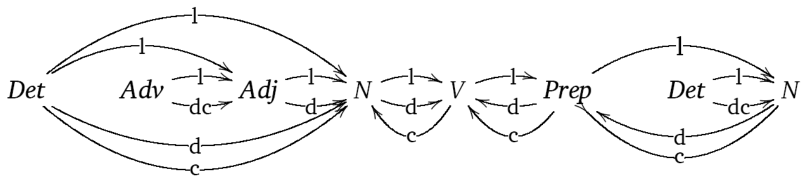


Fig. 5.4 Example of a property graph from Blache (2016:212).

This graph is complete because all the nodes are connected and it has sub-graphs. Sub-graphs are those constituents that are connected with each other. As an example, $\{Det, Adj, N\}$ is a complete sub-graph while $\{Det, Adv, Adj\}$ is not, because all its elements are not fully linked. However, PG constructions are flexible because they allow overlapping within types of construction. The theory considers the following constructions to be of major interest, labeled as *maximal complete sub-graphs* (Blache, 2016:212):

- $Adv - Adj$: Adjectival construction.
- $Det - Adj - N$: Nominal construction.
- $N - V$: Subject/verb construction.
- $V - Prep$: Verb/indirect object construction.
- $Prep - N$: Prepositional construction.
- $Det - N$: Nominal construction.

PGs used the notion of construction and the other notions discussed above from construction grammars to develop MarsaGram (Blache et al., 2016). This tool automatically extracts the constructions and constituents of a language using algorithms based on frequency values. It is extremely useful for extracting the properties of a language considering constructions, dependencies and frequencies.

5.3 Heads in Property Grammars

In older versions of PGs, the concept of *head* can be found as the property of *obligation* in Balfourier et al. (2002); Blache (2000, 2006); Blache and Azulay (2002); Blache et al.

(2003); Blache and Prost (2008); Guénot and Blache (2005). Blache (2016) discusses why PG should not include the notion of *head* in its framework.

The notion of *head* assumes that syntactic information is determined hierarchically. Consequently, priority is given to identifying this property over the others. This clashes with the core idea of PGs that *a property must be identified immediately*. Additionally, we may encounter formal problems by assuming the notion of head such as in *constituency-based grammars*, *dependency grammars* and lexicalized approaches similar to HPSG.

The first problem these approaches have to solve is to determine what the head of a structure is. The typical example for this problem is determining whether a specifier is the head of a phrase rather than the traditional core element such as a noun, a verb and so on. If the head is changed the whole structure will change, so the linguistic analysis will be different. Abney (1987) proposed that the Noun Phrase was in fact a Determiner Phrase (DP) in the X-bar theory, since the determiner is the head of that phrase. The DP hypothesis was accepted by Chomsky (1995) and Stowell (1989). These arguments are still being discussed by linguists such as Jackenddoff (2002) and Hudson (2004:39), who claims that "D does depend on N".

The second problem of these *head dependent* theories is the serious difficulties for parsing constructions without heads. Blache (2016:203) claims that this is an important limitation for describing utterances, since we can easily find constructions without heads in natural language: 1) elision of an NP: *"John sets the red cube down and takes the black"*; 2) constructions without verbal head: *"First trip, New York"*.

PGs avoid such discussions since they do not accept the head as a mandatory hierarchical notion or the formal problem of determining a *head*. PGs provide alternative tools for dealing with the strict notion of *head*. These tools are the property of *requirement* ($x \Rightarrow y$) and the property of *dependency* ($x \rightsquigarrow_{dep_i} y$).

This first property assumes that two elements must co-occur, which means that both elements require each other without any hierarchical relation, i.e. NOUN $\Rightarrow$ DET. This property, together with others, such as *exclusion* or *precedence*, enable PGs to propose a *decentralized* approach. Therefore, there is no need to express with heads what can be (better) expressed with co-occurrence.

Likewise, rather controversially, the property of *dependency* means that one *head* element is dependent on the other *non-head*. However, this notion of *head* regarding the property of dependency is not fully equivalent to its homonym constraint in *dependency grammars*. In PGs, this property merely focuses on *defining a relation between two words*, which respects the core idea of Property Grammars. So, success in defining an input with a PG does not depend at all on whether dependency properties and their head are satisfied or violated.

Therefore, the violation of this property is not an impediment for defining a construction because the notion of head is relaxed to its mildest expression and affects just one linguistic element rather than a whole construction. In contrast, dependency approaches are fully head-based since a head is parsed before its dependent elements, and is hierarchically placed before the others.

<table>
<tr><td>level 1</td><td colspan="6">NOUN-root 1:1</td></tr>
<tr><td>level 2</td><td>DET-spec</td><td>head*</td><td colspan="4">VERB-mod:relcl 1:2</td></tr>
<tr><td>level 3</td><td>the</td><td>reporter</td><td>PRON-dobj</td><td colspan="2">NOUN-subj 1:3</td><td>head*</td></tr>
<tr><td>level 4</td><td colspan="2"></td><td>who</td><td>DET-spec</td><td>head*</td><td>attacked</td></tr>
<tr><td>level 5</td><td colspan="3"></td><td>the</td><td>senator</td><td></td></tr>
</table>

Fig. 5.5 Example of hierarchical dependencies in Universal Dependencies.

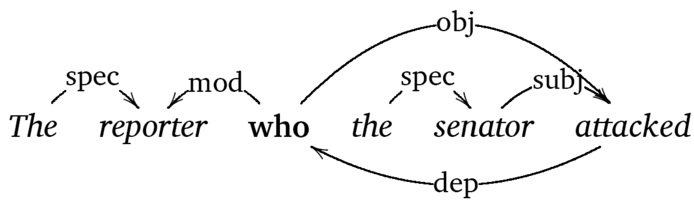


Fig. 5.6 Example of dependencies in PG from Blache (2016:197).

The comparison of Figure 5.5 with Figure 5.6 shows how different the relation between heads and dependencies are in PGs and in dependency grammars. PGs display all their properties on one level, while a head-based dependency approach needs five levels to explain all the syntactic information. Thus, heads rank the importance of each element in their hierarchy on the syntax of a construction. The numbers on each phrase tag the headed constructions, and disregard all other syntactic elements. The relative pronoun in the example is not considered as a head because does not dominate any other element.

In contrast, PGs make it possible to specify more dependencies between elements thanks to their non-hierarchical approach. They can represent how the verb in the relative construction also depends on the relative pronoun, while the relative pronoun depends on the verb as a direct object. As a result, PGs display more linguistic information between words than the dependency-head approach.

Consequently, dependencies in PG are parsed linearly rather than hierarchically and derivationally, thus avoiding the notion that *head* is mandatory for the syntactic description. PGs displace the notion of *head* as a capital element over the rest of the syntactic information; taking into account the dependency property as relevant as the others.

5.4 Parsing with Property Grammars

PGs propose a parsing method in Blache (2016:213-215). In this section, we provide a summary of the phases of the parsing operations proposed by PGs disregarding the formal approach. This operation consists of three successive phases:

- **Categorization**: In this first phase, all the categories that can be associated with each of the lexical items contained in the entry are activated and assigned a position. A set of categories is thus obtained.
- **Assignment**: In the second phase, all possible constructions that can describe the input are identified, which leads to a set of constructions.
- **Characterization**: Finally, in this phase, the properties of each construction are verified. A list of properties violated (P^-) and satisfied (P^+) is thus obtained.

PGs use to offer the **characterization** of a construction in a similar approach as in Figure 5.7:

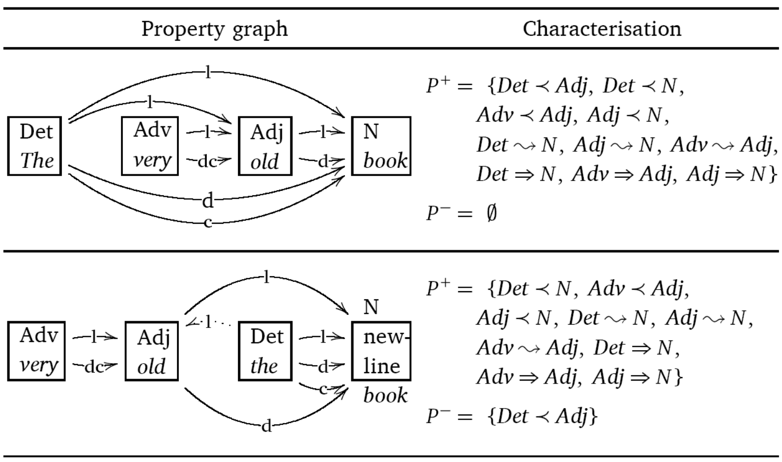


Fig. 5.7 Example of a characterization in PG from Blache (2016:215).

Since properties are represented independently, they are easily calculable. In this formalism, linguistic information is presented autonomously, which allows the system to describe any type of input. Therefore, regardless of the degree of grammaticality that an input presents, the system will provide a characterization with a list of satisfied and violated properties. Unlike other constraint grammars, PGs provide a lot of information about violations, as they tell us what kind of violation has occurred, where it happened, and why it is not grammatical.

Once again, this model can be used to analyze any structure regardless of its degree of grammaticality. These characteristics make these grammars a useful formalism for defining natural language productions, since they offer tools to the grammar for evaluating the constraints of an input and its degree of grammaticality.

5.5 Why Property Grammars for a Fuzzy Model?

PG is the perfect theory for a fuzzy grammar that takes into account degrees of grammaticality.

Firstly, PGs are represented on an axis that includes three elements: constraints with immediate descriptive capacity, the specific notion of construction, and the disconnection of linguistic elements from a hierarchical and derivational point of view. This makes it easier to identify the relationships between words and local language phenomena so that natural language processing can be described. These characteristics mean that PGs can describe any input and are ideal for responding to the challenge of representing degrees of grammaticality. This model seems to be able to provide a satisfactory explanation for the different degrees of grammaticality in Spanish syntax. Additionally, PGs provide MarsaGram with support for extracting a natural language grammar. This tool helps linguists to manage useful linguistic information such as constituencies, dependencies and properties.

Secondly, PG has considerable potential from a multi-modal point of view. Certainly, this theory has been mostly applied in the syntactic domain. However, Blache and Prévot (2010) and Blache et al. (2008) have already explored the possibility of a multi-modal annotation. This fits our multi-modal, fuzzy grammar approach perfectly, and paves the way for future work on degrees of grammaticality from a multi-modal perspective.

Part II

Fuzzy Property Grammars

Chapter 6

A Fuzzy Grammar for Degrees of Grammaticality

In this chapter, we show the basis of our fuzzy grammar. Firstly, section 6.1 provides a summary of what is already seen in Part I as a justification for this fuzzy grammar. Secondly, section 6.2 reveals the application of fuzzy logic tools, Fuzzy Natural Logic (Novák, 2015), for defining a grammar which takes into account degrees of grammaticality regarding linguistic competence. Finally, section 6.5 presents a system for evaluating grammaticality and its bases.

6.1 A Justification for a Fuzzy Grammar

In Part I, we highlight the need to propose formal models that consider grammaticality as a gradient property instead of the categorical view of grammaticality defended in theoretical linguistics. Since grammatical deviations are inherent to the spontaneous use of language, and speakers can decode them, linguistic analysis tools should account for different degrees of grammaticality.

6.1.1 Why a Fuzzy Grammar

In chapter 2, we stressed the nature of natural language. This work proposes a grammar which captures the idiosyncrasy of natural language. When people use natural language in natural conversation, it is prevalent to hesitate over what they are going to say, to abandon the discursive thread or to repeat words and phrases. Natural language is thus described as spontaneous, immediate and ambiguous, and it is often produced with grammar violations. These features mean a problem for traditional language analysis systems that, in general, reject sentences with errors. A grammar has traditionally been defined in linguistics from a categorical point of view (that is to say, an input is either grammatical or ungrammatical). However, throughout the history of linguistics, many authors have considered the possibility that a grammar accounts for non-canonical or non-grammatical productions, yielding to a gradual or "fuzzy" conception of language.

We have discussed in section 2.1 the reasons behind the Chomsky's distinction between competence and performance which has determined theoretical linguistics. Besides, in section 2.2, we have argued that the competence and performance distinction shows how the notion of competence is centered on the syntactic domain, disregarding rules from other domains for representing grammaticality. Because of this situation, we take into account a multi-modal perspective of what is a grammar, proposing the following definitions in section 2.4:

- Grammaticality is the value that represents how much satisfied is an input according to the linguistic knowledge (constraints) that defines the competence of a natural language grammar. This value is fuzzy since it takes into account numerous criteria that makes it vague. The vagueness of the fuzzy value of grammaticality is going to be determined by a value in terms of degrees giving a certain amount of satisfied or violated criteria based on a grammar with constraints.

- This definition might satisfy many linguistic approaches that distinguish between grammaticality and acceptability. We distinguish them as well but taking into account that every object which could be explained by rules or linguistic constraints is part of the linguistic competence.

6.1.2 Gradience and Grammaticality in Linguistics

A grammar must accept gradient phenomena, such as gradient grammaticality, according to the reality of natural language. In this regard, gradience is a well-known linguistic term that designates this conception. Aarts (2004b) defines gradience as a term to designate the spectrum of continuous phenomena in language, from categories at the level of grammar to sounds at the level of phonetics. However, we have shown, in chapter 3, that the most studied gradual phenomenon in the processing of natural language is given by gradient *acceptability* (not *grammaticality*); examples of this are: Aarts (2004a,b); Aarts et al. (2004), Chomsky (1965, 1975), Bolinger (1961), Ross (1972), Prince and Smolensky (1993), Keller (2000, 2006), Manning (2003), Legendre et al. (1990a).

Some linguists have endeavored to represent the degrees of acceptability with formal systems that fit the studies on linguistic gradience. Examples of these approaches are:

Standard Optimality Theory (OT) (Prince and Smolensky, 1993), Linear Optimality Theory (LOT) (Keller, 2006), Probability Theory (Manning, 2003), Harmonic Grammars (Legendre et al., 1990a). Nevertheless, most of these approaches take into account grammaticality by considering degrees of "ungrammaticality". Making use of the concept of "ungrammaticality" arises some problems for representing grammaticality because:

- Degrees of ungrammaticality imply a discrete approach for grammaticality: one linguistic expression is *grammatical* if it is optimal and if it satisfies all the constraints; whereas one linguistic expression is *ungrammatical* in terms of degree depending on its violated constraints. Therefore, grammaticality is discrete while *ungrammaticality* is gradient.
- Most of these approaches illustrate grammaticality as acceptability, using the notion of *grammaticality judgments*. This idea is argued in section 3.8 in which it is shown how these theories are representing acceptability rather than grammaticality. The main reason behind this claim is that a judgment always implies a subject such as a speaker. Therefore, a speaker is never able to judge grammaticality without being naturally conditioned by extra-linguistic elements. Consequently, in those approaches, a *grammaticality* judgment is always an *acceptability* judgment.

In the end, there is still not a framework which can represent degrees of grammaticality; partially, because there is not a grammar which can formally represent the degrees of grammaticality regarding just linguistic competence. In other words, there is not a formal framework yet which represents grammaticality strictly from a grammar perspective, without involving a speaker and extra-linguistic facts, as shown in Table 6.1. Because of that, we propose our model with fuzzy grammar for representing the degrees of grammaticality regarding linguistic competence.

Linguistic Value	Evaluation of	Object of Evaluation	Nature of the Object
Degrees of Grammaticality	**Competence**	**Constraints** (Knowledge of a Language)	Theoretical
Degrees of Acceptability	Performance	Grammatical objects -Grammaticality Extra-linguistic objects -Attitude -Psycholinguistic variables	Theoretical & Empirical

Table 6.1 Degrees of grammaticality and acceptability.

6.2 A Fuzzy Grammar with Fuzzy Logic

This section shows the fuzzy logic tools which have been used for defining a grammar. Firstly, the used formalism is presented. We present it here, rather than before, in order to provide handy checks to the reader concerning our fuzzy grammar. Secondly, we display our grammar and its mechanism.

6.2.1 Fuzzy Type Theory and Fuzzy Natural Logic

The basis of the Fuzzy Type Theory (FTT) and Fuzzy Natural Logic (FNL) in Novák (2015), which are the formalism used for defining our grammar, are presented here.

Fuzzy Type Theory (FTT), a higher-order fuzzy logic, was introduced by Novák (2005). Novák further introduced the program of Fuzzy Natural Logic (FNL) in Novák (2015) as the program for the development of a mathematical model of human reasoning that is based on the use of natural language. Novák (2016) points out the main expected contributions by this theory:

- Development of methods for construction of models of systems and processes on the basis of expert knowledge expressed in genuine natural language.
- Development of algorithms making computer to "understand" natural language and behave accordingly.
- Help to understand the principles of human thinking.

As a novelty, this thesis has applied FNL to a new contribution: *its application for representing degrees of grammaticality regarding linguistic competence*.

The formal background of Fuzzy Natural Logic is FTT (Novák, 2015). Taking into account that for applications in linguistics the most convenient is FTT with a Łukasiewicz algebra of truth values, we will in the sequel refer to it as Ł-FTT.

Let us summarize the basic concepts of FTT and FNL. For more details, we refer the reader to the above cited literature.

(a) The algebra of truth values is the standard Łukasiewicz MV_Δ-algebra

$$\mathscr{L} = \langle [0,1], \vee, \wedge, \otimes, \rightarrow, 0, 1, \Delta \rangle \tag{6.1}$$

where

$$\begin{aligned} \wedge &= \text{minimum}, & \vee &= \text{maximum}, \\ a \otimes b &= \max(0, a+b-1), & a \rightarrow b &= \min(1, 1-a+b), \\ \neg a &= a \rightarrow 0 = 1-a, & \Delta(a) &= \begin{cases} 1 & \text{if } a = 1, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

(b) The basic concept in FTT is that of a *type*. This is a special subscript (denoted by Greek letters) assigned to all formulas using which we distinguish kinds of objects represented by formulas. The atomic types are: ε representing elements and o representing truth values. In the semantics of FNL, it is the type ε assigned a set M_ε whose elements can be anything: people, objects, languages, etc.

(c) The type o (omicron) is the type of truth degree. In the semantics, it is assigned a set of truth values M_o which, in our case, is $M_o = [0,1]$.*) The degree of truth $a \in [0,1]$ may represent various degrees, for example the degree of grammaticality, complexity, etc.

(d) From basic types, we form complex ones $\beta\alpha$ where α, β are already formed types. For example, $o\varepsilon$, $\varepsilon\varepsilon$, $(o\varepsilon)\varepsilon$, $o\alpha$, etc. In the semantics, the complex types $\beta\alpha$ represent *functions*. Thus, each type $\beta\alpha$ is in the semantics assigned as set $M_{\beta\alpha}$ which is a set of functions $M_\alpha \longrightarrow M_\beta$.

(e) Formulas are formed of variables, constants (each of specific type), and the symbol λ. They are denoted by capital letters and assigned a type, i.e., A_α is a formula of type α. In the semantics, A_α is interpreted by some element from the set M_α.

(f) The formula $\equiv$ is the basic connective of *fuzzy equality*. In the semantics, for example, the formula $A_\alpha \equiv B_\alpha$ represents a truth degree of the (fuzzy) equality between the element interpreting A_α and the element interpreting B_α. More concretely, let $\mathscr{M}$ be a semantic interpretation of formulas. Then $\mathscr{M}(A_\alpha) \in M_\alpha$ is an element from the set M_α and similarly, $\mathscr{M}(B_\alpha) \in M_\alpha$ is another element from the same set M_α. Then

*) Note that the use of [ ] means any real number/degree between 0 and 1. That could be, e.g., 0.85512 and so on. Note that in classical logic we consider only two truth values, i.e., the set of truth values is $\{0,1\}$ which means that we consider either 0 (false) or 1 (true).

interpretation $\mathscr{M}(A_\alpha \equiv B_\alpha) \in [0,1]$ is a truth value of the equality $A_\alpha \equiv B_\alpha$ in the interpretation $\mathscr{M}$.

(g) Semantics of Ł-FTT is defined in a model (or frame), which is the system $\mathscr{M}$= <(M$_\alpha$, $\equiv_\alpha)_{\alpha \varepsilon Types}$ > where M$_\alpha$ is the set of elements of type $_\alpha$ and $\equiv_\alpha$ is a fuzzy equality on the corresponding set M$_\alpha$. In other words, explanation of the model consists of couples of sets (fuzzy sets) for all equality. For all infinite sets (M$_\alpha$) and fuzzy equality ($\equiv$) exists a type which is connected by the standard Łukasiewicz MV$_\Delta$-algebra. With respect to (a) - (f), M$_o$=[0,1], M$_\varepsilon$ is a set due to $\mathscr{M}$, M$_{\beta\alpha}$ is a set of functions due (d) and $\equiv_\alpha$ is interpretation of connective $\equiv$ due to f. Fuzzy equality $\equiv$ on a set M is a fuzzy relation $\equiv$: M x M $\rightarrow$ [0,1].

(h) A *fuzzy set* is a function $B : M \longrightarrow [0,1]$ where M is a set having the role of a universe. The function B is often called a *membership function*, i.e., a fuzzy set is identified with its membership function. From the point of view of Ł-FTT, a fuzzy set is obtained as an interpretation of a formula $A_{o\alpha}$ of type α. The universe of such a fuzzy set is then the set M_α.

(i) There are several logical connectives in Ł-FTT, namely $\vee$ (disjunction) that is interpreted in the Łukasiewicz algebra by the operation $\vee$ (maximum); $\wedge$ (conjunction) interpreted by $\wedge$ (minimum); & (strong conjunction) interpreted by the operation $\otimes$; $\Rightarrow$ (implication) interpreted by the operation $\rightarrow$; and the special unary connective Δ interpreted by the operation Δ. We introduce also $\neg$ (negation) interpreted by the operation $1 - a$ (cf. item (a)). Besides the logical connectives, also the quantifiers $\forall$ (general quantifier) interpreted by the operation of *infimum* and $\exists$ (existential quantifier) interpreted by the operation of *supremum* are introduced.

(j) The formula $\lambda x_\alpha \cdot B_\beta$ has the type $\beta\alpha$ and it is interpreted by a function $M_\alpha \longrightarrow M_\beta$. It says that "each element x_α of type α is assigned an element of type β after we substitute the former in the (interpretation of) the formula B_β".

(k) The fuzzy type theory has 17 logical axioms and 2 inference rules.

Fuzzy natural logic (FNL) is a mathematical theory that provides models of terms and rules that come with natural language and allow us to reason and argue in it. At the same time, the theory copes with the vagueness of natural language semantics. So far, it is a set of the following formal theories of Ł-FTT:

- A formal theory of evaluative linguistic expressions (Novák, 2007, 2008a).
- A formal theory of fuzzy IF-THEN rules and approximate reasoning (derivation of a conclusion) (Novák and Lehmke, 2006).
- A formal theory of intermediate and generalized quantifiers (Murinová and Novák, 2016; Novák, 2008b).

6.3 A Fuzzy Grammar Formalism for Grammaticality

As seen in section 4.5, the parallelisms between fuzzy sets and grammaticality are clear.

The basic idea in this dissertation is representing grammaticality as a fuzzy set. The initial idea was considering a grammar as a fuzzy set G which assigns a degree of truth L to each constraint in a linguistic input I; this degree of truth is assigned according to the importance of each constraint and according to the fact that if they have been satisfied or violated.

$$G\colon I \longrightarrow L \tag{6.2}$$

In subsection 4.3.2, it has been shown how vagueness is found when we want to group in X certain objects (x, y, z...) which have a certain property φ:

$$X = \{x \mid x \;\; has\, the\;\, property \;\; \varphi\} \tag{6.3}$$

The same has been done with our grammar. We group objects (i.e., x_γ) in a linguistic module (i.e., X_γ) if they have a certain property which is being grammatical, in other words, an object such as a syntactic constraint has the property of being grammatical when it is in the grammar.

As exposed in subsection 4.3.2, two types of objects can be found when dealing with vagueness: (1) *borderline objects* and (2) *typical* or *prototypical objects*. The same happens with our grammar which captures vagueness. The fuzzy grammar deals with:

- *Special/variability constraints*;
- and *prototypical constraints* (gold standard/canonical constraints).

The first type of object shows up problems for being grouped in a linguistic module since it is unclear if they have the property φ. The second refers to those objects which clearly presents the property φ . In such manner, the fuzzy grammar deals with vagueness taking into account the canonical constraints and comparing them to other candidates.

Types of objects	
In Fuzzy Logic	*In a Fuzzy Grammar*
Borderline objects	Variability Constraints
Typical/Prototypical objects	Canonical Constraints (Gold Standard)

Table 6.2 Types of objects between Fuzzy Logic and Fuzzy Grammar.

Consequently, a Fuzzy Grammar takes into account different types of rules/constraints: satisfied rules, violated rules, and variability (special) rules:

- **A satisfied rule** is a constraint of a grammar which is found reproduced in a linguistic input. For example, a grammar of Spanish dictates that the subject must precede the verb: $X_{[subj]} \prec V$. Therefore, an input such as *"La chica juega a fútbol"* satisfies these constraints since the subject *"la chica"* is preceding the verb *"juega"*.
- **A violated rule** is a constraint of a grammar which is found infringed in a linguistic input. For example, a grammar of Spanish dictates that the subject must precede the verb: $X_{[subj]} \prec V$. Therefore, an input such as *"juega a fútbol la chica"* violates this constraint since the subject *"la chica"* **is not** preceding the verb *"juega"*.
- **A variability rule** is a constraint which is triggered when a violation occurs; compensating the final value of grammaticality. For example, a grammar of Spanish dictates that an intransitive verb such as *"jugar"* requires a subject: $V \Rightarrow X_{[subj]}$. Therefore, an input such as *"juega a fútbol"* violates this constraint since there is no subject in the input for satisfying the requirement of the verb *"juega"*. However, the grammar of Spanish has a special rule which says that: in case a subject cannot find a verb, the subject is found on the morpheme of the verb, which takes into account the person and the number: *"jueg-a"*, the subject is found in *"-a"*. Once this special rule is satisfied, the final value of grammaticality would be better if we consider both violated and special rules than if we would calculate the grammaticality of the input concerning only the violated rule.

  In such manner, vague linguistic phenomena are captured since a grammar can describe inputs with borderline cases through special rules, rather than representing inputs which are either fully satisfied or violated.

This first idea of defining a grammar as a fuzzy set had some problems:

- What is a grammar?
- Where are the degrees of grammaticality exactly placed?
- How the interactions between satisfied, violated and special constraints work?
- How a fuzzy grammar approach is applied on a linguistic input?

In order to solve these questions, a more precise formula of what is a fuzzy grammar is proposed in this work. Setting up the formal basis for a grammar with constraints which can represent degrees of grammaticality.

6.4 The Fuzzy Grammar

A fuzzy grammar (FGr) is a fuzzy set ($\underset{\sim}{\subseteq}$) on the whole set of linguistic rules or constraints. These constraints define the linguistic knowledge of a speaker in every module (linguistic domain) such as the syntactic module, semantic module, pragmatic module and so on. We define a fuzzy grammar in a multi-modal sense. The multi-modal perspective will provide multiple advantages in the future such as providing a multi-modal value for both degrees of grammaticality and acceptability, implementing multi-agent systems for our grammar, applying auto-lexical approaches, and so on.

Definition 1. *A Fuzzy Grammar (FGr) is a fuzzy set on the Cartesian product of:*

$$FGr \underset{\sim}{\subseteq} Ph_\alpha \times Mr_\beta \times X_\gamma \times S_\delta \times L_\varepsilon \times Pr_\zeta \times Ps_\kappa \tag{6.4}$$

where:

- $Ph_\alpha = \{ph_\alpha \,|\, ph_\alpha$ *is a phonological constraint*$\}$ *is the set of the phonological constraints.*
- $Mr_\beta = \{mr_\beta \,|\, mr_\beta$ *is a morphological constraint*$\}$ *is the set of the morphological constraints.*
- $X_\gamma = \{x_\gamma \,|\, x_\gamma$ *is a syntactic constraint*$\}$ *is the set of syntactic constraints.*
- $S_\delta = \{s_\delta \,|\, s_\alpha$ *is a semantic constraint*$\}$ *is the set of semantic constraints.*
- $L_\varepsilon = \{l_\varepsilon \,|\, l_\varepsilon$ *is a lexical constraint*$\}$ *is the set of lexical constraints.*

- $Pr_\zeta = \{pr_\zeta \,|\, pr_\zeta$ *is a pragmatic constraint*$\}$ *is the set of pragmatic constraints.*
- $Ps_\kappa = \{ps_\kappa \,|\, ps_\kappa$ *is a prosodic constraint*$\}$ *is the set of prosodic constraints.*

We calculate the value of grammaticality regarding a fuzzy grammar by aggregating membership degrees for all constraints of the grammar. Nevertheless, the grammar does not violate or trigger anything by itself. It needs another element which triggers the constraints on a grammar. In order to do that, the constraints that define the knowledge of a grammar have to be contrasted with another set of constraints of an *input* that we call *dialect*.

The dialect is considered here also as a set of constraints of an *input* (d_η), that is all the constraints that are in a dialect's or language's *output*.

Definition 2. *The set of constraints in a dialect is* $D_\eta = \{d_\eta \,|\, d_\eta$ *is a dialect constraint*$\}$.

In this regard, **every dialect** is considered as a **language**. The dialect might display constraints which are satisfied in a fuzzy grammar but violated regarding another fuzzy grammar. Consequently, the degree of grammaticality is determined by the fuzzy grammar that evaluates a dialects' input.

Example 1. *A grammar of "Cambridge English" would easily perceive as violation constraints from "Irish English" inputs regarding the phonological module. A grammar of "Castilian Spanish" would probably perceive as violations constraints found in a "Mexican Spanish" inputs. Thanks to this notion of a dialect set, we can provide a degree of grammaticality regarding a specific natural language grammar.*

Definition 3. *A Fuzzy Grammar which evaluates a dialect input in terms of degrees is:*

$$FGr \equiv \lambda d_\eta \lambda ph_\alpha \lambda mr_\beta \lambda x_\gamma \lambda s_\delta \lambda l_\varepsilon \lambda pr_\zeta \lambda ps_\kappa \cdot (Ph_{(o\eta)\alpha} ph_\alpha)d_\eta \wedge (Mr_{(o\eta)\beta} mr_\beta)d_\eta \wedge (X_{(o\eta)\gamma} x_\gamma)d_\eta \wedge (S_{(o\eta)\delta} s_\delta)d_\eta \wedge (L_{(o\eta)\varepsilon} l_\varepsilon)d_\eta \wedge (Pr_{(o\eta)\zeta} pr_\zeta)d_\eta \wedge (Ps_{(o\eta)\kappa} ps_\kappa)d_\eta \tag{6.5}$$

A FGr has a *fuzzy equality* ($\equiv$) regarding every value from each linguistic module. The modules are operated with $\wedge$. All the modules have a degree of truth which is triggered by either the satisfied or the violated set of constraints of a linguistic module in a dialect.

Example 2. *The syntactic module is taken as an example to explain how this formula works:*

$$(X_{(o\eta)\gamma} x_\gamma)d_\eta \tag{6.6}$$

The formula in Equation 6.6 is based in the following reasoning: a function such as $X : X_\gamma \times D_\eta \to M_o$:

- X means syntax of a grammar.
- X_γ is a set of syntactic constraints of a grammar.
- D_η is a set of constraints from the input's dialect.
- M_o is a set of truth values which represents the degree of grammaticality.

Therefore, X (which means syntax of a grammar) relates a set of syntactic constraints of a grammar (X_γ) with each constraint from the dialect's input (D_η). Every constraint of the syntactic set of constraints of an input will match a constraint in a dialect. Every matched constraint will be linked to a degree in [0,1].

The representation of this matching can be expressed by $X_\gamma \rightarrow (D_\eta \rightarrow M_o)$ which illustrates how the degrees of grammaticality are calculated strictly by the sets of linguistic constraints of a grammar which define the linguistic competence of a dialect. Therefore, the degrees of grammaticality in $X_\gamma \rightarrow (D_\eta \rightarrow M_o)$ are grasped **without** involving any judgment from any speaker.

Both the syntactic module of a grammar and a dialect's input are related in terms of degree.

6.4.1 Constraint Interaction regarding a Fuzzy Grammar

The grammar matches constraints from a linguistic module towards a constraint from an input:

- If the constraint is satisfied by the dialect, the degree will be represented by a degree of the satisfied constraint ($\mathrm{M_o} \equiv [0,1]$), according to the calculus regarding our evaluation system.
- If the dialect violates the constraint, the degree will be 0.
- In case a constraint is found violated on the dialect, as a consequence of the violation, the grammar *could* trigger another constraint to be matched in the dialect. The new triggered constraint will match the constraint found violated by the dialect. If now the first violated constraint and the new triggered one coincide, they will be matched with a new degree of grammaticality ($\mathrm{M_o} \equiv [0,1]$).

Example 3. *We understand C_x as a specific constraint.*

C_1, C_2, C_3, $C_4 \in X_\gamma$ *is an example of constraints that define the syntax of our fuzzy grammar.*

C_a, C_b, C_c, $C_d \in D_\eta$ *is an example of constraints that define an input in a dialect.*

$X(C_1, C_a) = 0.9$

$X(C_2, C_b) = 0.8$

$X(C_3, C_c) = 0$

$X(C_4, C_c) = 0.5$

In the syntax of a fuzzy grammar X, every canonical constraint (such as C_1) from one set of the linguistic module (such as syntax) is matched to its equivalent canonical constraint in the dialect (such as C_a). M_o characterizes the match between the rules of both sets with a degree of membership. A fuzzy grammar might found (or not) a constraint violated in a dialect's inputs. In this sense, we find degrees of grammaticality in both sets according to one fuzzy grammar.

Example 4. *In X (C_3, C_c) = 0 and X (C_4, C_c) = 0.5, a constraint in a dialect's input trigger two constraints in the set of constraints of the syntax of a FGr. In this case, C_3 is the prototypical object or gold standard constraint which has been violated in the dialect (C_c). While C_4 is the borderline object or special/variability constraint which assigns another degree in case the new constraint is satisfied in the dialect's input.*

The fuzzy grammar can operate degrees of grammaticality between modules taking into account linguistic constraints. The operations would be done using the minimum $\wedge$. Minimum is an operation which chooses the smaller value concerning two values.

Example 5. *Considering two linguistic modules such as syntax and semantics:*

$$(X_{(o\eta)\gamma}x_\gamma)d_\eta \wedge (S_{(o\eta)\delta}s_\delta)d_\eta \tag{6.7}$$

Due to a hypothetical satisfaction and the violation of constraints, the syntax ($(X_{(o\eta)\gamma}x_\gamma)d_\eta$) has a value of grammaticality of 0.5 and the semantics ($(S_{(o\eta)\delta}s_\delta)d_\eta$) a value of 0.4, therefore:

$$0.5 \wedge 0.4 = 0.4. \tag{6.8}$$

At this point, operating with minimum seems the best option since we do not know how a multi-modal perspective would react within this approach. Our (fuzzy) Property Grammar has only been dealing with Spanish syntax. Therefore, the multi-modal interactions are unknown. Consequently, it is assumed that the final value of grammaticality would be equal to the worst module. As a result, its value will prevail over the others. Nevertheless, this is just a hypothesis which will be clarified in future work.

Definition 4. *A fuzzy grammar operates as:*

$$FGr = \{^a/\_< Ph_\alpha, Mr_\beta, X_\gamma, S_\delta, L_\varepsilon, Pr_\zeta, Ps_\kappa >,^b/\_< ... >,^c/\_< ... >\} \tag{6.9}$$

where:

- ^{a}, ^{b}, ^{c} *are membership degrees (degrees of truth) of the corresponding elements in the angle brackets.*
- *The elements in the angle brackets are the modules of the grammar that matched with the elements of the dialect's input as well to a set of degrees.*

Example 6. *If we extract hypothetical degrees from* ^{a}*:*

$$Ph_\alpha = 1,\ Mr_\beta = 0.2,\ X_\gamma = 0.8565,\ S_\delta = 0.72,\ L_\varepsilon = 0.77,\ Pr_\zeta = 1,\ Ps_\kappa = 0.97. \tag{6.10}$$

and we operate with minimum $\wedge$ *it would have the following result:*

$$^a = 1 \wedge 0.2 \wedge 0.8565 \wedge 0.72 \wedge 0.77 \wedge 1 \wedge 0.97 = 0.2 \tag{6.11}$$

In this sense, the degree of grammaticality of both the FGr and a linguistic module will always be relying on the relation between the identified constraints in a dialect module and its degrees. The grammatical knowledge (competence) of a set takes into account the variables in a grammar in terms of degrees (if an input is satisfied or violated and its degree), but the degree of grammaticality of an input can only be triggered by a dialect's input concerning a grammar. Therefore, the degree of grammaticality is always related to the set of constraints which define the knowledge of a language of a fuzzy grammar.

Consequently, the more constraints that are satisfied in a grammar by a given input, the more grammatical it will be. Therefore, a given input has a high value of grammaticality according to its grammar (and not according to the speaker's perception). A given input which respects the structures and the rules of a grammar will have a high grammaticality value. A given input which triggers many violations will either display special constraints for a grammar or simply exhibiting rules which do not belong to the grammar which is evaluating the input of a dialect.

6.4.2 A Fuzzy Grammar Computed with Words

A theoretical $IF - THEN$ approach could be proposed. The fuzzy $IF - THEN$ is a tool which represents approximate reasoning concerning how we acknowledge the word. We do not say *"I like your meal a 0.845"*, we say things such as *"this meal is excellent"*. Fuzzy logic has the particularity of allowing us to compute with words numerical values as shown in subsection 4.4.1.

In this manner, we compute the following numbers with a linguistic expression to provide a model which can express the degrees of grammaticality more naturally:

- (1-0.8)[†)] *significantly*;
- (0.8-0.5) *quite*;
- (0.5-0) *barely*.

Therefore, we take into account approximate fuzzy reasoning such as:

- *IF an input* is *significantly* satisfied *THEN the value of grammaticality* is *high*.
- *IF an input* is *quite* satisfied *THEN the value of grammaticality* is *medium*.
- *IF an input* is *barely* satisfied *THEN the value of grammaticality* is *low*.

Similarly, we could express with the same values that:

- *IF the value of grammaticality* is *high THEN an input* is *significantly grammatical*.
- *IF the value of grammaticality* is *medium THEN an input* is *quite grammatical*.
- *IF the value of grammaticality* is *low THEN an input* is *barely grammatical*.

Such reasoning of grammaticality is much closer to natural language than saying *an input is 0.756*. We will implement such fuzzy reasoning when we apply our system for the evaluation of the degrees of grammaticality for a linguistic input for suggesting a more natural evaluation. These boundaries have been decided theoretically, and they are a proof-of-concept for the employment of computing grammaticality with words rather than numbers.

6.4.3 The Structure of a Word Defined with Fuzzy Logic

The structure of *word* through a fuzzy grammar has been defined. Thanks to this definition, we can link the fuzzy grammar with its degrees of grammaticality, together with the notion of *word* and a Property Grammars.

[†)] This number concerns to a final degree of grammaticality.

A word is understood here as an element which is multi-modal and constrained by the set of constraints of its structure and its meaning. The notion of *word* has always been a problem in linguistics. Most of the frameworks manage this concept, and yet it is difficult to define it without incurring in some problems. This investigation tested the use of fuzzy logic in order to define *what is a word*. One of the most challenging parts of defining the concept of word is regarding its *meaning*. We define a word just by considering its *structure*; saving the challenge of formalizing its meaning for further research.

This proposal lay down the formal basis for defining a word when a Property Grammars is going to be applied. Property Grammars are based on both the notion of word and category. Representing the structure of a word with fuzzy logic allow us to introduce the exact moment in which the properties take place for their evaluation. Therefore, this formalization takes place as a significant representation for both this work and a Fuzzy Property Grammars. Besides, this approach might show the potential of Fuzzy Natural Logic (FNL) as a tool for theoretical linguistics.

Regarding the formula of the grammar, FNL allows us to define *what is a word* in its structural complexity, acknowledging the following linguistic modules: phonology (*Ph*), morphology (Mr) and syntax (X).

Definition 5. *A word is a sequence of sounds: ($_\alpha$).*

Definition 6. *A message (m) is a sequence of sounds and every word is a part of this sequence of sounds: (m_α).*

Definition 7. *The sounds need to be phonologically correct according to the phonological constraints of a language. Thus, the degree of truth ($_o$) of the set of the phonological constraints (Ph) occurs already here.*

$$Ph_{o\alpha}m_\alpha \tag{6.12}$$

Now, these sounds are assigned with a degree of grammaticality in phonology, and they are identified with the phonemes and their phonological constraints.

In case that we are dealing with written language, we are going to assume that each written symbol (letter or letters) is supported by a sequence of phonemes which are related to the grammar which we are dealing with.

Definition 8. *The phonemes are understood as symbols with linguistic sign (morphemes) when they are found in the morphological module. These symbols with linguistic sign are represented by $_\beta$. Thus, the morphological module (Mr) allows to the phonemes $_\alpha$ to be* *understood as symbols with linguistic sign or morphemes $_\beta$ once they go through the set of the morpho-phonological module (MPh):*

$$MPh_{\beta\alpha}m_\alpha \tag{6.13}$$

Definition 9. *The phonemes assumed as morphemes (Mr_β) can be evaluated through the set of the constraints of the morphology component:*

$$Mr_{o\beta} \tag{6.14}$$

Definition 10. *The constraints of morphology understand a word as a compilation of morphemes. Morphemes are categorized in order to be syntactically available ($_\gamma$). This step is represented in the following way:*

$$(Mr_{\gamma\beta}(MPh_{\beta\alpha}m_\alpha)). \tag{6.15}$$

In words: a sequence of sounds (m_α) which passed through the constraints of phonology is a morpheme understood as a morpho-phonological element ($MPh_{\beta\alpha}$), which becomes categorized ($Mr_{\gamma\beta}$) and available to be constrained in a construction by its syntactic constraints. In essence, the combination of the sounds that have crossed the morphological module is now understood as morphemes $MPh_{\beta\alpha}$. The combination of the morphemes that are understood as only one part of speech defines the morphological module of a word. Once various morphemes have been assigned to one linguistic category or part of speech, we represented it with $Mr_{\gamma\beta}$.

Definition 11. *The syntactic properties can be triggered once a word has been categorized. The whole process can be expressed in one formula which takes into account the mention formulas above: $Mr_{\gamma\beta}(MPh_{\beta\alpha}m_\alpha)\&MPh_{\beta\alpha}m_\alpha$.*

$$(Mr_{\gamma\beta}(MPh_{\beta\alpha}m_\alpha)\&(Mr_{o\beta}(MPh_{\beta\alpha}m_\alpha)) \tag{6.16}$$

Definition 12. *The structure of a word has its final degree of grammaticality once the morphological categories (γ) have a degree of truth by the satisfaction or the violation of the set of constraints of the syntax (X_o) in a construction of a language. These regards allow us to understand a word as a pairing of its constraints plus the syntactic constraints of the word within the structure of a construction.*

$$X_{o\gamma}(Mr_{\gamma\beta}(Mr_{o\beta\alpha}m_\alpha)mr_\beta) \tag{6.17}$$

Definition 13. *The structure of a word is defined as follows:*

$$X_{o\gamma}(Mr_{\gamma\beta}(Mph_{\beta\alpha}m_{\alpha}))\&(Mr_{o\beta}(MPh_{\beta\alpha}m_{\alpha}))\&Ph_{o\alpha}m_{\alpha} \quad (6.18)$$

Following this definition, the word is understood as the sequence of sounds in a message, and its phonological constraints ($Ph_{o\alpha}m_{\alpha}$), through the association of those sounds with the morphemes ($Mr_{\beta\alpha}$ m_{α}), to the morphemic construction of the word once they are tagged by categories or part of speech ($Mr_{\gamma\beta}$). The structure and the features of the word will be understood in a multi-modal perspective by the satisfaction or the violation of its features as well as its features in a construction of a language. The degree of grammaticality of the set of constraints of the syntax of a word will be according to the violation or the satisfaction of its set of the syntactic constraints ($X_{o\gamma}$).

6.5 Evaluating Grammaticality in a Fuzzy Grammar

This section displays several mechanisms which allow the evaluation of the degrees of grammaticality. It is important to point out the following: it is assumed that, before the evaluation of grammaticality in the syntactical module, either the tagging or the categorization has been done according to the lexical entry of our grammar, just as shown in $X_{o\gamma}(Mr_{\gamma\beta}(Mph_{\beta\alpha}m_{\alpha}))$. Therefore, *we assume an ideal categorization from our grammar concerning a construction*.

6.5.1 Constraint Behaviour for a Fuzzy Property Grammars

Differently from PGs, in order to identify degrees of grammaticality in linguistic constructions, we must first identify the most *prototypical objects* for each construction in a grammar. These prototypical objects would be called *canonical constraints* or *canonical properties*. On the other hand, we need to identify the *borderline objects* for each construction as well. These would be called *variability constraints* or *variability properties*. According to this, a list with a definition for each constraint behavior is provided[‡)]:

a **Syntactic Canonical Properties**: These are the properties which define the gold standard of the Fuzzy Grammar. These are strictly the most representative constraints, based on both their frequency of occurrence and some theoretical reasons. These properties are represented with the type $_{\alpha}$: C_{α} is a canonical property.

b **Syntactic Violated Properties**: These properties are canonical properties which has been violated regarding a linguistic *input* or a dialect. Pointing out the violation of

[‡)]From now, Greek symbols are not related with previous sections.

a canonical property is necessary in order to trigger the related syntactic variability properties (if it is needed). These properties are represented with the type $_\beta$: C_β is a violated property.

c **Syntactic Variability Properties**: These properties are the core of this framework. These are triggered in the fuzzy grammar **only** when a violation is identified in an *input*. Therefore, *these are borderline cases in-between a violation and a canonical*. They explain linguistic variability concerning a fuzzy grammar. When a variability property is satisfied, it triggers a new value over the violated constraint improving its degree of grammaticality. These properties are represented with the type $_\gamma$: C_γ is a variability property.

6.5.2 Syntactic Variability and *xCategory*

The syntactic variability properties need of another significant effect for triggering variability properties. That is the notion of *xCategory*.

Definition 14. *A xCategory is* ***a feature*** *which specifies that a certain category is displaying a syntactic fit from another category, for example, a determiner with a syntactic fit of a pronoun. All the xCategories are marked with a x before a prototypical category, i.e. for a pronoun: xPRON. The properties of a xCategory are going to be placed in the description of the prototypical category.*

Example 7. *Consider the following sentence: "El rojo es mi coche favorito" ("The red is my favourite car"). The determiner "El" is categorized as a determiner, consequently some violations are triggered, i.e.* $DET \prec NOUN$ *and* $DET \Rightarrow NOUN$*. The violations are not erased; in fact, the PG detects the violations. However, once these violations are triggered, a Fuzzy Property Grammar finds a variability constraint in Spanish grammar which links these violations taking into account new constraints for a new fit* $DET_{[xPRON]}$*. If the* DET *satisfies the variability constraints from its new fit, the degree of grammaticality will be higher in comparison to its violation.*

Therefore, thanks to the notion of *xCategory* we specify a violation and, at the same time, we describe a fuzzy phenomenon such as a determiner performing as a borderline pronoun with its degree of grammaticality.

If we would consider that a category with a new *syntactic fit* changes its category, we would admit that there is no violation if the new fit satisfies all the new constraints. Additionally, we would be very discrete, because we would be admitting that a category has to be *either this or either that*. Therefore, in our framework, the *process* of categorization is discrete, but *a category* can be involved in fuzzy features regarding a structure. Consequently, we capture better the fuzzy phenomena involving a category.

6.5.3 Defining Syntactic Variability in a Fuzzy Property Grammar

The following formula represents the tool by which a violated property triggers a variability property:

Definition 15. *A variability constraint defined in a grammar occurs when a category in a construction has a violated constraint C_β, which is a negated canonical constraint $\neg(C_\alpha)$, and implies a variability constraint C_γ.*

$$\begin{gathered} A\, Category\, In\, a\, Construction: \\ C_\beta : \neg(C_\alpha) \implies C_\gamma \end{gathered} \tag{6.19}$$

Example 8. *A determiner (DET), in a subject construction (SUBJ), a violation property (C_β) has been triggered because a canonical property (C_α), i.e. DET $\prec$ NOUN, has been unsatisfied ($\neg$): $\neg(C_\alpha)$, i.e. DET in $SUBJ : C_\beta : \neg(DET \prec NOUN)$. This violation implies ($\implies$) that the input can still trigger another constraint of the grammar, a variability constraint (C_γ), i.e. $DET_{[xPRON]\gamma_{1\vee 2}}$.*

$$\begin{gathered} A\, Determiner\, In\, a\, Subject\, Construction: \\ C_\beta : \neg(DET \prec NOUN) \implies DET_{[xPRON]\gamma_{1\vee 2}} \end{gathered} \tag{6.20}$$

In other words, in Equation 6.20, syntactic variability properties are triggered once a Determiner in Subject Construction violates ($\neg$) the property $DET \prec NOUN$; therefore, the input has to satisfy the properties found in the syntactic variability properties of the *xPRON* ($xPRON_\gamma$) either the first one (1) or the second one (2) ($[xPRON]\gamma_{1\vee 2}$). The symbol $\implies$ is used for pointing out that the syntactic variability properties are true only when both elements (the violation and the variability property) co-occur at the same time.

If these variability properties are satisfied, a degree of grammaticality will be provided regarding the value of the satisfied variability property. On the other hand, if this new condition is not satisfied, the violation will remain with the value of 0.

Therefore, we find degrees of grammaticality regarding a violated property when a variability property is satisfied. In other words, *a variability property is a property triggered by a violated property*. Both properties are part of a fuzzy grammar, they need each other to be true, and they provide a gradient value for a grammatical violation, which means that **a**

variability property is fuzzy since it is a borderline constraint of a grammar triggered by a violation.

6.5.4 Numerating Constraints in a Fuzzy Grammar

Our fuzzy grammar proposes a system for enumerating and placing constraints. This system approaches our framework for computational applications. By enumerating the constraints, we can better represent their matches and interactions for representing their degrees of grammaticality; as shown in subsection 6.4.1. With our table design for placing constraints, we pursue to standardize the definition of a fuzzy grammar with constraints.

Every constraint has a number by which it can be identified. In Table 6.3, all the enumeration is specified.

Numerating Constraints						
C	x		y		z	
	POS		Constructions		Number of Constraint	
	0	DET -determiner-	0	Subject *Subj*	α	1-∞
	1	ADJ -adjective-	1	Verbal *Verb*	β	1∞
	2	NOUN	2	Direct Object *Dobj*	γ	1/1.1-∞
	3	PROPN -proper noun-	3	Indirect Object *Iobj*		
	4	PRON -pronoun-	4	Modifier *Mod*		
	5	VERB	5	Specifier *Spec*		
	6	ADV -adverb-	6	Coordinate *Conj*		
	7	CONJ -coordinate conjunction-	7	Subordinate *SubC*		
	8	SCONJ -subordinate conjunction-				
	9	ADP -preposition-				

Table 6.3 Numerating constraints.

- Our Part of speech nomenclature (POS) for categorization in Table 6.3 is understood as:
  - *DET*: determiner;
  - *ADJ*: adjective;
  - *NOUN*: noun;
  - *PROPN*: proper noun;

- *PRON*: pronoun;
- *VERB*: verb;
- *ADV*: adverb;
- *CONJ*: coordinate conjunction;
- *SCONJ*: subordinate conjunction;
- and *ADP*: preposition.

- Our construction nomenclature in Table 6.3 is understood as:
  - *Subj*: Subject construction;
  - *Verb*: Verbal construction;
  - *DObj*: Direct Object construction;
  - *IObj*: Indirect Object construction;
  - *Mod*: Modifier construction;
  - *Spec*: Specifier construction;
  - *Conj*: Coordinate construction;
  - *SubC*: Subordinate construction.

- Our number of constraints nomenclature in Table 6.3 is understood as:
  - for canonical properties (α), from 1 to the maximum displayed (∞) in the property table of a category;
  - for violated properties (β), from 1 to the maximum displayed (∞) in the property table of a category;
  - for variability properties (γ), from 1 to the maximum displayed in a property table of a category. Additionally, variability properties also define the properties of a *xCategory* by 1.1.

Example 9. *Firstly, we identify a constraint by its category with a number. For example, a property C of a word categorized as a noun ($_2$) would be:*

$$C_2 \tag{6.21}$$

Secondly, we identify the construction where the property takes place. A property C of a word categorized noun ($_2$) in a subject construction ($_0$) would be:

$$C_{20} \tag{6.22}$$

Finally, we point out the behavior of the constraint and its number together with its category and construction. Following the example of a noun: a property C of a NOUN ($_2$)*, in a Subject Construction* ($_0$)*, triggering the first canonical property of this construction:* α_1 $NOUN \Rightarrow DET$*, would be represented as:*

$$C_{20\alpha_1} \tag{6.23}$$

Example 10. *Numerating violated properties of a construction follow the same pattern as for the canonical ones: Following the example of a noun: a property C of a NOUN* ($_2$)*, in a Subject Construction* ($_0$)*, triggering a violation of the first canonical property of this construction:* β_1 $NOUN \Rightarrow DET$*, would be represented as:*

$$C_{20\beta_1} \tag{6.24}$$

Example 11. *Numerating variability properties of a construction follow the same patter as for the canonical ones: a property C of a NOUN* ($_2$)*, in a Modifier Construction* ($_4$)*, triggering its first variability property* γ_1*:* β_1 $\neg(NOUN_{[mod]} \Rightarrow ADP_{[spec]}) \implies NOUN_{[xADJ]}$*, would be represented as:*

$$C_{24\gamma_1} \tag{6.25}$$

Example 12. *Numerating the variability properties of a xCategory is firstly specified following the same pattern as the variability ones. Additionally, we add a second number which will specify the variability property satisfied in the xCategory.*

Considering the first variability property of an xADJ: γ_1 $NOUN \prec ADJ$*, a property C of a NOUN* ($_2$)*, in a Modifier Construction* ($_4$)*, triggering its first variability property* γ_1*:* β_1 $\neg(NOUN_{[mod]} \Rightarrow ADP_{[spec]}) \implies NOUN_{[xADJ]}$*, which satisfies the first variability property of an xADJ:* γ_1 $NOUN \prec X_{[xADJ]}$*, would be represented as:*

$$C_{24\gamma_{1.1}} \tag{6.26}$$

6.5.5 Placing Constraints in a Fuzzy Grammar

We provide an example in Table 6.4 for a Fuzzy Property Grammar in order to clarify how both a fuzzy property grammar description is done, and its constraint interactions.

Category (Cat) and Construction	
First Specification of Features for a Category	
α_1	$Cat property Cat \wedge Cat \wedge Cat$
α_2	$Cat property Cat$
α_3	$Cat Property Cat$
Second Specification of Features for a Category	
α_4	$Cat property Cat \vee Cat \vee Cat$
α_5	$Cat property Cat$
α_6	$Cat property Cat$
Variability Properties	
γ_1	$\beta_1 \neg(Cat property Cat) \implies Cat property Cat$
γ_2	$\beta_1 \neg(Cat property Cat) \implies Cat property Cat$
γ_3	$\beta_3 \implies Cat property Cat$
xCategory and Construction	
γ_1	$Cat_{[xCat]} property Cat$
γ_2	$Cat_{[xCat]} property Cat$

Table 6.4 Representation of a Fuzzy Property Grammar description for a category.

Table 6.4 show the following characteristics:

(a) α, β, γ: It assigns to each property a behavior and a number regarding the category in a construction.

(b) *Specifications*: They can specify features for each category. This trait is handy for those categories which have sub-categories, just like the pronouns. We could specify some properties for relative pronouns ($PRON_{[rel]}$), and others for personal pronouns ($PRON_{[pers]}$) and so on.

(c) $\wedge$: This symbol is understood as *and*. It allows defining a category and its properties concerning many different categories (or features) at the same time. Therefore, all the elements must be satisfied, or it will trigger a violation. This property prevents over-satisfaction, since it groups many categories under the same property. The over-satisfaction mainly occurs concerning the exclusion property. Exclusion property used to involve many categories.

Example 13. *A proper noun exclude categories such as a determiner, a noun, and a pronoun:* $PROPN \otimes DET \wedge NOUN \wedge PRON$.

If we would define an exclusion separately when one of these excluded categories would occur with the PROPN, such as {DET-PROPN}, the property grammar will label as satisfied the exclusions of the noun and the pronoun: β $PROPN \otimes DET$, α $PROPN \otimes NOUN$, α $PROPN \otimes PRON$.

We want to evaluate one property one time. Therefore, we are interested in accepting the exclusion property as satisfied just if all the categories are excluded since the over satisfaction regarding the other categories would tell us that exclusion is (paradoxically) satisfied and violated at the same time. Therefore, we define exclusion with $\wedge$*, triggering a violation in case any of its categories are not excluded.*

(d) $\vee$: This symbol is understood as *or*. It allows defining a category and its property concerning many different categories (or features) at the same time. One of the elements regarding $\vee$ must satisfy the specified property, or a violation will be triggered. This property prevents over-violation.

Example 14. *A transitive verb requires a subject such as a noun, or a proper noun or a pronoun:* $VERB_{[trans]} \Rightarrow NOUN_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$.

With $\vee$*, we specify that one requirement is enough to satisfy this property. If we described this property separately, when a verb would satisfy the requirement of a noun, the property grammar would trigger as a violation that the verb is not satisfying the requirement of a proper noun and a pronoun both as subjects:* α $VERB \Rightarrow NOUN_{[subj]}$, β $VERB \Rightarrow PROPN_{[subj]}$, β $VERB \Rightarrow PRON_{[subj]}$.

(e) *xCategory*: It allows to specify the properties for the feature *xCat* within the prototypical category.

Example 15. *A Pronoun in Subject construction (PRON in $Subject$ $Construction$), we specify the properties of a xPRON in Subject Construction:*

$PRON$ in $Subject$ $Construction$	
$PRON_{[personal]}$	
α_1	$PRON \otimes NOUN \wedge PROPN$
α_2	$CatpropertyCat$
Variability Properties	
γ_1	$\beta_1 \neg(PRON \otimes NOUN) \implies CatpropertyCat$
γ_2	$\beta_1 \neg(PRON \otimes PROPN) \implies CatpropertyCat$
γ_3	$\beta_2 \implies CatpropertyCat$
$xPRON$ in $Subject$ $Construction$	
γ_1	$Cat_{[xCat]}propertyCat$
γ_2	$Cat_{[xCat]}propertyCat$

Table 6.5 Description of an *xCategory* in its prototypical category.

(f) The violation of a property which define different relations with $\vee$ and $\wedge$ can trigger different variability properties.

Example 16. *A pronoun in subject construction excludes a noun and a proper noun: $PRON \otimes NOUN \wedge PROPN$. This property is α_1.*

Let's consider a violation of no exclusion between the PRON and the NOUN. This violation is β_1 and it can be specified as: $\beta_1 \neg(PRON \otimes NOUN)$.

In case a fuzzy grammar has a variability rule for this context, the fuzzy grammar can specify a variability rule with a new relation between categories regarding the specific violation of β_1 with: γ_1: $\beta_1 \neg(PRON \otimes NOUN) \implies CatpropertyCat$.

On the other hand, if the violation would be $PRON \otimes PROPN$, this violation would trigger a new variability property γ_2.

However, because the violation of no exclusion between the PRON and the PROPN is still the violation of α_1, the fuzzy grammar defines this violation with β_1. Nevertheless, the violation of the canonical β_1 is defined differently in γ_2 since the variability property has been triggered for a different context: $\beta_1 \neg(PRON \otimes PRON)$.

Therefore, the fuzzy grammar can specify a variability rule with a new relation between categories regarding the specific violation of α_1 with either γ_1: $\beta_1 \neg(PRON \otimes NOUN) \implies CatpropertyCat$ or γ_2: $\beta_1 \neg(PRON \otimes PROPN) \implies CatpropertyCat$.

If a property is not defined with $\wedge$ *or* $\vee$*, for example* α_2*, its variability can be defined just with the reference of its violation (*β_2*) without the need of describing the violation with* $\neg$*:* γ_3*:* $\beta_2 \implies Cat property Cat$.

6.5.6 Constraints and Grammaticality Values

Our system which calculates grammaticality values takes into account the following claims exposed in subsection 3.9.3.

(1) The value of grammaticality has to be theoretical. Sometimes, working with a corpus taking into account frequencies presents paradoxes. Frequency cannot always model neither grammaticality or a grammar. As an example, a Spanish corpus with frequencies such as Universal Dependencies strangely provides examples of personal pronouns being used as subjects (almost not at all). If we built grammaticality being guided just by frequencies, we would be forced to say that, in Spanish, the use of the personal pronoun as a subject is non-grammatical. However, no native Spanish speaker would consider such a thing. This reason is why frequencies are a tool for the linguist who combines such data with theoretical reasons. Such paradoxes are one of the main reasons why we cannot model grammaticality *only* under the spectrum of frequencies. These considerations are discussed in subsection 3.9.3 and section 7.3.

(2) A fuzzy grammar should take into account grammaticality values from a multi-modal perspective.

(3) A value of grammaticality regarding probabilities cannot model vagueness.

In this work, we have considered some of the objective theoretical notions for modeling gradient data (section 3.4) in order to evaluate grammaticality from a theoretical point of view. Those notions are the following:

(a) **Context effects**: We have extracted the properties according to its frequency and by applying theoretical notions such as the concept of *markedness*. A value just based on frequencies is avoided, in favor of a value based on a combination of frequencies plus the notion of *markedness* among other theoretical reason according to context effects. In such manner:

- A theoretical canonical value is understood as 1 (α=1).
- A violated value is understood as 0 (β=0).
- A variability value is understood as a 0.5 (γ=0.5).

(b) **Cumulativity, ganging up effect, constraint counterbalance and positive ganging up effect**. A Property Grammar takes into account different constraint behavior (both violated and satisfied) and the multiple repetitions of both a single *violation* or *various violations* for calculating degrees of grammaticality. It also considers the multiple repetitions of both a single *satisfaction* or *various satisfied* properties for calculating degrees of grammaticality.

(c) **Density**. This notion weights each constraint regarding the number of constraints that defines a category. In our approach, density works a bit different. Density weights each constraint according to the number of constraints of a category in the construction of an input that **are triggered** (either satisfied or violated).

(d) **Multi-modal values**. This work is focused on calculating grammaticality in the syntactic module. Consequently, in this work, a multi-modal value of grammaticality is not provided. This challenge is saved for further research.

The PG is the tool which manages most of the fuzzy details for evaluating grammaticality: it sorts out the types of properties, and its behavior, the property interactions, its context effects, and it can easily deal with both cumulativity and ganging up effects for their both positive and violated values.

Density values definitely fit the framework of a PGs since those are based on part-of-speech. Therefore, it is necessary to provide tools for extracting the density value for each category. A density value is entirely theoretical which is ideal for using this notion as a weight for representing the degrees of grammaticality regarding linguistic competence.

6.6 Word's Density and Degrees of Grammaticality

In this section, the formulas for representing grammaticality regarding an input for a Fuzzy Grammar are displayed.

Definition 16. *Each category is a word which has a whole full value of grammaticality:*

$$word = 1 \tag{6.27}$$

We acknowledge the above because there are not enough theoretical reasons to objectively estimate one word over the other. In the end, if all the constraints of a word are fully satisfied, the word would have a value of grammaticality of 1.

Definition 17. *The canonical value of each constraint of a word (Cn_w) is the value of a canonical property (1) divided by all the triggered constraints of a word (C_δ).*

$$Cn_w = \frac{1}{C_\delta} \tag{6.28}$$

Example 17. *A NOUN which triggers* 4 *constraints will assign a value of* 0.25 *for each of its canonical constraints.*

$$0.25\, Cn_w = \frac{1}{C_4} \tag{6.29}$$

Definition 18. *The variability value of each constraint of a word (Vab_w) is the value of a variability property (0.5) divided by all the triggered constraints in an word (C_δ).*

$$Vab_w = \frac{0.5}{C_\delta} \tag{6.30}$$

Example 18. *A NOUN which triggers* 4 *variability constraints will assign a value of* 0.125 *for each of its variability constraints.*

$$0.125\, Vab_w = \frac{0.5}{C_4} \tag{6.31}$$

Definition 19. *The final grammaticality value of a word (VG_w) is the addition of all the canonical values of each constraint of a word (All_{Cn_w}) plus all the variability values of each constraint of a word (All_{Vab_w}) divided by the value of a word (Equation 6.27).*

$$VG_w = \frac{(All_{Cn_w} + All_{Vab_w})}{word} \tag{6.32}$$

Example 19. *A NOUN which triggers* 4 *constraints with an assigned value of* 0.25 *and it satisfies* 2 *canonical constraints, it will have a grammatical word value of* 0.5. *If the same NOUN would satisfy 2 canonical constraints and 2 variability constraint, it would have a grammatical word value of 0.75.*

$$0.75\, VG_w = \frac{(0.5\, All_{Cn_w} + 0.25\, All_{Vab_w})}{1 word} \tag{6.33}$$

6.6.1 Extracting the Grammaticality Values from an Input

Once we acknowledge 1) the values of the constraints for each type of constraint in a category and 2) the final grammaticality value of each word, we can extract the grammaticality value of an *input*. Note that we mention *input* because this formula is made for evaluating any utterance, construction or linguistic input in which their words can be identified.

Definition 20. *The value of grammaticality VG is the result of dividing all the final grammatical values of each word (All_{VG_w}) with all words in an input (All_δ):*

$$VG = \frac{(All_{VG_w})}{All_\delta} \tag{6.34}$$

Example 20. *Considering an hypothetical sentence with hypothetical constraints such as in Table 6.6:*

$$\{DET_{[spec]} = 1\; NOUN_{[subj]} = 1\; NOUN_{[mod]} = 0.75\; VERB_{[intrans]} = 1\; ADJ_{[mod]} = 1\; ADV_{[mod]} = 0.33\} = 5.08 \tag{6.35}$$

$$0.846\; VG = \frac{(5.08\; All_{VG_w})}{All_6} \tag{6.36}$$

	$DET_{[spec]}$	$NOUN_{[subj]}$	$N_{[mod]}$	$VERB_{[intrans]}$	$ADJ_{[mod]}$	$ADV_{[mod]}$
	4/4	5/5	2/4	3/3	4/4	1/3
Cn_w and Vab_w	$C_{05\alpha_1} = 0.25$ $C_{05\alpha_2} = 0.25$ $C_{05\alpha_3} = 0.25$ $C_{05\alpha_4} = 0.25$	$C_{20\alpha_1} = 0.2$ $C_{20\alpha_2} = 0.2$ $C_{20\alpha_3} = 0.2$ $C_{20\alpha_4} = 0.2$ $C_{20\alpha_5} = 0.2$	$C_{24\alpha_1} = 0.25$ $C_{24\alpha_2} = 0.25$ $C_{24\gamma_3} = 0.125$ $C_{24\gamma_{4.1}} = 0.125$	$C_{51\alpha_1}$ $C_{51\alpha_2}$ $C_{51\alpha_3}$	$C_{14\alpha_1} = 0.25$ $C_{14\alpha_2} = 0.25$ $C_{14\alpha_3} = 0.25$ $C_{14\alpha_4} = 0.25$	$C_{64\alpha_1} = 0.33$ $C_{64\beta_2} = 0$ $C_{64\beta_3} = 0$
VG_w	1	1	0.75	1	1	0.33
VG	0.846					

Table 6.6 Example of an input with values of grammaticality and its constraints.

Table 6.6 shows how an application of all the formulas for extracting grammaticality values would be. It is worth to point out how all the constraints are enumerated and specified with its behavior.

Equation 6.32 is very flexible since we can evaluate any input or linguistic construction with it. We do not need to evaluate just phrases, full expressions and so on. For example, we can consider as an *input* the full construction in Table 6.6, the value of such is 0.846. However, if we are interested in finding out the VG of another construction, such as $\{V_{[intrans]}$ $ADJ_{[mod]}$ and $ADV_{[mod]}\}$, we just need to apply the same formula for these two words, and it reveals that their VG is 0.776.

$$0.776\,VG = \frac{(2.33\,All_{VG_w})}{All_3} \tag{6.37}$$

Besides, the relation between violation and variability can be seen in $N_{[mod]}$; where its violated constraints have a value of 0.125 rather than 0, as it would happen with other approaches. Therefore, the value of the violated constraint has a degree according to its borderline case. Contrarily, the violated constraints in $ADV_{[mod]}$ are violated without the possibility of any compensatory value because, in this case, the grammar does not acknowledge any variability constraint for it.

6.7 Fuzzy Grammar and its Evaluation: A Summary

The proposed formalization of a fuzzy grammar and its evaluative system provides the following benefits:

(1) Fuzzy Grammar is an *alternative representation* for the basis of the theoretical controversy regarding the linguistic competence and performance.

(2) It acknowledges the grammar as an external entity able to evaluate inputs in terms of degree (without involving speakers' judgments).

(3) It has a *universal* and *multi-modal* character. The formula of fuzzy grammar might be applied to any language.

(4) The formula of *word* helps us to acknowledge the moment of the calculation of the degrees of grammaticality concerning the syntax of a language.

(5) It takes into account the typical notions of gradient data for representing grammaticality.

(6) It provides an innovative system which illustrate *different fuzzy phenomena*, introducing the notions of *variability properties* and *xCategory*.

(7) It acknowledges *grammaticality as a vague concept* with borderline constraints.

(8) It provides a system for a *positive calculation for fuzzy degrees of grammaticality*.

(9) Our approach can evaluate any kind of *input* which takes into account the notion of *word*.

The fuzzy grammar and the systems proposed leave room for uncertainty and predictive tools for a full explanation of natural language inputs. In future work, we will test what happens if we adopt grammatical weights for each constraint according to its occurrences and probabilities — expecting that they would model degrees of acceptability within a fuzzy grammar.

Apart from this, we believe that this grammar proposal is innovative in contrast to other approaches which had treated the gradient phenomena in linguistics, such as optimality theory frameworks, and other grammars with constraints. The reason why our fuzzy grammar works it is thanks to the acknowledgment of those variability properties and borderline *fits* which are known as *xCategory*.

These notions came up as a consequence of building a Fuzzy Property Grammar for Spanish syntax with the corpus of Universal Dependencies Treebank. The Fuzzy Grammar would be just a theoretical formula without a grammar of constraints supporting it. These reasons show why it is so important to present at the same time the fuzzy grammar and its evaluation system together with the extracted Property Grammar.

Chapter 7

Extracting Constraints for a (Fuzzy) Spanish Property Grammar

This chapter presents the techniques used for extracting the properties of Spanish syntax from a representative data corpus (Universal Google Dataset 2.0) by combining two tools:

1. Universal Dependency Spanish Treebank (Spanish UD), a corpus with dependency treebanks of Spanish language;
2. and *MarsaGram*, a corpus tool based on frequencies which uses Universal Dependencies for inducing properties automatically.

Firstly, we show the used methods for extracting a Property Grammar (PG) for the Spanish language. Secondly, we provide an example of the extraction regarding the subject construction in Spanish language (standard Spanish/Castilian Spanish). The extraction method is determined by all the theoretical basis for a fuzzy grammar; which have been presented in chapter 6. Finally, section 7.13 will display the Property Grammar for Spanish language, which has been built in this dissertation as a proof of concept of the grammatical basis for a fuzzy grammar.

7.1 Spanish Universal Dependency Treebank

Universal Dependency Spanish Corpus Treebank is obtained from the Universal Google Dataset (version 2.0). It consists of 16,006 tree structures and 430,764 tokens and is built from newspaper articles, blogs, and consumer reviews. The parsed sentences are the data that MarsaGram will use in order to extract properties for a PG automatically.

The Spanish Universal Dependency Treebank provides dependency relations, heads, parts of speech (POS) and phrases. Figure 7.1 is an example of the whole linguistic information regarding a dependency treebank.

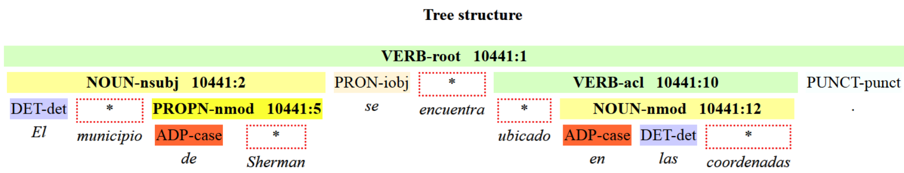


Fig. 7.1 An example of a dependency treebank from Spanish UD.

Guidelines for Universal Dependencies (UD) can be found in https://universaldependencies.org/guidelines.html. We are using their nomenclature during the process of the extraction of the Spanish properties. Even though most of them are somewhat intuitive, we suggest to check the part of speech (POS) tags for morphology https://universaldependencies.org/u/pos/index.html, and the syntactic dependencies https://universaldependencies.org/u/dep/index.html.

Spanish UD takes into account more categories and dependency relations than the ones considered in our Property Grammar (Figure 5.1, Table 6.3). In our work, we will adapt the Spanish UD nomenclature to the one proposed in our PG (Table 6.3) for both POS and dependencies.

7.1.1 Adjusting POS from UD to our Property Grammar

Spanish UD have the following tags shared with our PG (Table 6.3):

- *DET* (determiners);
- *ADJ* (adjectives);
- *NOUN* (nouns);
- *PROPN* (proper noun);
- *PRON* (pronoun);
- *VERB* (verbs);
- *ADV* (adverbs);
- *CONJ* (coordinate conjunction);
- *SCONJ* (subordinate conjunction);

- *ADP* (preposition).

However, in Spanish UD, we can also found as POS:

(1) *AUX*: *Auxiliary particles*. These are typically part of a verbal construction. For example, in Figure 7.2, *"ha"* is an auxiliary verb from an auxiliary phrase *"ha sido"*, which is part of a verbal construction such as: *"ha sido actualizado"* (*has been actualized*). In our PG, we will consider that the auxiliary particles are already included in the final word which will be tagged as a verb category. This is because we assume that the previous morphological processes (illustrated in subsection 6.4.3) already put-altogether the different parts of a word which are going to be considered as one single category, specifying the passive voice (if needed): $\{ha, sido, actualizado\} = VERB_{[pass]}$.

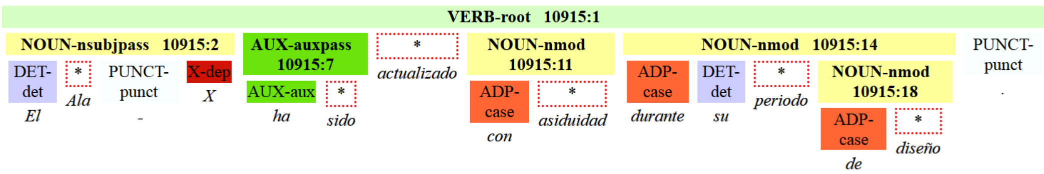


Fig. 7.2 UD treebank with AUX.

(2) *NUM*: *Numerals*. These are usually performing as a determiner or as an adjective. During the corpus extraction we will try to sort out in which category we place them better. In our PG, the numeral is going to be considered as a feature. In Figure 7.3, *"primera partida"* (*first game* would be defined as $ADJ_{[num]} \rightsquigarrow_{mod} NOUN$. In Figure 7.4, *"Dos cohortes"* will be considered as $DET_{[num]} \rightsquigarrow_{spec} NOUN$.

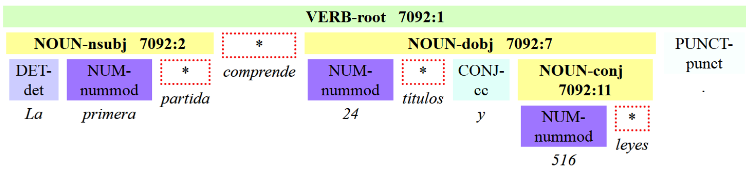


Fig. 7.3 Example of a NUM which will be considered as a $ADJ_{[num]}$.

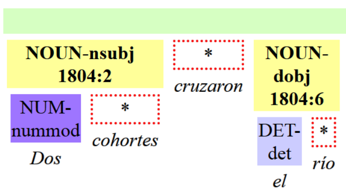


Fig. 7.4 Example of a NUM which will be considered as a $DET_{[num]}$.

(3) *PART*, *X*, *PUNCT*, *SYM*. *Parts, unknown, punctuation and symbols*. These tags are dismissed. We consider them irrelevant for the extraction of our PG. Figure 7.5 illustrates these tags.

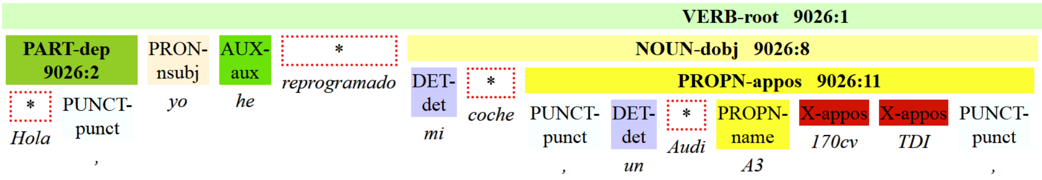


Fig. 7.5 Example with dismissed tags: *PART*, *X*, *PUNCT*.

7.1.2 Adjusting Spanish UD to our PG's Dependencies

MarsaGram has extracted the properties utilizing constituents and dependencies. This task has been possible because the dependency relations have been taking into account as a guide for establishing constructions. UD creates treebanks of phrases based on dependencies. Since a phrase is a type of construction, we extract the basic constructions from these dependency phrases.

The difference between the dependency phrases from UD and the constructions in our work relies on the fact that we define a construction in terms of properties and one of these properties is the dependency property. In contrast, dependency phrases from UD are constructions based solely in the notion of dependency, which we acknowledge as just one of our properties.

A construction is generally defined as a pair of structure and meaning. In our work, dependencies are one of those elements which define structure and meaning in syntax. For example, considering the dependency of subject in *"The boy is swimming"*, the dependency of subject is in fact determined by:

1) A syntactic meaning which is "the subject who is performing an action."

2) A legit structure which performs such syntactic meaning (*DET* and *NOUN*).

UD displays different kinds of dependency relations between the tagged POS. In our work, we simplified these relations grouping those similar dependencies (if not the same) under a label regarding the dependency relations from PG illustrated in Figure 5.1.

(a) **Subject dependencies**: In Spanish UD, subject dependency is represented by:

- *nsubj*: nominal subject, which is the most frequent subject dependency in our corpus.
- *nsubjpass*: nominal subject passive, which is the second most frequent subject dependency in our corpus.
- *csubj*: clausal subject, which is the less frequent subject dependency in our corpus.

A *csubj* is a subject which introduces a subordinate clause. In Spanish UD, *csubj* is typically performed by a verb in infinitive such as *"reservar"* in Figure 7.6. In our PG, we group all these dependencies under the label of *subject dependency*.

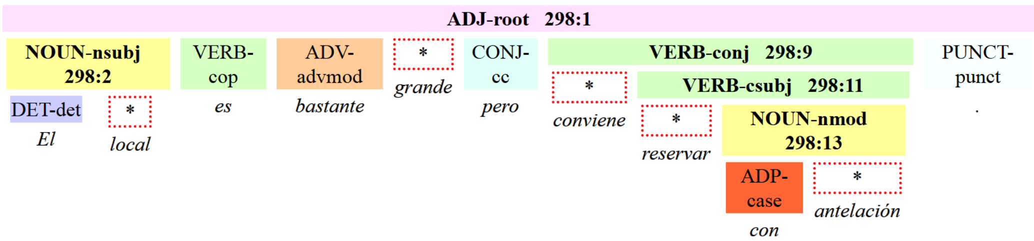


Fig. 7.6 Example of a *csubj*.

(b) **Verbal dependencies**: In Spanish UD, the verbs are tagged with the dependencies of *root*, *cop*, *aux* and *aux:pass*.

- In Spanish UD, the verbs are mostly specified with *root* dependency. This dependency specifies the category which is in the highest position of the hierarchy of a sentence. For those sentences with copulative verbs, *root* is a dependency for nominal elements and adjectives, because UD considers that the typical *attributive function* is the *head* of the construction rather than the copulative verb.
- The copulative verb is tagged with *cop* dependency. This is showed in Figure 7.6 with the verb *"es"* as *cop* and the adjective *"grande"* as *root*.
- The other verbal dependencies are both *aux* and *aux:pass*, which specifies the dependency form either an auxiliary or a passive auxiliary towards a verb.

In our PG, the dependencies of *root*, *aux* and *aux:pass* are avoided. The *attributive function* is going to be specified with the dependency of *comp* which indicates the most general relation between a head and an object: $NOUN \rightsquigarrow_{comp} VERB_{[cop]}$.

(c) **Direct object dependency**: In Spanish UD, the direct object dependency is *dobj*. Spanish UD has other types of complement similar to *dobj* such as *ccomp* and *xcomp*:

- The *ccomp* is a clausal complement, which is usually a subordinate phrase with a direct object function. In Figure 7.7, we have an example of *ccomp* with the clause *"que perderá peso político"*.
- On the other hand, *xcomp* is a subordinate phrase with a direct object function, just like *ccomp*, but without an explicit subject. In Spanish UD, *xcomp* is an open clausal complement, and it is usually performed by a verb in infinitive introducing a clause of a direct object such as in Figure 7.8 with *"ver quien golpea a Spike"*.

In our PG, we group all these dependencies under the label of *direct object dependency*.

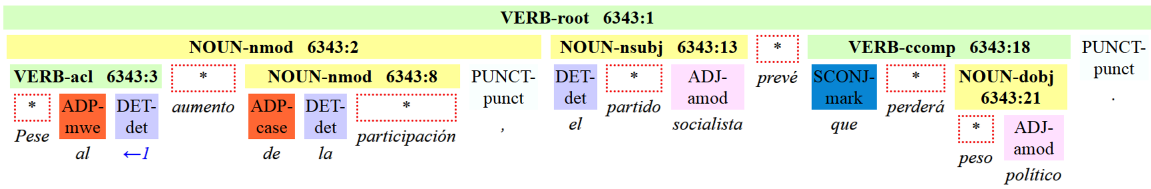


Fig. 7.7 Example of a *ccomp*.

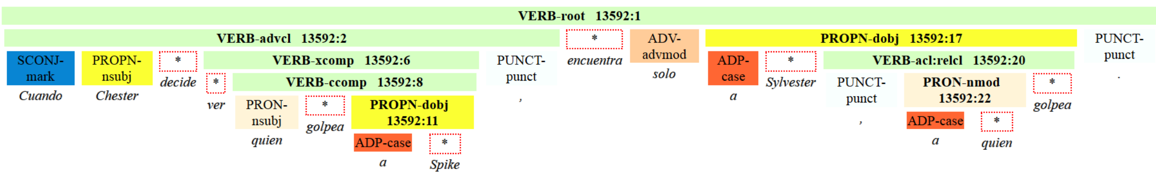


Fig. 7.8 Example of a *xcomp* and *ccomp*.

(d) **Indirect object dependency**: In Spanish UD, the indirect object dependency is *iobj*. In our PG, we group this dependency under the same label of *indirect object dependency*.

(e) **Specifier dependency**: In Spanish UD, we don't explicitly find specifiers. Specifications are typically understood as the relation between $DET - NOUN$. However, the Spanish UD provides:

- A specific dependency for determiners which is *det*.
- A special dependency for adpositions which is *case* (case marking).

Although, both *det* and *case* are treated like elements which typically introduce a nominal phrase. In this manner, these dependencies are specifying nominal elements. In our PG, we agree with this perspective, and we group both *det* and *case* under the label of *specifier dependency*.

(f) **Conjunction dependency**. In Spanish UD:

- *cc* is the dependency for coordinate conjunctions. The dependency of *cc* is labeling the coordinate conjunctions.
- *conj* is the dependency for a coordinate phrases. The dependency of *conj* labels the whole coordinate phrase such as in Figure 7.9; where *"y"* has a dependency of *cc* and the whole conjunction phrase of *"ocupó toda la ciudad"* is labeled with *conj*.

In our PG, the dependency of *conj* specifies two cases:

1) The conjunction is dependent of its principal: $CONJ_{[cc]} \rightsquigarrow_{conj} X_{[main]}$.
2) The coordinated element is dependent of its conjunction: $X_{[conj]} \rightsquigarrow_{conj} CONJ_{[cc]}$.

Following the case of Figure 7.9:

1) *y* ($CONJ_{[cc]}$) $\rightsquigarrow_{conj}$ *trasladó* ($VERB_{[main]}$).
2) *ocupó* ($VERB_{[conj]}$) $\rightsquigarrow_{dep}$ *y* ($CONJ_{[cc]}$).

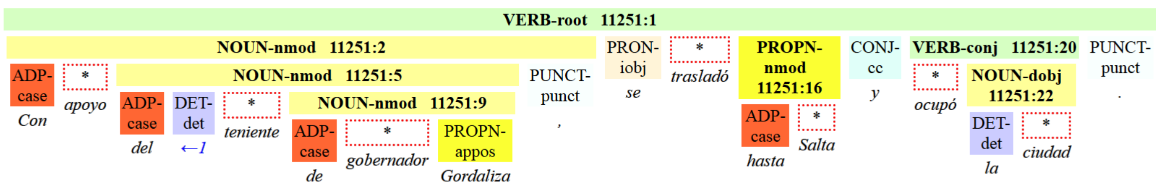


Fig. 7.9 Example of *cc* and *conj*.

(g) **Subordinate (conjunction) dependency**. In Spanish UD:

- The dependency *mark* (marked) introduce a subordinate conjunction.
- *acl*:*relcl* is a dependency which labels a clausal modifier of a noun with a relative. Therefore, a typical relative clause noun modifier.

Both *mark* and *acl*:*relcl* are illustrated in Figure 7.10 in which *"que"* is *mark* and *"que lleva su nombre"* is the relative clause with *acl*:*relcl* dependency. The controversy here is that typically a relative clause is introduced by a relative pronoun, and not by a conjunction. Actually, in the presented sentence, *"que"* can be substituted by *"la cual"* which is a proof that it is a pronoun. Besides, the pronoun *"que"* has a function of a subject and it can be substituted by its antecedent (*"la fábrica"*) in the sentence: *"la fábrica de cerveza que"*/*"[la fábrica] lleva su nombre"*. At the same time, the relative clause has a dependency as a modifier of its antecedent, and yet, the verb of the relative clause is dependent on the relative pronoun which introduces it. In our PG, because of the idiosyncrasy of the relative pronouns, we specify the following relations:

1. its dependency in the subordinate construction;
2. its dependency towards its referent;
3. the dependency of the subordinate verb towards the relative pronoun which introduces the subordinate.

Following this last example:

1) $que(PRON_{[rel]}) \rightsquigarrow_{subj} lleva(VERB_{[rel]})$;
2) $que(PRON_{[rel]}) \rightsquigarrow_{mod} fábrica\ (NOUN_{[ref]})$;
3) $lleva\ (VERB_{[rel]}) \rightsquigarrow_{dep} que(PRON_{[rel]})$.

It is true that in Spanish a conjunction such as *"que"* can introduce a subordinate. In Figure 7.11, *"se espera **que** escriba literatura que gane premios"* is an example of a subordinate introduced by a conjunction. In our PG, this case would be modeled in two steps as well:

1. the conjunction is dependent of its main;
2. the subordinate verb is dependent of the subordinate conjunction:
   1) $(que)SCONJ_{[sconj]} \rightsquigarrow_{conj} (se\ espera)\ VERB$;
   2) $(escriba)VERB_{[sconj]} \rightsquigarrow_{dep} (que)SCONJ$.

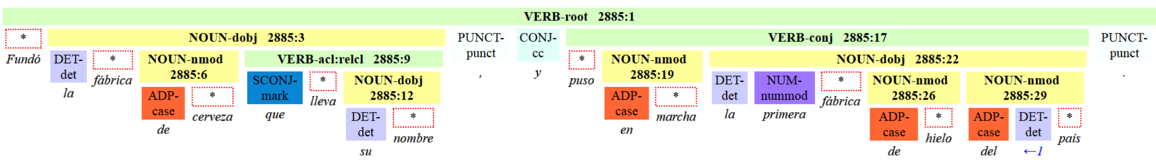


Fig. 7.10 Example of *mark* and *acl*:*relcl*.

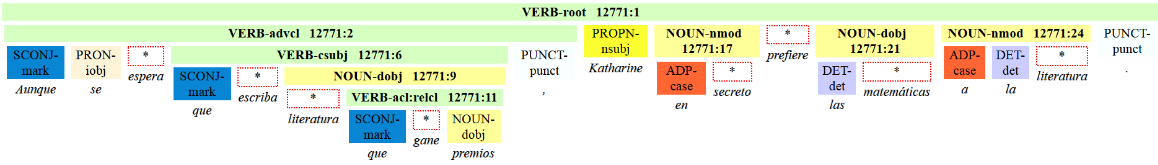


Fig. 7.11 Example of a subordinate conjunction.

(h) **Modifier dependency**. In Spanish UD, there are many specific dependencies for each POS regarding modifiers: *nmod*, *appos*, *name*, *nummodd*, *amod*, *advmod*, *neg*, *acl*, *advcl*.

- *nmod* dependency specifies a nominal modifier, usually a noun introduced by a preposition.
- *appos* dependency is a nominal modifier without any specifier such as a bare noun modifying another noun.
- *name* dependency is typically an apposition of a proper noun modifying another proper noun.
- *nummod* dependency is a numeral modifying a noun.
- *amod* dependency specifies that an adjective is modifying a noun.
- *advmod* dependency illustrates the dependency between an adverb modifying an adjective.
- The negative adverb *no* typically performs the *neg* dependency.
- *acl* dependency is an adjectival clause which modifies a noun. Usually, this clause is introduced by a participle verb.
- *advcl* is an adverbial clause modifier. In our PG, all these dependencies are grouped under the same label of *modifier dependency*.

(i) **Other dependencies.** In Spanish UD, we also find dependencies such as:

- *parataxis* (a dependency which tags a parenthetical clause or a clause after a ":" or a ";" without any explicit coordination, subordination, or argument relation with the head word).
- *punct* (punctuation).
- *mwe* (multi-word expression).
- *compound* (word composition).

– *dep* (unspecified dependency).

These dependencies are generally dismissed for their low frequency. We are not interested in *punct*. The *mwe* and *compound* dependencies are avoided because we assume that the multi-word expressions are already parsed as one single category when we are analyzing a construction with our properties.

Taking into account what we have said, Table 7.1 contrasts constructions and its expected dependencies with the Spanish UD.

Universal Dependencies Treebank in Spanish									
Constructions	Subject Construction	Verbal Construction	Direct Object Construction	Indirect Object Construction	Modifier Construction	Specifier Construction	Conjunctive Construction	Subordinative Conjuctive Construction	Others
Dependencies in UD	nsubj, nsubjpass, csubj	root, cop, aux, aux:pass	dobj ccomp xcomp	iobj	nmod, appos, name nummod, amod, advmod, neg acl, advcl	det, case	cc, conj	mark, acl:relcl	Compound, mwe, parataxis, punct, dep
Dependencies in FPG	subj	dep, comp, aux	obj	iobj	mod	spec	conj, dep	conj, dep	Not considered

Table 7.1 Contrasting constructions, UD dependencies and PG dependencies.

7.2 Characteristics of MarsaGram for the Extraction

Spanish syntactic properties have been extracted automatically by applying MarsaGram to the Universal Dependency Spanish Corpus Treebank (Blache et al., 2016). MarsaGram was developed in the *Laboratoire Parole et Langage* in Aix-en-Provence. As we have already said, this corpus is obtained from the Universal Google Dataset (version 2.0) and consists of 16,006 tree structures and 430,764 tokens and is built from newspaper articles, blogs, and consumer reviews. MarsaGram extracts 7535 rules (constructions) from this Spanish treebank plus 42235 properties.

The Spanish Universal Dependency Treebank provides dependency relations, heads and parts of speech. Whereas MarsaGram ranks each set of constituents with their dependencies by frequency, automatically deducing the most extended constructions and properties, which will be reviewed by the linguist. In this way, this method combines three main types of linguistic information –dependencies, constituency, and syntactic constraints– for building a PG.

MarsaGram has essential advantages for linguistic review; that is, it can analyze and simultaneously extract constituency and dependencies by their frequency. These traits allow us to define and characterize Spanish constructions and their properties using an objective data criterion. Once a proper linguistic review has been conducted, this linguistic information is used to define both gradient relations and fuzzy phenomena in syntax.

The advantages that MarsaGram has for linguistic research are the following*):

- The corpus allows us to work with categories and their dependencies and find dependency phrases: noun phrases, adjective phrases, prepositional phrases, and so on. Once we select the category that we want to check, the corpus provides all the constituents together with their dependencies and the number of occurrences of the elements that appear in the phrase (weighted according to their frequency). So, we can check the most important/frequent categories for each construction in terms of dependencies. Figure 7.12 shows the general interface of MarsaGram, which gives this information. The symbol * means *head*. In Figure 7.12 * is a $VERB$ because we are looking at $VERB - root$ constructions. In Figure 7.13, * is an ADJ because we are looking at $ADJ - amod$ constructions.

*) This 8-page paper is recommended for further information concerning MarsaGram and its technical details: https://hal.archives-ouvertes.fr/hal-01462181/document.

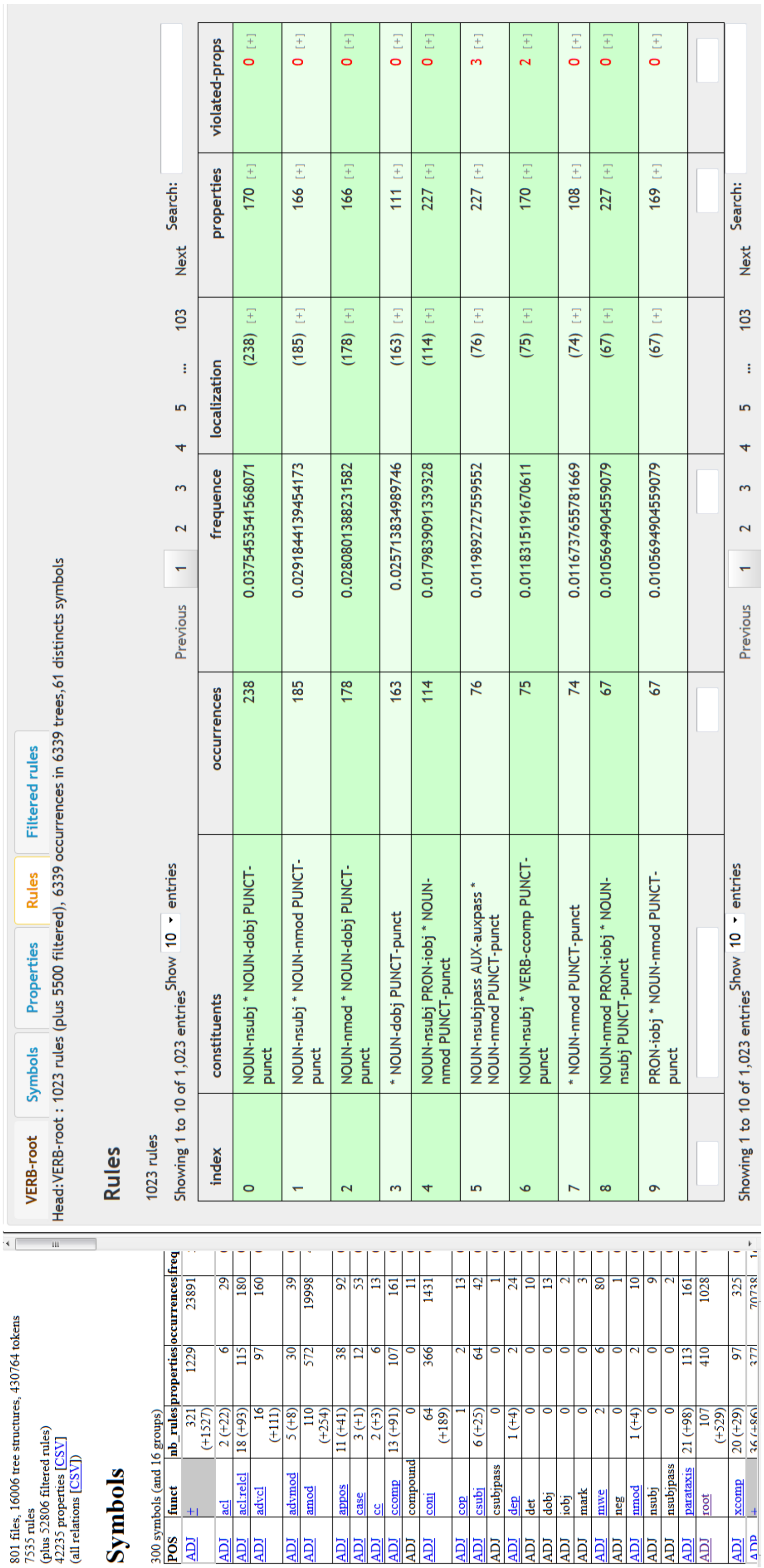
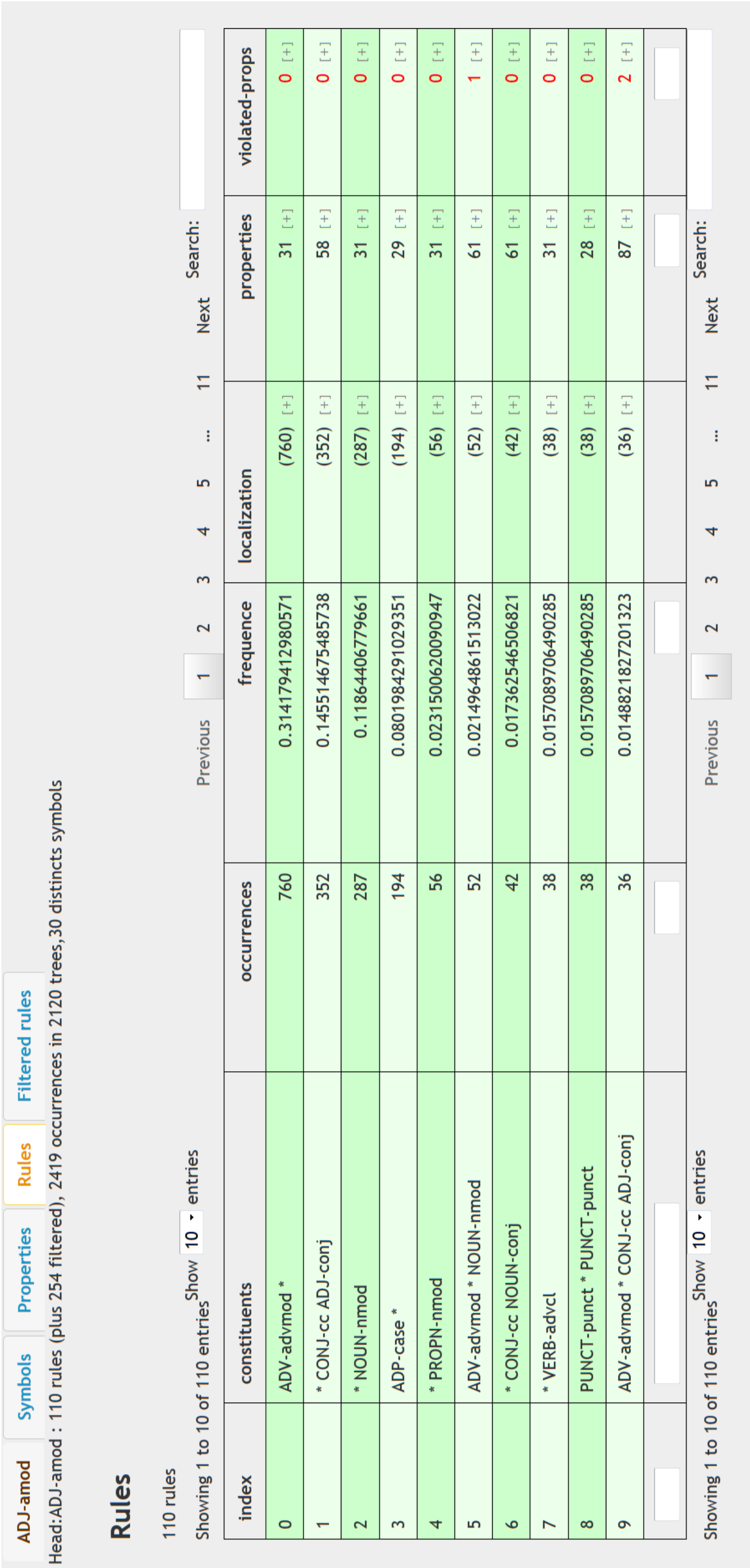


Fig. 7.12 MarsaGram interface with its constructions, heads, dependencies, and frequencies.

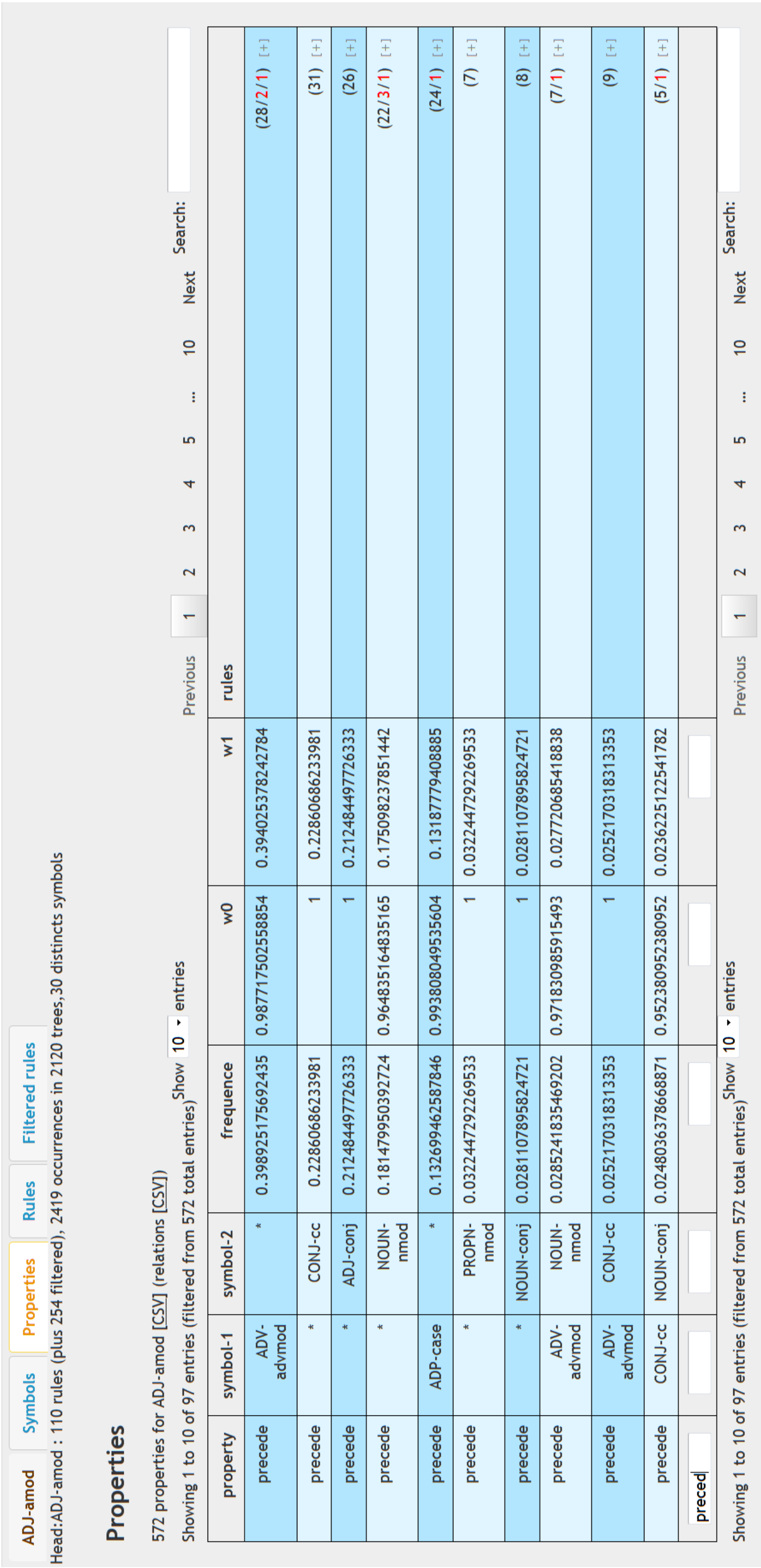
- We can apply the notion of construction (Goldberg, 1995) to the pair of constituency plus dependencies which appear in the RULES section, i.e. a subject construction is a subject dependency-constituency-phrase, a direct object construction is extracted from a direct object dependency-constituency-phrase and so on. Therefore, we can see which constituents take part in the most common syntactic constructions of Spanish since we operate with an objective statistical frequency number. In this way, constructions in RULES with an index close to 0 will be the most recurrent and, probably, the most canonical; it will also be shown if the most remaining ones are either exceptions or are the result of an error during the parsing process. As an example, Figure 7.13 shows the rank of adjective constructions with *amod* dependency.



index	constituents	occurrences	frequence	localization	properties	violated-props
0	ADV-advmod *	760	0.314179412980571	(760) [+]	31 [+]	0 [+]
1	* CONJ-cc ADJ-conj	352	0.145514675485738	(352) [+]	58 [+]	0 [+]
2	* NOUN-nmod	287	0.11864406779661	(287) [+]	31 [+]	0 [+]
3	ADP-case *	194	0.0801984291029351	(194) [+]	29 [+]	0 [+]
4	* PROPN-nmod	56	0.0231500620090947	(56) [+]	31 [+]	0 [+]
5	ADV-advmod * NOUN-nmod	52	0.0214964861513022	(52) [+]	61 [+]	1 [+]
6	* CONJ-cc NOUN-conj	42	0.017362546506821	(42) [+]	61 [+]	0 [+]
7	* VERB-advcl	38	0.0157089706490285	(38) [+]	31 [+]	0 [+]
8	PUNCT-punct * PUNCT-punct	38	0.0157089706490285	(38) [+]	28 [+]	0 [+]
9	ADV-advmod * CONJ-cc ADJ-conj	36	0.0148821827201323	(36) [+]	87 [+]	2 [+]

Fig. 7.13 MarsaGram ranking adjective constructions with modifier-dependency.

- MarsaGram provides two weights based on the frequency of each property. $w0$ is a weight that depends on the number of times a property has been violated, while $w1$ is a numerical value of the importance of a property in the corpus. This value is correlative to the frequency of a property. Therefore, a property that has never been violated ($w0$ as 1) but which has a low numerical value in the corpus ($w1$ as 0.001) means that it is either residual or an exception. A property with a high value of importance ($w1$), together with a high value of satisfaction ($w0$), is a significant property which the speakers tend to respect. In Figure 7.14 precedence values of $w0$ and $w1$ in $[ADV - advmod]$ and $[*ADJhead]$ are very high. Therefore, this is an important property in Spanish, since it is often satisfied and often used. Nevertheless, the property of precedence of the adjective in $[CONJ - cc]$ and $[NOUN - conj]$ is almost always satisfied and less present. It can be deduced that this property is secondary to the general property of precedence between the adjective and the conjunction. Thus, we can ignore it as a general property of the adjective construction. So, these values help us build a PG, although they are not necessary for grammaticality, according to the discussion in subsection 3.9.3.


ADJ-amod | Symbols | Properties | Rules | Filtered rules

Head:ADJ-amod : 110 rules (plus 254 filtered), 2419 occurrences in 2120 trees,30 distincts symbols

Properties

572 properties for ADJ-amod [CSV] (relations [CSV])

Showing 1 to 10 of 97 entries (filtered from 572 total entries) Show 10 entries

Previous 1 2 3 4 5 … 10 Next Search:

property	symbol-1	symbol-2	frequence	w0	w1	rules
precede	ADV-advmod	*	0.398925175692435	0.987717502558854	0.394025378242784	(28/2/1) [+]
precede	*	CONJ-cc	0.22860686233981	1	0.22860686233981	(31) [+]
precede	*	ADJ-conj	0.212484497726333	1	0.212484497726333	(26) [+]
precede	*	NOUN-nmod	0.181479950392724	0.964835164835165	0.175098237851442	(22/3/1) [+]
precede	ADP-case	*	0.132699462587846	0.993808049535604	0.131877779408885	(24/1) [+]
precede	*	PROPN-nmod	0.0322447292269533	1	0.0322447292269533	(7) [+]
precede	*	NOUN-conj	0.0281107895824721	1	0.0281107895824721	(8) [+]
precede	ADV-advmod	NOUN-nmod	0.0285241835469202	0.971830985915493	0.0277206854188838	(7/1) [+]
precede	ADV-advmod	CONJ-cc	0.0252170318313353	1	0.0252170318313353	(9) [+]
precede	CONJ-cc	NOUN-conj	0.0248036378668871	0.952380952380952	0.0236225122541782	(5/1) [+]
preced						

Showing 1 to 10 of 97 entries (filtered from 572 total entries) Show 10 entries

Previous 1 2 3 4 5 … 10 Next Search:

Fig. 7.14 MarsaGram extracting the property of precedence of adjective constructions with *amod* dependency.

- The properties of linearity, co-occurrence, exclusion, and uniqueness have been *automatically* extracted by MarsaGram. However, particular care needs to be taken with the *exclusion* property (or it should be disregarded) since it seems that the algorithm over induces *exclusion* regarding a category for every other category which does not appear in a construction. MarsaGram makes it possible to check every property extracted in the context of the real sentence as shown in Figure 7.15.

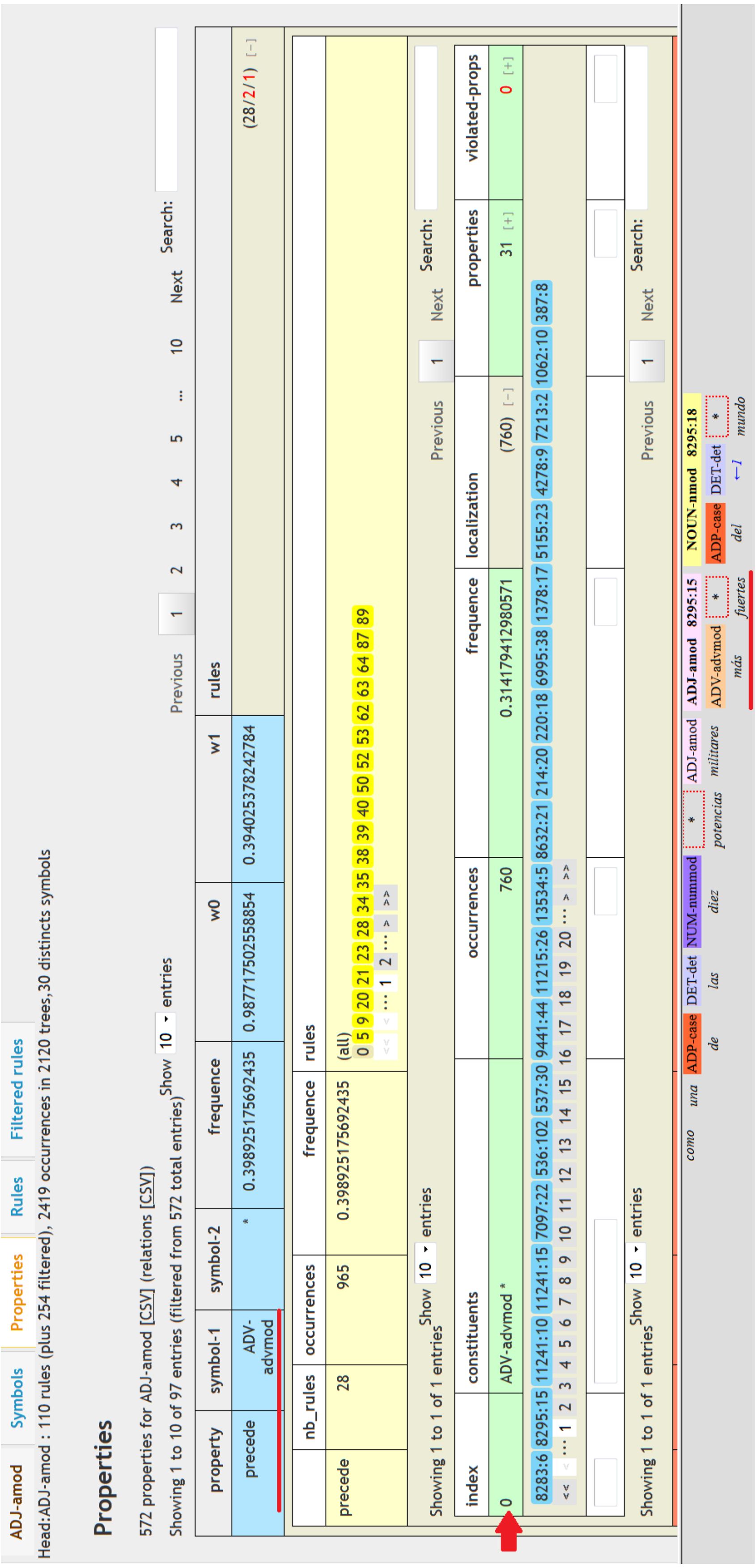


Fig. 7.15 Example of evaluating properties with its real sentence.

- In some cases, universal dependencies get the wrong tag, such as in Figure 7.16. For this reason, in general, it is better to *always* review the properties for each specific construction, its dependencies, and the actual sentence altogether, without implicitly trusting the automatically extracted ones. Therefore, some properties need to be justified with additional theoretical reasons rather than just frequencies.

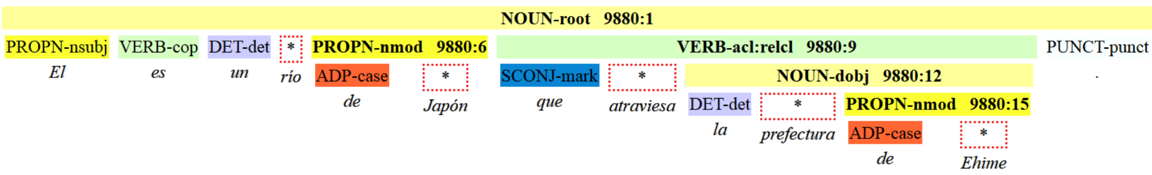


Fig. 7.16 Example of a wrong tag. *"El"* (a determiner) tagged as a proper noun.

- Universal dependencies select as a phrase something that is not in very few cases. As an example, MarsaGram finds the rule $^{*}ADJ \prec ADV$ because of Figure 7.17, where *"maya aún"* is clearly not a clause.

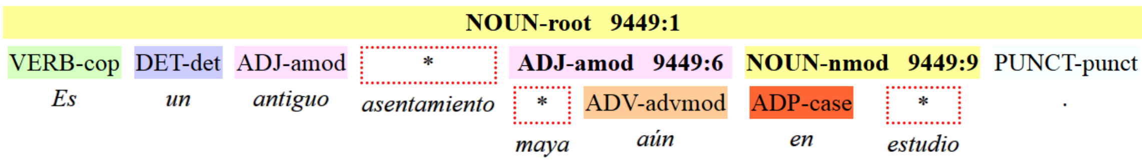


Fig. 7.17 Example of a wrong parsing in UD.

- It is difficult to extract rules or properties for single elements with MarsaGram and UD. For example, if we want to check a PRON (alone) as a subject such as ***"Este** es mi cuarto"*, we can't do it checking PRON-nsubj because *"Este"*, a *PRON* (alone) as a subject, is not a clause. In order to check so, we have to do it manually. In this case, we would need to check a *PRON* (alone) as a subject on the rules and properties extracted from the *Verb* as *root* of a sentence. This is illustrated in Figure 7.18. These limitations of MarsaGram forces us to extract properties through category, not words. However, we provide an approach based on syntactic generalities, which is based on the notion of syntactic category, without going deeper on each word details. Therefore, MarsaGram is the perfect tool for our task. Thanks to the extraction of properties regarding syntactic categories, it is possible to represent linguistic variability and its grammaticality taking into account the different properties displayed by each category according to its different fits.

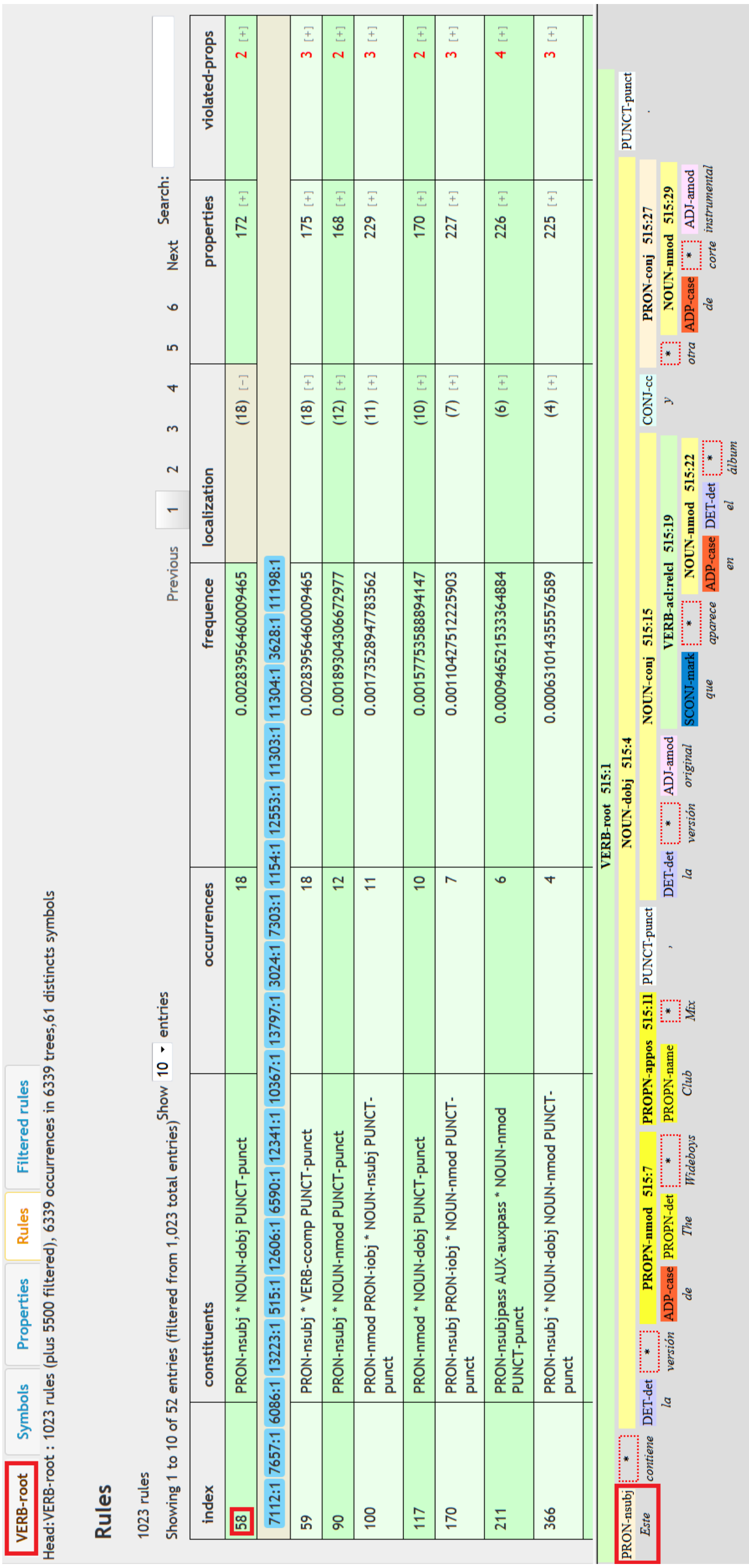


Fig. 7.18 Example of a non-intuitive extraction.

7.3 Using Occurrences for Extracting Constraints

In our work, we used occurrences for extracting linguistic constraints and clarifying how certain typical part of speech can perform as another type of linguistic category. However, *occurrences and its recurrence are a trail* while constraints are justified by a mixture of theoretical reasons such as *markedness* and *context effects*. The theoretical traits can be observed as the properties that allow us to classify different structures as canonical or intermediate (less grammatical but still grammatical) constructions. It is not a matter that a given structure can possibly occur; it is a matter of that a structure either intermediate or canonical shares specific linguistic features that relate them.

Generally, linguistic occurrences allow us to justify if a set of constraints of a structure are belonging to a specific grammar. A canonical constraint belongs to a grammar because it behaves by being displayed (occurring) in a recurrent amount of times given a construction.

A historical example can be found in Spanish and Italian. Proto-Spanish speakers decided to use Latin in its accusative case, while proto-Italian speakers decided to use Latin in its nominative case. That is why Spanish typically does plurals with *-s*, and Italian typically does plurals with *-i*. These phenomena might happen primarily as a random fact, and then it started to be quite probable that someone did plurals in either way, until it was truth that each language did plurals in one way or the other. The occurrence of use was a reflection of that each constraint is belonging to each grammar.

Another clear example is the fact that adjectives precede nouns in English. We can be sure that this precedence is a constraint because in English the adjective always occurs before the noun. However in Spanish, adjective's precedence is not that clear, since the noun can also precede the adjective, so it can be either way such as: *"El fuego rojo"* (*the fire red*) and *"El rojo fuego"* (*the red fire*). We can compare the behavior of the linguistic elements in other constructions, the displayed traits in these constructions, and its occurrences in a representative corpus as a guide to assure which construction is more canonical than the other and why $ADJ \prec N$ is still grammatical.

However, frequencies cannot always model neither grammaticality or a grammar. As an example, Spanish UD does not illustrate examples of personal pronouns being used as subjects (almost not at all). If we were building our model being guided by frequencies, we would have to admit that, in Spanish, using a personal pronoun as a subject is non-grammatical. However, no native Spanish speaker would consider such a thing. Sometimes, working with a corpus taking into account frequencies presents such paradoxes. These reasons show why frequencies are a tool for the linguist who combines such data with theoretical facts. Such paradoxes are one of the main reasons why we cannot model grammaticality under the spectrum of frequencies. Grammaticality needs theoretical reasons plus objective data in order to be modeled with weights.

In summary, the fuzzy degree of grammaticality that we display does not rely on occurrences neither frequencies; it relies on theoretical reasons supported by occurrences. These theoretical reasons are always looking for representing the behavior of a linguistic input in a specific grammar. Because of combining tools with frequencies and theoretical reasons, the relation between an input and a grammar can be modeled. A linguistic input is vague since it has some constraints which we do not know if they have been satisfied or violated. Once we *defuzzify* a linguistic input, we are going to acknowledge how true is the belonging of that input to a grammar; therefore, how grammatical a specific linguistic input is regarding a grammar.

7.4 Considerations before the Extraction of a Spanish PG

The process of extraction is based on the following facts:

- This method aims to extract syntactic properties of Spanish language (standard register, Castilian dialect) through constructions.

- The extracted grammar aims to *define the general cases of Spanish syntax*. It pursues to give a satisfactory explanation for representing degrees of grammaticality in terms of properties of the Spanish language.

- Involving semantic properties is avoided. This extraction is focused on collecting *just* syntactic properties. It is acknowledged that this is a limitation since semantic phenomena completely determine some syntactic phenomena. We dismiss those cases which cannot be (apparently) solved through syntactic properties. We neither try to model particular cases regarding each word and special cases. The extraction has been made in terms of syntactic categories.

- UD tags one object differently. For example, a word such as *"este"* can be tagged as *NOUN*, *ADJ*, *PRON*, *DET*... It depends on the context. In order to find the borderline cases for each category, we treat all words discretely. Once the canonical category for a word is found, we can apply on it such notions as *syntactic fit*, *coercion* or *xCategory*. Therefore, we want to extract *one category for each word*. The fuzzy membership of a word is explained in terms of both features and properties once there is enough data to acknowledge a *xCategory*.

- A word is understood as an entry in the lexicon. UD tags every element, i.e. *"al igual que"* is tagged as *ADP DET ADJ CONJ*. Otherwise, most of the Spanish grammatical approaches consider this structure as one word; one entry in the lexicon, a *CONJ*. During the extraction, most of the *locutions* are also considered as one category. Multi-word expressions and compound words are treated as one single category.

- We know that a case is part of a variability constraint because of three main reasons: (1) it displays a special structure which has some invariable parts; (2) the new structure displays a new meaning which is different from its canonical; (3) this structure is frequent but never as frequent as its canonical. Therefore, a variability constraint is usually depending on both triggering new meanings without the need of creating new words for a language and recurrence.

- The extracted grammar is a proposal for a future application of the extracted properties towards a fuzzy grammar; providing a formal model through the evaluation of degrees of grammaticality.

- The extracted grammar is not pursuing to give an exhaustive explanation for **all** the cases of Spanish syntax nor of particular constructions. The extracted grammar is a proof-of-concept: (1) to illustrate the potential of a Property Grammar for defining the Spanish language; and (2) to illustrate the viability of a fuzzy grammar with a Property Grammar.

- Even though UD displays many dependencies, we are focused on describing our PG with the dependencies proposed by this theory (Table 7.1).

- We follow the notions exposed in chapter 5 and chapter 6. We specify each category through features and their properties. We follow Table 6.4 for defining properties taking into account both categories and constructions. We enumerate the properties according to Table 6.3.

- It is essential to be aware of the use of α, β, γ: It assigns to each property a behavior and a number regarding the category in a construction. α designs canonical properties, β designs violated ones, γ defines the variability properties.

- It is worth to specify that we will use $\wedge$. This symbol is understood as *and*. It allows defining a category and its property concerning many different categories (or features) at the same time. Therefore, all the elements must be satisfied, or it will trigger a violation.

- It is worth to specify that we will use $\vee$. This symbol is understood as *or*. It allows defining a category and its property concerning many different categories (or features) at the same time. One of the elements regarding $\vee$ must satisfy the specified property, or a violation will be triggered.

7.5 First Steps with the Corpus: The Situation

The first clue given by the corpus is the frequency of appearance of each of the categories in Table 7.2.

POS	NB RULES / FILTERED RULES	PROPERTIES	OCURRENCIES	CORPUS FREQUENCY
NOUN	1769 (+1070)	177	77.925	18.09%
ADP –Adposition/Preposition–	26 (+40)	86	70.738	16.42%
DET –Determiner–	9 (+27)	45	60.465	14.04%
PUNTC –Punctuation–	1 (+6)	0	47.448	11.01%
VERB	2437 (+6387)	157	40.950	9.51%
PROPN –Proper Noun–	670 (+1467)	140	40.506	9.40%
ADJ –Adjective–	358 (+1070))	166	23.891	5.55%
CONJ –Coordinating conjunction–	16 (+24)	52	13.787	3.20%
PRON –Pronoun–	146 (+351)	118	13.552	3.15%
ADV –Adverb–	72 (+124)	117	12.510	2.90%
NUM –Numeral–	116 (+211)	116	11.834	2.75%
SCONJ –Subordinating conjunction–	16 (+6)	67	8059	1.87%
AUX –Auxiliary–	15 (+32)	42	6033	1.40%
X –Non-classified–	94 (+263)	114	1952	0.45%
SYM –Symbol–	44 (+73)	133	1077	0.25%
PART –Particle–	1 (+4)	2	37	0.01%

Table 7.2 Frequency of occurrence of categories in the corpus.

According to the corpus, the NOUN is the most dominating category in terms of occurrences. *ADP* and DET follows its domination in terms of occurrences. Given the usual connection between these categories, we can think that in most situations in which a DET and a *ADP* appear, they introduce a NOUN. Therefore, the number of occurrences of a NOUN should be equal to or greater than the number of occurrences of these two elements. The corpus fulfills this logic. The verb is the fourth most frequent element.

These four elements make up 58.06% of the corpus. Consequently, the $NOUN$ is the most present element in Spanish language. In short, these four elements, $NOUN$, DET, *ADP*, and $VERB$, perform half of the Spanish grammar.

Otherwise, all the nominal elements, *NOUN*, *PROPN*, and *PRON* represent the 30.64% of the corpus. In such terms, extracting the properties of the nominal elements, especially the *NOUN*, is a priority since they are the most frequent elements.

On the other hand, each category appears much more frequently regarding some dependencies over the others. In such a way, we could say that each category is more involved in some dependencies over others because those most frequent ones are the canonical dependencies of each category.

7.6 Why an Extraction of the Subject Construction?

Categories and Dependencies.									
subj		*dobj*		*iobj*		*mod*		*root*	
CAT	FREC %	CAT	FREC %	CAT	FREC %	CAT	FREC %	CAT	FREC %
NOUN	2.32%	NOUN	2.49%	PRON	1.54%	ADP+NOUN	14.70%	VERB	2.75%
PROPN	0.92%	PRON	0.45%	NOUN	0.13%	ADJ	4.64%	NOUN	0.61%
PRON	0.35%	PROPN	0.18%	PROPN	0.05%	NUM	2.48%	ADJ	0.24%
NUM	0.02%	NUM	0.01%			ADV	2.17%	PRON	0.06%
						PROPN	0.62%	PROPN	0.04%
						NOUN	0.03%	NUM	0.01%

Table 7.3 Categories and dependencies.

In Table 7.3, we can recognize which are the most prototypical categories for each dependency.

The nominal elements are clearly present in most of the dependencies such as in *subj*, *dobj*, *iobj*. However, the *VERB* stands out for having the highest frequency as *root*. As expected, the verb is the element which articulates many other constructions around it. The *VERB* has a close relationship with the subject construction which is, at the same time, one of the most interesting ones regarding the nominal elements.

Therefore, if we associate dependencies and constructions, we have a clue of those categories which are going to have a critical role for the grammaticality of each construction:

- The categories of the *NOUN*, *PROPN*, and *NOUN* are the ones which most perform the categories of subject construction, direct object construction, indirect object construction.
- The *ADJ*, *ADV*, and *NOUN* (with a preposition) are the categories which mostly introduce a modifier construction.

- The *VERB* is essential for representing verbal constructions, together with all those other constructions that have requirement relations with it: subject, direct object, and indirect object constructions.

Nevertheless, from all these constructions, the subject construction in Spanish is probably the one where we can find much more linguistic interrelations. A subject construction requires a nominal element, which usually requires a determiner. At the same time, a noun in subject construction can have other nominal modifiers introduced by a preposition. It can also have adjective modifiers. Additionally, it bears the most significant dependency towards the verb. We can find verbs without the need of a direct object, but it is less likely to find verbs without requiring a subject construction. Besides, the subject construction can be complemented by both a conjunctive construction or a subordinate one. Moreover, regarding future work, the properties of the Spanish subject construction are the perfect base for a complete Fuzzy Property Grammars, since it would be straightforward to adapt the nominal properties from the subject to other nominal constructions such as direct or indirect object. Therefore, the subject construction is both the most representative construction of Spanish language and the less isolated one.

For these theoretical reasons, we have extracted a Spanish Property Grammar to define the subject construction in Spanish. The subject construction bears many property relations, and it is the perfect proof-of-concept of the potential of the system that we propose in this work: the combination of both a fuzzy grammar and a Property Grammar for Spanish language.

The subject construction is nominal dependent as shown in Table 7.4. We acknowledge this by checking the Spanish UD corpus which reveals that just the *NOUN*, *PROPN*, and *PRON* can be categories for a subject construction. The category of *NUM* with *nsubj* dependency is considered as a part of the *NOUN*. The numerals in Spanish, which are found with a subject dependency, are mostly ordinal numerals and those considered as nouns such as *"segunda, primero"*.

Subject Construction		
Category	Frequency in Corpus	Frequency as Subject
NOUN	2.32%	64.62%
PROPN	0.92%	25,62%
PRON	0.35%	9.47%

Table 7.4 Frequency within subject construction.

If we compare the three nominal elements in Table 7.4, it can be seen that the category of the *NOUN* mostly performs the subject. *PROPN* and *PRON* can introduce a subject construction but with a lower frequency.

Nevertheless, the Subject construction is closely related to the properties of the *VERB*. Therefore, we need first to describe the *VERB* in terms of properties.

Once this is accomplished, we will extract the properties from the categories which are mandatory for a subject construction in Spanish: *NOUN*, *PROPN*, and *PRON*.

As a consequence of these descriptions, other categories will be partially described *DET*, *ADJ*, and *ADP*.

In summary, it would be expected that the properties from the subject construction are not going to be very much different from those which will be displayed in other nominal constructions. Therefore, the extraction of the subject construction is an excellent choice for a proof-of-concept since many other constructions are nominal-dependent such as direct object construction, indirect object construction, and modifier construction with nominal categories.

7.7 A Property Grammar for the *VERB*

We check the verb in order to look for the most representative structures of the Spanish language. The verb is a category which is known for having arguments. An argument is understood as those structures that a verb requires in order to create a sentence with complete sense. In our description concerning the properties of the verb, we specify some of the complements of the verb and its relations in the same description of the verb.

Some of these properties would be better placed in other descriptions of constructions. However, in this work, we include them in the description of the verb since we will not explain in depth the verbal complement constructions in our proof-of-concept of a Fuzzy Property Grammar.

First of all, we check the structures in $VERB - root$. The verbal category is the most frequent one regarding this dependency. By checking this category with this dependency, we extract the most frequent sentence structure in Spanish, as shown in Table 7.5.

In Table 7.5, we can see the ten most frequent sentence structures of Spanish language. The peculiarities of a verb mostly determine these structures, that is if the verb is transitive, intransitive, copulative, passive, or impersonal. We will extract the properties for each verb concerning its idiosyncrasy.

INDEX	RULES	OCCURRENCES	FREQ
0	NOUN-nsubj * NOUN-dobj PUNCT-punct	238	0.0375
1	NOUN-nsubj * NOUN-nmod PUNCT-punct	185	0.0291
2	NOUN-nmod * NOUN-dobj PUNCT-punct	178	0.0280
3	* NOUN-dobj PUNCT-punct	163	0.0257
4	NOUN-nsubj PRON-iobj * NOUN-nmod PUNCT-punct	114	0.0179
5	NOUN-nsubjpass AUX-auxpass * NOUN-nmod PUNCT-punct	76	0.0119
6	NOUN-nsubj * VERB-ccomp PUNCT-punct	75	0.0118
7	* NOUN-nmod PUNCT-punct	74	0.0116
8	NOUN-nmod PRON-iobj * NOUN-nsubj PUNCT-punct	67	0.0105
9	PRON-iobj * NOUN-nmod PUNCT-punct	67	0.0105
10	NOUN-nmod NOUN-nsubj * NOUN-dobj	63	0.0099

Table 7.5 Most frequent constructions in $VERB - root$.

7.7.1 Transitive Verb

Index 0 indicates the most frequent verbal structure in Spanish language: *Subject-Verb-Object*. The verbs which perform these structures are transitive. We cannot check the type of verbs automatically; therefore, we had to navigate through the sentence manually for acknowledging the type of verbs that we find in this structure.

MarsaGram deduces that a verb requires both a noun as subject, and a noun as a direct object: $V \Rightarrow N_{[subj]}$; $V \Rightarrow N_{[dobj]}$. Besides, MarsaGram points out this first property is respected 2823 times and violated 2 times; while the property of requirement between the verb and the direct object is respected 2303 times and violated 13 times. Regarding this case, we found already one of those paradoxes when we work with frequencies. Other categories perform this requirement as well, but in a lower frequency, as shown in Table 7.6.

Properties of $VERB - root$	Satisfied/Violated times			
$V \Rightarrow X_{[subj]}$	$N_{[subj]}$ 2823s/2v	$PROPN_{[subj]}$ 748s/2v	$PRON_{[subj]}$ 114s/3v	
$V \Rightarrow X_{[dobj]}$	$N_{[dobj]}$ 2303s/13v	$PROPN_{[dobj]}$ 81s/0v	$PRON_{[dobj]}$ 69s/0v	$V_{[inf][dobj]}$ 201s/2v

Table 7.6 Requirements of the $VERB - root$.

Regarding requirement of transitive verbs, index 6 takes into account subordinate clauses (*SubC*) with a $dobj$ function: "*El aceite de motor hace que la gasolina se adhiera a cualquier superficie*" (4445:1)[†)]. In this manner, a subordinate clause can satisfy the requirement of the verb: $V \Rightarrow SubC_{[dobj]}$.

MarsaGram extracts as well the linearity property. The transitive verb prefers that the subject precedes the verb: $N_{subj} \prec V_{[trans]}$. The precedence of the subject before the verb is

[†)] These numbers concern to the position of the sentence in the UD treebank corpus.

satisfied 2766 times and violated 57 in the whole $VERB-root$. Besides, it shares with the transitive the uniqueness of the subject $Uniq$: $X_{[subj]}$, and the precedence of the $PRON_{[iobj]}$.

The verb have to precede either a noun in direct object or a proper noun with this property: $V_{[trans]} \prec N_{[dobj]} \vee PROPN_{[dobj]}$. In index 19, a verb in infinitive can be a direct object too, which is going to be preceded by the transitive verb: $V_{[trans]} \prec V_{[inf][dobj]}$: "*La nueva normativa permite disfrutar de este servicio*" (9321:1).

In contrast, the pronoun as a direct object will precede the verb: $PRON_{[dobj]} \prec V$. We do not specify the type of pronoun because we assume that the morpho-lexical part of the grammar will apply rules which restrict the pronouns that can perform a $dobj$ before our properties are checked in syntax. The pronouns which can generally perform a $dobj$ in Spanish are *"lo, la, los, las"* y *"se"* in some cases (*reflexive* and *reciprocal*). The type of pronoun which should be used would be determined by semantic reasons taking into account the referents. For example in *"He ayudado a Juan"* the pronoun would be *"lo"* because *"Juan"* is a *he*: *"Lo he ayudado"*. On the other hand, in *"he ayudado a Ana"*, the pronoun would be *"la"* because "Ana" is a *she*: *"la he ayudado"*.

We observed that the pronoun as an indirect object has to precede the transitive verb: $PRON_{[iobj]} \prec V_{[trans]}$. In index 4, almost all the $PRON_{[iobj]}$ are *"se"*: "*Su actividad se desarrolla en el verano*" (10477:1). The relation between both pronouns with $dobj$ and $iobj$ is going to be clarified while extracting the properties of the pronoun. This reason is why we do not specify more properties regarding $iobj$.

MarsaGram deduces precedence between the verb and a coordinate conjunction structure: $V \prec CONJ$. The construction in which a verb precedes a coordinate construction is ranked with index 24, and it occurs 31 times. There are more ranked structures with *CONJ*, but these are not very frequent in comparison with others. Although, this relation is possible and we define it.

The transitive verbs demand the property of uniqueness for both subjects and direct objects: $Uniq$: $X_{[subj]}$; $Uniq$: $X_{[dobj]}$. We do not specify the categories for uniqueness since the most important thing is the no repetition of these both functions and constructions.

We do not specify dependency relations between elements because we assume dependency from elements such as $subj$ and $dobj$ towards the verb construction, and not the other way around. With the property of requirement, we balance the needs and the hierarchy between arguments and verb.

To summarize, the properties of a transitive verb are described below.

Verb in Verbal Construction

$Verb_{[trans]}$

* α_1: $V_{[trans]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
* α_2: $V_{[trans]} \Rightarrow N_{[dobj]} \vee PROPN_{[dobj]} \vee PRON_{[dobj]} \vee V_{[inf][dobj]} \vee SubC_{[dobj]}$
* α_3: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[trans]}$
* α_4: $V_{[trans]} \prec N_{[dobj]} \vee PROPN_{[dobj]} \vee PRON_{[dobj]} \vee V_{[inf][dobj]} \vee SubC_{[dobj]}$
* α_5: $PRON_{[dobj]} \prec V_{[trans]}$
* α_6: $PRON_{[iobj]} \prec V_{[trans]}$
* α_7: $V_{[trans]} \prec CONJ_{[conj]}$
* α_8: $Uniq: X_{[subj]}$
* α_9: $Uniq: X_{[dobj]}$

Thanks to the use of $\vee$, we can specify better the relations between part-of-speech and its properties. For example, the requirement of a transitive verb would not trigger a violation if this requirement is either satisfied by a noun or a proper noun or a pronoun.

7.7.2 Intransitive and Copulative Verbs

Regarding the Table 7.5, the structure from index 1 represents the structure of intransitive and copulative verbs.

The intransitive verb does not require a direct object. Therefore, it excludes such construction and categories with this function like $PRON_{[dobj]}$: $V_{[intrans]} \otimes X_{[dobj]}$.

The intransitive verb requires a subject, usually a noun. Despite of the fact of this preference, the same categories which can be a subject for the transitive verb can be a subject for the intransitive verb: $V_{[intrans]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$.

Similarly, the intransitive verb prefers that the subject precedes the verb: $N_{subj} \prec V_{[intrans]}$. This precedence is satisfied 2766 times and violated 57 in the whole $VERB-root$. Besides, it shares with the transitive verb the uniqueness of the subject $Uniq$: $X_{[subj]}$, and the precedence of the $PRON_{[iobj]}$.

In index 7, we can see examples like: "*Nació en un pueblecito llamado Canabal*" (5642:1). This precedence, $V_{[intrans]} \prec X_{[mod]}$, is satisfied 3249 times in index 7 and violated 49 times. We assume that the intransitive verb precedes its modifier.

The copulative verb demands as well the precedence of the subject. Similarly to the transitive and the intransitive verb: $X_{[subj]} \prec V_{[cop]}$. The copulative verb has a modifier as an argument rather than a direct object. We can find many of these sentences in index 1, in which a noun introduced by a preposition is performing the modifier required by the copulative verb: "*La densidad de población era de 55,06 hab/km²*" (13936:1).

In $VERB-root$, this modifier is a noun introduced by a preposition. However, in $ADJ-root$ and $NOUN-root$, we find copulative verbs which depend on either an adjective or a noun which is the head of the sentence under the Spanish UD regard: $ADJ-root$: "*El trato de las empleadas es bueno por lo general*" (8604:1); $NOUN-root$: "*El nuevo Ford B-MAX es un monovolumen de pequeño tamaño*" (2956:1). A pronoun and a proper noun can also be a modifier which satisfies the requirement of a copulative verb. Both categories are found in Spanish UD as a *root* of a copulative verb: $PRON-root$: "*Nicolet fue uno de los pocos franceses que permanecieron en Canadá*" (7729:1); $PROPN-root$: "*La ciudad más cercana a Ist es Zadar*" (129:1). The frequency of these latter cases is very low: 0.06% in the case of the *PRON* and 0.04% in the case of the *PROPN*. However, we accept it since they do not trigger any awkward expression.

Therefore, we specify that the copulative verb requires a modifier, either a nominal element or an adjective: $V_{[cop]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[PRON]} \vee ADJ_{[mod]}$.

The copulative verb precedes the modifier frequently. In $ADJ-root$, this precedence occurs 483 times and it is never violated. In $NOUN-root$, it occurs 1095 times and it is violated 109 times. These 109 violations occur because UD assumes that the root is the subject when the copulative has a subordinate clause introduced by a subordinate conjunction or an infinitive as a modifier: "*La diferencia es que estos cobran grandes dietas*" (8916:1); "*El trato es importar vehículos*" (2837:1). In such manner, we can recognize two properties:

1. The copulative verb precedes its required modifier: $V_{[cop]} \prec X_{[mod]}$.
2. A subordinate construction can satisfy as well the requirement of the copulative verb: $V_{[cop]} \Rightarrow N_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$.

Because of the semantic idiosyncrasy of the copulative verb, which barely exhibits any meaning, we specify that the copulative verb is dependent of its modifier: $V_{[cop]} \leadsto_{dep} X_{[mod]}$.

The copulative verbs display the same behavior as the transitive regarding the properties of requirement and uniqueness towards a subject. Nevertheless, we have found an exception. When the copulative verb has an adjective as a modifier, and a subordinate clause as subject, both the copulative verb and the adjective precede the subordinate clause as a subject. This structure has an index 2 in $ADJ-root$: "*Es evidente que no se cumplen las normativas sanitarias vigentes*" (14064:1). Therefore: $V_{[cop]} \wedge ADJ_{[mod]} \prec SubC_{[subj]}$. This subject structure for the copulative verb excludes the possibility of having a subordinate clause as a modifier of the copulative: $SubC_{[subj]} \otimes SubC_{[mod]}$.

Below, we present the Spanish properties for intransitive and copulative verbs in verb construction.

Verb in Verb Construction

$Verb_{[intrans]}$

- α_{10}: $V_{[intrans]} \otimes X_{[dobj]}$
- α_{11}: $V_{[intrans]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
- α_{12}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[intrans]}$
- α_{13}: $V_{[intrans]} \prec X_{[mod]}$
- α_{14}: $PRON_{[iobj]} \prec V_{[intrans]}$
- α_{15}: $Uniq$: $X_{[subj]}$

$Verb_{[cop]}$

- α_{16}: $V_{[cop]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \vee SubC_{[subj]}$
- α_{17}: $V_{[cop]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
- α_{18}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[cop]}$
- α_{19}: $V_{[cop]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
- α_{20}: $V_{[cop]} \wedge ADJ_{[mod]} \prec SubC_{[subj]}$.
- α_{21}: $SubC_{[subj]} \otimes SubC_{[mod]}$
- α_{22}: $V_{[cop]} \rightsquigarrow_{dep} X_{[mod]}$.
- α_{23}: $Uniq: X_{[subj]}$

7.7.3 Verbs in Passive

The passive voice is found in index 5. The verbs which can perform a passive structure are transitive. For this reason, the verbs in passive share α_1, α_3, α_6 with the transitive verbs. We specify a verb in passive voice with its auxiliary particles with the following feature: $V_{[pass]}$.

Firstly, we acknowledge the requirement of an agent introduced by "*por*". A verb in passive voice occurs, in index 5, 76 times; displaying prepositional clauses with different kinds of preposition and, specially, with the preposition *"por"*; which typically introduces an agent for the passive voice in Spanish: "*sus otros fueron adquiridos por famosos personajes*" (1372:1). Therefore: $V_{[pass]} \Rightarrow N_{[mod:por]} \vee PROPN_{[mod:por]} \vee PRON[mod:por]$. The verb in passive precedes these modifier constructions.

However, there are many cases in which the verb in passive still has a nominal structure introduced by a preposition which is not *"por"*. For this reason, we introduce a variability property in which in case the verb in passive is not satisfied by its initial requirement, a second requirement of a nominal structure introduced by preposition is demanded: $\neg(V_{[pass]} \Rightarrow N_{[mod:por]} \vee PROPN_{[mod:por]} \vee PRON[mod:por]) \implies V_{[pass]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON[mod]$. In this manner, we point out that a verb in a passive structure is dependent on that modifier introduced by a preposition in order to have full grammaticality. For example, the sentence "*El veredicto fue emitidio en una corte*" (8495:1), would have a better value of grammaticality if it would be: *"El veredicto fue emitido por un juez en una corte"*. On the other hand, the degree of grammaticality would be slightly worst on *"el veredicto fue emitido"*.

Besides, there should be just one construction with a nominal introduced by the preposition *"por"*: $V_{[pass]}$: $Uniq$: $X_{[mod:por]}$. Consequently, we constrain the presence of two nominal agents for one verb in passive voice. An example of the violation of this case of uniqueness: *"El veredicto fue emitido por un juez por un jurado"*.

Verb in Verb Construction

$Verb_{[passive]}$

- α_{24}: $V_{[pass]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
- α_{25}:$V_{[pass]} \Rightarrow N_{[mod:por]} \vee PROPN_{[mod:por]} \vee PRON_{[mod:por]}$
- α_{26}: $V_{[pass]} \rightsquigarrow_{mod} N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
- α_{27}: $Uniq$: $X_{[mod:por]}$
- β_{25}: $\neg(V_{[pass]} \Rightarrow N_{[mod:por]} \vee PROPN_{[mod:por]} \vee PRON_{[mod:por]}) \implies V_{[pass]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$

On the other hand, we find in Spanish UD structures with verbs without a subject. In index 8 and 9, we have sentences with the use of the *"se"* as a part of the verb. The use of this particle allows to Spanish speakers to display much variability in their expressions. Spanish idiosyncrasy especially allows doing passives or impersonal sentences. In this work, we do not study this kind of variability deeply. We just propose a couple of properties in order to differentiate two uses of this verb: the passive use known as *pasiva refleja* ($V_{[se:auxpass]}$) and the use of impersonal voice with *"se"*, *"se" impersonal* ($V_{[se:auximp]}$).

Verb in Verb Construction

$Verb_{[se:auxpass]}$

- α_{28}: $V_{[se:auxpass]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
- α_{29}: $V_{[se:auxpass]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
- α_{30}: $V_{[se:auxpass]} UniqX_{[subj]}$

$Verb_{[se:auximp]}$

- α_{31}: $V_{[se:auximp]} \prec X_{[mod]}$
- α_{32}: $V_{[se:auximp]} \otimes X_{[subj]}$

Other uses of *"se"* would be determined depending on if they are a direct or an indirect object. Universal dependencies have severe problems in order to tag this particle, and it usually tags it as $iobj$.

Another use of an impersonal sentence in Spanish is the use of verbs in the third person with both transitive and intransitive verbs. Index 7 is an example of this: "*Abren todos los mediodías y viernes y sábado por la noche*" (8331:1); "*Huele a leche de montaña*" (1397:1). These sentences can allow both subject and a direct object according to the features of their verb:

- *"Los dueños abren la tienda todos los mediodías y viernes y sábado por la noche"*
- *"El queso huele a leche de montaña"*

However, even though syntactically these last structures would be more canonical, the pragmatical and semantic traits for transmitting an intentional impersonal feature to the verb would be erased in case the full sentence would be produced. Therefore, the omission or use of the subject is determined in favor of the intended meaning of a sentence.

In order to express generality in Spanish, it is necessary to remove the subject. We cannot represent this characteristic of the Spanish language just by the syntactic level. Such pragmatic considerations are not going to be formalized here; although, these are necessary for a multi-modal grammar.

In this work, since we deal with syntax, we propose a variable rule for all those verbs (impersonal or not) that do not have the requirement of a subject. That is if the requirement between a verb and a subject is not satisfied, the subject of the verb would be understood from the verbal suffix which displays the person and the number of the verb: $\neg(V \Rightarrow X_{[subj]}) \implies X_{[subj]}\ in\ V_{[morph:pers:num]}$. The canonical properties which would trigger this variability property if it would be violated are: α_1, α_{11}, α_{16}, α_{29}.

In such manner, with this rule, we represent the frequent use of a verb which does not have an explicit subject in Spanish language. Besides, we highlight that this is a violation; however, it does not trigger a low level of grammaticality. This trait contrasts the treatment of degrees of grammaticality within a property grammar with fuzzy properties with a discrete grammar which would need to say that just one or the other is grammatical.

Therefore, we represent with this variable property the fact that a grammar for the Spanish language admits phrases without an explicit subject which are very natural and frequent among the speakers of this language.

7.7.4 Variability Properties in Verbal Construction

The most relevant variability property regarding the verb construction is found in:

$$\neg(V \Rightarrow X_{[subj]}) \implies X_{[subj]}\ in\ V_{[morph:pers:num]}$$

However, after working with Spanish UD and MarsaGram, we have detected that, generally, when the subject is missing, a nominal construction introduced by a preposition is either preceding or after the verb. In other words, the verb is never alone. Therefore, we introduce a variability property which demands that in case a subject is missing the verb requires a nominal modifier:

$$\neg(X_{[subj]} \Rightarrow V) \implies V \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$$

The typical cases with this variability property in our corpus can be found in the construction with index 7 in the $VERB - root$. In case they would be violated, the properties triggered by this variability property are: α_1, α_{11}, α_{16}, and α_{29}.

We illustrate the application of the variability properties on an example in natural language in Table 7.7.

Case: $V_{[intrans]}$ triggering Variability Properties				
Sentence	Nació	en	un	pueblo
CAT	$V_{[intrans]}$	$ADP_{[spec]}$	$DET_{[spec]}$	$NOUN_{[mod]}$
Properties	α_{10}: $V_{[intrans]} \otimes X_{[dobj]}$ β_{11}: $V_{[intrans]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$ γ_1: $\beta_{11} \implies X_{[subj]}$ *in* $V_{[morph:pers:num]}$ γ_2: $\beta_{11} \implies V \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$ β_{12}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[intrans]}$ α_{13}: $V_{[intrans]} \prec X_{[mod]}$ α_{15}: $Uniq$: $X_{[subj]}$			
Grammaticality	Cn_w=0.2 Vab_w=0.1 VG_w=0.8			

Table 7.7 Example of variability properties in $V_{[intrans]}$.

Table 7.7 display the application of both variability properties γ_1 and γ_2. We did not consider the canonical properties α_{14} because there is no $PRON_{[iobj]}$ for activating this property. Besides, this property would be placed in the construction of $PRON_{[iobj]}$ in our property grammar, since it is considered that it is the *PRON* which should trigger it and not the verb. However, since we are providing a proof-of-concept for the subject construction in Spanish, we included some properties regarding verbal complements in the $VERB$.

The property α_{15} is satisfied because we found a unique subject in γ_1 in the morpheme of the verb *"nació"*.

The value of grammaticality of $V_{[intrans]}$ has been calculated using the formulas presented in section 6.6. Since we cannot be sure about the different weights between categories, we acknowledge the same weight of 1 for all types of words. Therefore, we apply Equation 6.27 to $V_{[intrans]}$.

$$V_{[intrans]} = 1 \tag{7.1}$$

The value of each canonical constraint of $V_{[intrans]}$ (Cn_w) is calculated dividing all the triggered canonical properties both satisfied (α) and violated (β) (5) by our standard value of a canonical property (1). We assume that all canonical properties have the same density, which means that no canonical property is more important than another. By assuming this value of 1, we provide a theoretical value for each canonical constraint strictly from the perspective of the syntactic module, without involving our weights with frequencies, avoiding in such manner the paradoxes of weighting canonical properties by probabilities as seen in subsection 3.9.3 and section 7.3. The canonical value of each property in Table 7.7 is 0.2

$$0.2\, Cn_w = \frac{1}{C_5} \tag{7.2}$$

The violated properties are weighed as 0. We calculate the value of the variability properties triggered as a consequence of the violated canonical properties. The variability properties triggered in $V_{[intrans]}$ are:

- γ_1: $\beta_{11} \implies X_{[subj]}$ *in* $V_{[morph:pers:num]}$: the grammar extracts the subject from the gender and the number of the morpheme *-ó* from the verb *naci-**ó***.

- γ_2: $\beta_{11} \implies V \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$. In γ_2, the operation $\vee$ helps us to avoid over-violation pointing out that just one context is enough in order to satisfy this property. In Table 7.7, γ_2 is satisfied because the $V_{[intrans]}$ can require a $N_{[mod]}$.

If $V_{[intrans]}$ would not satisfy the variability properties, its violations would remain as 0.

In Equation 7.3, the value of a variability property is calculated by dividing the value of a variability property (0.5) by all triggered satisfied (α) and violated (β) constraints in $V_{[intrans]}$ (*nació*).

$$0.1\, Vab_w = \frac{0.5}{C_5} \tag{7.3}$$

We calculate the value of grammaticality of $V_{[intrans]}$ by applying Equation 6.32. For calculating VG_w, we take into account all the Cn_w plus all the Vab_w. In Table 7.7:

- The satisfied properties (α) which keep the value as canonical properties ($0.2\, Cn_w$) are 3 out 5, we calculate All_{Cn_w} as 0.6.

- The satisfied variability properties (γ) which keep the value as variability properties ($0.1\, Vab_w$) are 2 out 2, we calculate All_{Vab_w} as 0.2.

These both values are divided by the value of $V_{[intrans]}$ which is 1. The final value of grammaticality of $V_{[intrans]}$ in Table 7.7 is calculated as 0.8 as showed in Equation 7.4.

We point out that Cn_w and Vab_w are **not** values of grammaticality. These solely provide the value of a constraint relying on both its behavior and its satisfaction or violation. Once all the values for each constraint are calculated the fuzzy grammar can compute degrees of grammaticality regarding all the values of all the triggered constraints. Therefore, the value of grammaticality can *only* be computed by VG_w (regarding a word) or by VG (regarding an input).

$$0.8\, VG_w = \frac{(0.6\, All_{Cn_w} + 0.2\, All_{Vab_w})}{1\, V_{[intrans]}} \tag{7.4}$$

In this manner, we could compute the value of grammaticality of this example by words regarding subsection 6.4.2.

- The value of grammaticality of the word *nació* is 0.8. An *input* is *significantly* satisfied when its value is estimated between 1-0.8.
- *IF* an input is *significantly* satisfied *THEN* the value of grammaticality is *high*.
- *IF* the value of grammaticality is *high THEN* an input is *significantly grammatical*.

Therefore, the word *nació* in *nació en un pueblo* is *significantly grammatical*.

To sum up, we provide in Table 7.8, all the properties of the Verb in Verb Construction, including the variable properties.

$Verb$ in $Verbal\ Construction$
$Verb_{[transitive]}$
α_1: $V_{[trans]\vee[pass]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
α_2: $V_{[trans]} \Rightarrow N_{[dobj]} \vee PROPN_{[dobj]} \vee PRON_{[dobj]} \vee V_{[inf][dobj]} \vee SubC_{[dobj]}$
α_3: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[trans]\vee[pass]}$
α_4: $V_{[trans]} \prec N_{[dobj]} \vee PROPN_{[dobj]} \vee PRON_{[dobj]} \vee V_{[inf][dobj]} \vee SubC_{[dobj]}$
α_5: $PRON_{[dobj]} \prec V_{[trans]}$
α_6: $PRON_{[iobj]} \prec V_{[trans]\vee[pass]}$
α_7: $V_{[trans]} \prec CONJ_{[conj]}$
α_8: $Uniq$:$X_{[subj]}$
α_9: $Uniq$: $X_{[dobj]}$
$Verb_{[intrans]}$
α_{10}: $V_{[intrans]} \otimes X_{[dobj]}$
α_{11}: $V_{[intrans]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
α_{12}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[intrans]}$
α_{13}: $V_{[intrans]} \prec X_{[mod]}$
α_{14}: $PRON_{[iobj]} \prec V_{[intrans]}$
α_{15}: $Uniq$: $X_{[subj]}$
$Verb_{[cop]}$
α_{16}: $V_{[cop]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \vee SubC_{[subj]}$
α_{17}: $V_{[cop]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
α_{18}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[cop]}$
α_{19}: $V_{[cop]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
α_{20}: $V_{[cop]} \wedge ADJ_{[mod]} \prec SubC_{[subj]}$.
α_{21}: $SubC_{[subj]} \otimes SubC_{[mod]}$
α_{22}: $V_{[cop]} \rightsquigarrow_{dep} X_{[mod]}$.
α_{23}: $Uniq$: $X_{[subj]}$
$Verb_{[passive]}$
α_{24}: $V_{[pass]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
α_{25}: $V_{[pass]} \Rightarrow N_{[mod:por]} \vee PROPN_{[mod:por]} \vee PRON_{[mod:por]}$
α_{26}: $V_{[pass]} \rightsquigarrow_{mod} N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
α_{27}: $Uniq$: $X_{[mod:por]}$
$Verb_{[se:auxpass]}$
α_{28}: $V_{[se:auxpass]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
α_{29}: $V_{[se:auxpass]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
α_{30}: $Uniq$: $X_{[subj]}$
$Verb_{[se:auximp]}$
α_{31}: $V_{[se:auximp]} \prec X_{[mod]}$
α_{32}: $V_{[se:auximp]} \otimes X_{[subj]}$
Variability Properties
γ_1: $\beta_{1\vee 11\vee 16\vee 29} \implies X_{[subj]}\ in\ V_{[morph:pers:num]}$
γ_2: $\beta_{1\vee 11\vee 16\vee 29} \implies V \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
γ_3: $\beta_{25} \implies V_{[pass]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$

Table 7.8 Table of Spanish properties of verb in verbal construction.

7.8 A Property Grammar for the $NOUN$

$NOUN\ in\ Subject\ Construction$
α_1: $DET_{[spec]} \prec NOUN_{[subj]}$
α_2: $NOUN_{[subj]} \Rightarrow DET_{[spec]}$
α_3: $Uniq$: $DET_{[spec]}$
α_4: $NOUN_{[subj]} \otimes ADV \wedge PRON \wedge V_{[inf]}$
α_5: $NOUN_{[subj]} \rightsquigarrow_{subj} V$
Variability Properties
γ_1: $\beta_2 \implies NOUN_{[subj]} \Rightarrow NOUN_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
$xNOUN\ in\ Subject\ Construction$
γ_1: $DET_{[spec]} \prec X_{[xNOUN]}$
γ_2: $X_{[xNOUN]} \Rightarrow DET_{[spec]}$
γ_3: $Uniq$: $DET_{[spec]}$
γ_4: $X_{[xNOUN]} \otimes ADV_{[mod]} \wedge PRON \wedge V_{[inf]}$
γ_5: $X_{[xNOUN]} \rightsquigarrow_{subj} V$

Table 7.9 Spanish properties of $NOUN$ in subject construction.

In Table 7.9, we show the properties of the noun in subject construction.

The precedence of $DET \prec NOUN$ occurs in MarsaGram in subject construction 8749 times, and it is never violated. These data is why α_1 represents this linearity between the determiner and the noun in the subject construction. The second property α_2 defines the requirement relation of the $NOUN$ over the DET. This relation is based on the high frequency of co-occurrence of these two elements given by its precedence.

MarsaGram deduces that the determiner cannot be repeated with a noun in subject construction. Under such data, we represent the no repetition of the determiner with α_3 $Uniq : DET$. This property occurs 8678 times in the noun as a subject dependency. It is just violated 71 times.

Before accepting the property of uniqueness in α_4, we have been thinking about some controversies that are present in the determiner concerning the noun. Other which would be considered as determiners too can appear with the noun in the nominal construction. Such elements use to be ordinal numeral adjectives or undefined adjectives (which are known as *artículos indefinidos* in Spanish). An example of this would be:

- Ordinal numeral adjective: "*La primera partida comprende 24 títulos y 516 leyes*" (7092:2).

- *Otros*: "*Su otro hijo cumplía condena*" (127:46). In such case, this word would be considered as an adjective, which modifies the noun. Differently from the majority of adjectives, this adjective can precede the noun.

Similarly, we consider some other controversial determiners such as *"todos"* as adjectives:

- *Todos*: "*Todos los ingresos se asignaron a la corona*" (11255:2). In such case, this word would be considered as an adjective, which precedes the determiner article. Both definite or indefinite. This would be specified in the adjective construction as: $ADJ_{[todos/as]} \prec DET_{[art]}$.

We treat these kinds of elements as adjectives because we want to avoid a theoretical explanation of the subject in Spanish which take into account big determiner constructions. Some grammars consider these adjectives as determiners, together with other undefined adjectives. Our grammar is looking for a nominal based approach, since we have seen, in the Spanish UD corpus, that Spanish is a nominal based language. The *NOUN* is the most frequent category, and it is present in almost all constructions. Following this reasoning, we prefer to fix a balance around the *NOUN* category. We are describing its relations with both adjectives and determiners. We acknowledge the semantic characteristic of these both ordinal and undefined adjectives such as *"otros"* as attributes which complete the meaning of the noun, more as nominal modifiers, than specifiers. However, thanks to our notion of the *xCategory* we can establish different syntactic fits for them when it is necessary.

The property α_4 bans the co-occurrency between the noun and an adverb modifying it. As well, it excludes the possibility of being in co-occurrence with an infinitive or a pronoun. We do not exclude the possibility of being modified by a *PROPN* because there are undoubtedly many cases in which this happens. However, not all nouns can be modified by a *PROPN*: "*Su hija María Helena Martins creó en 1997 el Centro de Estudios de Literatura y Psicoanálisis Cyro Martins*" (1065:2). As commented before, we are not able to extract these semantic particularities right now; it would need methods in distributional semantics. This task is saved for future work.

Finally, we specify that the noun in the subject construction has a subject dependency relation with a verb. The last canonical property is $\alpha 5$ which defines the dependency relation between this construction and a verb. Because these relations have been already described in the verb, we do not do that here. It is the verb which requires if the subject is demanded or not. In such manner, we avoid some paradoxes such as a noun with a satisfied property of requiring a verb, together with an impersonal verb which is excluding the subject.

The noun as a *xCategory* is specified with the same properties as the canonical noun. Any category which would be specified to be in a noun context should satisfy these properties.

The most frequent *xNOUN* is the *PROPN* in a noun *fit*. In our work, the PROPN is deeply involved with this *xCategory*. This situation is going to be represented more precisely in section 7.10.

Other categories rather than a proper noun can perform a *xNOUN*. For example, some adverbs can be an *xNOUN* such as in: *"El pero de la profesora estaba claro"* or *"El no de Pedro Sánchez"*. However, the categories or words which are *xNOUN* should be specifically noted.

It has to be noted that we have not considered the adjectives as a *xNOUN*. The omission of a noun regarding the adjectives has been solved by treating the determiner as a pronoun.

The presented variability property is a consequence of cases in which we find sentences in which neither the co-occurrence and the precedence between the noun and determiner are satisfied; such as in index 12: "*Contribuciones a la fundamentación de una metafísica volitivotáctil (2006) son sus aportaciones más notables*" (1653:2).

Probably, the lack of *DET* is due to some expressive need in order to trigger another meaning. In this case, for example, this noun is not preceded by a determiner because "*Contribuciones a la fundamentación de una metafísica volitivotáctil*" is the name of a publication. In order to provide more specificity to a noun, or for displaying proper noun semantics, the determiner is erased.

However, from the regard of a noun construction, the canonical construction would be: *"Las/Sus contribuciones a la fundamentación de una metafíisica volitivotáctil (2006) son sus aportaciones más notables"*. Since we are focused on the syntactic level, such variability should be marked with a violation. It should be noted that in all the examples where the *DET* does not occur, the noun precedes a modifier. This mark indicates that the noun needs to be supported either by a specifier or a modifier. However, there is no case in which the subject position will be a noun alone. Therefore, the lowest degree of grammaticality should be given with the noun without specifier or modifier. Because of these reasons, we added: γ_1: $\beta_2 \implies NOUN_{[subj]} \Rightarrow NOUN_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} SubC_{[mod]}$. Taking into account that, in case that the requirement of the *DET* is violated, the noun requires a modifier construction.

Case: $NOUN_{[subj]}$ triggering Variability Properties			
Sentence	Funcionarios	del estado	sufrieron las pérdidas
CAT	$NOUN_{[subj]}$	$ADP_{[spec]}$ $DET_{[spec]}$ $NOUN_{[mod]}$	$V_{[trans]}$ $DET_{[spec]}$ $NOUN_{[dobj]}$
Properties	β_1: $DET_{[spec]} \prec NOUN_{[subj]}$ β_2: $NOUN_{[subj]} \Rightarrow DET_{[spec]}$ γ_1: $\beta_2 \implies NOUN_{[subj]} \Rightarrow NOUN_{[mod]}$ α_4: $NOUN_{[subj]} \otimes ADV \wedge PRON \wedge V_{[inf]}$ α_5: $NOUN_{[subj]} \leadsto_{subj} V$		
Grammaticality	Cn_w=0.25 Vab_w=0.125 VG_w=0.625		

Table 7.10 Variability properties of $NOUN_{[subj]}$.

In Table 7.10, we illustrate an example in natural language of the variability property γ_1 in $NOUN_{[subj]}$

- The value of the word *"funcionarios"* as a $NOUN_{[subj]}$ is estimated with the theoretical value of 1: $NOUN_{[subj]}$= 1.
- The value of each canonical property of $NOUN_{[subj]}$ (Cn_w) is calculated dividing all the triggered canonical properties both satisfied (α) or violated (β) (4) by our standard value of a canonical property (1). The canonical value of each property in Table 7.10 is 0.25.
- The value of a variability property (Vab_w) is calculated by dividing the value of a variability property (0.5) by all triggered satisfied (α) and violated (β) constraints in $NOUN_{[subj]}$ (*"funcionarios"*). The variability value of each property in Table 7.10 is 0.125.
- $NOUN_{[subj]}$ cannot either satisfy or violate α_3 because any determiner has appeared. Therefore, our property grammar cannot evaluate its uniqueness in α_3. In this manner, the property is not triggered.
- $NOUN_{[subj]}$ satisfies 2 canonical properties out of 4. We calculate All_{Cn_w} as 0.5.
- $NOUN_{[subj]}$ satisfies 1 variability properties out of 1: γ_1: $\beta_2 \implies NOUN_{[subj]} \Rightarrow NOUN_{[mod]}$. We calculate All_{Vab_w} as 0.125.

$$0.625\, VG_w = \frac{(0.5\, All_{Cn_w} + 0.125\, All_{Vab_w})}{1\, N_{[subj]}} \tag{7.5}$$

If we didn't take into account this variability property, the value of grammaticality of $NOUN_{[subj]}$ would be 0.5.

$$0.5\,VG_w = \frac{(0.5\,All_{Cn_w} + 0\,All_{Vab_w})}{1\,N_{[subj]}} \tag{7.6}$$

In this manner, we could compute the value of grammaticality of this example with words concerning that a value between 0.8-0.5 is understood as *quite grammatical*, and a value of 0.5-0 is understood as *barely grammatical*.

- The value of grammaticality of the word *funcionarios* in subject construction in Equation 7.5 is 0.625. The *input* is *quite grammatical*.
- However, the value of grammaticality of $NOUN_{[subj]}$ in Equation 7.6 is 0.5. The input displays a borderline case between being *quite grammatical* and *barely grammatical*. Because of our Fuzzy Property Grammar took into account such variability property as γ_1 in $NOUN_{[subj]}$, we can provide a more fine-grained value such as the one displayed in Equation 7.5

7.9 Adjective and Nominal Modifiers

$ADJ\ in\ Modifier\ Construction$
α_1: $NOUN_{[subj]} \prec ADJ_{[mod]}$ α_2: $ADJ_{[mod][num:ord]\vee[undef]} \prec NOUN_{[subj]}$ α_3: $ADJ_{[mod]} \rightsquigarrow_{mod} NOUN_{[subj]}$ α_4: $ADJ_{[mod][todos/as]} \prec DET_{[spec][art:def:pl]\vee[poss:pl]\vee[demons:pl]} \rightsquigarrow_{spec} NOUN_{[subj]}$ α_5: $ADJ_{[mod][todo/a]} \prec DET_{[poss:s]\vee[demons:s]} \rightsquigarrow_{spec} NOUN_{[subj]}$ α_6: $ADJ_{[mod][todo/a]} \prec DET_{[spec][art:undef:s]} \rightsquigarrow_{spec} NOUN_{[mod]}$ α_7: $ADV_{[mod][casi]} \rightsquigarrow_{mod} ADJ_{[mod][todo/a/os/as]}$
Variability Properties
γ_1: $\beta_1 \implies ADJ_{[mod][semantics]}$
$xADJ\ in\ Modifier\ Construction$
γ_1: $NOUN_{[subj]} \prec X_{[xADJ]}$ γ_2: $X_{[xADJ]} \rightsquigarrow_{mod} NOUN_{[subj]}$

Table 7.11 Spanish properties of *ADJ* in modifier construction.

The precedence property deduced by MarsaGram between the noun, and the adjective ($NOUN \prec ADJ$) occurs in noun as *nsubj* 1703 times, and it is never violated. However, this property falls in contradiction with rule 8 in *nsubj*: $\{DET - det\ ADJ - amod\ NOUN - nsubj\}$, which occurs 142 times.

The precedence of the noun towards the adjective is assumed as a canonical property. Most of the adjective in rule 8 are adjectives which must precede the noun such as cardinals and undefined adjectives. We can find other kind of adjectives such as *"nuevo"* or *"importante"* which they would be odd in case of being preceded by a noun: *"Este nuevo concepto engloba a todas aquellas personas con una cultura, raza e historia común"* (7043:1) and *"Esta importante recolección genera un importante impacto económico"* (6684:2). For these special cases, it is the lexical domain which will provide the specific canonical properties for each word which needs them.

In rule 8, we find adjectives which trigger different meaning regarding if they precede the noun or not. For example, in *"Una popular historia dice que María Antonieta, reina de Francia, al intentar escapar de Francia y de sus problemas, se embarca hacia Estados Unidos"* (6308:7). *"Una popular historia"* means *A well-known story*, however, *"una historia popular"* means *a story from the common people*. An adjective preceding a noun display different both semantics and form such as the cases of *"gran/grandes"* in (12193:6) and *"buen/buenos"* in (2808:1)).

In this manner, we theoretically assume that the precedence of the adjective towards the noun in Spanish is marked by stressing either the canonical meaning or a new meaning for the adjective:

- *"El talentoso croata podría dar un paso hacia delante por su propósito de cambiar de equipo"* (1166:1); the canonical meaning of *"talentoso"* is stressed.
- *"Los grandes medios cuentan la parte más naïf de twitter"* (12193:1); the adjective *"grandes"* triggers a new meaning when precedes the noun. In this case, it means *the most important/principal media*: *"Los medios más importantes cuentan la parte más naïf de twitter"*.

Therefore, theoretically, and because of rule 8, we display a special property γ_1: $\beta_1 \Longrightarrow ADJ_{[mod][semantics]}$. This variability property is an example of interrelation in our multi-modal grammar. We theoretically assume that the violation of $NOUN \prec ADJ$ is due to semantic reasons. In Spanish, the precedence of the adjective towards the noun in a subject construction stress the semantics of the adjective. Since we did not extract the semantic properties of Spanish, we do not know which ones would be triggered. However, we point out that in case this violation occurs, this violation should satisfy semantic properties probably based on results from distributional semantics analysis. While we were extracting the properties of the subject, we extracted some particular cases in which the adjectives precede the noun.

In particular, the ordinal numerals ($ADJ_{[num:ord]}$) and some undefined adjectives like *otros* ($ADJ_{[undef]}$).

We specify in α_2 that the adjectives which are ordinal numerals such as *"primero, segundo, tercero"*, and other undefined determiners such as *"otros"* have to precede the noun in case they appear in the subject construction.

One of the controversies around the adjective properties is the tag of the adjective. We label an adjective as undefined as a general case, without being too precise since the undefined adjectives need of in-depth research with automatic tools in order to classify them. Probably, with methods in distributional semantics.

It would be interesting to have a look at the type of adjectives and classify them, in order to know if they canonically precede the noun or if they are preceded by it. Since MarsaGram extracts properties through clauses, we cannot check such particularities in this corpus. This task is saved for future work.

The third property specified (α_3) is the dependency between the adjective in a modifier construction and the noun.

The properties α_4, α_5, α_6, and α_7 are related with the use of the word *"todo"*. This word is one of the elements that allow a very long subject in Spanish: *"casi todos los otros chicos apuestos estaban en el otro lado de la clase"*.

"Todo/a/os/as" can appear as a pronoun. However, when this occurs, it excludes its relation towards a noun in the construction of a subject.

"Todo/a/os/as" has been considered an adjective regarding its relations as a modifier of a noun in subject construction. This adjective must precede a determiner which is the specifier of a noun.

"Todo/a/os/as" is an indefinite adjective which presents problems for its classification. This adjective can only have a plural form when appear as a modifier of a noun in subject construction which is specified by either a plural definite article (*"los/las"*) or a plural possessive determiner (*"sus"*) or a plural demonstrative determiner (*"estos/estas"*). These data indicate that when it appears in plural with a noun as subject, it must be given the property α_4.

On the other hand, this adjective can only present a singular form when it is a modifier of a noun in subject construction which is specified by either a singular possessive determiner (*"sus"*) or a singular demonstrative determiner(*"estos/estas"*). These data indicate that when it appears in singular with a noun as subject, it must be given the property α_5.

Besides, the adjective *"todo/a"* presents a singular form when is a modifier of a noun in a modifier construction which is specified by an indeterminate singular article (*"un/una"*):

"Juan es todo un caballero". These data indicate that when it appears in singular with a noun as a modifier, it must be given the property α_6.

In the UD dependency corpus, *"Todo/a/os/as"* just accepts one adverbial modifier *"casi"*. Therefore, we specify in $\alpha 7$ that the adverb *"casi"* can be dependent on this adjective. In case this adverb would not appear, it would not trigger a violation.

These properties α_4, α_5, α_6 and α_7 has been extracted mostly theoretical when working with UD dependencies. The extraction of the rules by category from MarsaGram display difficulties to examine the relation of particular words such as examining all the cases regarding *"todo/a/os/as"* independently as a word which precedes a nominal construction. *"Todo/a/os/as"* presents quite a lot of variability regarding precedence and requirements for this word. We provide some order concerning this word with our properties α_4, α_5, α_6; however, this word demands a more in-depth analysis of its properties through tools specialized in extracting distributional frequencies.

Regarding the properties of the noun, we have not pointed out that the noun should precede any modifier. We neither have specified in the noun that some modifiers must precede the noun such as we did in the *ADJ* modifier construction with α_2, α_4, α_5. This lack of specificity regarding the modifiers and the noun of a subject includes modifier structures such as *PRON*, *PROPN*, *SubC*, *ADJ*, and others.

These properties, which take into account the noun in the subject and its modifiers, are not introduced in the description of the noun as a subject for the following reasons:

- If they were introduced in the *NOUN*, they would be violating such properties as $NOUN \prec ADJ$ when this precedence was not fulfilled; such case it is not true.
- The noun does not depend on this restriction of precedence; the noun should not activate them.
- It is adjective which triggers this relationship of precedence and, therefore, these constraints should be included in the category that triggers them.

In this manner, the precedence between the noun and the modifying adjective are satisfied or violated only when the adjective appears on the structure. The category of the adjective is the one that detonates these properties and the one that checks if they are satisfied or violated. In Table 7.12 and Table 7.13, we show these particularities with an example in natural language.

In Case $NOUN \prec ADJ$ is in $NOUN_{[subj]}$			
Sentence	El	chico	corre
CAT	$DET_{[spec]}$	$NOUN_{[subj]}$	$V_{[intrans]}$
Properties		α_1: $DET_{[spec]} \prec NOUN_{[subj]}$ α_2: $NOUN_{[subj]} \Rightarrow DET_{[spec]}$ α_3: $Uniq$: $DET_{[spec]}$ α_4: $NOUN_{[subj]} \otimes ADV \wedge PRON \wedge V_{[inf]}$ β_5: $NOUN_{[subj]} \prec ADJ_{[mod]} \vee X_{[mod]} \vee CONJ_{[conj]}$ β_6: $ADJ_{[mod][num:ord]\vee[undef]} \prec NOUN_{[subj]}$ α_7: $NOUN_{[subj]} \rightsquigarrow_{subj} V$	
Grammaticality		Cn_w=0.142 VG_w=0.714	

Table 7.12 Example of including $NOUN \prec ADJ$ in $NOUN_{[subj]}$.

In Case $NOUN \prec ADJ$ is in $ADJ_{[mod]}$			
Sentence	El	chico	corre
CAT	$DET_{[spec]}$	$NOUN_{[subj]}$	$V_{[intrans]}$
Properties		α_1: $DET_{[spec]} \prec NOUN_{[subj]}$ α_2: $NOUN_{[subj]} \Rightarrow DET_{[spec]}$ α_3: $Uniq$: $DET_{[spec]}$ α_4: $NOUN_{[subj]} \otimes ADV \wedge PRON \wedge V_{[inf]}$ α_5: $NOUN_{[subj]} \rightsquigarrow_{subj} V$	
Grammaticality		Cn_w=0.2 VG_w=1	

Table 7.13 Example of not including $NOUN \prec ADJ$ in $NOUN_{[subj]}$.

Regarding Table 7.12 and Table 7.13, they show a grammatical example in Spanish. The value of grammaticality of $NOUN_{[subj]}$ has been calculated in both cases using the formulas presented in section 6.6. Since we cannot be sure about the different weights between categories, we acknowledge the same weight of 1 for all types of words. Therefore, we apply Equation 6.27 to $NOUN_{[subj]}$.

$$NOUN_{[subj]} = 1 \tag{7.7}$$

Secondly, we reveal the canonical value of each constraint of $NOUN_{[subj]}$ (Cn_w) by our standard value of a canonical property (1). We assume that all canonical properties have the same density, which means that no canonical property is more important than another.

By assuming this value of 1, we provide a theoretical value for each canonical constraint strictly from the perspective of the syntactic module, without involving our weights with frequencies, avoiding in such manner the paradoxes of weighting canonical properties by probabilities as seen in subsection 3.9.3 and section 7.3. This canonical value of 1 is divided by all the triggered constraints of a the $NOUN_{[subj]}$ (C_{δ}), both satisfied (α) and violated (β). In Table 7.12, 7 constraints has been triggered. In Table 7.13, 5 constraints has been triggered. The two additional constraints triggered in Table 7.12 are those constraints which define the precedence between the noun and the adjective. In Equation 7.8 and Equation 7.9, the value of the canonical constraints for Table 7.12 and Table 7.13 are respectively calculated following Equation 6.28.

$$0.142\,Cn_w = \frac{1}{C_7} \tag{7.8}$$

$$0.2\,Cn_w = \frac{1}{C_5} \tag{7.9}$$

The violated properties are weighed as 0. Because any variability property is triggered, we already calculate the value of grammaticality of $NOUN_{[subj]}$ by applying Equation 6.32. For calculating VG_w, we take into account all the canonical weights. In Table 7.12, the satisfied properties (α) which keep the value as canonical properties ($0.142\,Cn_w$) are displayed in Equation 7.8. Since all its satisfied constraints are 5 out 7, we calculate All_{Cn_w} as 0.714. Because there are no variability properties triggered, the value of all variability properties of the $NOUN_{[subj]}$ is 0: All_{Vab_w} as 0. These both values are divided by the value of $NOUN_{[subj]}$ which is 1. The final value of grammaticality of $NOUN_{[subj]}$ in Table 7.12 is calculated as 0.714 as showed in Equation 7.10.

$$0.714\,VG_w = \frac{(0.714\,All_{Cn_w} + 0\,All_{Vab_w})}{1\,NOUN_{[subj]}} \tag{7.10}$$

In Table 7.13, the satisfied properties (α) which keep the value as canonical properties ($0.2\,Cn_w$) are displayed in Equation 7.9. Since all its satisfied constraints are 5 out 5, we calculate All_{Cn_w} as 1. Because there are no variability properties triggered, the value of all variability properties of $NOUN_{[subj]}$ is 0: All_{Vab_w} as 0. This both values and are divided by the value of $NOUN_{[subj]}$ which is 1. The final value of grammaticality of $NOUN_{[subj]}$ in Table 7.13 is calculated as 1 as showed in Equation 7.11

$$1\,VG_w = \frac{(1\,All_{Cn_w} + 0\,All_{Vab_w})}{1\,NOUN_{[subj]}} \tag{7.11}$$

In such manner, if we would specify the precedence of the noun towards a modifier in $NOUN_{[subj]}$, it would trigger a lower degree of grammaticality than what it should be. The properties β_5: $NOUN_{[subj]} \prec ADJ_{[mod]} \vee X_{[mod]} \vee CONJ_{[conj]}$ and β_6: $ADJ_{[mod][num:ord] \vee [undef]} \prec NOUN$ trigger a violation since the noun is not preceding any adjective, nor modifier, nor conjunction, nor an adjective numeral ordinal and undefined. Because of these violations, a grammatical sentence has 0.284 less of grammaticality. Therefore, because we placed these precedence relations in the *ADJ* modifier construction, these properties would only be triggered when an adjective appears —allowing a better calculation for our degrees of grammaticality just as in Table 7.13.

The nominal modifier constructions introduced by a preposition present a similar case regarding specification and modifier structures. In Table 7.14, we present both the properties of a nominal modifier and the properties of an ADP as a specifier.

$NOUN\ in\ Modifier\ Construction$
α_1: $NOUN_{[mod]} \Rightarrow ADP_{[spec]}$ α_2: $X_{[subj]} \prec NOUN_{[mod]}$ α_3: $NOUN_{[mod]} \Rightarrow DET_{[spec]}$ α_4: $Uniq$: $DET_{[spec]}$ α_5: $NOUN_{[mod]} \otimes ADV \wedge V_{[inf]}$ α_6: $NOUN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$
Variability Properties of $NOUN\ in\ Modifier\ Construction$
γ_1: $\beta_1 \implies Noun_{[xADJ]}$
$ADP\ in\ Specifier\ Construction$
α_1: $ADP_{[spec]} \prec X_{[mod]}$ α_2: $ADP_{[spec]} \Rightarrow X_{[mod]} \vee V$ α_3: $ADP_{[spec]} \rightsquigarrow_{spec} NOUN_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]}$

Table 7.14 Spanish properties of both *NOUN* in modifier and *ADP* in specifier construction.

The noun as a modifier shares some properties with the noun in subject structure. These shared properties in noun as a modifier are: α_3, α_4, α_5.

The particularities of this construction of noun modifier are:

- It specifies that the nominal element requires the preposition as specifier: $NOUN_{[mod]} \Rightarrow ADP_{[spec]}$.
- The precedence between the element which is going to be modified, in this case an unspecified subject: $X_{[subj]} \prec NOUN_{[mod]}$. This unspecified subject might include whatever a proper noun or a pronoun if it would be the case.

- The modifier dependency between the noun as modifier and the unspecified subject: $NOUN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$.

Additionally, in this *NOUN in Modifier Construction*, we have explained the performance of the noun in apposition as a *xCategory*, specifically, a *xADJ*. The noun in apposition occurs in 1461 in the corpus. However, with this property, we want to point out that a noun as a modifier without a preposition is not as grammatical as if an adjective would be used instead of the noun. We want to capture the variability nature of the language in which: 1) the noun is more grammatical as a modifier if a preposition introduces it as it occurs 39808 times in the corpus as *nmod*; 2) the use of a *NOUN* in apposition for modifying another noun is not radically wrong, or right. It is grammatical but to a certain degree.

Therefore, if a nominal element would be a modifier without a *mod construction*, this element would be considered as a *xADJ*, and it would have to satisfy the properties of the adjective.

The preposition is treated as a specifier since we consider that its job is introducing a clause rather than modifying or providing other syntactical functions. Therefore, we specified *ADP* in a specifier construction with the following properties:

- α_1: $ADP \prec X_{[mod]}$: the preposition has to precede an unspecified modifier construction. This could be a *NOUN*, *PROPN* or a *PRON*.

- α_2: $ADP \Rightarrow X_{[mod]} \vee V$: the preposition requires a modifier construction or a verb. This representation is important since many times, regarding the constructions close to the verb, it is complicated to acknowledge if the preposition is chosen by the noun (for example) or by a verb. In Spanish, we can find many verbs which are lexicalized with a preposition such as: *"acusar a/de"*; *"aspirar a"*; *"excluir de/a"*, and so on. Therefore, in case the verb would determine the preposition, the modifier construction would still satisfy its requirement and its precedence properties which are α_1 and α_2.

- α_3: $ADP \rightsquigarrow_{spec} NOUN_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]}$: This property specifies that the preposition has a dependency relation with a noun or a proper noun, a pronoun, or an adjective. We include the adjective here because an adjective can appear as a modifier construction introduced by a preposition without compromising its properties. Additionally, we do not represent that the preposition would specify the verb in case of requirement ($V \Rightarrow ADP$) because the specification would always go towards the modifier constructions, even if such preposition belongs to a verb: $V \Rightarrow ADP$ and $ADP_{[spec]} \rightsquigarrow_{spec} X_{[mod]}$ and not $V \Rightarrow ADP$ and $ADP_{[spec]} \rightsquigarrow_{spec} V$.

The precedence between the *DET* and the prepositions do not overlap with the descriptions of the *NOUN* modifier construction and the *ADP* specifier construction. The name in a modifier structure makes explicit that an article must precede this one. *ADP*, in modifier structure, defines that the preposition must precede a name in modifier structure. Therefore, it precedes the category and its structure. In this way, the precedence of $ADP \prec DET \prec NOUN$ can converge.

Besides, as mentioned before, this relation between the preposition and the modifier in Table 7.14 also represents how a noun as a modifier requires the preposition. If the noun as modifier does not satisfy the requirement of the preposition, the noun will exercise a role in apposition similar to an adjective. Therefore, if this requirement between the preposition and the noun as a modifier is violated, the noun will perform such as a *xADJ* in modifier construction.

In Table 7.15, we show an example in natural language which displays the variability property of the noun as modifier construction in Table 7.14.

Case: $NOUN_{[mod]} \Rightarrow ADP_{[spec]}$ in $NOUN_{[mod]}$				
Sentence	El	hombre	robot	corre
CAT	$DET_{[spec]}$	$NOUN_{[subj]}$	$NOUN_{[mod]}$	$V_{[intrans]}$
Properties		α_1: $DET \prec NOUN$ α_2: $NOUN \Rightarrow DET$ α_3: $Uniq$: DET α_4: $NOUN \otimes ADV \wedge V_{[inf]}$ α_5: $NOUN \rightsquigarrow_{subj} V$	β_1: $NOUN_{[mod]} \Rightarrow ADP_{[spec]}$ β_2: $DET_{[spec]} \prec NOUN_{[mod]}$ β_3: $NOUN_{[mod]} \Rightarrow DET_{[spec]}$ α_4: $Uniq$: $DET_{[spec]}$ α_5: $NOUN_{[mod]} \otimes ADV \wedge V_{[inf]}$ α_6: $NOUN_{[mod]} \rightsquigarrow_{mod} NOUN_{[subj]}$ γ_1: $\beta_1 \implies Noun_{[xADJ]}$ $xADJ_{[mod]}$ γ_1: $NOUN_{[subj]} \prec NOUN_{[xADJ]}$ γ_2: $NOUN_{[xADJ]} \rightsquigarrow_{mod} NOUN_{[subj]}$	
Grammaticality		Cn_w=0.2 VG_w=1	Cn_w = 0.166 Vab_w=0.083 VG_w=0.581	

Table 7.15 *NOUN* in modifier construction as a *xADJ*.

Regarding Table 7.15, it shows an example of the grammaticality value of a word as a *xCategory* in Spanish. The value of grammaticality of *"robot"* $NOUN_{[mod]}$ has been calculated using the formulas presented in section 6.6. Again, we point out that since we cannot be sure about the different weights between categories, we acknowledge the same weight of 1 for all types of words. Therefore, we apply Equation 6.27 to $NOUN_{[mod]}$ in Equation 7.12.

$$NOUN_{[mod]} = 1 \tag{7.12}$$

On one hand, we reveal the canonical value of each constraint of $NOUN_{[mod]}$ (Cn_w) by our standard value of a canonical property (1). This canonical value of 1 is divided by all the triggered constraints of a the $NOUN_{[mod]}$ (C_δ), both satisfied (α) and violated (β). In Table 7.15, 6 constraints (either α or β) has been triggered. The variability constraints don't count as an additional triggered constraint since it is a consequence of a violated constraint as seen in Equation 6.19. In Equation 7.13, the value of the canonical constraints for $N_{[mod]}$ in Table 7.15 is calculated following Equation 6.28.

$$0.166\, Cn_w = \frac{1}{C_6} \tag{7.13}$$

On the other hand, we calculate the value of the variability properties triggered as a consequence of the violated canonical properties. The variability triggered properties of the $NOUN_{[mod]}$ are 1: $\gamma_1{:}\beta_1 \implies NOUN_{[xADJ]}$. We do not take into account the satisfied variability properties from the *xADJ* since its satisfaction evaluates the possibility of being a $NOUN_{[mod]}$ with a *xADJ* feature. The variability constraints in the word *"robot"* as $NOUN_{[mod]}$ just count as one because the variability rule in $NOUN_{[mod]}$ specifies that it triggers all the constraints in *xADJ*. Therefore, in case some of the variability constraints would not be satisfied, this variability would not be applied. That is why we separate the constraints which count on calculating grammaticality and the others that just matter for including the variability rule of the $NOUN_{[mod]}$, γ_1. In other words, if the $NOUN_{[mod]}$ couldn't satisfy the variability properties of the *xADJ*, a $NOUN_{[mod]}$ could not take into account the calculation of the variability property $\gamma_1{:}\beta_1 \implies NOUN_{[xADJ]}$.

In Equation 7.14, the value of a variability property is calculated by dividing the value of a variability property (0.5) by all triggered all satisfied (α) and violated (β) constraints in $NOUN_{[mod]}$ (*"robot"*).

$$0.083\, Vab_w = \frac{0.5}{C_6} \tag{7.14}$$

Finally, we calculate the value of grammaticality of $NOUN_{[mod]}$ (VG_w) by applying Equation 6.32. For calculating VG_w, we take into account all the canonical weights plus all the variability weights of the word *"robot"* as a $NOUN_{[mod]}$. In Table 7.15, the satisfied properties (α) which keep the value as canonical properties ($0.166\, Cn_w$) are displayed in Equation 7.13. Since all its satisfied constraints are 3 out 6 we calculate All_{Cn_w} as 0.498. In Table 7.15, the variability properties (γ) which keep the value as variability properties ($0.083\, Vab_w$) are displayed in Equation 7.14. Since all its satisfied constraints are 1 out 1 we calculate All_{Vab_w} as 0.083. The final value of grammaticality of $NOUN_{[mod]}$ in Table 7.15 is calculated as 0.581 as showed in Equation 7.15.

$$0.581\, VG_w = \frac{(0.498\, All_{Cn_w} + 0.083\, All_{Vab_w})}{1\, NOUN_{[mod]}} \tag{7.15}$$

We can compute these values with words following the representation of our Fuzzy Grammar applying approximate reasoning in subsection 6.4.2:

- IF an input is *significantly* satisfied (1-0.8) $THEN$ the value of grammaticality is *high*.
- IF an input is *quite* satisfied (0.8-0.5) $THEN$ the value of grammaticality is *medium*.
- IF an input is *barely* satisfied (0.5-0) $THEN$ the value of grammaticality is *low*.

Similarly, we could define inputs taking into account the same relations of value with the following expressions:

- IF the value of grammaticality is *high* $THEN$ an input is *significantly grammatical*.
- IF the value of grammaticality is *medium* $THEN$ an input is *quite grammatical*.
- IF the value of grammaticality is *low* $THEN$ an input is *barely grammatical*.

When applying this reasoning to our example, it reveals that the variability property raised the value of grammaticality of $NOUN_{[mod]}$ enough to be better considered in our Fuzzy Grammar:

- Our Fuzzy Grammar considers a value of 0.581 as *an input quite satisfied*; therefore, its value of grammaticality is *medium*. The input is *quite grammatical*.
- In contrast, the value of the input *"robot"* as a $NOUN_{[mod]}$ would be 0.498 without our variability property. In this manner, this input without our variability property in our fuzzy grammar would be computed as *an input which is barely satisfied*. Therefore, its value of grammaticality would be *low*. The input would be *barely grammatical*.
- We recognize that another combination of words for $NOUN_{[subj]}$ and $NOUN_{[mod]}$ such as *"El hombre paz"* or *"El cielo hombre"* would have the same value of grammaticality regarding the syntactic domain. However, it would not have the same grammaticality value regarding the whole multi-modal Fuzzy Grammar. We state that the combination of $NOUN_{[subj]}$ and $NOUN_{[mod]}$ is syntactically possible to a certain degree. However, its degree of grammaticality regarding the other domains (such as semantics) would rely on the satisfaction or the violation of its properties in such modules. Consequently, the final value of grammaticality of two identical syntactic structures might be different in a Fuzzy Property Grammar when we calculate grammaticality of an utterance regarding all the properties in all their domains.

Therefore, it seems that the most important properties for a noun as a modifier are α_1, α_2 and α_6. Thanks to Table 7.15 we recognize that:

- It matters where the properties are placed (see Table 7.12 and Table 7.13).
- The benefits of working with *xCategories* for a more precise calculation of the degrees of grammaticality. Thanks to the use of an *ADP* as a specifier, we can explain how the noun can perform in an *adjective fit*, in a construction as a modifier. Once the requirement of an *ADP* for a *NOUN* as a modifier is violated, the property grammar recognizes that noun has a role/feature of an adjective. Therefore, the noun is still a noun but with a new feature which defines its borderline case. Without this feature of an *xCategory*, the degree of grammaticality of $NOUN_{[mod]}$ would be 0.498 and *barely satisfied* instead of 0.581 and *quite satisfied*.
- The precedence relation of a noun in subject and modifiers works better if it is placed in the modifier constructions.
- Taking into account *ADP* as a specifier allows great flexibility since we can establish properties for any word which is going to be a modifier construction introduced by a preposition.
- Defining the preposition in a specific construction allows us to better define its dual nature in Spanish between verbal constructions and modifier constructions.

7.10 A Property Grammar for the *PROPN*

PROPN in Subject Construction
α_1: $PROPN_{[subj]} \otimes DET \wedge NOUN \wedge PRON \wedge ADJ \wedge ADV \wedge SCONJ \wedge V_{[non]}$ α_2: $PROPN_{[subj]} \rightsquigarrow_{subj} V$
Variability Properties
γ_1: β_1 $\neg(PROPN_{[subj]} \otimes DET_{[spec]}) \implies PROPN_{[xNOUN]}$ γ_2: β_1 $\neg(PROPN_{[subj]} \otimes V_{[non]}) \implies PROPN_{[subj]} \Rightarrow V_{[part]} \Rightarrow X_{[mod]}$ γ_3: β_1 $\neg(PROPN_{[subj]} \otimes PRON) \implies PROPN_{[subj]} \Rightarrow PRON_{[rel]}$

Table 7.16 Spanish properties of *PROPN* in subject construction.

In Table 7.16 we find the properties of the Spanish *PROPN*. Property α_2 indicates the dependency relation of a *PROPN* as a subject towards the verb. Nevertheless, the property which stands out is the *PROPN* excluding many elements as shown in property α_1.

However, we find exceptions regarding this exclusion which are pointed out.

We found in Spanish UD that the *PROPN* can precede or be preceded by some modifiers in exceptional cases. Firstly, it allows the precedence of *PROPN* towards an ordinal numeral. This relation is mostly found in index 19: "*Francisco I manifiesta un verdadero afecto por el viejo hombre*" (13009:2).

Similarly, a proper noun can be preceded by an adjective in case we would treat either *"san"* or *"santa"* as an adjective: *"San Diego"* (11616:2).

However, in this work, we treat the numeral element and the adjective *"san"* or *"santa"* as part of the proper noun itself. These cases seem an exception in which both elements together are a part of the proper noun itself. Therefore, we consider them as a part of the *PROPN* in the lexical entry. Consequently, *"Jaime"*, *"Jaime I"* and *"San Jaime"* are three different lexical entries with three different meanings in the lexicon of our fuzzy grammar.

On the other hand, the exclusions of the *PROPN* are pretty clear. *PROPN* almost does not appear with any modifiers next to it except another *PROPN*. That is why any precedence is specified since it does not matter which *PROPN* is preceding to each other. This is one of the particularities of this category and the reason why it is tagged many times under the dependencies of *name* and *appos*: "*Javier Plano (Buenos Aires, 1979) es un artista argentino contemporáneo*" (4440:2); "*Pancho Domínguez, alcalde de Querétaro, encabeza un mitin de apoyo al aspirante*" (1387:2).

Such characteristic would present the *PROPN* in a modifier construction different from the one shown in Table 7.17.

PROPN in Modifier Construction
α_1: $X_{[subj]} \prec PROPN_{[mod]}$ α_2: $PROPN_{[mod]} \otimes DET \wedge NOUN \wedge PRON \wedge ADJ \wedge ADV \wedge V_{[non]}$ α_3: $PROPN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$ α_4: $PROPN_{[mod]} \Rightarrow ADP_{[spec]}$
Variability Properties
γ_1: $\beta_2\ \neg(PROPN_{[mod]} \otimes DET_{[spec]}) \implies PROPN_{[xNOUN]}$ γ_2: $\beta_2\ \neg(PROPN_{[mod]} \otimes V_{[non]}) \implies PROPN_{[mod]} \Rightarrow V_{[part]} \Rightarrow X_{[mod]}$ γ_3: $\beta_2\ \neg(PROPN_{[mod]} \otimes PRON) \implies PROPN_{[mod]} \Rightarrow PRON_{[rel]}$

Table 7.17 *PROPN* in modifier construction.

PROPN as a modifier does not require an *ADP*, also neither of other properties such as articles. This characteristic does not mean that the *PROPN* cannot be preceded or introduced by an *ADP*. Because we have described the *ADP* as a specifier, we can allow to the *ADP* construction as specifier require a *PROPN* as a modifier construction while we do not require this preposition on the definition of the properties of the *PROPN* in a modifier construction.

The importance of these properties regarding the *PROPN* can be found in the variability properties. In γ_2, it is specified that if a proper noun is in co-occurrence with a non-finite verb, it is required that this non-finite verb is a participle which it requires a modifier construction: $V_{[part]} \Rightarrow X_{[mod]}$. This property has almost no precedence in the corpus; just 39 occurrences are found in subject construction. However, from the theoretical point of view, these occurrences are valid, since it is frequent in Spanish to use such structure, however, it is not canonical: "*Francisco Sánchez conocido por el nombre artístico de Paco el Barbero fue un guitarrista*" (3858:2). This variability property could be found as well in *PROPN* in modifier construction: "*El hijo de Francisco Sánchez conocido por el nombre artístico de Paco el Barbero fue un guitarrista*".

The property γ_3 has a low frequency as well (28), but, in the same way, as with γ_2, these structures are valid in Spanish; however, they seem a bit forced or old fashion. Because of this undefined boundary, we define them with a variability property; pointing out that if the *PROPN* violates the exclusion with the *PRON*, the *PROPN* requires that such pronoun has to be a relative pronoun: "*Charles Duncan, que interpretaba a Spit, dejó la producción*" (7826:2). This variability property could be found as well in *PROPN* in modifier construction: "*El sucesor de Charles Duncan, que interpretaba a Spit, dejó la producción*"

Finally, we consider that the most important property is γ_1, which reflects the mechanism of a fuzzy approach working together with a property grammar. In general, Spanish speakers would say that *PROPN* excludes the use of a determiner. However, we can find many cases in which this is not entirely true.

We know that a determiner precedes some *PROPN* in exceptional contexts, however, from the formal point of view of the syntactic properties, these *PROPN* are triggering different semantic aspects than if the determiner would not precede them. The comparison between structures can be seen in Table 7.18.

PROPN CONSTRUCTIONS		*NOUN* CONSTRUCTIONS WITH *PROPN* SEMANTICS	
0	PROPN-name *	1	DET-det *
3	* CONJ-cc PROPN-conj	2	DET-det * PROPN-nmod
4	PROPN-name PROPN-name *	6	DET-det * PROPN-appos
5	* PROPN-amod	12	PROPN-det *
7	PROPN-name * NOUN-appos	15	DET-det * PROPN-name
8	* NOUN-appos	18	DET-det * PROPN-amod PROPN-nmod
9	* PROPN-appos	20	DET-det PROPN-name *
10	* PROPN-nmod		
13	PROPN-name * PROPN-appos		
14	PROPN-amod *		
16	* NUM-nummod		
17	PROPN-name PROPN-name * PROPN-appos		
19	* PROPN-name		

Table 7.18 List of structures of PROPN with determiner and its index.

In Table 7.19, we expose a table with determiners preceding a PROPN and specifying which kind of constructions these are. The indexes in Table 7.19 (i.e. I.1) are related to the indexes shown in Table 7.18. The *PROPN* such as *"Mike"*, *"Javier"*, *"Cristina"* have been chosen as the typical *PROPN* forms which almost always respect the *PROPN* properties in Table 7.16. Thus, these *PROPN* are never preceded by a *DET*.

TYPE:	*NOUN*	NOTES
Acronym:	CNT (I.1/37:9)= Confederación Nacional del Trabajo	Nominal Construction with *PROPN* Semantics
	RIRA (I.1/13046:20)= Real Irish Republican Army	Nominal Construction with *PROPN* Semantics
Nicknames:	La Victoria (I.1/828:32)	Nominal Construction with *PROPN* Semantics
	El Misericordioso (I.12/36:31)	Nominal Construction with *PROPN* Semantics
	El Llanero Solitario (I.15/5436:2)	Nominal Construction with *PROPN* Semantics
	El Ché Guevara (I.20/1985:14)	Nominal Construction with *PROPN* Semantics (Ché means boy, dude)
Brands:	El Sportprinz (I.1/13046:20)= El coche Sportprinz	***PROPN* with *NOUN* semantics (means car)**
	El Audi A3 (I.6/13983:2) El coche Audi A3	***PROPN* with *NOUN* semantics (means car)**
Companies:	La Fox (I.1/4638:2)	***PROPN* with *NOUN* semantics (means TV channel)**
Sports Club:	El Manchester (I.1/2957:19)= El equipo de fútbol de la ciudad de Manchester (Manchester United)	***PROPN* with *NOUN* semantics (means football team)**
	El FC Barcelona (I.12/56:2)= El club de fútbol de la ciudad de Barcelona	***PROPN* with *NOUN* semantics (means football team)**
Others:	The Guardian (I.12/1109:5) El Guardián	Nominal Construction with *PROPN* semantics
	El Hard Rock Cafe (I.20/3035:2) El café Hard Rock	Nominal Construction with *PROPN* semantics
	La Challenge League (I.20/3035:2) La Liga Desafío	Nominal Construction with *PROPN* semantics
Tribes:	Los Mlarbi (I.1/4639:2)	***PROPN* with *NOUN* semantics (means the tribe of Mlarbi)**
Public places and artistic objects	La Fontana Moro (I.2/13134:2)	Nominal Construction with *PROPN* semantics
	La Universidad Técnica de Dresde (I.18/1863:2)	Nominal Construction with *PROPN* semantics
Family names:	Los Lincoln (I.1/11355:6)= La familia de los Lincoln	***PROPN* with *NOUN* semantics (means the Lincoln family)**
Distinctive titles:	El Virrey Luis de Velasco (I.18/1863:2)	Nominal Construction with *PROPN* semantics ("El virrey"= The viceroy)

Table 7.19 List of *PROPN* with determiner.

Nevertheless, certain *PROPN* are preceded by a *DET* when a diacritic meaning is triggered. That is: a *PROPN* with a *NOUN* semantics. For example, *"el Manchester"* (*the football team, not the city*); *"los Lincoln"* (*the Lincoln family not the president*). A similar phenomenon occurs with the *NOUN* *"Victoria"* (*PROPN*); *"la Victoria"* (*The Victory*); *"Esperanza"* (*PROPN*); *"la Esperanza"* (*The hope*).

Here is where variability occurs since a particular structure (which is not fully grammatical but it is violated) is triggered in order to provide a recursive expressive resource. This variability of the *PROPN* and its structures are frequent, but they are never as frequent as the canonical ones (*PROPN* non-preceded by a *DET* in a 69.99%). Therefore, the most crucial element that syntactically differentiates a *NOUN* from a *PROPN* is the precedence of the *DET* even though, semantically, it could not coincide with the syntax modality.

Acronyms (*CNT* (*"La Confederación Nacional del Trabajo"* [The National Confederation of Labor]), nicknames (*"El Llanero Solitario"* [Lone Ranger]), pieces of art (*"La Fontana del Moro"*), distinctive names (*"El Virrey [Viceroy] Luis de Velasco"*), etc. are examples of *NOUN* with a semantic meaning typically from a *PROPN*. They refer to specific meanings, but both properties and original substantives are a NOUN. Thus they are not considered as a *PROPN* from a syntactic point of view.

In this manner, we specify that *PROPN* violates the property $PROPN \otimes DET$. However, the violation triggers new variability properties and a new fit with *xNOUN*. Therefore, the *PROPN* would need to satisfy the properties from *xNOUN*.

According to the spirit of our fuzzy grammar, the relation between the *PROPN* and *NOUN* in Spanish are the perfect example of categories which can be found in fuzzy boundaries such as a *PROPN* being syntactically performed like a *NOUN*. We do not discriminate discretely the fact that a determiner must precede a *PROPN* or not. We acknowledge the variability of the language, and we define it, specifying that the new structure might be not as grammatical as the canonical one, however, it responds to other domains: semantic and pragmatical modules.

In Table 7.20, we provide an example in natural language of a *PROPN* with a feature of an *xNOUN* and its grammaticality degree with the sentence: *"El nuevo Volskwagen está equipado con la última tecnología"*.

Case: $PROPN_{[subj]}$ as $PROPN_{[xNOUN]}$				
Sentence	El	nuevo	Volkswagen	está equipado con la uĺtima tecnología
CAT	$DET_{[spec]}$	$ADJ_{[mod]}$	$PROPN_{[subj]}$	$V_{[cop]}$
Properties		β_1: $NOUN_{[subj]} \prec ADJ_{[mod]}$ γ_1: $\beta_1 \implies ADJ_{[mod][semantics]}$ α_3: $ADJ_{[mod]} \rightsquigarrow_{mod} NOUN_{[subj]}$	β_1 $PROPN_{[subj]} \otimes DET$ γ_1: β_1: $\implies PROPN_{[xNOUN]}$ α_2: $PROPN_{[subj]} \rightsquigarrow_{subj} V$ $PROPN_{[xNOUN]}$ γ_1: $DET_{[spec]} \prec PROPN_{[xNOUN]}$ γ_2: $PROPN_{[xNOUN]} \Rightarrow DET_{[spec]}$ γ_3: $Uniq$: $DET_{[spec]}$ γ_4: $PROPN_{[xNOUN]} \otimes ADV_{[mod]} \wedge PRON \wedge V_{[inf]}$ γ_5: $PROPN_{[xNOUN]} \rightsquigarrow_{subj} V$	
Grammaticality	VG_w=1	Cn_w=0.5 Vab_w=0.25 VG_w=0.75	Cn_w=0.5 Vab_w=0.25 VG_w=0.75	
	VG=0.833 *significantly grammatical*			

Table 7.20 Example of *PROPN* as a *xNOUN*

Regarding Table 7.20, it shows an example of the grammaticality value of a *PROPN* as a *xCategory* (*xNOUN*) in Spanish. The word *"volskwagen"* is canonically a *PROPN* which means the *"volskwagen company"*. As a *PROPN*, the word *"volskwagen"* satisfy the canonical properties of Table 7.16 when it displays its meaning as a proper noun. Sentences such as *"Volskwagen invierte en I+D"* and *"Volskwagen sacará un nuevo modelo en 2020* are an example of this word as a canonical *PROPN*.

However, as in Table 7.19, a *PROPN* can be preceded by a determiner when it is marked with a meaning typically understood as a *NOUN*. In Spanish, the proper nouns of car companies display the meaning of *"coche"* (*car*) (a typical *NOUN*) when a determiner precedes them. Consequently, the canonical sentence concerning the *PROPN* in Table 7.20 would be: *"El nuevo coche de Volskwagen está equipado con la última tecnología"* (*the new car of Volskwagen is equipped with the latest technology*).

In this manner, our Fuzzy Grammar represents the variability phenomenon of a typical category such as a *PROPN* performing with a *NOUN* feature such as a *xNOUN*. This phenomenon is represented through variability properties, and the $PROPN_{[xNOUN]}$ have to satisfy the variability properties of *xNOUN* in order to be considered with the *xCategory* feature.

Besides, in Table 7.20, we also illustrate an example of the variability property γ_1 of an *ADJ* in a modifier construction (Table 7.11). We assume that the satisfaction or violation of this property relies on the satisfaction or the violation of properties in the semantic domain. In Table 7.20, we have assumed that those semantic properties are satisfied; allowing to the syntactic property γ_1 in the adjective modifier to be satisfied.

In short, because of these variability properties, our Fuzzy Property Grammar can calculate the degree of grammaticality of Table 7.20 taking into account solely linguistic constraints.

The value of grammaticality of *Volskwagen* $PROPN_{[subj]}$ has been calculated using the formulas presented in section 6.6.

- We acknowledge the same weight of 1 for all types of words.

$$PROPN_{[subj]} = 1 \tag{7.16}$$

- We reveal the canonical value of each constraint of $PROPN_{[subj]}$ (Cn_w) by our standard value of a canonical property (1). This canonical value of 1 is divided by all the triggered constraints of a the $PROPN_{[subj]}$ (C_δ), both satisfied (α) and violated (β). In Table 7.20, 2 constraints (either α or β) has been triggered.

$$0.5\, Cn_w = \frac{1}{C_2} \tag{7.17}$$

- The variability constraints does not count as an additional triggered constraint since it is a consequence of a violated constraint. The variability constraints in the word *Vol-*

skwagen as $PROPN_{[subj]}$ just count as one because the variability rule in $PROPN_{[subj]}$ specifies that it triggers all the constraints in $xNOUN$. Therefore, in case some of the variability constraints would not be satisfied, this variability would not be applied. *Volskwagen* as a $PROPN_{[xNOUN]}$ satisfies all the properties of the $xNOUN$, therefore, the variability property $\gamma_1 : \beta_1 \implies PROPN_{[xNOUN]}$ is satisfied. The value of a variability property is calculated by dividing the value of a variability property (0.5) by all triggered all satisfied (α) and violated (β) constraints in $PROPN_{[subj]}$ (*Volskwagen*)).

$$0.25\, Vab_w = \frac{0.5}{C_2} \tag{7.18}$$

- We calculate the value of grammaticality of $PROPN_{[subj]}$ (VG_w) by applying Equation 6.32. For calculating VG_w, we take into account all the canonical weights plus all the variability weights of the word *Volskwagen* as a $PROPN_{[subj]}$.

$$0.75\, VG_w = \frac{(0.5\, All_{Cn_w} + 0.25\, All_{Vab_w})}{1\, PROPN_{[subj]}} \tag{7.19}$$

- The values in the word *“nuevo”* as an $ADJ_{[mod]}$ share the same grammaticality value since 1 canonical property out of two has been satisfied, and the violated canonical property triggered one variability property which has been satisfied too.

$$0.75\, VG_w = \frac{(0.5\, All_{Cn_w} + 0.25\, All_{Vab_w})}{1\, ADJ_{[mod]}} \tag{7.20}$$

- We can calculate the final value of grammaticality of the subject construction of Table 7.20 as a whole input with Equation 6.32:

$$\{DET_{[spec]} = 1\, ADJ_{[mod]} = 0.75\, PROPN_{[subj]} = 0.75\} = 2.5 \tag{7.21}$$

$$0.833\, VG = \frac{(2.5\, All_{VG_w})}{All_3} \tag{7.22}$$

We can compute these values with words following the representation of our Fuzzy Grammar applying approximate reasoning in subsection 6.4.2:

- Our VG in the subject construction of Table 7.20 is 0.833.
- *IF* an input is *significantly* satisfied (1-0.8) *THEN* the value of grammaticality is *high*.

- *IF* the value of grammaticality is *high THEN* an input is *significantly grammatical*.
- Therefore, the input *"El nuevo Volskwagen"* is *significantly grammatical*.

When applying this reasoning to our example, it reveals that the variability property raised the value of grammaticality of both $PROPN_{[subj]}$ and $ADJ_{[mod]}$ enough to be better considered in our Fuzzy Grammar. Without these variability properties, both inputs would be evaluated as 0.5:

- Our Fuzzy Grammar considers a value of 0.5 as a borderline case of *an input quite satisfied* (0.8-0.5) or *an input barely satisfied* (0.5-0). Therefore, their degree of grammaticality would be vaguely *medium* and *low*. The input would be both *quite grammatical* and *barely grammatical*.

7.11 A Property Grammar for the *PRON*

PRON in Subject Construction
α_1: $PRON_{[subj][demons]\vee[rel]\vee[pers]} \otimes ADP \wedge ADJ \wedge ADV \wedge DET \wedge PRON$
α_2: $PRON_{[subj][lo]} \prec ADJ_{[mod]}$
α_3: $PRON_{[subj][lo]} \Rightarrow ADJ_{[mod]}$
α_4: $PRON_{[subj][lo]} \otimes ADP \wedge DET \wedge ADV$
α_5: $PRON_{[subj]} \rightsquigarrow_{subj} V$
Variability Properties
γ_1: $\beta_1\ \neg(PRON_{[subj]} \otimes DET) \implies DET_{[el]} \prec PRON_{[subj][yo]}$
γ_2: $\beta_1\ \neg(PRON_{[subj]} \otimes ADJ) \implies ADJ_{[mod][solo]\vee[mismo]}$
xPRON in Subject Construction
γ_1: $X_{[xPRON]} \Rightarrow ADJ_{[mod]} \wedge \{X_{[xPRON]} \prec ADJ_{[mod]}\}$
γ_2: $X_{[xPRON]} \Rightarrow \{NOUN_{[mod]} \vee PROPN_{[mod]}\} \wedge \{X_{[xPRON]} \prec ADP_{[de]}\}$

Table 7.21 Spanish properties of *PRON*.

Regarding the *PRON* in subject construction, the canonical *PRON* can be found as those elements which can appear alone in the subject construction; that is, without any specifier or modifier. The most frequent *PRONs* which appear alone in subject construction are neutral demonstratives ($PRON_{[demons]}$), personal pronouns ($PRON_{[pers]}$), and relative pronouns ($PRON_{[rel]}$). However, these last ones use to perform as a subject in relative constructions. In Table 7.22, we display the occurrences of the different types of pronoun, which appear alone in subject construction. Many structures have been manually checked in order to see what is the behavior of these pronouns for two reasons:

- Because of the low frequency of the *PRON* alone in the corpus, we cannot rely on the extracted frequencies at first sight.
- Because MarsaGram does not extract occurrences for those elements which appear alone without being the head of a clause.

We have checked the pronoun as a subject in sentences headed by a verb ($VERB - root$), in relative sentences ($VERB - acl{:}recl$), in conjunctive constructions ($VERB - conj$), in sentences with copulative verbs ($NOUN - root$ and $ADJ - root$).

TYPE OF *PRON*	**Verb** [*root*]	**Verb** [*acl:relcl*]	**Verb** [*conj*]	**Noun** [*root*]	**Adjective** [*root*]	Occurrences Total: 112	%
Demonstrative	26	2	4	12		44	39,2%
Quien		16	2			18	16,07%
Que		16	2			18	16,07%
Personal Pronoun	6	3	3	1		13	11,06%
Ambos	3	1	2	1		7	6,25%
Uno		2	1	1		4	3,5%
Otros	1		1			2	1,78%
Todos	1					1	0,89%
Algunos	1					1	0,89%
Nadie	1					1	0,89%
Muchos				1		1	0,89%
Cuales				1		1	0,89%
Usted					1	1	0,89%
Todos (as *ADJ*)			2				

Table 7.22 Frequencies of the *PRON* alone.

We take into account the total number of the occurrences in which we can find a *PRON* alone and we establish a new frequency of occurrence:

- Demonstrative: 39.20%;
- Relative: 32.14%;
- Personal pronouns: 11.60%;
- *"Ambos"* [*both*]: 6.25%;
- *"Uno"* [*One/Oneself*]: 3.50%;

- *"Otros"* [*others*]: 1.78%;
- Indefinite determiner: <1%.

Therefore, neutral demonstrative, relatives and personal pronouns are the most representative ones because of their gradient frequency. They exclude any complement or specifier; otherwise, a violation will be triggered. The personal pronoun is considered as canonical since it is the typical *PRON*. Their lack of frequency can be explained in order to avoid a repetition of the grammatical person, because Spanish already has the person in its verbal morphology, while in English the PRON is strictly necessary because the verbal morphology does not contain information related to the grammatical person: *"Yo estudio"* [*I study*]. On the other hand, the 34.35% of the precedence between *DET* and the *PRON* is governed by pronominal locutions such as *"el cual; la cual; los cuales; las cuales; cada uno; el que; la que; lo que"*. Tremblay (2012) suggests that these *locutions* are processed as one word. Therefore, a *PRON* should, in this case, be considered to be formalized in the lexicon as one word instead of two separate words.

In Table 7.23, we present an extract of the corpus with some constructions with *PRON* as part of the subject construction. It exemplifies some cases in which this category performs a subject excluding any other element.

PRON (without mods) [*nsubj*] in *VERB* [*root*]		
TYPE	NUMBER	OCCURRENCES
Personal Pronoun	58/6590:1, 59/203:1, 59/2328:1, 59/5249:1, 90/6223:1, 90/7532:1	6
"**Él** será un amigo para Daisuke y un enemigo transformado en Krad" (6590:1). "**Yo** creo que tienen un buen mecánico, y que no es nada caro" (203:1).		
Demonstratives as pronouns	58/515:1, 58/12341:1,58/10367:1, 58/13797:1. 58/3024:1, 58/7303:1, 58/1154:1, 58/12553:1. 59/13217:1, 59/7820:1, 59/173:1, 59/267:1, 59/3118:1, 59/6188:1, 59/12355:1, 59/4987:1, 59/6550:1, 90/13314:1, 90/5080:1, 90/6944:1, 90/4153:1, 170/2509:1, 170/3032:1, 211/2934:1, 211/6482:1, 211/8389:1.	26
"**Este** contiene la versión de The WideBoys Club Mix, la versión original que aparece en el álbum y otra de corte instrumental" (515:1). "**Eso** hizo que el pequeño grupo de agentes de la Seguridad del Estado que habitualmente las acompaña para evitar incidentes, fuera incrementado a un fuerte cordón policial en las marchas siguientes" (7820:1). "**Ello** provocó que Olmedo se dirija hacia Perú mostrando su desacuerdo con Simón Bolivar" (6550:1).		
Ambos	58/12606:1, 90/14055:1, 170/8606:1.	3
"**Ambos** comparten una sucia y oscura celda en una prisión brasileña" (12606:1).		
Todos	58/11303:1	1
"**Todas** tienen control de estabilidad y airbag de rodillas para el conductor" (11303:1)		
Algunos	59/344:1	1
"**Algunos** creen que el incremento de la participación de los gobiernos locales en crear la política conducirá a una mayor propiedad de los programas de préstamos, por lo tanto a una mejor política fiscal" (344:1).		
Otras	90/14097:1	1
"**Otras** consisten en un cuartelado de Castilla, Sicilia (sólo con el águila imperial), Aragón, y León" (14097:1)		
Nada	170/7589:1	1
"**Nada** se sabía del virrey Lord Eglin, quien se encontraba de viaje e incomunicado, por lo que miembros del gobierno de Calcuta impulsaron una retirada general" (7589:1)		

Table 7.23 Examples from UD dependencies of *PRON* as subject.

Besides, while checking in Spanish UD the *PRON* as subject taking into account clauses, we found out the recurrence of the pronoun *"lo"* modified by an adjective as a subject construction. This reason why we have specified the properties α_2, α_3, α_4 with $PRON_{[lo]}$, indicating that these properties are applied just for this pronoun. Therefore, the properties of this pronoun are satisfied when:

- $PRON_{[subj][lo]} \prec ADJ_{[mod]}$: a pronoun lo precedes an adjective as a modifier.
- $PRON_{[subj][lo]} \Rightarrow ADJ_{[mod]}$: a pronoun lo requires an adjective as a modifier.
- $PRON_{[subj][lo]} \otimes ADP \land DET \land ADV$: a pronoun lo excludes an APD, and a DET and an ADV.

Finally, the last canonical property that we define is the dependency relation between the pronoun and the verb.

Regarding the violated constraints and its variability properties, variability could trigger a diacritic meaning such as *"el yo"* (generic meaning: *philosopical I*) marked in γ_1. However, the representation of this case in the corpus is meaningless in terms of occurrences. Nonetheless, it has been considered that γ_1 is the typical exception regarding the co-occurrence between a pronoun and a determiner. In such manner, γ_1 specifies that: in case the property α_1 is violated by a *DET* occurring with a *PRON*, this violation β_1, which is defined as, $\neg(PRON \otimes DET)$, requires that the violated pronoun must be the personal pronoun of *"yo"* and this personal pronoun must be preceded by the determiner *"el"*.

On the other hand, γ_2 is modeling an exception as well. In γ_2, we define that in case the exclusion between the pronoun and the adjective is violated (β_1 $\neg(PRON \otimes ADJ)$), this adjective must be either *"solo"* or *"mismo"*. Some examples concerning these exceptions can be found in Table 7.24. In this figure, there are also cases in which the demonstrative pronoun is preceding the adjective *"último"*. We did not reflect this in our definition of the pronoun as a subject since it seems a case even more particular than the others.

PRON$_{[nsubj]}$ with *ADJ*$_{[amod]}$ [* *ADJ* − *amod*]			
Index 3	STRUCTURE	NUMBER	OCCURRENCES (26)
3.1.(1)	**Personal Pronoun** + *ADJ*	3020:4, 2328:17.	2
	"Ahora actualizamos **nosotros** solos los contenidos de nuestra página web sin pasar por ello" (3020:4).		
3.1.(2)	**Personal Pronoun** + *ADJ*$_{[undef]}$	71822:, 13223:2, 2328:17, 312:22, 1626:24, 14091:21, 2675:27, 12729:72, 4164:53	9
	"Atenea le entregó directamente a Pegaso después de haberlo domado **ella** misma" (7188:22). "**Él** mismo define su método literario como psicología-ficción" (13223:2). "Ella piensa que los límites en la vida solo no[s] los ponemos **nosotros** mismos" (2328:17). "Finalmente decidí buscar los datos de la oficina física de Chrono Express para dirigirme **yo** mismo" (312:22). "**Ellos** mismos me los han fabricado a un precio muy interesante" (14091:21). "Después de mirar bastantes peluquerías dí con este estilista y me encantó, sobre todo porque te atiende **él** mismo no como otros que ponen ayudantes que no saben y chicas sin experiencia que están estudiando peluquería" (2675:27).		
3.2.	**Demonstrative**+*ADJ*	415:18, 9693:28, 144:46, 9637:4, 5231:2, 3760:2	6
	"El Ogro sube y ve a la joven mientras que después **esta** asustada corre pidiendo ayuda con Mike y Jessica" (415:18). "**Esto** último no menoscaba la parte dramática" (9693:28). "Si **éste** último no actuaba conjuntamente en esa solicitud durante un plazo de 60 días, el incremento entraría en efecto automáticamente" (9637:4). "**Esto** último no fue del agrado del Rey Enrique II de Inglaterra" (5231:2). "**Estas** últimas se hacen patentes en la amplia presencia de instrumentos como la guitarra acústica o la flauta" (3760:2).		
3.3.	**Pronoun**$_{[undef]}$+ADJ	8586:38, 7704:16, 5062:41, 9970:8	4
	"Tendría que pasar **algo** muy extraño para que deje este club" (8586:38). "**Algo** extraño ocurrió en el mundo del periodismo deportivo en México" (7704:16). "Ocurrió **algo** muy extraño: Mausolo entregó dinero y regalos al espartiata y dejó el asedio" (5062:41). "A mí me pasó **algo** parecido pero no llegué a pedir que me midieran ya que pedían dinero solo por ir" (9970:8).		
3.4.	**Lo**+*ADJ*	121:23, 8959:2, 261:14	3
	"Este saber, metafísico, comenzó cuando el espíritu humano se hizo consciente de que **lo** real sin más no es lo que nos ofrecen los sentido, sino lo que se aprehende con el pensamiento" (121:23). "**Lo** más notable era una sala con zócalos de mármol cuyo techo estaba pintado por Vicente López más dos cuadros pintados por Ribera" (8959:2). "La relación mieloeritroide varía entre 10:1 y 50:1 (**lo** normal es de 2:1 a 4,5:1)" (261:14).		
3.5.	**Uno** de [los+*ADJ*]	6846:18, 10150:13	2
	"Existen diversos procesos para la destilación del aire líquido, pero **uno** de los más comunes y empleados con mayor frecuencia por la industria es el ciclo de dos columnas de Carl von Linde" (6846:18).		

Table 7.24 Examples from UD dependencies of *PRON* with adjectives.

As we have been defending, this PG is recognizing that variability properties also works for explaining words which undergo a partial transition in terms of part of speech. These transitions concern fuzzy boundaries in parts of speech. Thus, we would assume that the word-class does not undergo a complete transition of membership but more of context. This phenomenon explains why other properties must be taken into account regarding variability. Several *DET* (especially articles and demonstratives) occur as *PRON* quite often, but never as often as they occur as a *DET* (articles: 73.10%; demonstratives: 10,44% in more than 4000 occurrences). If those *DET* ever appear as a *PRON*, this framework detects a violation in the first parsing since, canonically, a DET must precede $NOUN$. Once the violation is detected, a variability property will be triggered clarifying how it is possible to have a determiner without a $NOUN$: Determiner in Specifier Construction: $\neg(DET_{[spec]} \prec NOUN) \implies DET_{[xPRON]\gamma_1 \vee \gamma_2}$. It has to be specified that the variability rules that this new fit must satisfy are either γ_1 or γ_2 because both rules are contradictory themselves. In *xPRON*, γ_1 specifies that an unspecified element as *xPRON*, $X_{[xPRON]}$, requires an adjective and that $X_{[xPRON]}$ has to precede this adjective. We define these constraints all together since they need each other to satisfy the *xPRON*. On the other hand, it is specified in γ_2 that an unspecified element as *xPRON*, $X_{[xPRON]}$, requires a modifier construction, and that $X_{[xPRON]}$ has to precede an *ADP* which must be *"de"*. Therefore, $X_{[xPRON]}$ has to precede the modifier construction introduced by the preposition *"de"*.

In such a case, the determiner is still a determiner, but, at the same time, it is performing a *pronoun fit*. The satisfaction or the violations of the new constraints from the *xCategory* will determine the final degree of grammaticality.

The transitions are possible because of the parts of speech share properties from a multi-modal point of view. In fact, DET and *PRON* share the semantic property of requiring a referent. The canonical determiner needs a $NOUN$ as a real referent while the *PRON* needs a nominal referent in order to make complete semantic sense. If a DET acting like a *PRON* can find a suitable referent, the degree of grammaticality would be significantly reduced in a multi-modal value of gradience and grammaticality.

Otherwise, if a complete transition were to occur, it should be treated in the lexicon as a new lexical entry. For example, *"La roja"* (*the red [one]*) which refers to the Spanish football team selection is considered as a full $NOUN$ but a lexicalized exception. Hence, since in Spanish syntax a DET must precede a $NOUN$, not an adjective.

7.12 Frequencies in the *DET*

$DET\ in\ Specifier\ Construction$
α_1: $DET_{[spec]} \prec NOUN$
α_2: $DET_{[spec]} \Rightarrow NOUN$
α_3: $Uniq$: DET
α_4: $DET_{[spec]} \leadsto_{spec} NOUN$
Variability Properties
γ_1: $\beta_1\ \neg(DET_{[spec]} \prec NOUN) \implies DET_{[xPRON]\gamma_1 \vee \gamma_2}$

Table 7.25 Spanish properties of DET.

The description of the properties of the determiner has been left for the end because it was necessary to represent both *NOUN* and *PRON* before.

The main problem according to extracting the properties of the determiner lies under the fact that many determiners in Spanish language are categorized as undefined determiners. Such elements are named like that because they can function with different fits such as *adjective*, *pronoun*, and *determiner*. We cannot extract frequencies over these specific words since MarsaGram extracts frequencies regarding linguistic categories and clauses. We save the challenge of classifying the undefined adjectives for future work regarding the morphological domain of a fuzzy grammar.

Type of *DET*	Occurrences	%
Articles	2528	**73,1%**
Demonstratives	361	**10,44%**
Possesives	360	**10,41%**
Cuyo	41	1,2%
Algún/a/os/as	31	<1%
Dicho	25	
Otro/a/os/as	22	
*Lo+*ADJ*/PART ('mismo')	16	
Cada	15	
Muchos/as	14	
Ambos/as	9	
Interrogatives	8	
Ningún/a	7	
Varios/as	6	
Todo/a	3	
Cualquier	3	
Cierto/os	3	
Tal/es	2	

Table 7.26 Types of determiners in Spanish.

Since our primary objective is extracting the most general properties of Spanish in order to define categories, in Table 7.26, we have extracted the frequencies of the determiners over 3455 sentences from index 0 regarding the *NOUN* as $nsubj$.

Taking into account Table 7.26, we shall propose the description of the determiner in a specifier construction considering the most frequent cases, which are: *articles, demonstratives* and *possessives*. Nevertheless, with our properties, we can perfectly describe most of the undefined articles. The problem with these last ones is when it comes to link the *xPRON* features to the determiner.

As shown in Table 7.22, the demonstratives appear highly frequently as a pronoun together with other undefined determiners such as *"muchos"*. In Table 7.27, many typical determiners are places in a *xPRON* context. The advantage of acknowledging in Spanish a typical determiner such as *"el"* as a pronoun in these cases is that we are working in fact with language rather than with empty categories or cognitive explanations. We suggest that the determiners perform the typical function of a pronoun which is assigning a referential in this specific context. As the *xPRON* properties indicate, there is not very much flexibility in order to allow a transition of one fit to another. The marked context and the advantage of working with elements which are present in an utterance are the theoretical reasons for considering that the option of a *xPRON* is the most suitable for these cases.

$PRON_{[nsubj]}$ with $NOUN_{[nmod]}$ [* $NOUN - nmod$]				
CASES	STRUCTURE	NUMBER	OCCURRENCES (88)	
			IN CASE	TOTAL
1.1. Relative Pronoun	(a)+**cual** de las+*NOUN*	5346:12	1	1
1.2 $PRON_{[undef]}$	**Nada** de [lo esperado]	1373:3	1	35
	Ninguna de [*DET+NOUN*]	84:40	1	
	Ninguna de las cuales	8965:39	1	
	Ninguno de [*DET+NOUN*]	6669:13	1	
	Varios de [*DET+NOUN*]	5:31	1	
	Muchos de [*DET+NOUN*]	5815:2, 244:2, 6086:2, 6766:2, 8916:14, 1631:2, 1552:22, 677:26, 7161:20, 1131:2, 2640:2, 13153:2, 4910:7, 10957:2, 3628:2, 3462:21, 3473:12	17	
	Algunos en/de [*DET+NOUN*]	12875:61, 4430:8, 841:2, 111:2, 2061:2, 9535:2	6	
	Alguno de [*DET+NOUN*]	12279:15	1	
	Algunas de [*DET+NOUN*]	11887:2	1	
	Otra de [*DET+NOUN*]	3487:2, 12302:2	2	
	Cualquiera de [*DET+NOUN*]	5995:2	1	
	Pocos de [*DET+NOUN*]	11052:5	1	
	Algo de [*DET+NOUN*]	5934:8	1	
1.3. VIOLATION of $DET \prec NOUN$	**El** de [*DET+NOUN* or *PROPN*]	6651:2, 1123:6, 3921:7, 12291:24, 4313:2, 2446:13, 935:17	7	14
	La de [*DET+NOUN* or *PROPN*]	4552:26, 629:28, 1639:40, 11231:51	4	
	Los de [*DET+NOUN*]	5465:27, 1286:18, 13644:13	3	
1.4. $PRON_{[uno/a]}$	**Uno** de [*DET+NOUN*]	13207:2, 8437:6, 7712:2, 6114:2, 207:8, 210:6, 4942:2, 4954:29, 6474:25, 13101:17, 26:2, 8981:2, 12242:6, 12670:2,1307:10, 2449:4, 12164:2, 3586:2, 13178:2	19	37
	Una de [*DET+NOUN*]	5465:2, 7390:18, 12136:2; 12012:3, 12701:2, 1558:21, 27:5, 246:14, 491:4, 219:31, 4036:2, 13175:2, 1704:2, 804:2, 1210:2, 147:2, 954:12, 2911:13	18	

Table 7.27 Determiners as *xPRON* in Spanish.

We provide an example with natural language that displays this variability property from the determiner as *xPRON* in Table 7.28.

Case: $DET_{[xPRON]}$					
Sentence	Las	de	seda	son	mías
Category	$DET_{[spec]}$	$ADP_{[spec]}$	$NOUN_{[mod]}$	$V_{[cop]}$	$ADJ_{[mod]}$
Properties	β_1: $DET_{[spec]} \prec NOUN$ β_2: $DET_{[spec]} \Rightarrow NOUN$ α_3: $Uniq$: DET β_4: $DET_{[spec]} \rightsquigarrow_{spec} NOUN$ γ_1: $\beta_1 \implies$ $DET_{[xPRON]}\gamma_2$ $DET_{[xPRON][subj]}$ γ_2: $DET_{[xPRON]} \Rightarrow NOUN_{[mod]}$ $\wedge \{DET_{[xPRON]} \prec ADP_{[de]}\}$	α_1: $ADP_{[spec]} \prec X_{[mod]}$ α_2: $ADP_{[spec]} \Rightarrow X_{[mod]} \vee V$ α_3: $ADP_{[spec]} \rightsquigarrow_{spec} NOUN_{[mod]}$	α_1: $NOUN_{[mod]} \Rightarrow ADP_{[spec]}$ α_2: $X_{[subj]} \prec NOUN_{[mod]}$ β_3: $NOUN_{[mod]} \Rightarrow DET_{[spec]}$ α_4: $Uniq$: $DET_{[spec]}$ α_5: $NOUN_{[mod]} \otimes ADV \wedge V_{[inf]}$ α_6: $NOUN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$	α_{16}: $V_{[cop]} \Rightarrow PRON_{[subj]}$ α_{17}: $V_{[cop]} \Rightarrow ADJ_{[mod]}$ α_{18}: $PRON_{[subj]} \prec V_{[cop]}$ α_{19}: $V_{[cop]} \prec ADJ_{[mod]}$ α_{22}: $V_{[cop]} \rightsquigarrow_{dep} X_{[mod]}$ α_{23}: $Uniq$: $X_{[subj]}$	
Grammaticality	Cn_w=0.25 Vab_w=0.125 VG_w=0.375	Cn_w=0.333 VG_w=1	Cn_w=0.166 VG_w=0.833	Cn_w=0.166 VG$_w$=1	
	VG_w=0.736 *quite grammatical*				

Table 7.28 Example of FPG application for $DET_{[xPRON]}$.

Table 7.28 shows an example of the grammaticality value of a word *DET* with the feature of a *xPRON* in Spanish. The value of grammaticality of *"las"* $DET_{[spec]}$ has been calculated using the formulas presented in section 6.6. Again, we point out that since we cannot be sure about the different weights between categories, we acknowledge the same weight of 1 for all types of words. Therefore, we apply Equation 6.27 to $DET_{[spec]}$ in Equation 7.12.

$$DET_{[spec]} = 1 \tag{7.23}$$

On one hand, we reveal the canonical value of each constraint of $DET_{[spec]}$ (Cn_w) by our standard value of a canonical property (1). This canonical value of 1 is divided by all the triggered constraints of a the $DET_{[spec]}$ (C_δ), both satisfied (α) and violated (β). In Table 7.28, 4 constraints (either α or β) have been triggered. The variability constraints doesn't count as an additional triggered constraint since it is a consequence of a violated constraint as seen in Equation 6.19. In Equation 7.24, the value of the canonical constraints for $DET_{[spec]}$ in Table 7.28 is calculated following Equation 7.24.

$$0.25\, Cn_w = \frac{1}{C_4} \tag{7.24}$$

On the other hand, we calculate the value of the variability properties triggered as a consequence of the violated canonical properties. The variability triggered properties of the $DET_{[spec]}$ are 1: $\gamma_1{:}\beta_1 \implies DET_{[xPRON]_{\gamma_2}}$. We do not take into account the satisfied variability properties from the *xPRON* since its satisfaction evaluates the possibility of being a *DET* with a *xPRON* feature. The variability properties of *xPRON* in Table 7.21 specify that an *xPRON* has to satisfy either ($\vee$) γ_1 or γ_2. In this case, *DET* satisfies $xPRON_{\gamma_2}$. If the *DET* could not satisfy the variability properties of any of the *xPRON*, a *DET* could not take into account the calculation of the variability property $\gamma_1{:}\beta_1 \implies xPRON_{\gamma_2}$.

In Equation 7.25, the value of a variability property is calculated by dividing the value of a variability property (0.5) by all triggered all satisfied (α) and violated (β) constraints in $DET_{[spec]}$ (*"las"*).

$$0.125\, Vab_w = \frac{0.5}{C_4} \tag{7.25}$$

Finally, we calculate the value of grammaticality of $DET_{[spec]}$ (VG_w) by applying Equation 6.32. For calculating VG_w, we take into account all the canonical weights plus all the variability weights of the word *las* as a $DET_{[spec]}$. In Table 7.15, the satisfied properties (α) which keep the value as canonical properties (0.25 Cn_w) are displayed in Equation 7.24. Since all its satisfied constraints are 1 out 4 we calculate All_{Cn_w} as 0.25. In Table 7.15, the variability properties (γ) which keep the value as variability properties (0.083 Vab_w) are displayed in Equation 7.25. Since all its satisfied constraints are 1 out 1 we calculate All_{Vab_w} as 0.125. The final value of grammaticality of $DET_{[spec]}$ in Table 7.28 is calculated as 0.375 as showed in Equation 7.26.

$$0.375\, VG_w = \frac{(0.25\, All_{Cn_w} + 0.125\, All_{Vab_w})}{1\, DET_{[spec]}} \tag{7.26}$$

We can compute these values with words following the representation of our Fuzzy Grammar applying approximate reasoning in subsection 6.4.2:

- IF an input is *barely* satisfied (0.5-0) $THEN$ the value of grammaticality is *low*.
- IF the value of grammaticality is *low* $THEN$ an input is *barely grammatical*.

When applying this reasoning to our example, it reveals that the variability property could not raise enough the value of grammaticality of $DET_{[spec]}$ in our Fuzzy Grammar:

- Our Fuzzy Grammar considers a value of 0.375 as *an input barely satisfied*; therefore, its value of grammaticality is *low*.

Nevertheless, in this Table 7.28, we can see how, once the $DET_{[xPRON]}$ is identified, it becomes a valid target for being a subject $DET_{[xPRON][subj]}$. The change of the fit is possible because DET satisfies the variability properties of *xPRON*. Besides, the change of the fit allows satisfying the properties from the other words since a *xPRON* allows to satisfy dependencies towards $DET_{[xPRON]}$ just like if the DET was a *PRON*. The $N_{[mod]}$ and the $V_{[cop]}$ would have more violated properties without a *xPRON* with a subject feature. These violated properties in case we would disregard *xPRON* with a subject feature are:

$in\, NOUN\, Modifier\, Construction$

- $\alpha_2\ X_{[subj]} \prec NOUN_{[mod]}$:
- $\alpha_6\ NOUN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$

  These properties would be violated because there would not be a subject preceding $NOUN_{[mod]}$ to satisfy it. Neither a subject to fulfilling the dependency of $NOUN_{[mod]}$.

  Here we have one of those cases in which the $N_{[mod]}$ should not require a DET. Our Fuzzy grammar, as it is now, cannot distinguish in which cases a $NOUN_{[mod]}$ should require a DET or not. We assumed this requirement property because the $NOUN_{[mod]}$ use to require DET in general. However, we need further research on this requirement property in order to acknowledge in which cases this requirement has to be satisfied or not. Probably, the solution to this problem would be found through a multi-modal grammar which takes into account the semantic and the pragmatic module all-together with the syntactic properties.

in VERB Verbal Construction

– α_{16}: $V_{[cop]} \Rightarrow PRON_{[subj]}$
– α_{18}: $PRON_{[subj]} \prec V_{[cop]}$

These properties would be violated because $V_{[cop]}$ couldn't satisfy its requirement of a subject. Besides, α_{18} wouldn't find a subject preceding $V_{[cop]}$ for satisfying its property.

We calculate the value of grammaticality (VG_w) of *"seda"* $NOUN_{[mod]}$ (VG_w) by applying Equation 6.32. For calculating VG_w, we take into account all the canonical weights plus all the variability weights of the word *"seda"* as a $NOUN_{[mod]}$. In Table 7.28, the satisfied properties (α) which keep the value as canonical properties are calculated as: $Cn_w = 0.166$. Since all its satisfied constraints are 5 out 6 we calculate All_{Cn_w} as 0.833. The word *"seda"* as a $NOUN_{[mod]}$ doesn't display any variability property. The final value of grammaticality of this $NOUN_{[mod]}$ in Table 7.28 is calculated as 0.833 as showed in Equation 7.27.

$$0.833\,VG_w = \frac{(0.833\,All_{Cn_w} + 0\,All_{Vab_w})}{1\,NOUN_{[mod]}} \tag{7.27}$$

In case $NOUN_{[mod]}$ wouldn't have a word with a subject feature in the subject construction, it would violate two properties as we have seen before. Since all its satisfied constraints would be 3 out 6 we calculate All_{Cn_w} as 0.498 and its value of grammaticality would be 0.498 as shown in Equation 7.11:

$$0.498\,VG_w = \frac{(0.498\,All_{Cn_w} + 0\,All_{Vab_w})}{1\,NOUN_{[mod]}} \tag{7.28}$$

Therefore, thanks to the representation of $DET_{[xPron][subj]}$, we can compute with words the $NOUN_{[mod]}$ as:

- Our Fuzzy Grammar considers a value of 0.833 as *an input significantly satisfied*; therefore, its value of grammaticality is *high*. The input is *significantly grammatical*.
- In contrast, the value of the input *"seda"* as a $NOUN_{[mod]}$ would be 0.498 without our variability property in $DET_{[xPron][subj]}$. In this manner, the input *"seda"* $NOUN_{[mod]}$

without our variability property in our fuzzy grammar would be computed as *an input which is barely satisfied*. Therefore, its value of grammaticality would be *low*. The input is *barely grammatical*.

In a manner, the word *son* $VERB_{[cop]}$ represents a similar case. We calculate the value of grammaticality (VG_w) of *"son"* $V_{[cop]}$ (VG_w) by applying Equation 6.32. In Table 7.28, the satisfied properties (α) which keep the value as canonical properties are calculated as: $Cn_w = 0.166$. Since all its satisfied constraints are 6 out 6, we calculate All_{Cn_w} as 1. Therefore, VG_w in *son* $V_{[cop]}$ is 1.

In case $V_{[cop]}$ wouldn't have a word with a subject feature in the subject construction, it would violate two properties. Since all its satisfied constraints would be 4 out 6, we calculate All_{Cn_w} as 0.664 and its value of grammaticality would be 0.664 since it doesn't trigger any variability property as shown in Equation 7.11:

$$0.664\, VG_w = \frac{(0.664\, All_{Cn_w} + 0\, All_{Vab_w})}{1\, V_{[cop]}} \qquad (7.29)$$

Therefore, thanks to the representation of $DET_{[xPron][subj]}$, we can compute with words the $V_{[cop]}$ as:

- Our Fuzzy Grammar considers a value of 1 as *an input significantly satisfied*; therefore, its value of grammaticality is *high*. The input is *significantly grammatical*.

- In contrast, the value of the input *"son"* as a $V_{[cop]}$ would be 0.664 without our variability property in $DET_{[xPron][subj]}$. In this manner, it would be computed as *an input which is quite satisfied*. Therefore, its value of grammaticality would be *medium*. The input is *quite grammatical*.

Consequently, we calculate the final value of grammaticality of the subject construction considering all the elements which take place in it. We apply the formula showed in Equation 6.34. We take into account all the VG_w values dividing them by all the words in the construction: $\{DET_{[spec]} = 0.375\, ADP_{[spec]} = 1 NOUN_{[mod]} = 0.833\}$. The addition of all the VG_w of this construction is 2.208.

$$All_{VG_w}\, \{DET_{[spec]} = 0.375\, ADP_{[spec]} = 1\, NOUN_{[mod]} = 0.833\} = 2.208 \qquad (7.30)$$

We divide this amount by all the words which appear in the evaluated construction: $\{DET, ADP, NOUN\} = 3$ *words*. The result of the grammaticality value of the construction is 0.736 as seen in Table 7.28 and in Equation 7.31.

$$0.736\,VG = \frac{(2.208\,All_{VG_w})}{All_3} \tag{7.31}$$

Therefore, the subject construction in Table 7.28 can be computed with words as:

- The value of grammaticality of the subject construction *"las de seda"* is 0.736. This input is *quite satisfied*, therefore, its value of grammaticality is *medium*. The input is *quite grammatical*.

To summarize, the final value of grammaticality of the whole input *"Las de seda son mías"* displays a very different grammaticality value regarding the variability properties display in it in contrast with if they would not be represented. We take into account all the VG_w values in the input dividing them by all the words in the construction: $\{DET_{[spec]} = 0.375$, $ADP_{[spec]} = 1$, $NOUN_{[mod]} = 0.833$, $V_{[cop]} = 1$, $ADJ_{[mod]} = 1\}$. The addition of all the VG_w in this construction is 4.208.

$$\mathrm{All}_{VG_w}\{DET_{[spec]} = 0.375\; ADP_{[spec]} = 1\; NOUN_{[mod]} = 0.833\; V_{[cop]} = 1\; ADJ_{[mod]} = 1\} = 4.208 \tag{7.32}$$

We divide this amount by all the words which appear in the evaluated construction: $\{DET, ADP, NOUN, V, ADJ\} = 5\ words$. The result of the grammaticality value of the construction is 0.841.

$$0.841\,VG = \frac{(4.208\,All_{VG_w})}{All_5} \tag{7.33}$$

However, in case of disregarding the variability properties, the values of grammaticality of all words would be the following ones:

$$\mathrm{All}_{VG_w}\;\{DET_{[spec]} = 0.25\; ADP_{[spec]} = 1\; NOUN_{[mod]} = 0.498\; V_{[cop]} = 0.664\; ADJ_{[mod]} = 1\} = 3.412 \tag{7.34}$$

The final value of grammaticality of this construction would be 0.648 as shown in Equation 7.35.

$$0.648\,VG = \frac{(3.412\,All_{VG_w})}{All_5} \tag{7.35}$$

Therefore, the subject construction in Table 7.28 can be computed with words as:

- The final value of grammaticality of the input *"las de seda son mías"* is 0.841. This input is *significantly satisfied*, therefore, its value of grammaticality is *high*. The input is *significantly grammatical*.
- On the contrary, the final value of grammaticality of the input *"las de seda son mías"* would be 0.648 without considering the variability properties. This input would be *quite satisfied*; therefore, its value of grammaticality would be *medium*. The input would be *quite grammatical*.

The variability properties, together with the notions of both positive cumulativity and ganging-up effect, have considerably raised the final value of grammaticality from a 0.648 to a 0.841. The fuzzy grammar is displayed in this example as a system which can calculate positive values of grammaticality regarding linguistic competence without considering any acceptability judgment. Additionally, it can compute with words the values of grammaticality, which is a much more natural manner of expressing them than solely with numbers.

7.13 A Fuzzy Property Grammar for the Spanish Subject Construction

We have shown our method of extraction, and we have fit the Universal Dependency Treebank and MarsaGram to the necessities of this research. We have extracted a proof-of-concept of a Fuzzy Property Grammars applied to the Subject construction in Spanish. We have applied our Fuzzy Grammar to some natural language inputs for illustrating how it can represent inputs regarding linguistic competence and degrees of grammaticality, without taking into account performance, degrees of acceptability and degrees of ungrammaticality. Besides, we consider the two following claims which reflect the challenge of extracting a Property Grammar:

- It matters where the properties are placed. The extraction of the properties for the Spanish language is a sensible process for both working with the corpus and placing the properties in order to define a grammar which does not take clear grammatical cases with a value of grammaticality under than 1. However, the challenge is at the same time to do so without acknowledging fewer properties than the ones which should be defined.
- The benefits of working with *xCategories*. For example, thanks to the use of an *ADP* as a specifier, we can explain how the noun can perform as an *adjective fit*, in a

construction as a modifier. Regarding Table 7.15, the variability constraint counts as one because the variability rule in $NOUN_{[mod]}$ specifies that it triggers all the constraints in *xADJ*. Therefore, in case some of these constraints would not be satisfied, this variability would not be applied. That is why we separate the constraints which count on calculating grammaticality and the others that just matter for including the variability rule of the $NOUN_{[mod]}$, γ_1.

Our Fuzzy Property Grammar has many other unique idiosyncrasies:

- It displays three constraint behaviors: *satisfied*, *violated* and *variability constraints*.
- It defines the structure of a word in terms of degrees.
- It can compute with words the degrees of grammaticality.
- It extracts degrees of grammaticality regarding the category of a word with a positive value of grammaticality.
- It extracts degrees of grammaticality regarding both syntactic constructions and full-expressions.

In what follows, we present an overview of our extracted Fuzzy Property Grammar.

$NOUN\ in\ Subject\ Construction$
α_1: $DET_{[spec]} \prec NOUN_{[subj]}$ α_2: $NOUN_{[subj]} \Rightarrow DET_{[spec]}$ α_3: $Uniq$: $DET_{[spec]}$ α_4: $NOUN_{[subj]} \otimes ADV \wedge PRON \wedge V_{[inf]}$ α_5: $NOUN_{[subj]} \rightsquigarrow_{subj} V$
Variability Properties
γ_1: $\beta_2 \implies NOUN_{[subj]} \Rightarrow NOUN_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
$xNOUN\ in\ Subject\ Construction$
γ_1: $DET_{[spec]} \prec X_{[xNOUN]}$ γ_2: $X_{[xNOUN]} \Rightarrow DET_{[spec]}$ γ_3: $Uniq$: $DET_{[spec]}$ γ_4: $X_{[xNOUN]} \otimes ADV_{[mod]} \wedge PRON \wedge V_{[inf]}$ γ_5: $X_{[xNOUN]} \rightsquigarrow_{subj} V$

Table 7.29 Spanish properties of *NOUN* in subject construction.

$ADJ\ in\ Modifier\ Construction$
α_1: $NOUN_{[subj]} \prec ADJ_{[mod]}$ α_2: $ADJ_{[mod][num:ord] \vee [undef]} \prec NOUN_{[subj]}$ α_3: $ADJ_{[mod]} \rightsquigarrow_{mod} NOUN_{[subj]}$ α_4: $ADJ_{[mod][todos/as]} \prec DET_{[spec][art:def:pl] \vee [poss:pl] \vee [demons:pl]} \rightsquigarrow_{spec} NOUN_{[subj]}$ α_5: $ADJ_{[mod][todo/a]} \prec DET_{[poss:s] \vee [demons:s]} \rightsquigarrow_{spec} NOUN_{[subj]}$ α_6: $ADJ_{[mod][todo/a]} \prec DET_{[spec][art:undef:s]} \rightsquigarrow_{spec} NOUN_{[mod]}$ α_7: $ADV_{[mod][casi]} \rightsquigarrow_{mod} ADJ_{[mod][todo/a/os/as]}$
Variability Properties
γ_1: $\beta_1 \implies ADJ_{[mod][semantics]}$
$xADJ\ in\ Modifier\ Construction$
γ_1: $NOUN_{[subj]} \prec X_{[xADJ]}$ γ_2: $X_{[xADJ]} \rightsquigarrow_{mod} NOUN_{[subj]}$

Table 7.30 Spanish properties of *ADJ* in modifier construction.

$NOUN\ in\ Modifier\ Construction$
α_1: $NOUN_{[mod]} \Rightarrow ADP_{[spec]}$ α_2: $X_{[subj]} \prec NOUN_{[mod]}$ α_3: $NOUN_{[mod]} \Rightarrow DET_{[spec]}$ α_4: $Uniq$: $DET_{[spec]}$ α_5: $NOUN_{[mod]} \otimes ADV \wedge V_{[inf]}$ α_6: $NOUN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$
Variability Properties of $NOUN\ in\ Modifier\ Construction$
γ_1: $\beta_1 \implies Noun_{[xADJ]}$
$ADP\ in\ Specifier\ Construction$
α_1: $ADP_{[spec]} \prec X_{[mod]}$ α_2: $ADP_{[spec]} \Rightarrow X_{[mod]} \vee V$ α_3: $ADP_{[spec]} \rightsquigarrow_{spec} NOUN_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]}$

Table 7.31 Spanish properties of both *NOUN* in modifier and *ADP* in specifier construction.

PROPN in Subject Construction
α_1: $PROPN_{[subj]} \otimes DET \wedge NOUN \wedge PRON \wedge ADJ \wedge ADV \wedge SCONJ \wedge V_{[non]}$ α_2: $PROPN_{[subj]} \rightsquigarrow_{subj} V$
Variability Properties
γ_1: β_1 $\neg(PROPN_{[subj]} \otimes DET_{[spec]}) \implies PROPN_{[xNOUN]}$ γ_2: β_1 $\neg(PROPN_{[subj]} \otimes V_{[non]}) \implies PROPN_{[subj]} \Rightarrow V_{[part]} \Rightarrow X_{[mod]}$ γ_3: β_1 $\neg(PROPN_{[subj]} \otimes PRON) \implies PROPN_{[subj]} \Rightarrow PRON_{[rel]}$

Table 7.32 Spanish properties of *PROPN* in subject construction.

PROPN in Modifier Construction
α_1: $X_{[subj]} \prec PROPN_{[mod]}$ α_2: $PROPN_{[mod]} \otimes DET \wedge NOUN \wedge PRON \wedge ADJ \wedge ADV \wedge V_{[non]}$ α_3: $PROPN_{[mod]} \rightsquigarrow_{mod} X_{[subj]}$ α_4: $PROPN_{[mod]} \Rightarrow ADP_{[spec]}$
Variability Properties
γ_1: β_2 $\neg(PROPN_{[mod]} \otimes DET_{[spec]}) \implies PROPN_{[xNOUN]}$ γ_2: β_2 $\neg(PROPN_{[mod]} \otimes V_{[non]}) \implies PROPN_{[mod]} \Rightarrow V_{[part]} \Rightarrow X_{[mod]}$ γ_3: β_2 $\neg(PROPN_{[mod]} \otimes PRON) \implies PROPN_{[mod]} \Rightarrow PRON_{[rel]}$

Table 7.33 *PROPN* in modifier construction.

PRON in Subject Construction
α_1: $PRON_{[subj][demons]\vee[rel]\vee[pers]} \otimes ADP \wedge ADJ \wedge ADV \wedge DET \wedge PRON$ α_2: $PRON_{[subj][lo]} \prec ADJ_{[mod]}$ α_3: $PRON_{[subj][lo]} \Rightarrow ADJ_{[mod]}$ α_4: $PRON_{[subj][lo]} \otimes ADP \wedge DET \wedge ADV$ α_5: $PRON_{[subj]} \rightsquigarrow_{subj} V$
Variability Properties
γ_1: β_1 $\neg(PRON_{[subj]} \otimes DET) \implies DET_{[el]} \prec PRON_{[subj][yo]}$ γ_2: β_1 $\neg(PRON_{[subj]} \otimes ADJ) \implies ADJ_{[mod][solo]\vee[mismo]}$
xPRON in Subject Construction
γ_1: $X_{[xPRON]} \Rightarrow ADJ_{[mod]} \wedge \{X_{[xPRON]} \prec ADJ_{[mod]}\}$ γ_2: $X_{[xPRON]} \Rightarrow \{NOUN_{[mod]} \vee PROPN_{[mod]}\} \wedge \{X_{[xPRON]} \prec ADP_{[de]}\}$

Table 7.34 Spanish properties of *PRON*.

$DET\ in\ Specifier\ Construction$
α_1: $DET_{[spec]} \prec NOUN$ α_2: $DET_{[spec]} \Rightarrow NOUN$ α_3: $Uniq$: DET α_4: $DET_{[spec]} \rightsquigarrow_{spec} NOUN$
Variability Properties
γ_1: β_1 $\neg(DET_{[spec]} \prec NOUN) \implies DET_{[xPRON]\gamma_1 \vee \gamma_2}$

Table 7.35 Spanish properties of DET.

Verb in *Verbal Construction*
$Verb_{[transitive]}$
α_1: $V_{[trans]\vee[pass]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
α_2: $V_{[trans]} \Rightarrow N_{[dobj]} \vee PROPN_{[dobj]} \vee PRON_{[dobj]} \vee V_{[inf][dobj]} \vee SubC_{[dobj]}$
α_3: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[trans]\vee[pass]}$
α_4: $V_{[trans]} \prec N_{[dobj]} \vee PROPN_{[dobj]} \vee PRON_{[dobj]} \vee V_{[inf][dobj]} \vee SubC_{[dobj]}$
α_5: $PRON_{[dobj]} \prec V_{[trans]}$
α_6: $PRON_{[iobj]} \prec V_{[trans]\vee[pass]}$
α_7: $V_{[trans]} \prec CONJ_{[conj]}]$
α_8: $Uniq$:$X_{[subj]}$
α_9: $Uniq$: $X_{[dobj]}$
$Verb_{[intrans]}$
α_{10}: $V_{[intrans]} \otimes X_{[dobj]}$
α_{11}: $V_{[intrans]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
α_{12}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[intrans]}$
α_{13}: $V_{[intrans]} \prec X_{[mod]}$
α_{14}: $PRON_{[iobj]} \prec V_{[intrans]}$
α_{15}: $Uniq$: $X_{[subj]}$
$Verb_{[cop]}$
α_{16}: $V_{[cop]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \vee SubC_{[subj]}$
α_{17}: $V_{[cop]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
α_{18}: $N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]} \prec V_{[cop]}$
α_{19}: $V_{[cop]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]} \vee ADJ_{[mod]} \vee SubC_{[mod]}$
α_{20}: $V_{[cop]} \wedge ADJ_{[mod]} \prec SubC_{[subj]}$.
α_{21}: $SubC_{[subj]} \otimes SubC_{[mod]}$
α_{22}: $V_{[cop]} \rightsquigarrow_{dep} X_{[mod]}$.
α_{23}: $Uniq$: $X_{[subj]}$
$Verb_{[passive]}$
α_{24}: $V_{[pass]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
α_{25}: $V_{[pass]} \Rightarrow N_{[mod:por]} \vee PROPN_{[mod:por]} \vee PRON_{[mod:por]}$
α_{26}: $V_{[pass]} \rightsquigarrow_{mod} N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
α_{27}: $Uniq$: $X_{[mod:por]}$
$Verb_{[se:auxpass]}$
α_{28}: $V_{[se:auxpass]} \prec N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
α_{29}: $V_{[se:auxpass]} \Rightarrow N_{[subj]} \vee PROPN_{[subj]} \vee PRON_{[subj]}$
α_{30}: $Uniq$: $X_{[subj]}$
$Verb_{[se:auximp]}$
α_{31}: $V_{[se:auximp]} \prec X_{[mod]}$
α_{32}: $V_{[se:auximp]} \otimes X_{[subj]}$
Variability Properties
γ_1: $\beta_{1\vee 11\vee 16\vee 29} \implies X_{[subj]}$ in $V_{[morph:pers:num]}$
γ_2: $\beta_{1\vee 11\vee 16\vee 29} \implies V \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$
γ_3: $\beta_{25} \implies V_{[pass]} \Rightarrow N_{[mod]} \vee PROPN_{[mod]} \vee PRON_{[mod]}$

Table 7.36 Table of Spanish properties of verb in verbal construction.

This grammar has extracted the following cases:

- 32 canonical properties for 6 types of Verb construction. 3 variability properties for 6 types of verb construction.
- 5 canonical properties for the noun (*NOUN*) as subject, and 1 variability property.
- 7 canonical properties for the adjective (*ADJ*), and 1 variability property.
- 6 canonical properties for the noun (*NOUN*) as a modifier and 1 variability property.
- 3 canonical properties for the preposition (*ADP*) as a specifier.
- 2 canonical properties for the proper noun (*PROPN*) as subject and 3 variability properties.
- 4 canonical properties for the proper noun (*PROPN*) as a modifier and 3 variability properties.
- 5 canonical properties for the pronoun (*PRON*) and 2 variability properties.
- 4 canonical properties for the determiner (*DET*), and 1 variability property.
- 5 variability properties for *xNOUN*.
- 2 variability properties for the *xADJ*
- 2 variability properties for the *xPRON*.

Our Fuzzy Property Grammar has extracted a final amount of 68 canonical properties.

Our grammar is fuzzy because we extracted the following borderline cases: 15 variability properties in seven standard categories and 9 variability properties in three *xCategories*: *xNOUN*, *xADJ*, *xPRON*. Therefore, our Fuzzy Property Grammar has extracted the final amount of 24 variability properties.

Consequently, our Fuzzy Property Grammar has extracted a total amount of 92 properties. The variability properties represent 26,08% of our grammar. Hence, our Fuzzy Property Grammar can capture a 26,08% better the linguistic phenomena in natural language in contrast with a discrete grammar. Moreover, it can calculate degrees of grammaticality because of the variability properties which are display on it.

All these properties are a proof-of-concept of the *linguistic knowledge* that a speaker should have for being *competent* for the acknowledgment of the subject construction in the Spanish Language.

Part III

Conclusions

Chapter 8

Conclusions

We have started our work with the three research questions:

1. Is it possible to measure the degrees of grammaticality of a given linguistic input?
2. Which are the best formal tools to calculate the different levels of grammaticality?
3. Can we provide a fuzzy approach to the concept of grammaticality (not acceptability) that takes into account competence (but not performance)?

From these research questions, we have developed a hypothesis that we have tested in our work; that is, a formal model which combines fuzzy logic and a grammar with constraints can represent fuzzy degrees of grammaticality regarding linguistic competence.

The main aim of this work was to introduce a new linguistic model which can both represent and calculate degrees of grammaticality by considering the notion of grammaticality as a fuzzy phenomenon of the natural language. In order to achieve this main aim we have set up the following five research objectives.

1. To review and provide a critical analysis of the concept of grammaticality in linguistics.
2. To review and provide a critical analysis of the models of gradience over the notion of the concept of grammaticality.
3. To propose the bases for the definition of a fuzzy grammar to deal with degrees of grammaticality and their fuzzy features.
4. To extract/determine the linguistic properties that will define the linguistic inputs that take into account the degrees of grammaticality.

5. To provide a proof-of-concept of a fuzzy grammar with properties for Spanish syntax which can represent the fuzzy degrees of grammaticality.

In what follows we will give our conclusions for each of the objectives into which our thesis has been divided.

Regarding the first objective (*To review and provide a critical analysis of the concept of grammaticality in linguistics*), we can say the following.

The distinction of competence and performance from Chomsky has determined the scientific approaches in linguistics as well as the notions of grammaticality and acceptability.

Some trends in linguistics are focused on building grammars to generate language more than building grammars which can either define or both define and generate natural language. Thus, regarding this objective, underpinning a grammar from either a discrete or categorical point of view seems quite legitimate. However, the fact is that "the role of a generative grammar of a natural language is not merely to generate grammatical sentences of a language but to relate them to their logical forms" (Lakoff, 1970). Consequently, the role of any grammar is also to represent all the linguistic concepts and inferences that are made in natural language with adequate logical forms. Since humans decode both noisy text and grammatical deviations in natural language processing, any grammar needs to do the same. Thus, whether linguists want to explain how natural language processing works, they must explain how our linguistic competence permits us to understand either "wild" utterances or utterances with grammatical violations, together with linguistic concepts and inferences. For this reason, the linguist is forced to both face and deal with deviated utterances when a natural language grammar is being built. In order to do that, it is necessary to review the concepts of competence and performance and to propose both new definitions for them and for the notions of degrees of grammaticality and performance.

We conclude that it is essential to highlight two characteristics regarding the concept of grammaticality:

1. Grammaticality is the value that represents the extent to which an input is satisfied according to the linguistic knowledge in terms of constraints that defines the competence of a natural language grammar. Consequently, it is not a value which depends on the judgment of a speaker; it is a value which can be calculated within a grammar.

2. Grammaticality is a fuzzy-gradient value which must be explained in terms of degrees. The value is fuzzy because of all the criteria that are involved in it. It is a melting pot of linguistic constraints and linguistic modules in interaction which needs precise tools to separate everything in order to be able to calculate the value of every piece before the estimation of the final value.

Syntactic models should give up the notion of grammaticality as a discrete concept and adapt to the "real" use of language in which the grammaticality is a matter of degrees. Therefore, we argue for the definition of tools that formalize the degrees of grammaticality so we can analyze real data even if this contains "noisy text" which appears in social network language, as well in every spontaneous use of language.

If linguists want to explain natural language processing, they must explain why humans can understand utterances with grammatical violations. Why can the speakers extract meaning and why does the act of communication not fail given all the variability during an interaction. From this point of view, degrees of grammaticality matter for any natural language grammar.

If we want natural language processing techniques to be able to analyze ungrammatical inputs, they need to be shaped through a new concept of grammaticality already introduced in linguistics. Even though theoretical linguistics has established grammaticality as a categorical notion, many linguists have insisted on defending the gradual essence of language (Aarts, 2004a; Aarts et al., 2004; Bolinger, 1961; Lakoff, 1973). Gradient approaches are defined in several linguistic models such as: Harmonic Grammars (Legendre et al., 1990a); Optimality Theory (Prince and Smolensky, 1993); Probability Theory (Manning, 2003); Property Grammars (Blache, 2016); Gradient Grammars (Keller, 2000).

These examples state the interest in dealing with natural language as a non-discrete object. We argue that a formal theory of fuzzy grammar is possible and these frameworks have to be transferred to the parsing algorithms. Thus, we have to note that the primary objective is developing a formal framework which captures the idea of fuzzy set and is computationally accepted.

Regarding our second objective (*To review and provide a critical analysis of the models of gradience over the notion of the concept of grammaticality*), our main conclusions are the following.

Many approaches in gradience have a fundamental problem: they acknowledge the degree of grammaticality and degree of ungrammaticality as the same thing.

It is important to highlight that the degrees of *ungrammaticality* are not the same as the degrees of *grammaticality*. A degree of ungrammaticality implies that speakers are able to understand the unintelligible, which is a paradox. Furthermore, the assumption of the degree of ungrammaticality implies discrete grammaticality for the following reasons:

- Grammaticality does not have degrees of well-formedness, because an input without deviations is grammatical while an input with deviations is ungrammatical. Therefore, grammaticality is discrete.

- Ungrammaticality has degrees of ill-formedness because once an input has violated constraints, it has a degree of un-grammaticality. Therefore, ungrammaticality is gradient.

In terms of the present thesis, theoretically, as long as a linguistic object still satisfies certain constraints of a grammar, those satisfied constraints are grammatical; hence, the linguistic input is still grammatical concerning that grammar. Therefore, it is not true that an input has a degree of ungrammaticality since it has still satisfied grammatical constraints.

We recognize that the notion of ungrammaticality is a reinterpretation of the concept of acceptability because acceptability presents a broader spectrum since it depends on subjective values which are much more difficult to scale. The notion of grammaticality reaches its end when a linguistic input completely fails to satisfy the constraints of a grammar; that is, all its constraints have been violated.

We point out that the data from acceptability judgments cannot model grammaticality since that data provided on them is determined by both linguistic and extra-linguistic elements together with the subjective perspective of a speaker. Such features contrast with grammaticality which is a theoretical notion full-based on rule satisfaction without taking into account extra-linguistic considerations.

We conclude that including grammatical weights in our fuzzy grammar is controversial since our proposal gives the basis of a new approach in linguistics which is not fully developed. The grammatical weights pose some problems:

- We do not precisely know what the weights are, and which kind of calculations and elements are involved in determining them.

- Neither do we know the importance of the extra-linguistic features regarding the "grammatical weights".

- We cannot assure that one constraint is heavier than other through acceptability judgments since acceptability judgments are a value provided by a speaker who is involuntarily taking into account both linguistic and extra-linguistic objects.

- We do not know if the frequency is strictly related to the weight of grammaticality. Frequency could be modeling tolerance towards violated inputs. That is, a constraint which is equally essential regarding the linguistic competence is *perceived* as worst than another one because a speaker is used to process that violated constraint.

- A fuzzy grammar is a multi-modal approach; therefore, we cannot be sure of the weights of the constraints without taking into account all the linguistic domains.

Regarding our third objective (*To propose the basis for the definition of a fuzzy grammar to deal with degrees of grammaticality and their fuzzy features*), we made a distinction between what is uncertain and what is vague or fuzzy. A fuzzy mechanism would help us to deal with grammaticality since vagueness has (positive) degrees of truth regarding the membership of an object in a group.

Grammaticality IS a vague concept, so we have used a fuzzy grammar to define it.

The degree of grammaticality has a positive value since grammaticality is a vague object which determines the extent to which an input belongs to a grammar in terms of degrees of truth.

Consequently, an input is a vague object which is more or less true depending onthe number of linguistic elements that it satisfies and violates in a specific grammar.

We provide the following definitions which are the basis for a Fuzzy Grammar with constraints:

- A fuzzy grammar (FGr) is considered as a fuzzy set ($\underset{\sim}{\subseteq}$) on the whole set of linguistic rules or constraints. These constraints are the linguistic knowledge of the fuzzy grammar in every module (linguistic domain). Therefore, a fuzzy grammar is multi-modal.

- $X_\gamma \rightarrow (D_\eta \rightarrow M_o)$ is the formal representation of how a module from a fuzzy grammar takes into account degrees regarding linguistic competence and grammaticality, not performance. Linguistic competence is understood as all the constraints in the fuzzy grammar. The fuzzy grammar matches every constraint from the grammar with every constraint from a linguistic input. From this relation, both the constraints on the module and the constraints on the input match a degree of membership of being grammatical.

- We have modeled with approximate fuzzy reasoning linguistic expressions of grammaticality providing a more natural interpretation of the degrees of grammaticality: we express that an input is *significantly grammatical* when its degree is 1-0.8: an input is *quite grammatical* when its degree is 0.8-0.5, and an input is *barely grammatical* when its degree is 0.5-0.

- We have modeled the structure of a *word* with fuzzy logic by representing precisely the moment in which our property grammar takes its part. Property Grammars are a theory based on the concept of *word*; they rely on the notion of syntactic categories. Therefore, just after a word is categorized ($Mr_{\gamma\beta}$), our fuzzy grammar with properties can take part in order to describe its fuzzy properties. Once the linguistic definition of our property grammar regarding the linguistic category is done, a word can be assigned a degree of grammaticality regarding its syntactic degree of satisfaction: $X_{o\gamma}$.

- We have proposed a model which can extract values of grammaticality regarding a single word, a construction, or various sentences. Firstly, we acknowledge each category as a word which has a whole full value of grammaticality. Secondly, we can extract the grammaticality value of an *input* regarding its words. Our approach can evaluate any *input*. This system provides a *positive calculation for fuzzy degrees of grammaticality*.

Regarding our fourth objective (*To extract/determine the linguistic properties that will define the linguistic inputs for taking into account the degrees of grammaticality*), we can say the following.

Our proposal is based on the framework of Property Grammars. However, in order to determine and extract the linguistic properties for a fuzzy grammar, we have added some features. The following claims made it possible to determine the linguistic properties for taking into account the degrees of grammaticality:

- We have presented three constraint behaviors. These are necessary for adapting both fuzzy notions and a Property Grammar into a Fuzzy Property Grammar: *Syntactic Canonical Properties*: These are the properties which define the gold standard of the Fuzzy Grammar; *Syntactic Violated Properties*: These properties are canonical properties which have been violated regarding a linguistic *input* or a dialect; and *Syntactic Variability Properties*: these are borderline cases in-between a violation and a canonical use. They explain linguistic variability concerning a fuzzy grammar. When a variability property is satisfied, it triggers a new value over the violated constraint, thus improving its degree of grammaticality.

- We have offered a formulation of the process of variability constraints. We link variability constraints to violated constraints. When this match occurs, the variability properties are checked. If these are satisfied, a new degree of grammaticality will be provided

- We have provided the notion of a *xCategory* which specifies that a specific category is displaying a *syntactic fit* from another category. This concept is necessary for a fuzzy property grammar; it describes the fuzzy idiosyncrasy of some linguistic categories which can acquire different properties depending on their context in an utterance.

- We have also provided a standard for defining a category by properties and a numeration system in order to add computational value to the proposal.

- We have extracted the properties based on theoretical notions for modeling gradient data such as *context effects*, *cumulativity*, *ganging-up effect*, *constraint counterbalance* and *positive ganging-up effect*, *density* and *multi-modal values*.
- We have added two operations for Property Grammars for determining the properties:
  - $\vee$: This operation is understood as *or*. It defines a category and its property concerning many different categories (or features) at the same time. This operation avoids over-violation of properties. This operation defines that a category solely needs one element of the whole description in order to be satisfied.
  - $\wedge$: This operation is understood as *and*. It defines a category and its property concerning many different categories (or features) at the same time. Therefore, all the elements must be satisfied, or it will trigger a violation. This operation helps with over-satisfaction. Therefore, when various categories are involved in one property, and this property is violated, the other categories remain as inactive without over-satisfying the value of grammaticality.
- We accept that frequency cannot model grammaticality due to some paradoxes. If we built grammaticality solely on the basis of frequencies, we would be forced to say such things like "the use of the personal pronoun as a subject in Spanish is non-grammatical because his frequency in the corpus is almost 0". However, no native Spanish speaker would consider such a thing to be the case. The frequency of occurrence is a guide which helps with the theoretical decisions in order to select the properties for calculating the degrees of grammaticality.

Finally, regarding our fifth objective (*To provide a proof-of-concept of a fuzzy grammar with properties for Spanish syntax which can represent the fuzzy degrees of grammaticality*), we have provided a proof-of-concept of Spanish Subject construction with a property grammar which takes into account both vague boundaries and properties.

With this methodology, we have defined the properties of the verb construction; the noun, the proper noun and the pronoun in subject construction; the determiner and the preposition in a specifier construction; and the most frequent modifier constructions in Spanish subjects such as the noun and the proper noun as modifiers and the adjective.

These descriptions are fuzzy since we took into account as *borderline cases* 24 variability properties and 3 *xCategories*: *xNOUN*, *xADJ*, *xPRON*. These properties and features make possible the application of the formula of a fuzzy grammar. Furthermore, we have shown the applicability of these variability properties in certain cases to show how our Fuzzy Property Grammar calculates degrees of grammaticality.

Consequently, our Fuzzy Property Grammar has extracted a total amount of 92 properties. The variability properties represent 26,08% of our grammar. Hence, our Fuzzy Property Grammar can capture a 26,08% better the linguistic phenomena in natural language in contrast with a discrete grammar. Additionally, it computes a fine-grained positive value of grammaticality in terms of degrees.

All these properties are a proof-of-concept of the *linguistic knowledge* that a speaker should have for being *competent* for the acknowledgment of the subject construction in the Spanish Language.

The work developed in this dissertation allows us to answer our three research questions in the following way:

- **Is it possible to measure the degrees of grammaticality of a given linguistic input?**

  Yes, it is. However, since our proposal is just laying down the basis of a Fuzzy Property Grammar, we could not apply any computational method to acknowledge the performance of our proposal with a significant amount of sentences from different sources such as social media, blogs, and public reviews.

- **Which are the best formal tools to calculate the different levels of grammaticality?**

  We consider satisfactory the application of the following tools: Fuzzy Natural Logic proposed by Novák (2015); Property Grammars introduced by Blache (2016), the grammatical-fuzzy features proposed in this research in order to combine both fuzzy logic and grammar with constraints.

  Fuzzy Natural Logic and Property Grammars need each other to represent the degrees of grammaticality. This is why we call our system *Fuzzy Property Grammars*, because it could not exist without both of these theories.

- **Can we provide a fuzzy approach to the concept of grammaticality (not acceptability) taking into account competence (not performance)?**

  Yes. Combining fuzzy logic tools with grammar formalisms with constraints allow us to calculate grammaticality through the satisfaction of the constraints in a specific grammar.

Finally, we have verified our hypothesis. We can affirm that a formal model which combines fuzzy logic and grammar with constraints can represent fuzzy degrees of grammaticality regarding linguistic competence. The concept of grammaticality is a theoretical one, therefore, with a formal tool such as Fuzzy Natural Logic we would be able to define grammaticality regarding a grammar; without involving any acceptability trait on the formal definition.

Chapter 9

Future Work

The research developed in this dissertation opens new directions for future work. Firstly, the research would benefit from a more computational approach. The Fuzzy Property Grammar would need to be continuously tested by sentences in order to check which properties to dismiss and which ones need to be added. Secondly, it would be necessary to extract property grammars from other linguistic modules, particularly from semantics and pragmatics. Thirdly, a robust framework of a fuzzy grammar will have a direct impact on the successful creation of new formal tools from which several applications will benefit. New fuzzy formulas would help to define the notion of a word with its meaning. This would provide the significant benefit of structuring which constraints are needed for every linguistic domain.

Taking all this into account, many future benefits might come from the application of this new approach regarding the concept of degrees of grammaticality in a fuzzy grammar.

- **Representing Language Processing**: A fuzzy algorithm could explain the cognitive and psychological way that humans process language. In the future, after creating such an algorithm, we would collaborate with a psycho-linguistic team to develop test-type experiments to contrast our artificial model with the linguistic judgments that the speakers make naturally. This is necessary for the empirical verification of the formal model and to verify if the proposed algorithm is adapted to the cognitive reality of natural language and, therefore, works similarly to how humans process language.

- **Electronic Devices and Soft-wares** that require of dialogue systems (oral or written speech) for Artificial Intelligence. This point is translated into artificial assistants (i.e., Siri) that can have multiple applications: artificial web assistants; artificial assistants for driving; artificial administrative assistants, and others.

- **Web Data Mining**: This is possibly the area that receives the most investment nowadays. Social media language poses a significant challenge to parsing algorithms.

The problem of parsing the noisy language of social media opens a range of possibilities to natural language processing and leads linguists to reflect again, from a new perspective, on the problem of the notion of grammaticality. This new scenario requires interdisciplinary collaboration between different areas, such as linguistics, formal language theory, psycho-linguistics, computational linguistics, and so on. The collaboration between those disciplines can provide a new theoretical framework on fuzzy grammaticality that combines formalization, linguistic theory, and evidence from psycho-linguistics. The implementation of a fuzzy grammar on parsing algorithms could improve not only the analysis of web language but also the analysis of language in general. A new fuzzy algorithm based on linguistic theory would allow us to propose an innovative solution in this field.

Language Self-Learning Programs: A grammar taking into account degrees of grammaticality will have a direct impact on the creation of software for self-learning of languages. This is possible since the algorithm developed could analyze the text produced by the student and provide an assessment of the error on a scale of 1 to 10 by pointing out the grammatical rule violated and providing the optimal linguistic rule. This would also be useful in order to perform an automatic assessment concerning the level of competence a speaker has in a given language.

Computer tools for the Automatic Detection of Language pathologies: An algorithm that takes into account the degree of grammaticality of the sentences analyzed would automatically assess the degree of impairment of the patient's linguistic capacity. The program would associate the violated rules with a specific pathology. The program would assess the degree of damage concerning the severity of the language violations.

References

Aarts, B. (2004a). Conceptions of gradience in the history of linguistics. *Language Sciences*, 26(4):343–389.

Aarts, B. (2004b). Modelling linguistic gradience. *Studies in Language*, 28(1):1–49.

Aarts, B. (2007). *Syntactic gradience: The nature of grammatical indeterminacy*. Oxford University Press, Oxford.

Aarts, B., Denison, D., Keizer, E., and Popova, G., editors (2004). *Fuzzy grammar: A reader*. Oxford University Press, Oxford.

Abney, S. P. (1987). *The English noun phrase in its sentential aspect*. PhD thesis, Massachusetts Institute of Technology, Cambridge.

Ajdukiewicz, K. (1935). Die syntaktische konnexität. *Studia Philosophica*, 1:1–27.

Almela Pérez, R. (2003). Bases para una morfología continua del español. *ELUA. Estudios de Lingüística*, 17:57–79.

Altmann, G. T. and Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264.

Andersson, L.-G. and Trudgill, P. (1990). *Bad language*. Penguin Books, London.

Aristotle (1984). The Methaphysics. In Barnes, J., editor, *The Complete works of Aristotle: the revised Oxford translation*, volume 2, page 1552. Princeton University Press, Princeton, NJ.

Askedal, J. O., Roberts, I., and Matsushita, T. (2010). *Noam Chomsky and language descriptions*, volume 2. John Benjamins Publishing, Amsterdam.

Bader, M. and Häussler, J. (2010). Toward a model of grammaticality judgments. *Journal of Linguistics*, 46(02):273–330.

Baldwin, T., Cook, P., Lui, M., MacKinlay, A., and Wang, L. (2013). How noisy social media text, how diffrnt social media sources? In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, pages 356–364, Nagoya.

Balfourier, J.-M., Blache, P., and van Rullen, T. (2002). From shallow to deep parsing using constraint satisfaction. In *Proceedings of the 19th International Conference on Computational Linguistics*, volume 1, pages 1–7, Taipei.

Bard, E. G., Robertson, D., and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72(1):32–68.

Baroni, M. and Lenci, A. (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

Bever, T. G. (1975). Functional explanations require independently motivated functional theories. In Grossman, R. E., San, L. J., and Vance, T. J., editors, *Papers from the Parasession on Functionalism*, pages 580–609. Chicago Linguistic Society, Chicago.

Bever, T. G. and Carroll, J. M. (1981). On some continuous properties in language. *Advances in Psychology*, 7:225–233.

Blache, P. (2000). Property grammars and the problem of constraint satisfaction. In *Proceedings of ESSLLI 2000 workshop on Linguistic Theory and Grammar Implementation*, pages 47–56, Birmingham.

Blache, P. (2004). Property grammars: A fully constraint-based theory. In Christiansen, H., Skadhauge, P. R., and Villadsen, J., editors, *Constraint Solving and Language Processing*, volume LNAI 3438, pages 1–16. Springer, Berlin.

Blache, P. (2006). A robust and efficient parser for non-canonical inputs. In *Proceedings of ROMAND-06*, pages 27–32, Trento.

Blache, P. (2016). Representing syntax by means of properties: a formal framework for descriptive approaches. *Journal of Language Modelling*, 4(2):183–224.

Blache, P. and Azulay, D.-O. (2002). Parsing ill-formed inputs with constraint graphs. In Gelbukh, A., editor, *Computational Linguistics and Intelligent Text Processing*, volume LNAI 2276, pages 220–229. Springer, Berlin.

Blache, P. and Balfourier, J.-M. (2001). Property grammars: a flexible constraint-based approach to parsing. In *7th International Conference on Parsing Technologies, IWPT 2001*, Beijing.

Blache, P., Bertrand, R., and Ferré, G. (2008). Creating and exploiting multimodal annotated corpora: the toma project. In Kipp, M., Martin, J.-C., Paggio, P., and Heylen, D., editors, *Multimodal Corpora*, volume LNAI 5509, pages 38–53. Springer, Berlin.

Blache, P., Christiansen, H., Duchier, D., Dahl, V., and Villadsen, J. (2014). *Constraints and Language*. Cambridge Scholars Publishing, Newcastle upon Tyne.

Blache, P., Guénot, M.-L., and van Rullen, T. (2003). A corpus-based technique for grammar development. In *Proceedings of International Conference on Corpus Linguistics*, pages 124–131. ICCL.

Blache, P., Hemforth, B., and Rauzy, S. (2006). Acceptability prediction by means of grammaticality quantification. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 57–64. Association for Computational Linguistics.

Blache, P. and Prévot, L. (2010). A formal scheme for multimodal grammars. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 63–71. Association for Computational Linguistics.

Blache, P. and Prost, J.-P. (2008). A quantification model of grammaticality. In Villadsen, J. and Christiansen, H., editors, *Fifth International Workshop on Constraints and Language Processing (CSLP2008)*, volume 122, Computer Science Research Reports, pages 5–19. Roskilde University.

Blache, P., Rauzy, S., and Montcheuil, G. (2016). MarsaGram: an excursion in the forests of parsing trees. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, pages 2336–2342, Portorož, Slovenia.

Boersma, P. (2004). A stochastic OT account of paralinguistic tasks such as grammaticality and prototypicality judgments. *Rutgers Optimality Archive 648*.

Boersma, P. and Hayes, B. (2001). Empirical tests of the gradual learning algorithm. *Linguistic Inquiry*, 32(1):45–86.

Bolinger, D. L. M. (1961). *Generality: Gradience and the All-or-none*. Mouton Publishers, The Hague, 14 edition.

Borsley, R. D. and Kornfilt, J. (2000). Mixed extended projections. In Borsley, R. D., editor, *The nature and function of syntactic categories (Syntax and Semantics, vol. 32)*, pages 101–131. Brill, Amsterdam.

Bouchard, D. (1995). *The semantics of syntax: A minimalist approach to grammar*. University of Chicago Press, Chicago.

Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Featherston, S. and Sternefeld, W., editors, *Roots: Linguistics in Search of its Evidential Base*, volume 96, Studies in Generative Grammar, pages 75–96. Mouton de Gruyter, Berlin.

Bresnan, J. (2011). Linguistic uncertainty and the knowledge of knowledge. In Porter, R. and Reynolds, R., editors, *Thinking Reed. Centenial Essays by Graduates of Reed College*, pages 69–75. Reed College, Portland, OR.

Bresnan, J., Cueni, A., Nikitina, T., and Baayen, H. (2007). Predicting the dative alternation. In Boume, G., Kraemer, I., and Zwarts, J., editors, *Cognitive Foundations of Interpretation*, pages 69–94. Royal Netherlands Academy of Science, Amsterdam.

Bresnan, J. and Nikitina, T. (2003). The gradience of the dative alternation. *Ms., Stanford University, http://www-lfg. stanford. edu/bresnan/download. html*.

Bresnan, J. and Nikitina, T. (2009). The gradience of the dative alternation. In Uyechi, L. and Wee, L. H., editors, *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, pages 161–184. The University of Chicago Press, Chicago.

Buszkowski, W., Marciszewski, W., and Van Benthem, J. (1988). *Categorial Grammar*. John Benjamins Publishing, Amsterdam.

Bybee, J. and Moder, C. L. (1983). Morphological classes as natural categories. *Language*, 59(2):251–270.

Chersoni, E., Blache, P., and Lenci, A. (2016). Towards a distributional model of semantic complexity. In Brunato, D., Dell'Orletta, F., Venturi, G., François, T., and Blache, P., editors, *Proceedings of the Workshop on Computational Linguistics for Linguistic Complexity*, pages 12–22, Osaka. ACL.

Chersoni, E., Torrens Urrutia, A., Blache, P., and Lenci, A. (2018). Modeling violations of selectional restrictions with distributional semantics. In *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing*, pages 20–29, Santa Fe. ACL.

Chomsky, N. (1957). *Syntactic Structures*. Walter de Gruyter, Berlin.

Chomsky, N. (1961). Some methodological remarks on generative grammar. *Word*, 17(2):219–239.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge.

Chomsky, N. (1969). Quine's empirical assumptions. In Davidson, D. and Hintikka, J., editors, *Words and Objections*, pages 53–68. Springer, Berlin.

Chomsky, N. (1975). *The Logical Structure of Linguistic Theory*. Plenum Press, Aarhus.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Greenwood Publishing Group, Westport.

Chomsky, N. (1995). *The Minimalist Program*. MIT Press, Cambridge.

Chomsky, N. (1998). Some observations on economy in generative grammar. In Barbosa, P., Fox, D., Hagstrom, P., McGinnis, M., and Pesetsky, D., editors, *Is Best Good Enough? Optimality and Competition in Syntax*, pages 115–127. MIT Press, Cambridge.

Chomsky, N. (2002). *On Nature and Language*. Cambridge University Press, Cambridge.

Chomsky, N. and Halle, M. (1968). *The Sound Pattern of English*. ERIC. Education Resources Information Center, USA.

Chomsky, N., Huybregts, R., and van Riemsdijk, H. C. (1982). *Noam Chomsky on the generative enterprise: A discussion with Riny Huybregts and Henk van Riemsdijk*. Foris Publications, USA.

Corver, N. and van Riemsdijk, H. C. (2001). Semi-lexical categories. In Corver, N. and van Riemsdijk, H. C., editors, *Semi-lexical Categories: The Function of Content Words and the Content of Function Words*, pages 1–19. Mouton de Gruyter, Berlin.

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. Sage, Thousand Oak.

Crystal, D. (1967). English. *Lingua*, 17(1):24–56.

Curme, G. O. (1935). *A Grammar of the English Language, vol. II: Parts of Speech and Accidence*. DC Heath and Co, Boston.

Dahl, V. and Miralles, J. E. (2012). Womb grammars: Constraint solving for grammar induction. In Sneyers, J. and Frühwirth, editors, *Proceedings of the 9th Workshop on Constraint Handling Rules*, volume CW 624, report, pages 32–40, Leuven. Katholieke Universiteit Leuven.

Daneš, F. (1966). The relation of centre and periphery as a language universal. *Travaux Linguistiques de Prague*, 2:9–21.

Debusmann, R., Postolache, O., and Traat, M. (2005). A modular account of information structure in extensible dependency grammar. In Gelbukh, A., editor, *Computational Linguistics and Intelligent Text Processing*, pages 25–36. Springer, Berlin.

Duchier, D., Prost, J.-P., and Dao, T.-B.-H. (2011). A model-theoretic framework for grammaticality judgements. In de Groote, P., Egg, M., and Kallmeyer, L., editors, *Formal Grammar*, volume LNAI 5591, pages 17–30. Springer, Berlin.

Dunn, J. (2017). Computational learning of construction grammars. *Language and Cognition*, 9(2):254–292.

Eisenstein, J. (2013). What to do about bad language on the internet. In *Proceedings of the 2013 conference of the North American Chapter of the association for computational linguistics: Human language technologies*, pages 359–369, Atlanta. ACL.

Erk, K., Padó, S., and Padó, U. (2010). A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics*, 36(4):723–763.

Fanselow, G., Féry, C., Vogel, R., and Schlesewsky, M. (2006). Gradience in grammar. In Fanselow, G., Féry, C., Schlesewsky, M., and Vogel, R., editors, *Gradience in Grammar: Generative Perspectives*, pages 1–23. Oxford University Press, Oxford.

Ferretti, T. R., McRae, K., and Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language*, 44(4):516–547.

Fillmore, C. J. (1988). The mechanisms of “construction grammar”. In Axmaker, S., Jaisser, A., and Singmaster, H., editors, *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society*, volume 14, pages 35–55, Berkeley. Berkeley Linguistics Society.

Gazdar, G., Klein, E., Pullum, G. K., and Sag, I. A. (1985). *Generalized Phrase Structure Grammar*. Harvard University Press, Cambrige.

Gibson, E. and Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3):225–248.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago.

Goldberg, A. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5):219–224.

Goldberg, A. (2009). The nature of generalization in language. *Cognitive Linguistics*, 20(1):1–35.

Goldsmith, J. (1993). *The Last Phonological Rule: Reflections on Constraints and Derivations*. University of Chicago Press, Chicago.

Greenbaum, S. (1980). Syntactic frequency and acceptability. In Perry, T. A., editor, *Evidence and Argumentation in Linguistics*, pages 301–314. Walter de Gruyter, Berlin.

Greenberg, C., Sayeed, A., and Demberg, V. (2015). Improving unsupervised vector-space thematic fit evaluation via role-filler prototype clustering. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 21–31, Denver. ACL.

Grice, H. P. (1975). Logic and conversation. In Cole, P. and Morgan, J. L., editors, *Syntax and Semantics, Vol. 3, Speech Acts*, pages 41–58. Academic Press, New York.

Guénot, M.-L. and Blache, P. (2005). A descriptive and formal perspective for grammar development. In *Foundations of Natural-Language Grammar*, Edinburgh, United Kingdom.

Guy, G. (1997). Violable is variable: Optimality theory and linguistic variation. *Language variation and Change*, 9(3):333–347.

Hájek, P. and Novák, V. (2003). The sorites paradox and fuzzy logic. *International Journal of General Systems*, 32(4):373–383.

Hare, M., Jones, M., Thomson, C., Kelly, S., and McRae, K. (2009). Activating event knowledge. *Cognition*, 111(2):151–167.

Harris, R. A. (1993). *The Linguistics Wars*. Oxford University Press, Oxford.

Hayes, B. (1997). Gradient well-formedness in optimality theory. In *Unpub. Handout.* Department of Linguistics, University of California, Los Angeles.

Hayes, P. J., Mouradian, G. V., and George V (1981). Flexible parsing. *American Journal of Computational Linguistics*, 7(4):232–242.

Hockett, C. F. (1955). *A Manual of Phonology*. Waverly Press, Baltimore.

Huddleston, R. (1984). *Introduction to the Grammar of English*. Cambridge University Press, Cambridge.

Huddleston, R. and Pullum, G. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press, Cambridge.

Hudson, R. (2003). Gerunds without phrase structure. *Natural Language & Linguistic Theory*, 21(3):579–615.

Hudson, R. (2004). Are determiners heads? *Functions of Language*, 11(1):7–42.

Jackenddoff, R. (2002). *Foundations of Language*. Oxford University Press, New York, NY.

Jacobson, P. (2004). Direct Compositionality: Is there any reason why not? *Workshop on Linguistics and Philosophy, University of Michigan*.

Jacobson, P. (2007). Direct compositionality and variable-free semantics. In Barker, C. and Jacobson, P., editors, *Direct Compositionality*, pages 191–236. Oxford University Press, Oxford.

Jespersen, O. (1924). *The Philosophy of Grammar*. George Allen And Unwin Ltd, London.

Joos, M. (1958). *Readings in Linguistics. The Development of Descriptive Linguistics in America since 1925*. American Council of Learned Societies, New York.

Joshi, A. K., Levy, L. S., and Takahashi, M. (1975). Tree adjunct grammars. *Journal of Computer and System Sciences*, 10(1):136–163.

Kaplan, R. M. and Bresnan, J. (1982). Lexical-functional grammar: A formal system for grammatical representation. In Dalrymple, M., Kaplan, R. M., Maxwell, J. T., and Zaenen, A., editors, *Formal Issues in Lexical-Functional Grammar*, pages 29–130. CSLI Publications, Stanford.

Kaschak, M. P. and Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, 43(3):508–529.

Kasper, G. (1992). Pragmatic transfer. *Interlanguage Studies Bulletin (Utrecht)*, 8(3):203–231.

Kay, P. (2002). An informal sketch of a formal architecture for construction grammar. *Grammars*, 5(1):1–19.

Kay, P. and Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: the What's X doing Y? construction. *Language*, 75(1):1–33.

Keller, F. (1997). *Degrees of Grammaticality as Evidence for Suboptimal Linguistic Structures*. Universität Stuttgart, Fakultätsübergreifend/Sonstige Einrichtung.

Keller, F. (1998). Gradient grammaticality as an effect of selective constraint re-ranking. In Gruber, M. C., Higgins, D., Olson, K. S., and Wysocki, T., editors, *Papers from the 34th meeting of the Chicago Linguistic Society*, volume 2, pages 95–109. Chicago Linguistic Society, Chicago.

Keller, F. (2000). *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality*. PhD thesis, University of Edinburgh, Edinburgh.

Keller, F. (2002). Probabilistic learning algorithms and optimality theory. *Linguistic Inquiry*, 33:225–244.

Keller, F. (2006). Linear Optimality Theory as a model of gradience in grammar. In Fanselow, G., Féry, C., Schlesewsky, M., and Vogel, R., editors, *Gradience in Grammar: Generative Perspectives*, pages 270–287. Oxford University Press, Oxford.

Labov, W. (1973). The boundaries of words and their meanings. In Bailey, C. and Shuy, R., editors, *New Ways of Analyzing Variations in English*, pages 340–373. Georgetown University Press, Washington D.C.

Lakoff, G. (1970). Linguistics and natural logic. *Synthese*, 22(1-2):151–271.

Lakoff, G. (1973). Fuzzy grammar and the performance/competence terminology game. In Corum, C. W., Smith-Stark, T. C., and Weiser, A., editors, *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*. Chicago Linguistic Society, Chicago.

Lakoff, G. (1987a). Cognitive models and prototype theory. In Neisser, U., editor, *Emory Symposia in Cognition, 1. Concepts and Conceptual Development: Ecological and Intellectual Factors in Categorization*, pages 63–100. Cambridge University Press, Cambridge.

Lakoff, G. (1987b). *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press, Chicago.

Lau, J. H., Clark, A., and Lappin, S. (2014). Measuring Gradience in Speakers' Grammaticality Judgements. In *Proceedings of the Annual Meeting of the Cognitive Science Society, 36*, pages 821–826, Austin. Cognitive Science Society.

Lavie, A. (1996). *GLR*: A Robust Grammar-Focused Parser for Spontaneously Spoken Language*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh.

Lebani, G. E. and Lenci, A. (2018). A distributional model of verb-specific semantic roles inferences. In Poibeau, T. and Villavicenci, A., editors, *Language, Cognition, and Computational Models*, pages 118–158. Cambridge University Press, Cambridge.

Legendre, G., Miyata, Y., and Smolensky, P. (1990a). Harmonic Grammar: A formal multi-level connectionist theory of linguistic well-Formedness: An application. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, pages 884–891, Cambridge. Lawrence Erlbaum Associates.

Legendre, G., Miyata, Y., and Smolensky, P. (1990b). Harmonic grammar: A formal multi-level connectionist theory of linguistic well-formedness: Theoretical foundations. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, pages 388–395, Cambridge. Lawrence Erlbaum Associates.

Legendre, G., Miyata, Y., and Smolensky, P. (1991). Unifying syntactic and semantic approaches to unaccusativity: A connectionist approach. In Sutton, L. A., Johnson, C., and Shields, R., editors, *Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society*, volume 17, pages 388–395. Berkeley Linguistics Society.

Lesmo, L. and Torasso, P. (1984). Interpreting syntactically ill-formed sentences. In *Proceedings of the 10th International Conference on Computational Linguistics and 22nd annual meeting on Association for Computational Linguistics*, pages 534–539, Stanford. Association for Computational Linguistics.

Luria, A. R. (1980). *Fundamentos de Neurolingüística*. Toray-Masson, Barcelona.

Malouf, R. (1999). Verbal gerunds as mixed categories in head-driven phrase structure grammar. In Borsley, R., editor, *The Nature and Function of Syntactic Categories*, volume 32, pages 133–166. Brill, Leiden.

Malouf, R. (2000). *Mixed Categories in the Hierarchical Lexicon*. CSLI Publications, Stanford.

Manning, C. D. (2003). Probabilistic syntax. In Bod, R., Hay, J., and Jannedy, S., editors, *Probabilistic Linguistics*, pages 289–341. MIT Press, Cambridge.

Matsuki, K., Chow, T., Hare, M., Elman, J. L., Scheepers, C., and McRae, K. (2011). Event-based plausibility immediately influences on-line language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(4):913.

Matthews, P. (1981). *Syntax*. Cambridge University Press, Cambridge.

McCarthy, J. J. and Pater, J. (2016). *Harmonic Grammar and Harmonic Serialism*. Equinox, Sheffield.

McRae, K., Hare, M., Elman, J. L., and Ferretti, T. (2005). A basis for generating expectancies for verbs from nouns. *Memory & Cognition*, 33(7):1174–1184.

McRae, K., Spivey-Knowlton, M. J., and Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38(3):283–312.

Montague, R. (1970). Universal grammar. *Theoria*, 36(3):373–398.

Moure, T. (1994). *La Alternativa No-Discreta en Lingüística: Fundamentos Metodológicos y Proyección sobre los Conceptos Gramaticales*. PhD thesis, Universidade de Santiago de Compostela, Santiago de Compostela.

Muñoz Cuadros, R. (2006). Hacia una ¿nueva? lingüística: reflexiones sobre la llamada alternativa no-discreta. *Pandora: Revue d'Etudes Hispaniques*, 6:37–58.

Müller, G. (1999). Optimality, markedness, and word order in german. *Linguistics*, 37(5):777–818.

Müller, S. (2016). *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches.* Language Science Press, Berlin.

Murinová, P. and Novák, V. (2016). Syllogisms and 5-square of opposition with intermediate quantifiers in fuzzy natural logic. *Logica Universalis*, 10(2):339–357.

Neustupnỳ, J. V. (1966). On the analysis of linguistic vagueness. *Travaux Linguistiques de Prague*, 2:39–51.

Newmeyer, F. J. (1983). *Grammatical Theory: Its Limits and its Possibilities*. University of Chicago Press, Chicago.

Newmeyer, F. J. (1998). *Language Form and Language Function*. MIT Press, Cambridge.

Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language*, 79(4):682–707.

Novák, V. (2005). On fuzzy type theory. *Fuzzy Sets and Systems*, 149:235–273.

Novák, V. (2007). Mathematical fuzzy logic in modeling of natural language semantics. In Wang, P., Ruan, D., and Kerre, E., editors, *Fuzzy Logic – A Spectrum of Theoretical & Practical Issues*, pages 145–182. Elsevier, Amsterdam.

Novák, V. (2008a). A comprehensive theory of trichotomous evaluative linguistic expressions. *Fuzzy Sets and Systems*, 159(22):2939–2969.

Novák, V. (2008b). A formal theory of intermediate quantifiers. *Fuzzy Sets and Systems*, 159(10):1229–1246.

Novák, V. (2008c). Mathematical fuzzy logic: From vagueness to commonsese reasoning. In Kreuzbauer, G., Gratzl, N., and Hielb, E., editors, *Retorische Wissenschaft: Rede und Argumentation in Theorie und Praxis*, pages 191–223. LIT-Verlag, Wien.

Novák, V. (2015). Fuzzy natural logic: Towards mathematical logic of human reasoning. In Seising, R., Trillas, E., and Kacprzyk, J., editors, *Towards the Future of Fuzzy Logic*, pages 137–165. Springer, Berlin.

Novák, V. (2016). Fuzzy natural logic: Theory and applications. In *Fuzzy Sets and Their Applications FSTA 2016*, Liptovský Ján.

Novák, V. and Dvorák, A. (2011). Fuzzy logic: A powerful tool for modeling of vagueness. In *Logical Models of Reasoning with Vague Information*, pages 1–28, Cejkovice.

Novák, V. and Lehmke, S. (2006). Logical structure of fuzzy IF-THEN rules. *Fuzzy Sets and Systems*, 157:2003–2029.

Novák, V., Perfilieva, I., and Močkoř, J. (1999). *Mathematical Principles of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht.

Pereira, F. (2000). Formal grammar and information theory: Together again? *Philosophical Transactions of the Royal Society*, 358:1239–1253.

Pinker, S. (1999). *Words and Rules: The Ingredients of Language*. Widenfeld and Nicolson, London.

Pollard, C. and Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. University of Chicago Press, Chicago.

Prince, A. and Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University, New Brunswick.

Prost, J.-P. (2008). *Modelling Syntactic Gradience with Loose Constraint-based Parsing*. PhD thesis, Université de Provence-Aix-Marseille I; Macquarie University.

Prost, J.-P. and Lafourcade, M. (2011). Pairing model-theoretic syntax and semantic network for writing assistance. In *Proceedings of Constraints and Language Processing Workshop (CSLP 2011)*, pages 56–68, Karlsruhe.

Pullum, G. K. (1976). On the nonexistence of the verb-auxiliary distinction in English. *Nottingham Linguistic Circular*, 5(2):20–23.

Pullum, G. K. and Scholz, B. C. (2001). On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In de Groote, P., Morrill, G., and Retoré, C., editors, *Logical Aspects of Computational Linguistics*, volume LNAI 2099, pages 17–43. Springer, Berlin.

Quirk, R. (1965). Descriptive statement and serial relationship. *Language*, 41(2):205–217.

Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman, London.

Rickford, J. R. and Wasow, T. A. (1995). Syntactic variation and change in progress: Loss of the verbal coda in topic-restricting as far as constructions. *Language*, 71(1):102–131.

Rosch, E. (1973a). Natural categories. *Cognitive Psychology*, 4(3):328–350.

Rosch, E. (1973b). On the internal structure of perceptual and semantic categories. In Moore, T. E., editor, *Cognitive Development and Acquisition of Language*, pages 111–144. Academic Press, Amsterdam.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3):192.

Rosch, E. (1978). Principles of categorization. In Rosch, E. and Lloyd, B., editors, *Cognition and Categorization*, pages 27–48. Lawrence Erlbaum Associates, Hillsdale.

Ross, J. R. (1969a). Adjectives as noun phrases. In Reibel, D. A. and Schane, S. A., editors, *Modern Studies in English*, pages 352–360. Prentice-Hall, Englewood Cliffs.

Ross, J. R. (1969b). Auxiliaries as main verbs. In Todd, W., editor, *Studies in Philosophical Linguistics. Series I. Great Expectations*, pages 77–102, Evanston. Great Expectation.

Ross, J. R. (1972). The category squish: endstation hauptwort. In Peranteau, P. M., Levi, J., and Phares, G., editors, *Papers from the 8th Regional Meeting of the Chicago Linguistic Society*, pages 316–328. Chicago Linguistic Society, Chicago.

Ross, J. R. (1973a). A fake np squish. In Bailey, C.-J. N. and Shuy, R. W., editors, *New Ways of Analyzing Variation in English*, pages 96–140. Georgetown University Press, Washington D.C.

Ross, J. R. (1973b). Nouniness. In Kiparsky, P. and Fujimura, O., editors, *Three Dimensions of Linguistic Theory*, pages 137–328. TEC Company Ltd., Tokio.

Ross, J. R. (1974). Three batons for cognitive psychology. In Weimer, W. B. and Palermo, D. S., editors, *Cognition and the Symbolic Processes*, pages 63–124. Lawrence Erlbaum, Oxford.

Ross, J. R. (1987). Islands and syntactic prototypes. In Need, B., Schiller, E., and Bosch, A., editors, *Papers from the 23rd Annual Regional Meeting of the Chicago Linguistic Society (Part I: The General Session)*, pages 309–320, Chicago. Chicago Linguistic Society.

Ross, J. R. (2000). The frozeness of pseudoclefts: towards an inequality-based syntax. In Okrent, A. and Boyle, J., editors, *Papers from the Thirty-Sixth Regional Meeting of the Chicago Linguistic Society*, pages 385–426, Chicago. Chicago Linguistic Society.

Sadock, J. M. (1991). *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. University of Chicago Press, Chicago.

Sadock, J. M. (1998). Grammatical tension. In Gruber, M. C., Higgins, D., and Olson, K. S., editors, *Papers from the 34th Regional Meeting of the Chicago Linguistic Society. CLS 34: The Panels*, pages 179–198, Chicago. Chicago Linguistic Society.

Santus, E., Chersoni, E., Lenci, A., and Blache, P. (2017). Measuring thematic fit with distributional feature overlap. In *arXiv preprint arXiv:1707.05967*. Association for Computational Linguistics, Cornell University.

Schutze, C. T. (1996). *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. University of Chicago Press, Chicago.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press, Cambridge.

Smolensky, P. and Legendre, G. (2006). *The Harmonic Mind: From Neural Computation to Optimality-Theoretic Grammar*, volume 1: Cognitive architecture. MIT Press, Massachusetts.

Smolensky, P., Legendre, G., and Miyata, Y. (1992). *Principles for an integrated connectionist/symbolic theory of higher cognition*. Computer Science Technical Reports 575. University of Colorado at Boulder, Boulder.

Smolensky, P., Legendre, G., and Miyata, Y. (1993). Integrating connectionist and symbolic computation for the theory of language. *Current Science*, 64:381–391.

Sorace, A. and Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115(11):1497–1524.

Sternefeld, W. (1998). Reciprocity and cumulative predication. *Natural Language Semantics*, 6(3):303–337.

Stockwell, R. (1963). Review of "Generality, gradience, and the all-or-none" by Dwight L. Bolinger. *Language*, 39(1):87–91.

Stowell, T. (1989). Subjects, specifiers, and x-bar theory. In Baltin, M. R. and Kroch, A. S., editors, *Alternative Conceptions of Phrase Structure*, pages 232–262. University of Chicago Press, Chicago.

Taylor, J. (1995). *Linguistic Categorization: Prototypes in Linguistic Theory*. Oxford University Press, Oxford.

Tesnière, L. (1959). *Eléments de Syntaxe Structurale*. Klincksieck, Paris.

Tremblay, A. (2012). Empirical Evidence for an Inflationist Lexicon. *Yearbook of Phraseology*, 3(1):109–126.

Trillas, E. and Ríos, J. G. (1992). *Aplicaciones de la Lógica Borrosa*. CSIC Press, Madrid.

Vachek, J. (1966). On the integration of the peripheral elements into the system of language. *Travaux Linguistiques de Prague*, 2:23–37.

Van Riemsdijk, H. (1998). Categorial feature magnetism: The endocentricity and distribution of projections. *The Journal of Comparative Germanic Linguistics*, 2(1):1–48.

Van Riemsdijk, H. (2001). A far from simple matter syntactic reflexes of syntax—pragmatics misalignments'. In Kenesei, I. and Harnish, R. M., editors, *Perspectives on Semantics, Pragmatics, and Discourse. A Festschrift for Ferenc Kiefer*, pages 21–41. John Benjamins, Amsterdam.

Van Rullen, T., Blache, P., and Balfourier, J.-M. (2006). Constraint-based parsing as an efficient solution: Results from the parsing evaluation campaign EASy. In Calzolari, N., Choukri, K., Gangemi, A., Maegaard, B., Mariani, J., Odijk, J., and Tapias, D., editors, *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, pages 165–170, Genoa. European Language Resources Association (ELRA).

Vygotsky, L. S. (1964). Thought and language. *Annals of Dyslexia*, 14(1):97–98.

Wasow, T. (2007). Gradient data and gradient grammars. In *Proceedings of the 43rd the Annual Meeting of the Chicago Linguistic Society*, pages 255–271, Michigan. The Association for Computational Linguistics.

Wells, R. (1958). Is a structural treatment of meaning possible? In Eva, S., editor, *Proceedings of the Eight International Congress of Linguists*, pages 636–704, Oslo. Oslo University Press.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.

Zadeh, L. A. (1972). A fuzzy-set-theoretic interpretation of linguistic hedges. *Journal of Cybernetics*, 2(3):4–34.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8(3):199–249.

Zadeh, L. A. (1996). Fuzzy logic = computing with words. *IEEE Transactions on Fuzzy Systems*, 4:103–111.

Zadeh, L. A. (2004). Precisiated natural language (pnl). *AI Magazine*, 25(3):74–91.