An Approach to Measuring Complexity within the boundaries of a Natural Language Fuzzy Grammar

Adrià Torrens Urrutia

Universitat Rovira i Virgili, Tarragona, Spain
adria.torrens@estudiants.urv.cat

Abstract. This paper presents an approach to evaluate complexity by means of a Natural Language Fuzzy Grammar. Frequently, Linguistics has described a natural language grammar by means of discrete terms. However, a grammar can be explained in terms of degrees by following the concepts of linguistic gradience & fuzziness. Understanding a grammar as a fuzzy or gradient object allows us to establish degrees of grammaticality for every linguistic input. This shall be meaningful for linguistic complexity considering that the less grammatical an input is the more complex its processing will be. From this regard, an input's degree of complexity is always going to depend on its grammar. The bases of the natural language fuzzy grammar are shown here. Some of these are described by Fuzzy Type Theory. The linguistic inputs are characterized by constraints through a Property Grammar.

Keywords: Degrees of Grammaticality, Degrees of Complexity, Fuzzy Grammar, Local Complexity, Syntax

1 Introduction: What is Gradience & Fuzziness?

Fuzziness and gradience are pretty similar (if not the same). Gradience has appeared throughout the history of linguistics and can be 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"[1]. Some well-known studies approach gradience to lingusitic theory, such as Bolinger [2] or Keller [3]. However, it is in mathematics where we can find serious formal approaches to describe gradient relations, such as the gradient relation between *tall-short*, *big-small*. Nevertheless, the gradient phenomena in mathematics are called fuzzy phenomena and fuzzy logic is the right tool to formally describe these vague relations, which are also referred to as fuzziness. Zadeh's [4] [5] mathematical description of gradient phenomena is well known. He describes the variable semantic values of words, or fuzzy phenomena, in terms of degrees. However, Zadeh did not develop a formal linguistic framework to describe fuzziness in natural language grammar. A brief methodological description distinguishing between both terms is shown:

- A *Fuzzy system (fuzzy grammar)* is a formal framework which defines any kind of linguistic information in any context (as humans do). This framework is set through a flexible constraint system which describes a natural language grammar. These constraints are known as properties. They work as logical operators that represent grammatical knowledge. They are flexible because they can be violated or satisfied to different degrees.
- *Processing gradience* refers to our capacity to judge fuzziness through a scale of degrees. The degree represents how hard or soft the violation is. In fuzzy logic, this might be referred as truth values, but since we are talking about language, we are going to talk about linguistic gradience as the truth value of an object.

2 Grammaticality as a topic in Complexity

Nowadays the hypothesis of the "equi-complexity" is not as popular as in the 20th century. In fact, several authors such as Mc Worther [6] or Dhal [7] have challenged this concept. Besides, usually, two different types of complexity are distinguished: absolute complexity and relative complexity. The absolute complexity is defined as a theoretically-oriented approach which evaluates the complexity of a language-system in a whole sense. On the other hand, the relative complexity takes into account the users of the language to identify the difficulty of processing, learning or acquisition. Other authors such as Blache [8] and Lindstrom [9] distinguish between Global complexity, Local complexity, and Difficulty. *Global complexity* is the absolute perspective of complexity. It aims to provide a number to rank a language as a whole system by means of a degree of complexity. This level is purely theoretical and it does not depend on any kind of linguistic realization. Blache in [8] claims that "in Chomskyan terms, this level concern competence", while the local complexity and difficulty belongs to the performance. In contrast, the degree of *local complexity* and *difficulty* are correlated to relative complexity, which is always provided once an input is given. However, local complexity is connected to the linguistic structure and its rules, whereas difficulty is an aspect to take into account for both psycholinguistic approaches and cognitive aspects, which have a role in the complexity evaluation. Within this classification, some authors place grammaticality in difficulty since it is considered a phenomenon of a cognitive aspect from the performance stage. The fact that grammaticality has an important role in the linguistic performance as well as in psycholinguistic approaches is not denied. Nevertheless, in this work, grammaticality is placed as an aspect of the local complexity for two reasons:

- 1) Local complexity is structure-sentence based, and difficulty is speaker-based. In this approach, grammaticality has a tight relation with the structures and the rules of a given input. Consequently, grammaticality belongs to local complexity. However, it has an impact on the difficulty since: the more complex a structure is in terms of grammaticality, the more difficult to process will be.

- 2) The theoretical bases of the Natural Language Fuzzy Grammar allow us to explain grammaticality by means of the grammar of a language itself, independently from the judgment of the speaker. In this instance, grammaticality is strictly based on the rules of the local complexity.

2.1 Grammaticality as an element of Complexity

Linguistics has been highly influenced by the theoretical fragmentation of Competence - Performance from Chomsky's *Aspects* [10]. In general, grammaticality has been considered in two ways:

- *A categorical item*: since the competence is perfect, grammaticality can only be either satisfied or violated by means of the speaker or the receiver during the performance stage.
- *A matter of degrees*: grammaticality would be found as a part of an acceptability judgment. This regard considers that grammaticality is not equal to the whole value of an acceptability judgment, and yet it is an essential part which contributes to the total amount of the degree of acceptability from an input. As well as in the last case, here grammaticality belongs to the performance as well.

However, in the Natural Language Fuzzy Grammar (NFG) approach, the degree of grammaticality is something which is directly related to the grammar. Grammaticality in NFG does not necessarily come through the speaker, nor a performance. Once an input is given, the evaluation of the input is in contrast with the grammar of a language itself. The grammaticality value can be totally isolated from the acceptability judgment from either speaker or a receiver. Thus, in this regard, grammaticality is no longer only a psycholinguistic effect. It is also a direct consequence of a structure in relation to its grammar. In this sense, Grammaticality would play a role in the degree of relative complexity and local complexity. NFG might take into account the complexity of a linguistic structure and its features, such as: number of categories, number of words, number of rules in a structure and degree of grammaticality. In the following section, the base of the Fuzzy Grammar is going to be defined as well as described in a wider sense .

3 An approach to a Fuzzy Grammar with Fuzzy Descriptions for Complexity

First of all, a brief introduction in which the formalisms used are going to be described. Secondly, what is understood as a fuzzy grammar will be explained.

3.1 Brief Introduction to Fuzzy Natural Logic

Fuzzy Natural Logic (FNL) [11] has been used here to describe what a grammar is. FNL is a fuzzy logic system which uses a Fuzzy Type Theory (FTT), a high-order mathematical fuzzy logic, with a Łukasiewicz algebra of truth values (Ł-TT).

Semantics of FNL A few basic notions in relation to FNL are shown:

- FNL uses FTT. This is why this approach is based on describing elements (M) by types (ϵ). Every element has a type which represents the element M_ϵ. An element can be anything: people, objects, languages, etc.
- o: It is an essential type: *omicron.* It can be read as “degree of truth $\langle [0, 1]\rangle$ [1]”. The degree of truth might represent different approaches such as degree of grammaticality, degree of complexity, and so on.
- In order to obtain a truth value, we need to contrast two arguments following the next rule: A: U → L. In a Universe (A) an element (U) has a truth value (L). The rule can be extended in a way that many connected elements which share properties can have the same truth value. A: UxU →. L
- &: It represents the Łukaziewich conjunction and it can be read as an “and”.
- $\equiv$: is an essential connective in this approach. It can be read as “equivalent to” and it is called fuzzy equivalence/equality.
- λ: Can be read as “each”.

3.2 A Fuzzy Grammar structure to explain Degrees of Grammaticality & Complexity

A grammar is considered here as a set of rules. These rules allow the production of linguistic inputs. We show a simplified formula of a fuzzy grammar in a multimodal sense:

$$GR(ph, m, x, s, l, pr, ps)$$

.

A Grammar (GR) is equivalent to the phonetic rules (ph), plus the morphological rules (m), plus the syntax rules (x), plus the semantic rules (s), plus the lexical rules (l), plus the pragmatic rules (pr), plus the prosodic rules (pr). In the formal approach, every element would have its type and would be contrasted with a universe, in this case, a dialect (d with type η). In this regard every dialect would be considered as a language. From the contrast of all the rules of each module in the universe, we would obtain the value of grammaticality.

$$Grammar \equiv \lambda d_\eta \lambda ph_\alpha \lambda m_\beta \lambda x_\gamma \lambda s_\delta \lambda l_\epsilon \lambda pr_\zeta \lambda ps_\kappa \cdot (Ph_{(o\alpha)\eta} d_\eta) ph_\alpha \& (M_{(o\beta)\eta} d_\eta) m_\beta \&$$

$$(X_{(o\gamma)\eta} d_\eta) x_\gamma \& (S_{(o\delta)\eta} d_\eta) s_\delta \& (L_{(o\epsilon)\eta} d_\eta) l_\epsilon \& (pr_{(o\zeta)\eta} d_\eta) pr_\zeta \& (ps_{(o\kappa)\eta} d_\eta) ps_\kappa$$

The syntactic module is taken as an example to explain how this formula works. $(\mathrm{X}_{(o\gamma)\eta} d_\eta) x_\gamma$ means the syntactic module. Regarding this description, the syntax of a grammar has a type γ in a universe η. η is the type of universe. In this case, the universe is considered as a dialect. The syntax of this dialect

---

[1] Note that the use of [ ] means any number/degree between 0 and 1. That could be 0.85512 and so on. Additionally, [0,1] can represent a piece of infinite, and not necessarily real numbers [30-10^{10}] = [0,1]. The discrete approaches represent ⟨0,1⟩ as operators (without [ ]). In that way the use of ⟨0,1⟩ means “either 0 or 1”.

is defined in terms of linguistic rules or constraints. The type o demonstrates the degree of truth according to the value of grammaticality provided by the fuzzy grammar. Thus, the syntactic module (x_γ) is in a dialect (d_η) the degree of truth according to the satisfied/violated syntactic constraints of the dialect ($\mathrm{X}_{(o\gamma)\eta}$). 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 by the speaker's perception). A given input which respects the structures and rules of its dialect will have a high grammaticality value. A given input which triggers a lot of violations will display more complex structures since those are not structures which belong to the grammar evaluating the input. Therefore, the higher the value of grammaticality in an input, the lower the value of its complexity in a determinate grammar.

4 Property Grammars: A contraint-based theory for dealing with Fuzziness & Gradience

Regarding fuzzy grammar, Blache's [12], [13], [14] Property Grammars have been chosen as the formal theoretical framework in defining natural language fuzziness and variability. This theory combines a full-constraint framework of independent and flexible constraints (or properties), with syntactic dependencies under the notion of construction from Construction Grammars. Constructions have been described in terms of their properties. Property Grammars display several constraints in order to describe the syntactic relations between local language phenomena. However, here we focus on the following ones:

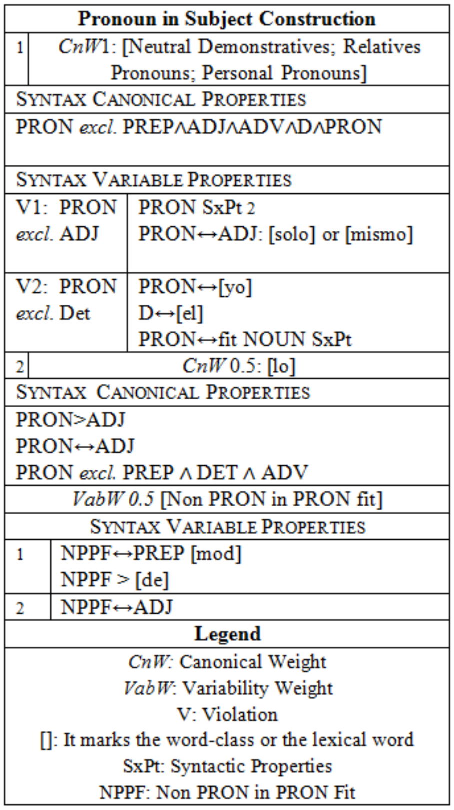
- *Linearity* ($>$): Precedence order between two elements. A precedes B.
- *Requirement* ($\leftrightarrow$): Co-occurrence between two elements: A requires B.
- *Exclusion* (excl.): A and B never appear in co-occurrence in the specified construction.

5 An example of Relative Complexity within the boundaries of a Fuzzy Grammar

Figure 1 is a sample of how gradient description of fuzziness and variability could work. We show the formal description of the PRON [pronoun]. Neutral Demonstrative, Relatives and Personal Pronouns are the canonical ones regarding our corpus (Universal Dependency Spanish Treebank Corpus 2.0). The most canonical structure is weighted as 1, a medium canonical is weighted as 0.5, a violation is weighted as -1 and recurrent variability has a 0.5 weight[2].

---

[2] Note that these weights illustrate a basic idea of gradience. They are not related to the real weights of gradience in Spanish syntax. A precise value of gradience for each weight in each set or construction will be established in the future. We emphasize that this is currently in progress.



Pronoun in Subject Construction	
1	*CnW*1: [Neutral Demonstratives; Relatives Pronouns; Personal Pronouns]
SYNTAX CANONICAL PROPERTIES	
PRON *excl.* PREP∧ADJ∧ADV∧D∧PRON	
SYNTAX VARIABLE PROPERTIES	
V1: PRON *excl.* ADJ	PRON SxPt 2 PRON↔ADJ: [solo] or [mismo]
V2: PRON *excl.* Det	PRON↔[yo] D↔[el] PRON↔fit NOUN SxPt
2	*CnW* 0.5: [lo]
SYNTAX CANONICAL PROPERTIES	
PRON>ADJ PRON↔ADJ PRON *excl.* PREP ∧ DET ∧ ADV	
VabW 0.5 [Non PRON in PRON fit]	
SYNTAX VARIABLE PROPERTIES	
1	NPPF↔PREP [mod] NPPF > [de]
2	NPPF↔ADJ
Legend	
CnW: Canonical Weight *VabW*: Variability Weight V: Violation []: It marks the word-class or the lexical word SxPt: Syntactic Properties NPPF: Non PRON in PRON Fit	

Fig. 1. Pronoun's Syntactic Properties in Subject Construction.

Our framework can describe inputs with grammatical violations and their variability. The fuzzy phenomenon is explained with a double analysis in the feature description:

- First Phase: *Syntax Canonical Properties*
- Second Phase: *Syntax Variable Properties*

Firstly, a normal parsing is applied. This parser describes the Syntactic Properties considering only the canonicals. The result of this parsing describes both satisfied and violated canonical properties. The canonical deviations with its violations will be defined in terms of properties. The value of the addition between CnW and VW will be divided by the Total amount of Part of Speech (TPS). A value of complexity in terms of grammaticality is provided here ($VG1$: Value of Grammaticality 1):

$$VG_1 = \frac{{}^{+}CnW^{-}VW}{TPS} \tag{1}$$

Straightaway, the parser runs for a second time, taking into account the violations and defining the Syntax Variable Properties. In case some Syntax Property is violated, such as V1 or V2, SVP are triggered. Their weight of violability is going to be mitigated in case the violation respects these new properties. If the new properties are not satisfied, variability is not going to have any effect here and VW would remain as before. After this second analysis, a new value will be provided (*VG2*: Value of Grammaticality 2) following the formula in (2). This formula refers to our capacity to evaluate a linguistic input as a gradient-fuzzy object.

$$VG_2 = \frac{(^{-}VW^{+}VabW)^{+}CnW}{TPS} \tag{2}$$

This system 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. The more transitions the more complex an input will be. Thus, we would assume that the word-class does not undergo a complete transition of membership, but more of context. This explains why other properties must be taken into account regarding variability. Several D [determiner] (especially articles and demonstratives) occur as PRON quite often, but never as often as they occur as a D (articles: 73.10%; demonstratives: 10,44% in more than 4000 occurrences). If those D ever appear as a PRON this framework detects a violation in the first parsing since, canonically, a D must precede N [Noun]. In the second parsing, the following SVP in the determiner will be triggered clarifying how it is possible to have a determiner without a N: Determiner: SVP: V1: D > N: D ↔ PRON *VabW* 0.5 NPPF 1 $\bigvee$ 2. Once this new fit is applied, and D is no longer considered a D but a PRON, the *VabW* impacts the *VW*, providing the above-mentioned a new grammaticality value (*VG2*).

Because the new fit in this case is a PRON, we describe their properties in the PRON. The same happens in V2 where PRON undergo a fit transition to the N syntactic properties and thus, their new properties are located in Noun Construction. In V1 occurs something similar but in a softer way, in which PRON undergo a transition to the properties of the canonical PRON case number 2 [lo].

6 Final remarks

Local Complexity is dependent on an input's rules and structure. Our Natural Fuzzy Grammar (NFG) takes into account what happens when a sentence has rules which are satisfied or violated. A given input has a value of grammaticality according to its grammar (and not by the speaker's perception). The more constraints that are satisfied, the more grammatical it will be. An input which triggers a lot of violations is going to display more variable rules in the fuzzy grammar (as it was shown in the example of the pronoun). The process of a double parsing for variability rules would increase the complexity of the given sentences. In this sense, the lower the value of grammaticality, the higher the value of complexity for a determinate grammar. Besides, the input with violations would probably be more ambiguous, as shown in the example of the pronoun. Therefore, yet more complex.

Some theories in complexity establish that the more rules there are in a sentence, the more complex a sentence is. Actually, in this proposed approach, the complexity of a sentence might be mitigated or reduced in case the grammar rules are satisfied.

7 Acknowledgement

This research has been supported by the Ministerio de Economía y Competitividad and the Fondo Europeo de Desarrollo Regional under the project number FFI2015-69978-P (MINECO/FEDER, UE) of the Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia, Subprograma Estatal de Generación de Conocimiento.

References

1. B. Aarts, "Conceptions of gradience in the history of linguistics," *Language Sciences*, vol. 26, no. 4, pp. 343–389, 2004.
2. D. L. M. Bolinger, *Generality: Gradience and the All-or-none.* Mouton & Company, 14 ed., 1961.
3. F. Keller, *Gradience in grammar: Experimental and computational aspects of degrees of grammaticality.* PhD thesis, Edinburgh: University of Edinburgh, 2000.
4. L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.
5. L. A. Zadeh, "A fuzzy-set-theoretic interpretation of linguistic hedges," *Journal of Cybernetics*, vol. 2, no. 3, pp. 4–34, 1972.
6. J. Mc Worther, "The world's simpliest grammars are creole grammars," *Linguistic Typology*, no. 5, 2001.
7. Ö. Dahl, *The growth and maintenance of linguistic complexity*, vol. 71. John Benjamins Publishing, 2004.
8. P. Blache, "A computational model for linguistic complexity.," *Biology, Computation and Linguistics.*, vol. 288, pp. 155–167, 2011.
9. E. Lindstrom, "Language complexity and interlinguistic difficulty," *Language Complexity: Typology, Contact, Change*, vol. 94, p. 217, 2008.
10. N. Chomsky, *Aspects of the theory of syntax.* Cambridge: MIT Press, 1965.
11. V. Novák, "Fuzzy natural logic: Towards mathematical logic of human reasoning," in *Towards the Future of Fuzzy Logic*, pp. 137–165, Springer, 2015.
12. P. Blache, "Property Grammars and the Problem of Constraint Satisfaction," *Proc. of ESSLLI 2000 workshop on Linguistic Theory and Grammar Implementation*, pp. 47–56, 2000.
13. P. Blache, "Property Grammars: A Fully Constraint-based Theory," *Constraint Solving and Language Processing*, vol. 3438, pp. 1–16, 2005.
14. P. Blache, "Representing Syntax by Means of Properties : a Formal Framework for Descriptive Approaches," *Journal of Language Modelling*, vol. 4, no. 2, pp. 183–224, 2016.